

Statistical Methods for Analyzing Extremes

Instructor: Whitney Huang



School of
**MATHEMATICAL AND
STATISTICAL SCIENCES**
Clemson University

✉: wkhuang@clemson.edu
↗: <https://whitneyhuang83.github.io/>



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

Golden, CO COLORADO SCHOOL OF
MINES , November 6 2024



samsi
NSF Duke NCSU UNC

NCAR

PACIFIC CLIMATE
IMPACTS CONSORTIUM

University
of Victoria



Agenda

Statistical Modeling: Context and Workflow

Univariate Extreme Value Theory

Probability Framework

Block-Maximum and Threshold-Exceedance Approaches

Temporal Dependence and Non-Stationary

Multivariate Extremes

Tail dependence

Dependence Modeling via Multivariate Extreme Value Models

Conditional Extreme Value Models

Spatial Extremes

Climate and Weather Spatial Effects

Bayesian Hierarchical Approach

Closing

Outline

Statistical Modeling: Context and Workflow

Univariate Extreme Value Theory

Probability Framework

Block-Maximum and Threshold-Exceedance Approaches

Temporal Dependence and Non-Stationary

Multivariate Extremes

Tail dependence

Dependence Modeling via Multivariate Extreme Value Models

Conditional Extreme Value Models

Spatial Extremes

Climate and Weather Spatial Effects

Bayesian Hierarchical Approach

Closing

Introduction to Statistical Modeling

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

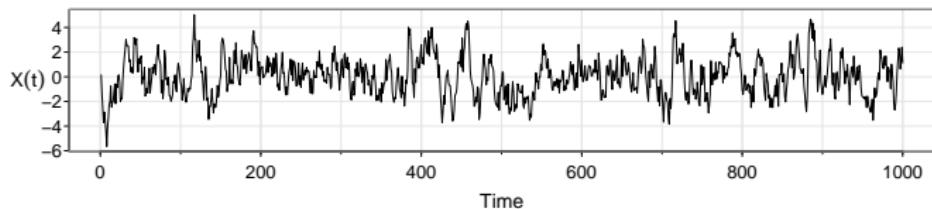
Purposes:

- ▶ 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):

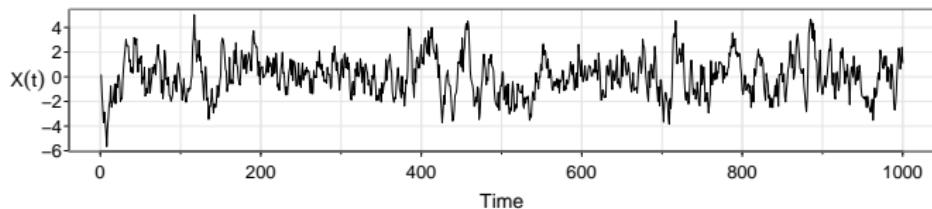
- Prediction: $\hat{x}_{t+1} = \hat{\phi} x_t$

- 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):

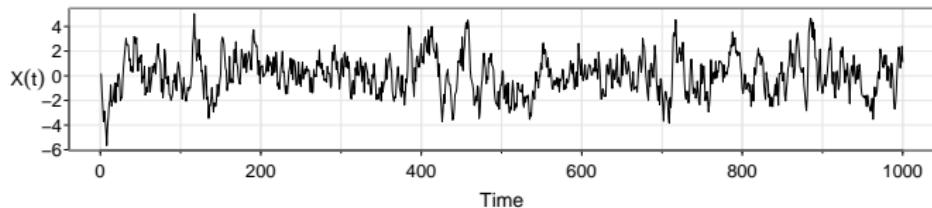
– Prediction: $\hat{x}_{t+1} = \hat{\phi} x_t$

– 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):

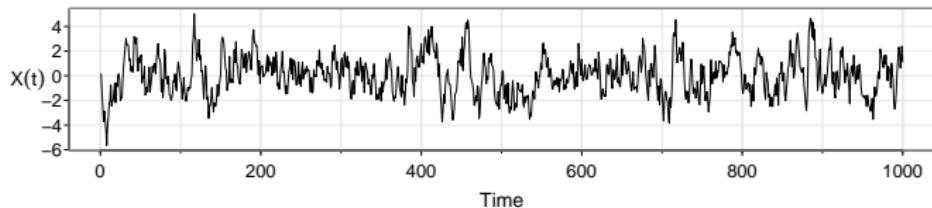
– Prediction: $\hat{x}_{t+1} = \hat{\phi} x_t$

– 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):

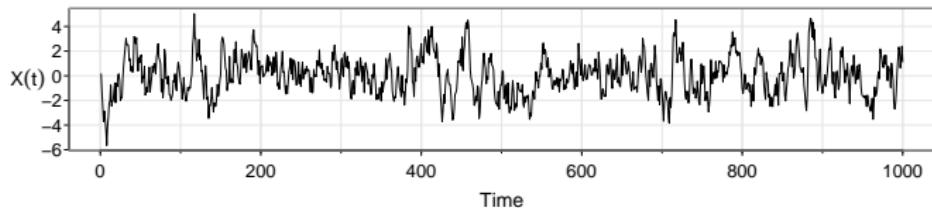
– Prediction: $\hat{x}_{t+1} = \hat{\phi} x_t$

– 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):

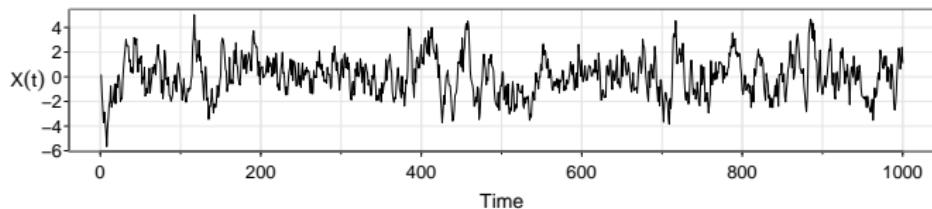
- Prediction: $\hat{x}_{t+1} = \hat{\phi} x_t$

- Simulation: $x_0 \sim P_0, \epsilon \sim P_\epsilon$, for $(t \text{ in } 1 : n)$ $x_{t+1} = \hat{\phi} x_t + \hat{\sigma} \epsilon_{t+1}$

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):

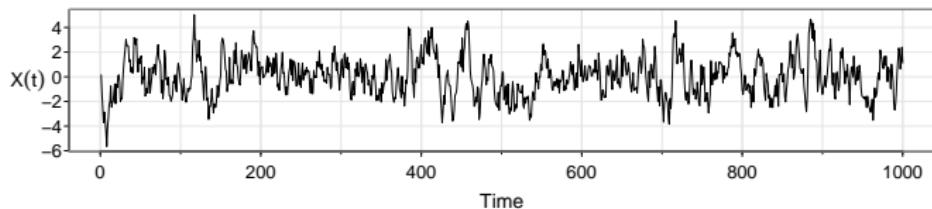
- Prediction: $\hat{x}_{t+1} = \hat{\phi} x_t$

- Simulation: $x_0 \sim P_0, \epsilon \sim P_\epsilon$, for $(t \text{ in } 1 : n)$ $x_{t+1} = \hat{\phi} x_t + \hat{\sigma} \epsilon_{t+1}$

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, ...

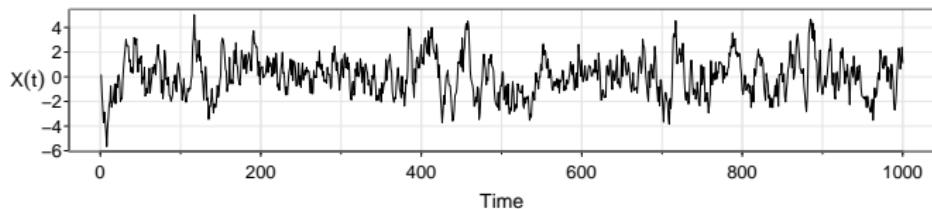
5. Inference, prediction, simulation AR(1):

- Prediction: $\hat{x}_{t+1} = \hat{\phi} x_t$
- Simulation: $x_0 \sim P_0, \epsilon \sim P_\epsilon$, for $(t \text{ in } 1 : n)$ $x_{t+1} = \hat{\phi} x_t + \hat{\sigma} \epsilon_{t+1}$

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, ...

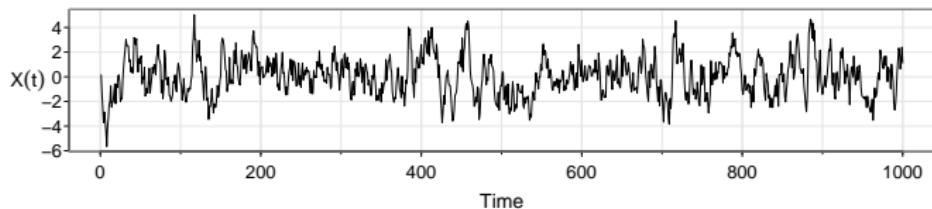
5. Inference, prediction, simulation AR(1):

- Prediction: $\hat{x}_{t+1} = \hat{\phi} x_t$
- Simulation: $x_0 \sim P_0, \epsilon \sim P_\epsilon$, for $(t \text{ in } 1 : n)$ $x_{t+1} = \hat{\phi} x_t + \hat{\sigma} \epsilon_{t+1}$

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,...

