Extremes: Univariate, Multivariate, and Spatial Processes

Dan Cooley
Department of Statistics
Colorado State University

Pan-American Studies Institute
Buzios, Brazil
June 23, 2014

Outline

"Generalized extreme value distributions. Generalized Pareto distribution. Max-stable processes. Composite likelihood."

Survey:

Knowledge	1	2	3	4	5
Spatial Extremes	11	15	7	2	1

1. Univariate
   - Analyzing a subset of extreme data.
   - Distributions for extremes.
2. Multivariate
   - What do we mean by tail dependence?
   - How multivariate extremes models capture tail dependence.
3. Spatial Processes
   - What is a maximal process?
   - Max-stable process models.

Why study extremes?

Although infrequent, extremes have large human impact.

Examples of extreme precipitation in Colorado:

Big Thompson, 1976

- 145 killed
- $41m damage

Ft Collins, 1997

- 5 killed
- $250m damage

N. Colorado, 2013

- 8 killed
- $?? damage

Fort Collins to Boulder $\approx$ 90 km.

Why study extremes?

Although infrequent, extremes have large human impact.

Goal of an extreme value analysis: to quantify the magnitude of a ~~worst-case~~ really-bad-case scenario. Often requires extrapolation.

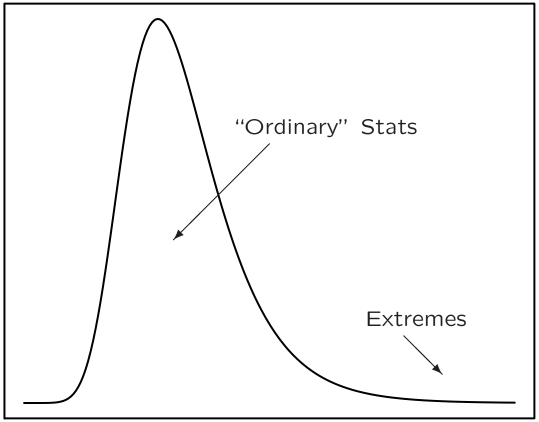
Application areas:

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

"Ordinary" vs Extreme Value Statistics

"Ordinary" Statistics: Describes main part of distribution.

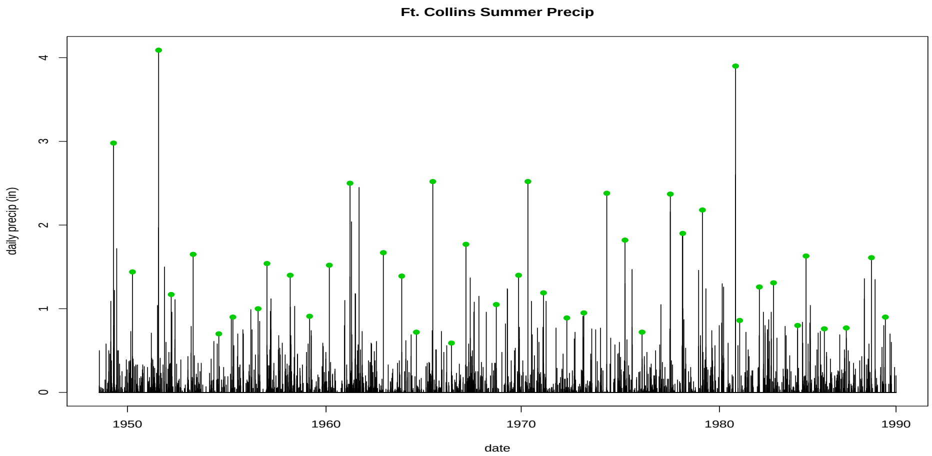
Extremes: Characterizes the tail of the distribution.



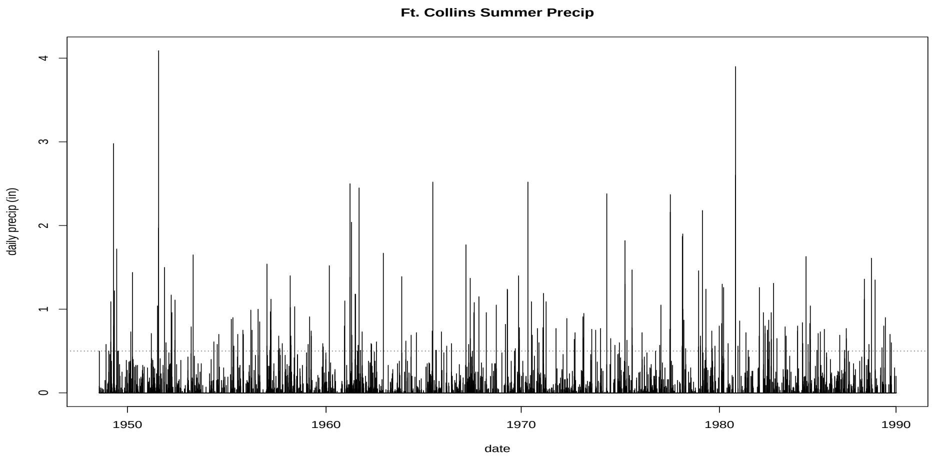
- Relies on asymptotic theory to provide models for the tail.
- Uses only the extreme observations to fit the model:
  Retain only a small percentage of data, discard the rest.

Approaches for Extracting Extreme Subset

Block Maxima (Generalized Extreme Value Dist)



Threshold Exceedances (Generalized Pareto Dist)



How unusual was the Fort Collins event?

Measured value for 1997 event: 6.18 inches.

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

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

Question requires extrapolation into the tail. Largest observation (1948-1990) is 4.09 inches.

Model the data in two ways:

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