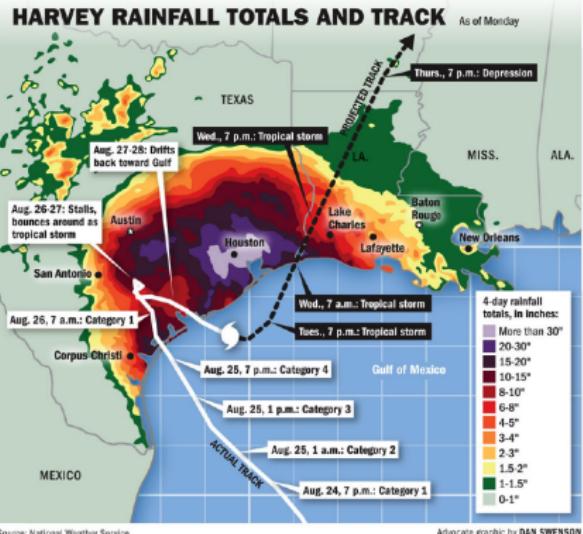
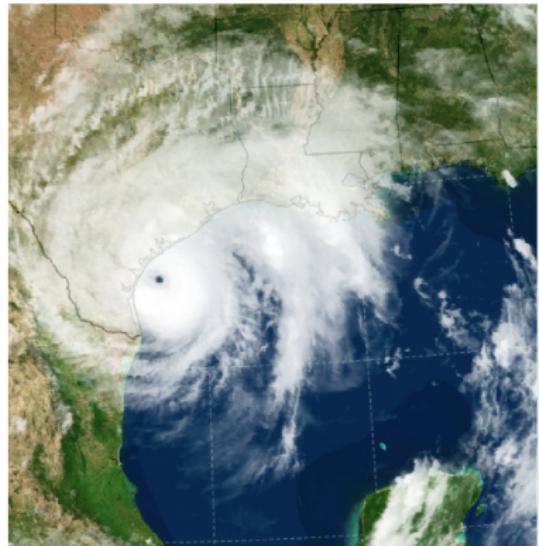
4. Evaluation of fitted model: QQ plot, out-of-sample validation, mean-squared error, scoring rules, ...

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 $(t \text{ in } 1 : n)$ $x_{t+1} = \hat{\phi} x_t + \hat{\sigma} \epsilon_{t+1}$

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

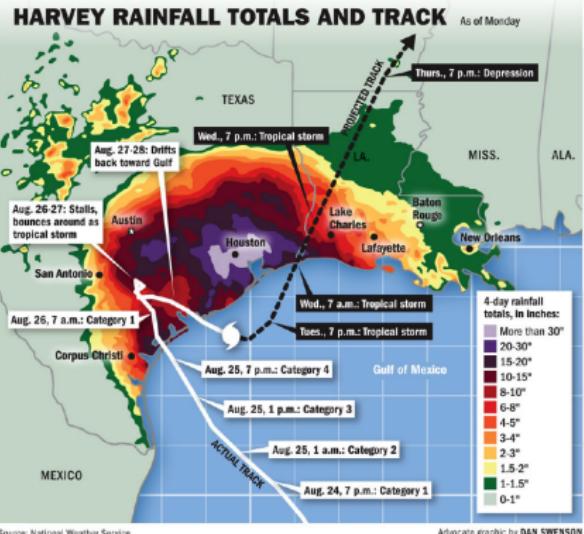
Extreme Rainfall During Hurricane Harvey



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)

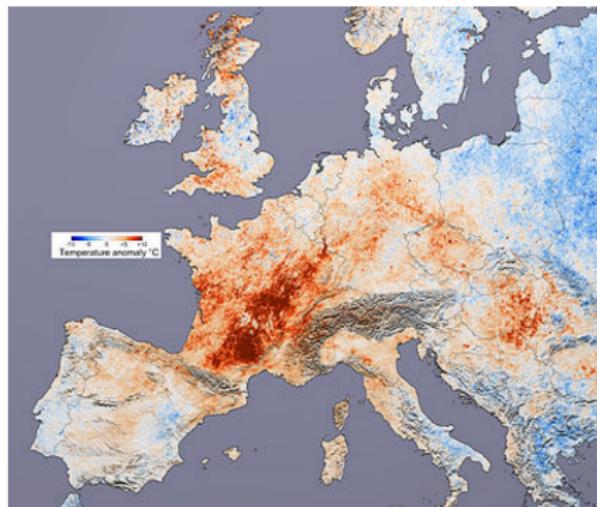
Extreme Rainfall During Hurricane Harvey



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)**

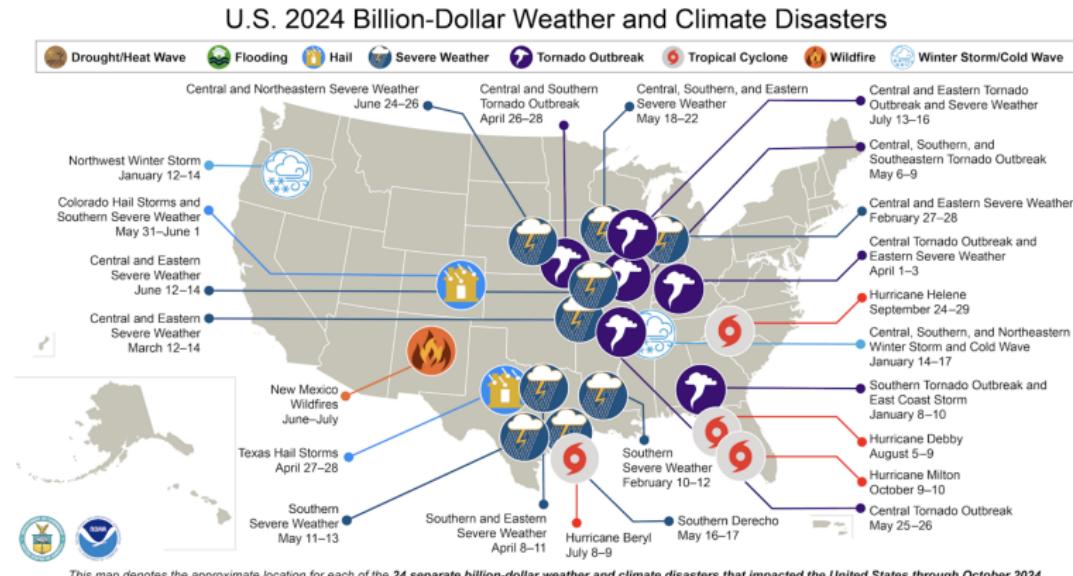
Some Examples of Weather/Climate Extremes



- ▶ **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.

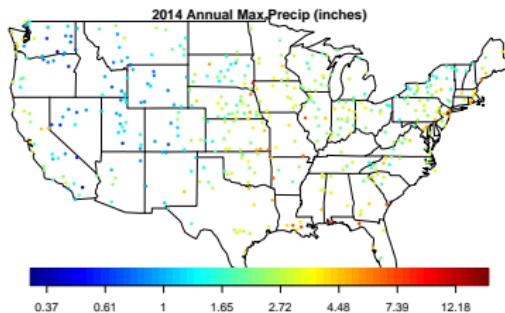


Source: National Oceanic and Atmospheric Administration

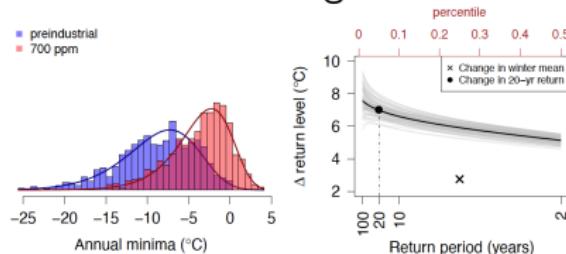
Goal: to quantify the tail behavior \Rightarrow may requires extrapolation.

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

Statistical Modeling: Context and Workflow

Univariate Extreme Value Theory

Probability Framework

Block-Maximum and Threshold-Exceedance Approaches

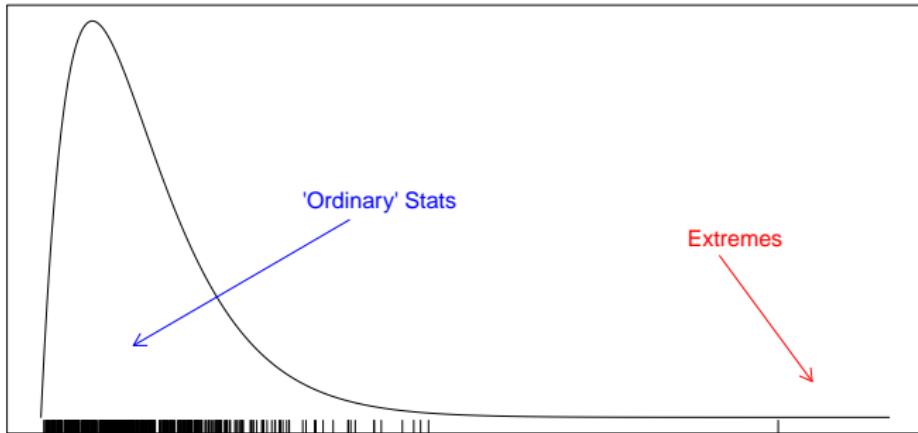
Temporal Dependence and Non-Stationary

Multivariate Extremes

Spatial Extremes

Closing

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

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

Remark

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

⇒ 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 $S_n = \sum_{i=1}^n X_i$

⇒ 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 $S_n = \sum_{i=1}^n X_i$

⇒ 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]

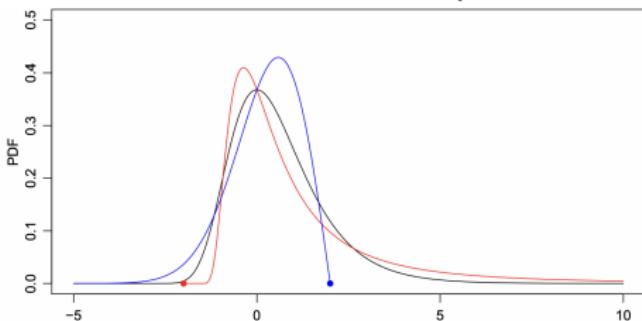
Define $M_n = \max\{X_1, \dots, X_n\}$ where $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} F$. If $\exists a_n > 0$ and $b_n \in \mathbb{R}$ such that, as $n \rightarrow \infty$, if

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

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

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

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

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

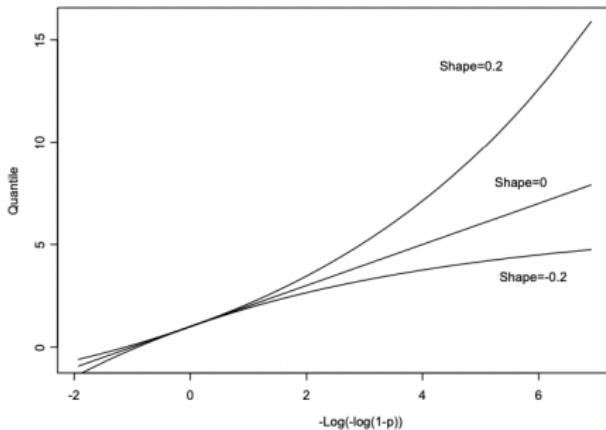
for some constants $a_k > 0$ and b_k

- ▶ 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

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



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

Outline

Statistical Modeling: Context and Workflow

Univariate Extreme Value Theory

Probability Framework

Block-Maximum and Threshold-Exceedance Approaches

Temporal Dependence and Non-Stationary

Multivariate Extremes

Spatial Extremes

Closing

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

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

Maximum Likelihood Estimation

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

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

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

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

The **maximum likelihood estimate (MLE)** is obtained by (numerically) maximizing the 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 methods for constructing confidence intervals (CIs):

- ▶ **Delta method**

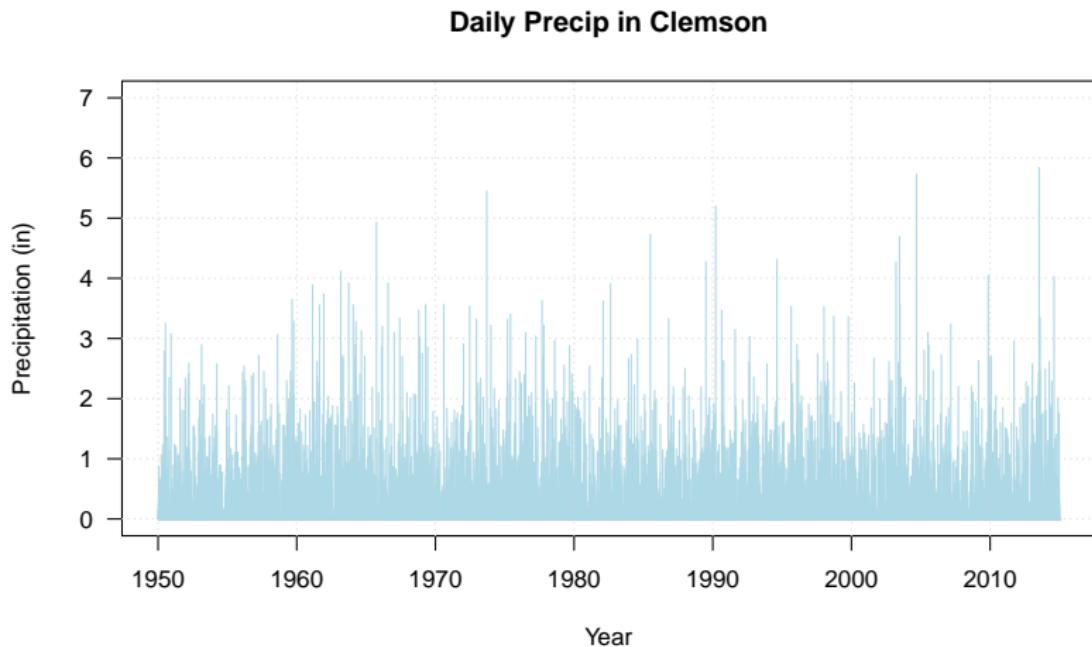
- ▶ +: easy to compute with a closed form expression
- ▶ -: symmetric confidence interval is not realistic (especially for long return levels)

- ▶ **Profile likelihood method**

- ▶ +: can allow for asymmetric confidence intervals
- ▶ -: need to compute numerically

Clemson Daily Precipitation [Data Source: USHCN²]

These data were compiled by [Will Kleiber](#), thanks Will!

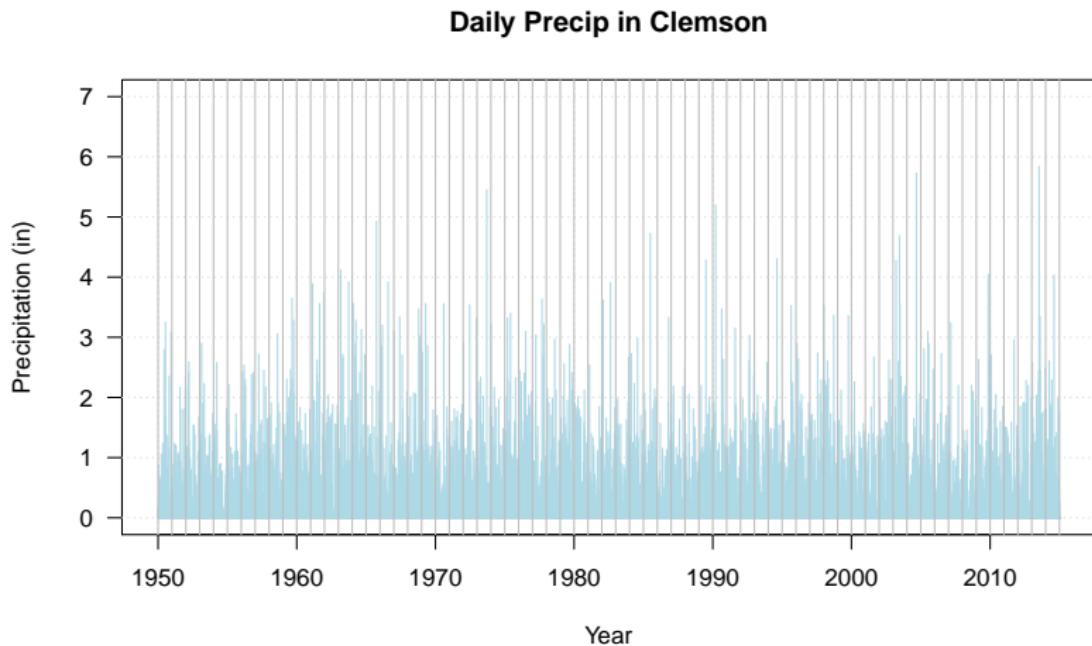


²United States Historical Climatology Network

(<http://cdiac.ornl.gov/epubs/ndp/ushcn/ushcn.html>)

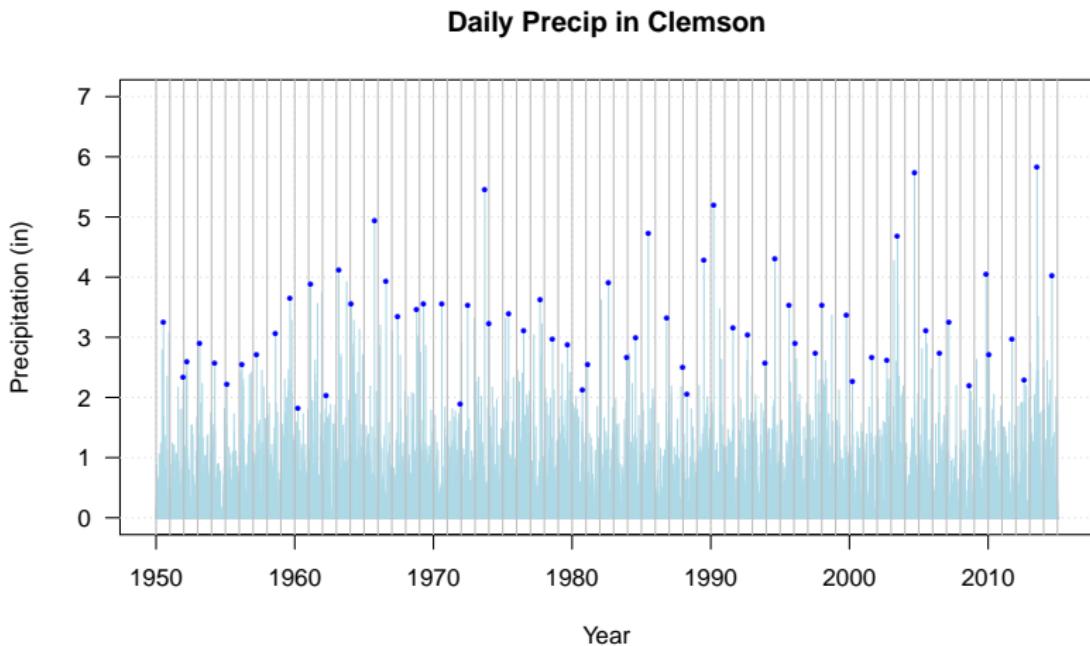
Block-Maximum Method [Gumbel 1958]

- Determine the **block size** and extract the block maxima



Block-Maximum Method [Gumbel 1958]

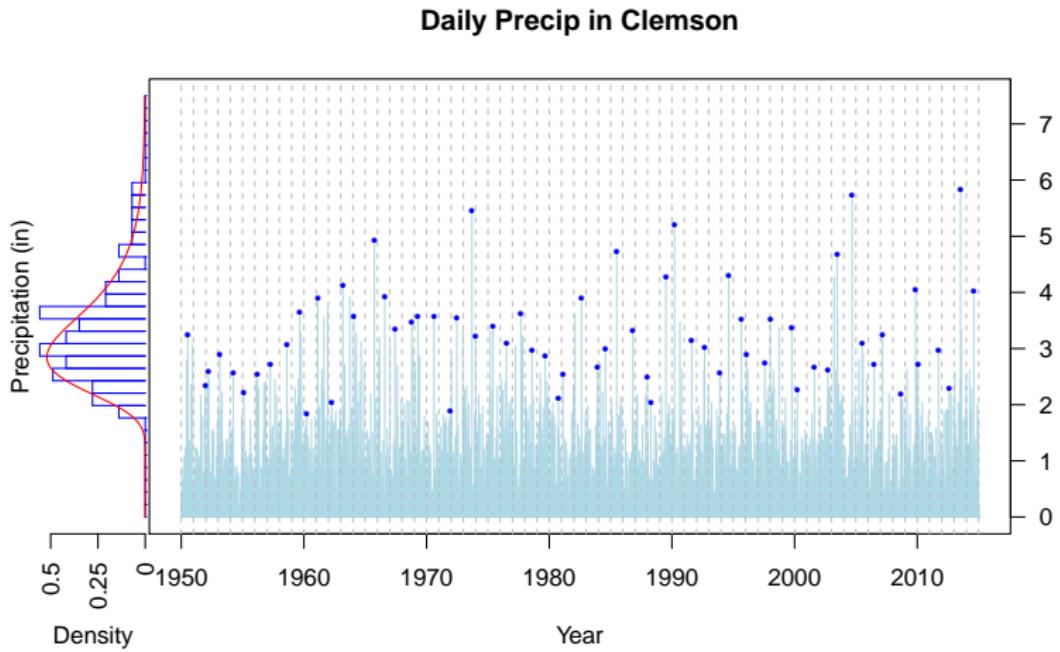
- Determine the block size and extract the block maxima



Block-Maximum Method [Gumbel 1958]

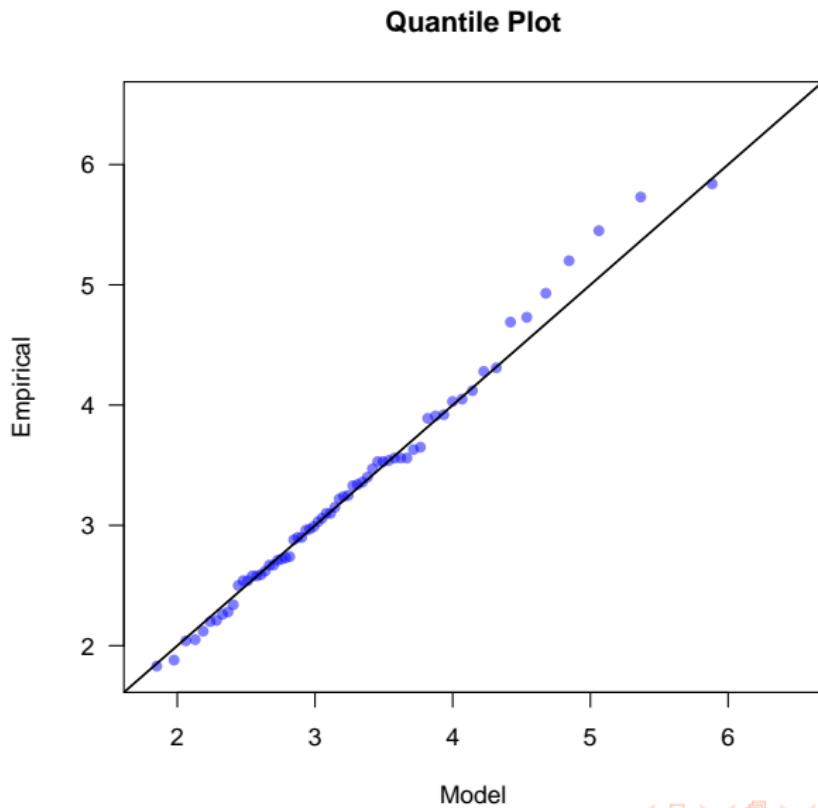
2. Fit the GEV to the maxima and assess the fit.

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



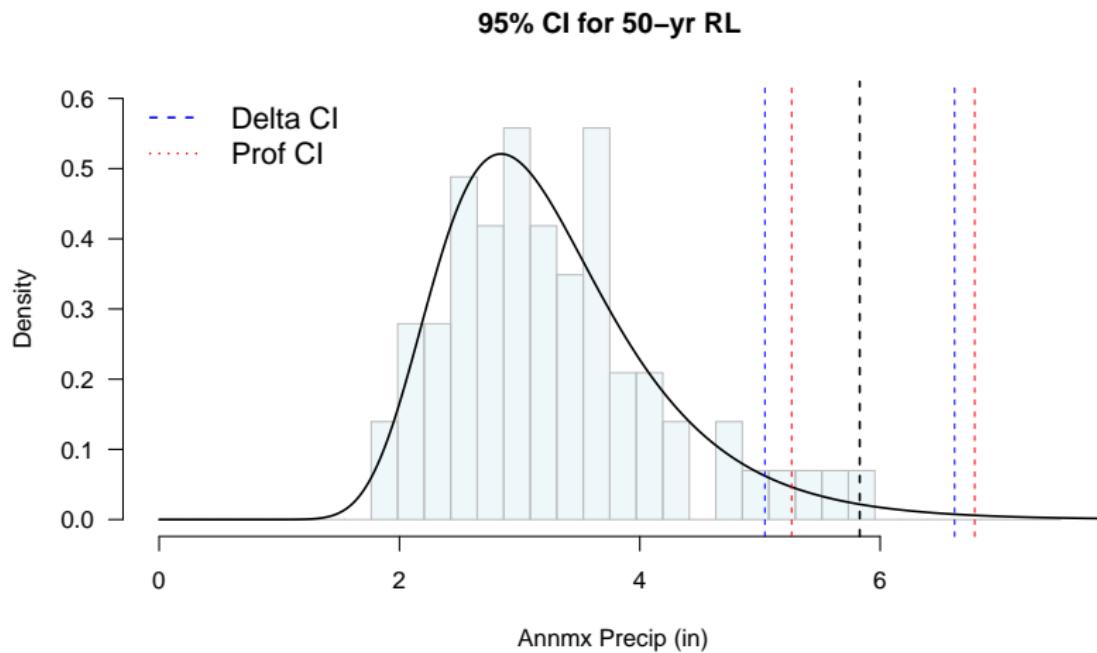
Block-Maximum Method [Gumbel 1958]

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

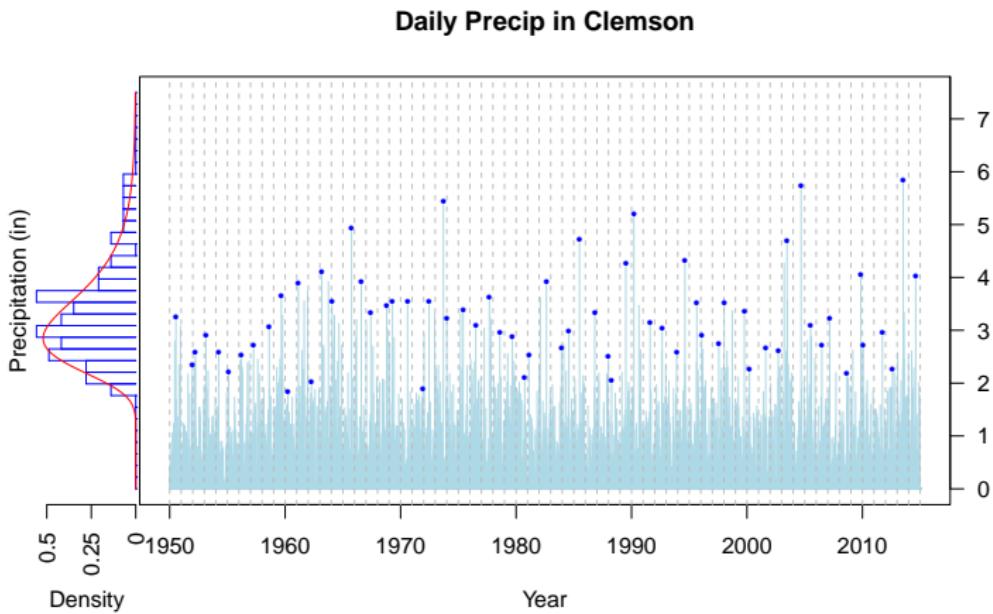


Block-Maximum Method [Gumbel 1958]

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



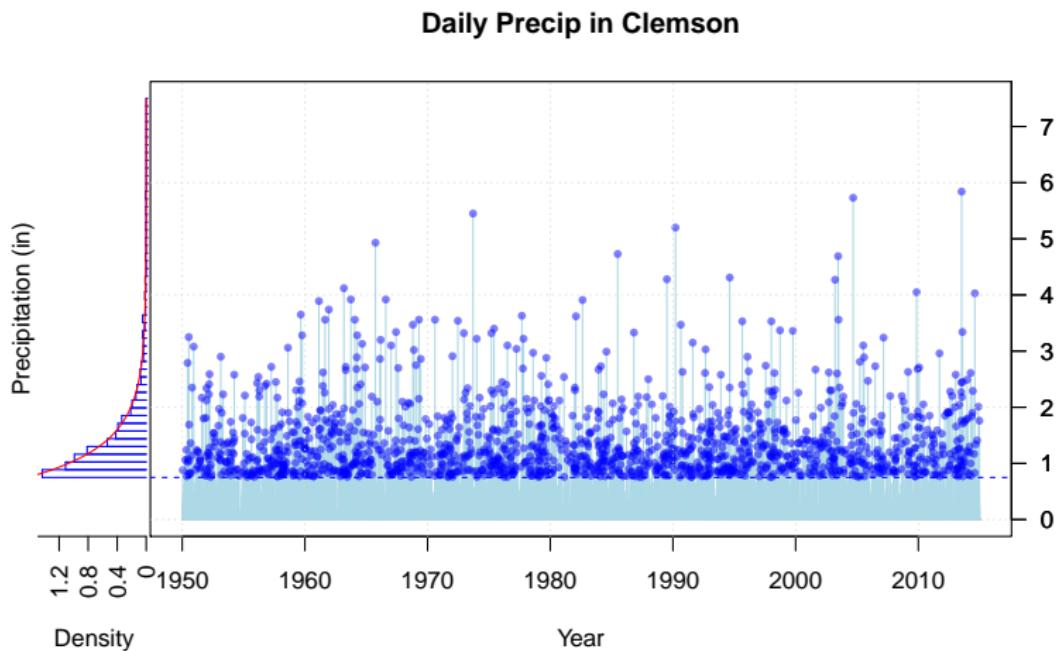
Can We Get More Data for Extreme Value Analysis?



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

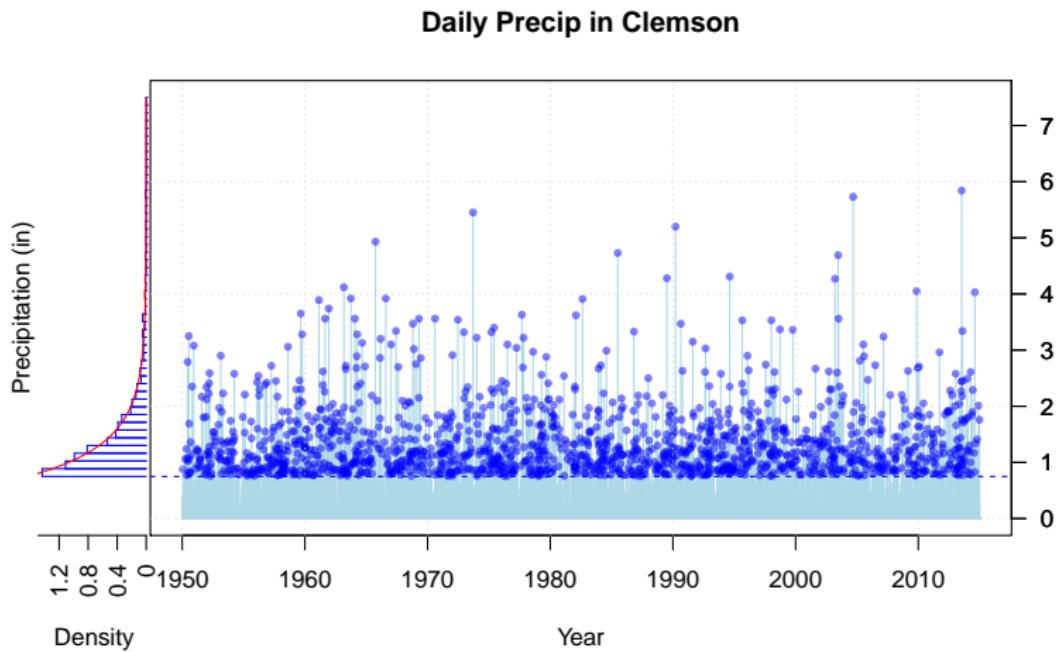
Threshold-Exceedance Method [Davison & Smith 1990]

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 $M_n = \max_{i=1,\dots,n} X_i$ (for a large n) can be approximated by a $\text{GEV}(\mu, \sigma, \xi)$, then for sufficiently large u ,

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

⇒ 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., $u \rightarrow x_F = \sup\{x : F(x) < 1\}$),

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

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

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

where

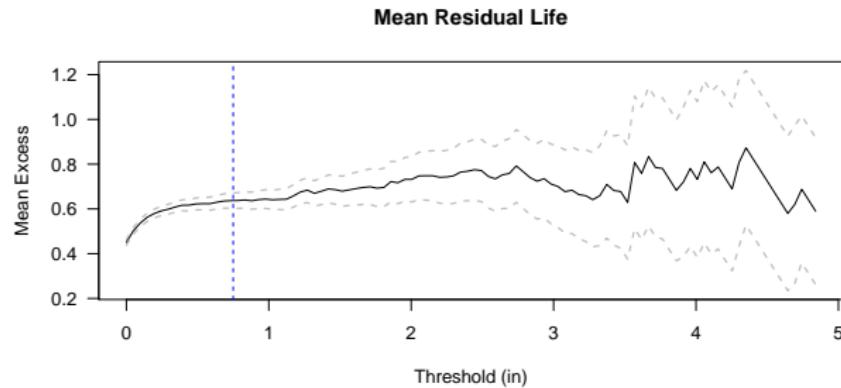
$$H_{\tilde{\sigma}, \xi}(y) = \begin{cases} 1 - (1 + \xi y / \tilde{\sigma})^{-1/\xi} & \xi \neq 0; \\ 1 - \exp(-y / \tilde{\sigma}) & \xi = 0. \end{cases}$$

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

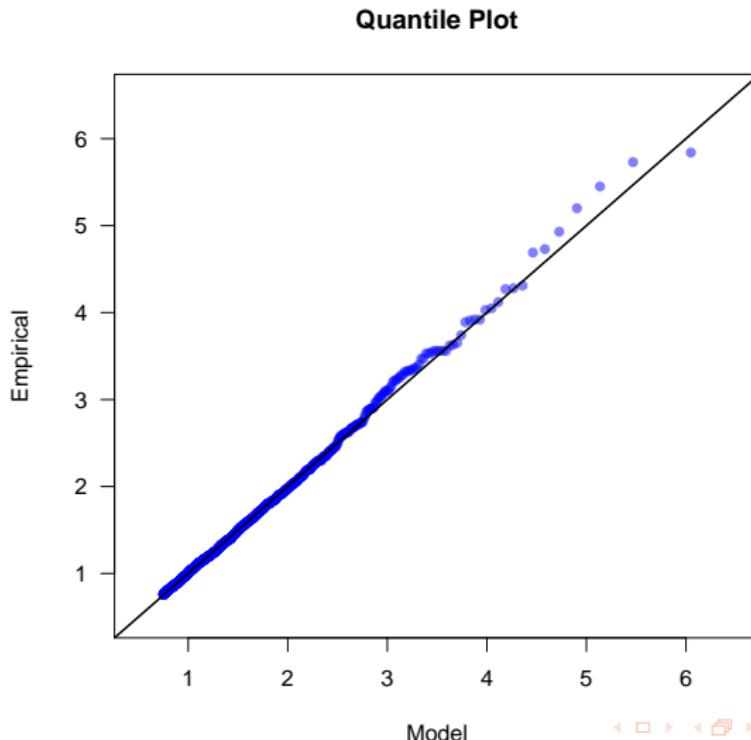


Task: To choose a u_0 s.t. the **mean residual life curve** starts to behave linearly for all $u > u_0$

Threshold-Exceedance Method [Davison & Smith 1990]

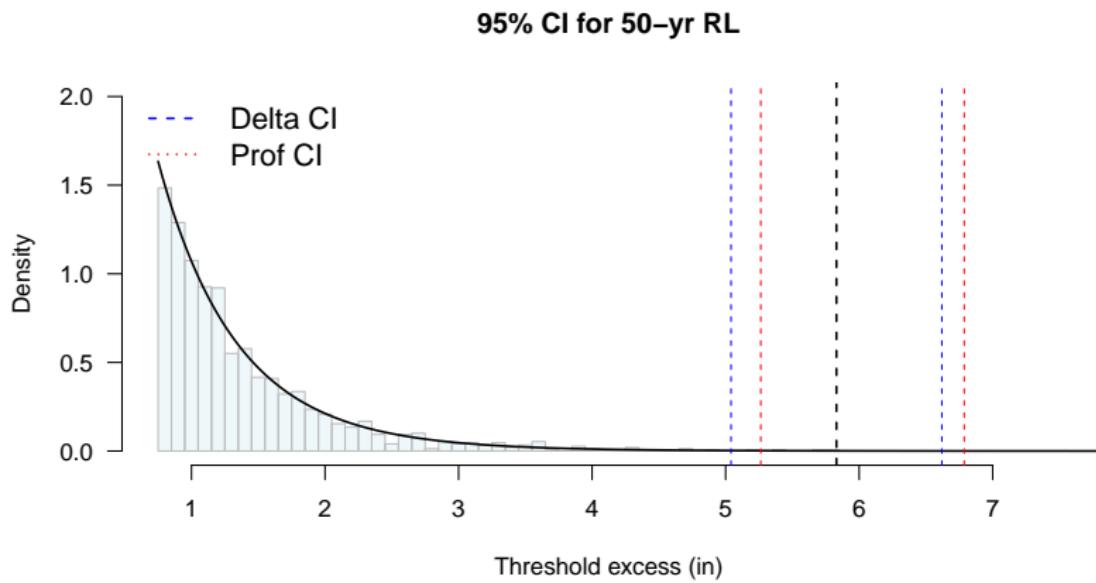
2. Fit the GPD to exceedances and assess the fit

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



Threshold-Exceedance Method [Davison & Smith 1990]

3. Perform inference for return levels, probabilities, etc



Outline

Statistical Modeling: Context and Workflow

Univariate Extreme Value Theory

Probability Framework

Block-Maximum and Threshold-Exceedance Approaches

Temporal Dependence and Non-Stationary

Multivariate Extremes

Spatial Extremes

Closing

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 😞

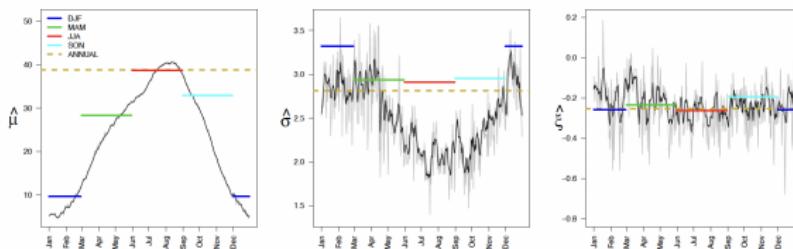
Modeling Threshold Exceedances

- ▶ **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

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



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

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

and 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 ⇒ **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**

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

Outline

Statistical Modeling: Context and Workflow

Univariate Extreme Value Theory

Multivariate Extremes

Tail dependence

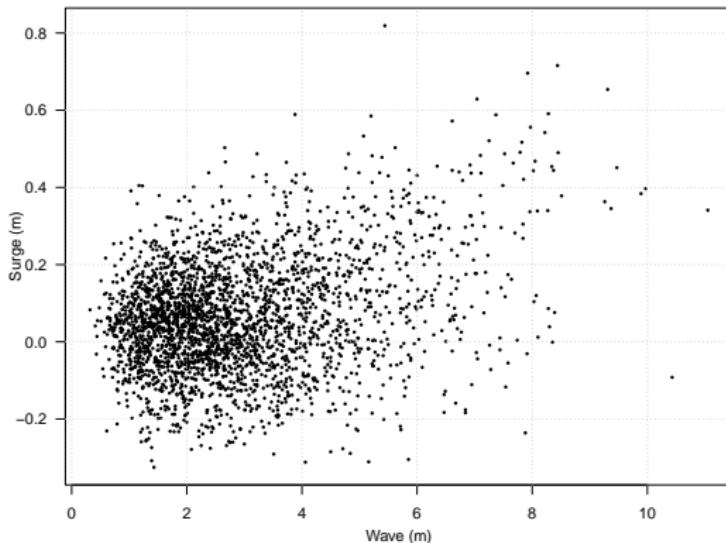
Dependence Modeling via Multivariate Extreme Value Models

Conditional Extreme Value Models

Spatial Extremes

Closing

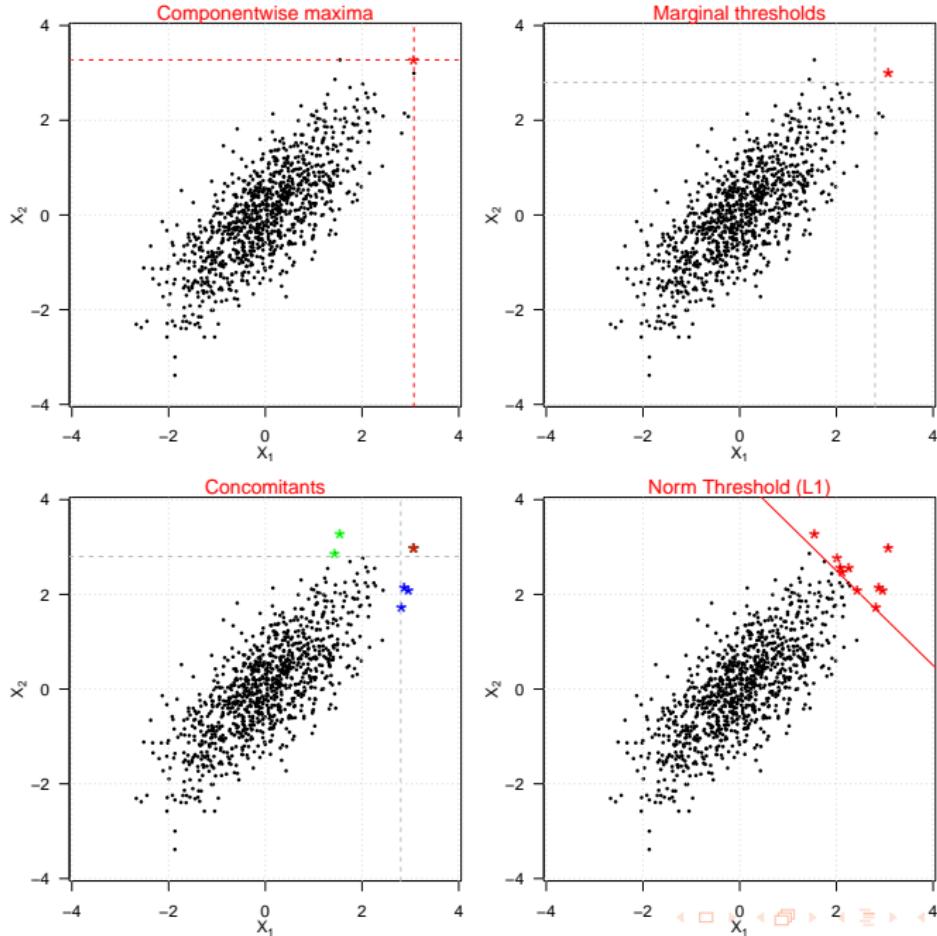
Wave–Surge Data at Newlyn, UK



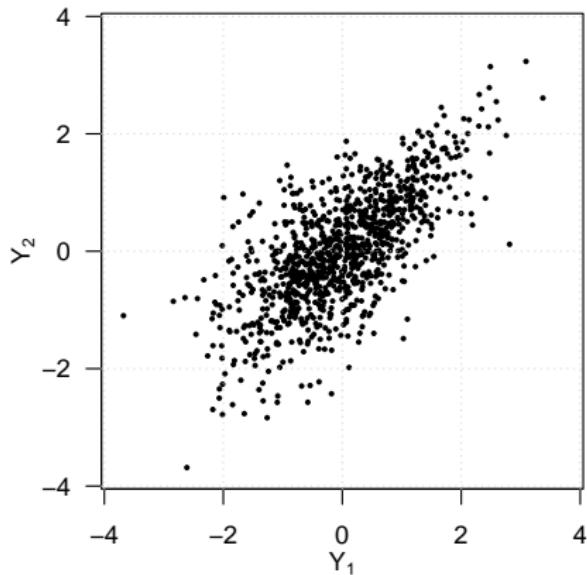
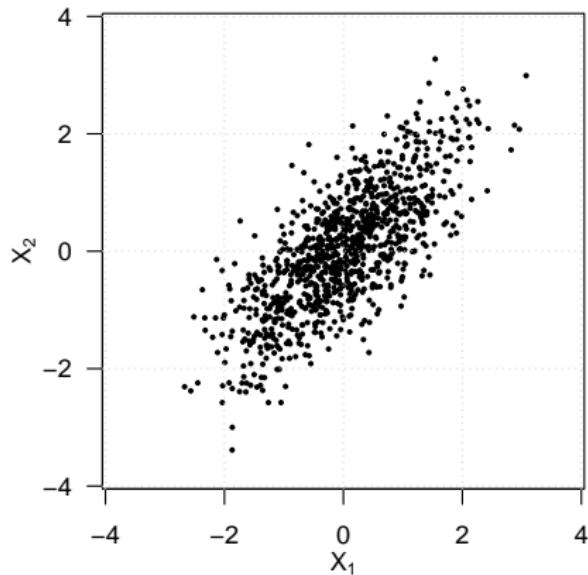
- ▶ **Scientific motivation:** many extremal problems are intrinsically multivariate
- ▶ **Statistical motivation:** uncertainty may be reduced by combining information from several sources

Challenge: what is “extreme” in two or more dimensions?

What is a Multivariate Extreme?

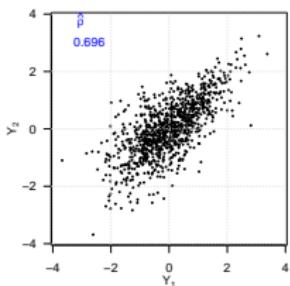
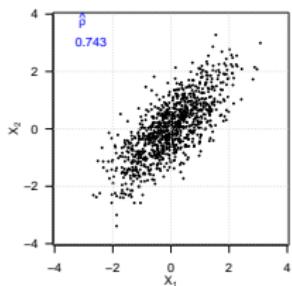


Describing Tail Dependence



A central aim of multivariate extremes is to describe **tail dependence**

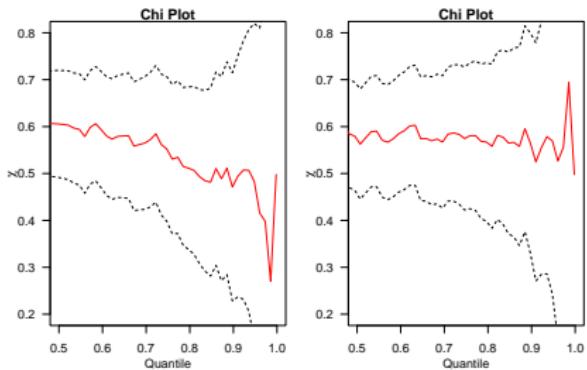
Correlation Coefficient and Tail Dependence Parameter



Correlation Coefficient

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

ρ measures “spread from center”
⇒ does not focus on extremes



Upper tail dependence parameter:

$$\begin{aligned}\chi &= \lim_{u \rightarrow 1} \chi(u) \\ &= \mathbb{P}(F_{X_1}(X_1) > u | F_{X_2}(X_2) > u)\end{aligned}$$

Outline

Statistical Modeling: Context and Workflow

Univariate Extreme Value Theory

Multivariate Extremes

Tail dependence

Dependence Modeling via Multivariate Extreme Value Models

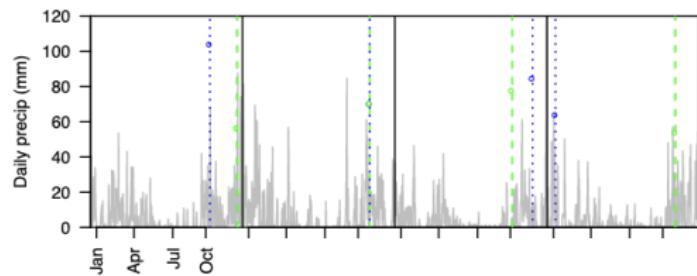
Conditional Extreme Value Models

Spatial Extremes

Closing

Componentwise Maxima

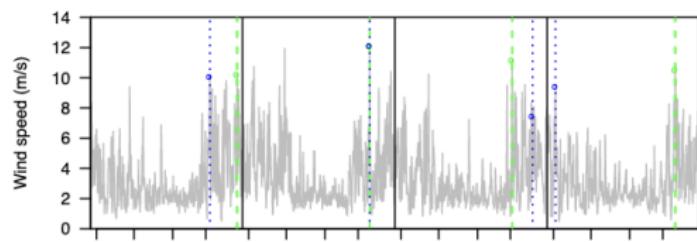
Let $\{X_i = (X_{1,i}, \dots, X_{d,i})\} \in \mathbb{R}^d$ are i.i.d. random vectors. Let's consider **componentwise maxima**: $M_n = (M_{1,n}, \dots, M_{d,n})^T$, where $M_{j,n} = \max_{i=1}^n X_{j,i}$



► M_n is not necessarily one of 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 $\mathbf{X}_1, \dots, \mathbf{X}_n \stackrel{\text{iid}}{\sim} \mathbf{F}$ where $\mathbf{X}_i \in \mathbb{R}^d$, if

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

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

- ▶ Each marginal (approximately) follows a GEV distribution (i.e., $M_j \approx \text{GEV}(\mu_j, \sigma_j, \xi_j)$)
- ▶ 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 ($\text{GEV}(1, 1, 1)$)

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

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

- ▶ Then

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

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 $m_2 > 0$, $\alpha \in (0, 1)$

- ▶ $\alpha \rightarrow 1$ corresponds to independent
- ▶ $\alpha \rightarrow 0$ 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 $\alpha = \beta \rightarrow 1$ 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

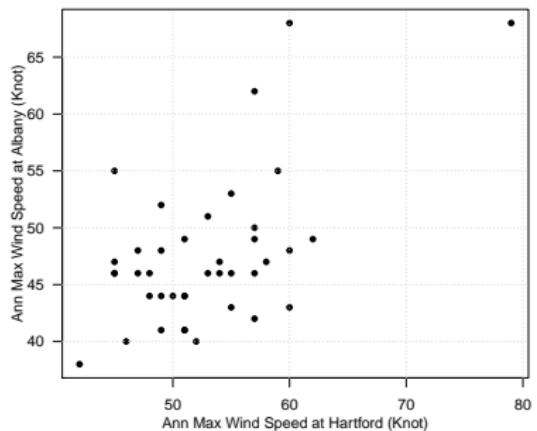
The probability density function is

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

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 maximize 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

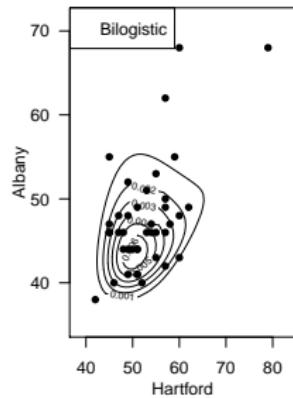
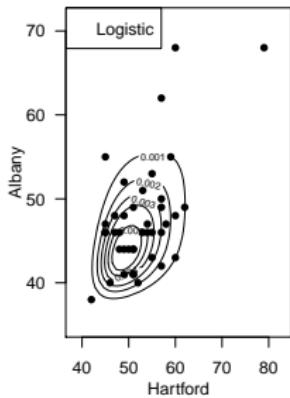


Marginal Parameter Estimates

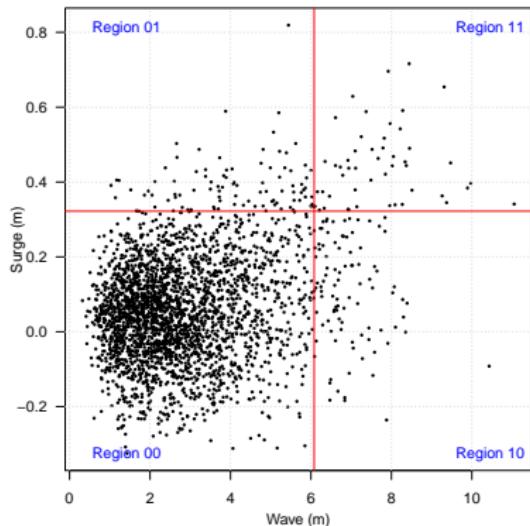
Model	Logistic	Bilogisitc
$\hat{\mu}_{\text{Hartford}}$	49.97 (0.87)	50.08 (0.84)
$\hat{\sigma}_{\text{Hartford}}$	5.04 (0.64)	5.07 (0.62)
$\hat{\xi}_{\text{Hartford}}$	0.01 (0.09)	-0.04 (0.05)
$\hat{\mu}_{\text{Albany}}$	44.58 (0.77)	44.64 (0.87)
$\hat{\sigma}_{\text{Albany}}$	4.34 (0.57)	4.40 (0.42)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

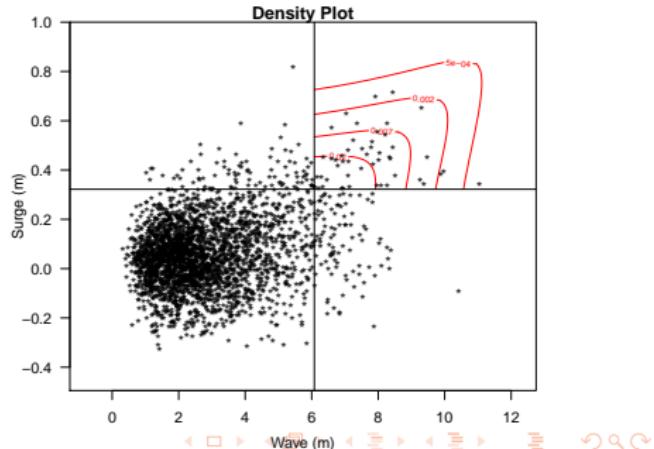


Bivariate Threshold Excess Model



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

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

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

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

Statistical Modeling: Context and Workflow

Univariate Extreme Value Theory

Multivariate Extremes

Tail dependence

Dependence Modeling via Multivariate Extreme Value Models

Conditional Extreme Value Models

Spatial Extremes

Closing

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 ,

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

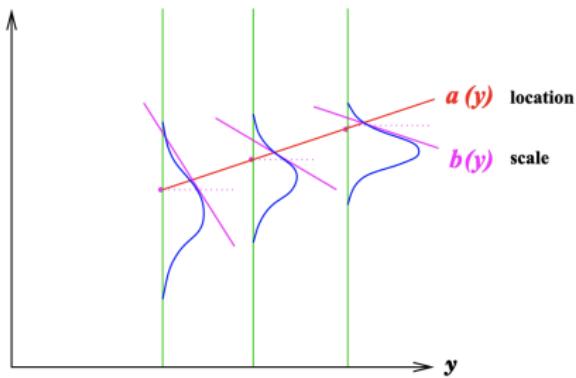
where $a(x) = \alpha x$ and $b(x) = x^\beta$, $\alpha \in [-1, 1]$, $\beta \in (-\infty, 1)$

A cartoon illustration of the CEV dependence modeling

Assume for large u ,

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

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



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

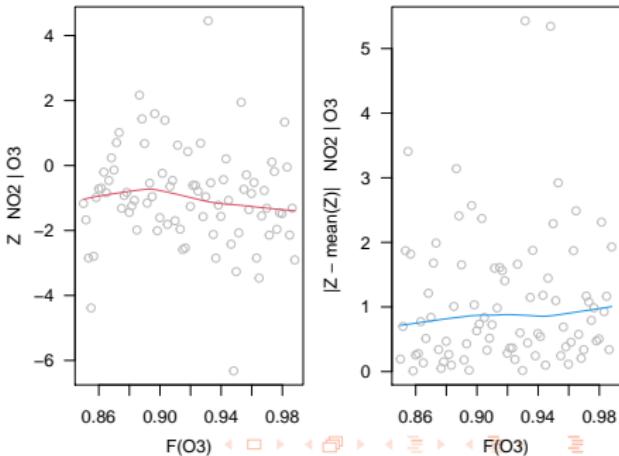
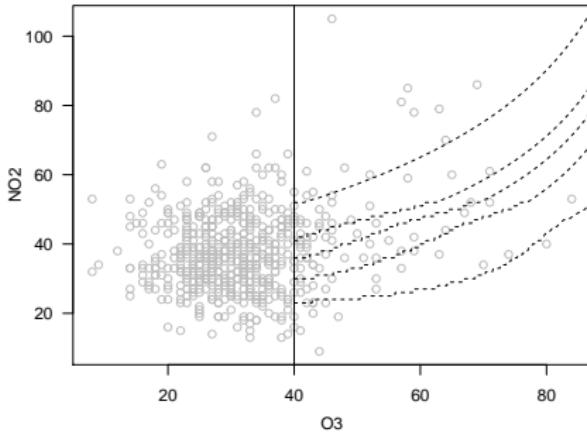
Source: Heffernan's slides given at the Interface 2008 Symposium

Modeling “Summer” O₃ and NO₂ [Heffernan & Tawn, 2004]

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

- Need to choose 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

- ▶ 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**.
- ▶ Conditional extreme value models are used for asymptotic independence cases.

Outline

Statistical Modeling: Context and Workflow

Univariate Extreme Value Theory

Multivariate Extremes

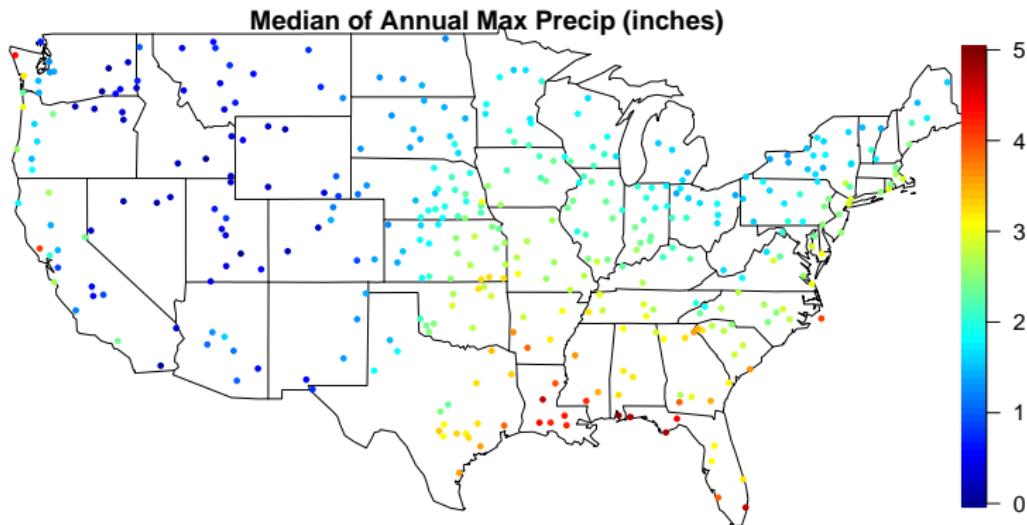
Spatial Extremes

Climate and Weather Spatial Effects

Bayesian Hierarchical Approach

Closing

Climate and Weather Aspects of Spatial Extremes

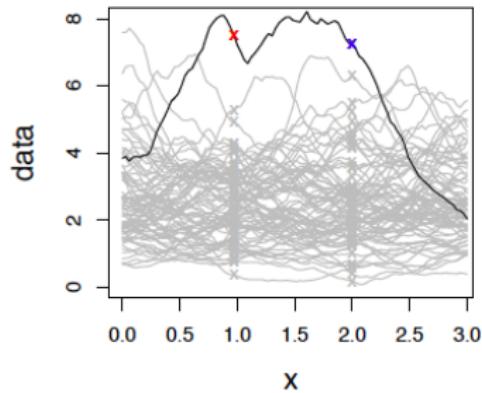
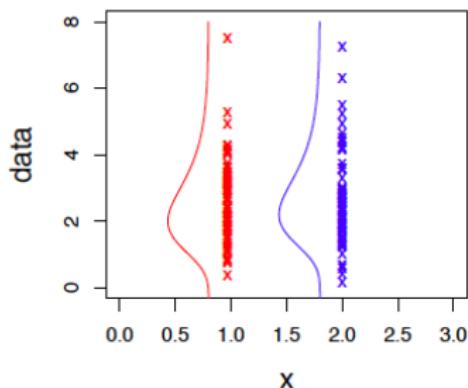


"Climate is what you expect, weather is what you get"

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

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 $s_1, \dots, s_k \in \mathcal{S}$, 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. $\text{GEV}(\mu(s), \sigma(s), \xi(s))$) 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

Statistical Modeling: Context and Workflow

Univariate Extreme Value Theory

Multivariate Extremes

Spatial Extremes

Climate and Weather Spatial Effects

Bayesian Hierarchical Approach

Closing

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

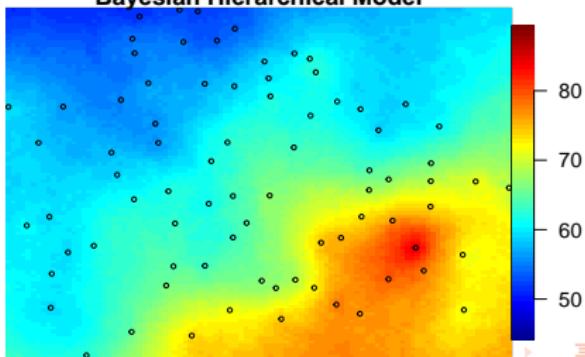
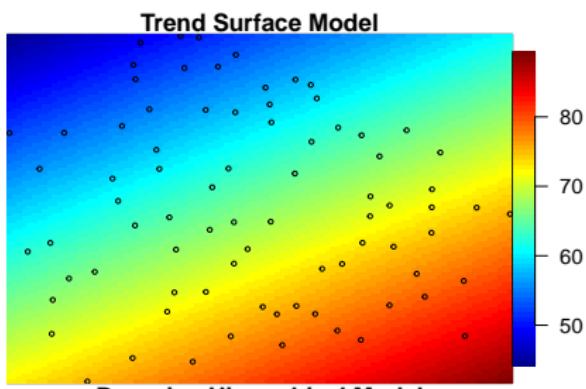
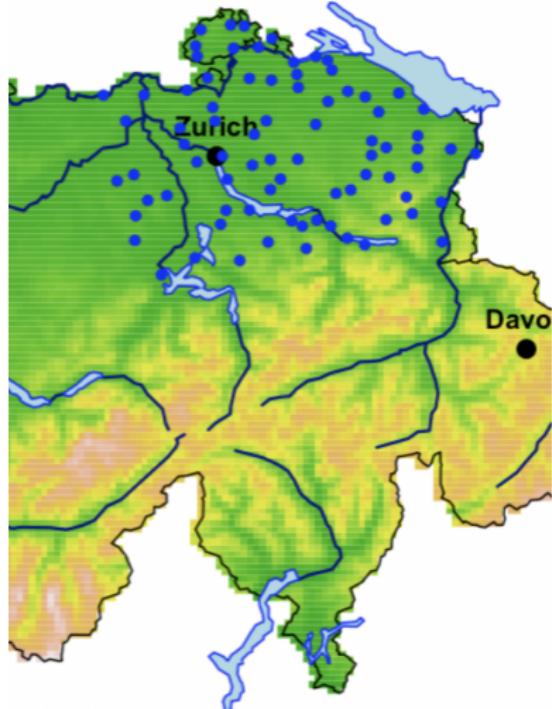
- ▶ **Data:** $M(\mathbf{s}_i) | (\mu(\mathbf{s}), \sigma(\mathbf{s}), \xi(\mathbf{s})) \stackrel{ind}{\sim} \text{GEV}(\mu(\mathbf{s}_i), \sigma(\mathbf{s}_i), \xi(\mathbf{s}_i))$, $\mathbf{s}_i \in \mathcal{S}$
- ▶ **Latent processes:** $(\mu(\mathbf{s}), \log(\sigma(\mathbf{s})))$ are modeled Gaussian Processes $\text{GP}(m_\theta(\cdot), K_\theta(\cdot, \cdot))$ 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 help 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

Statistical Modeling: Context and Workflow

Univariate Extreme Value Theory

Multivariate Extremes

Spatial Extremes

Closing

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 dependence 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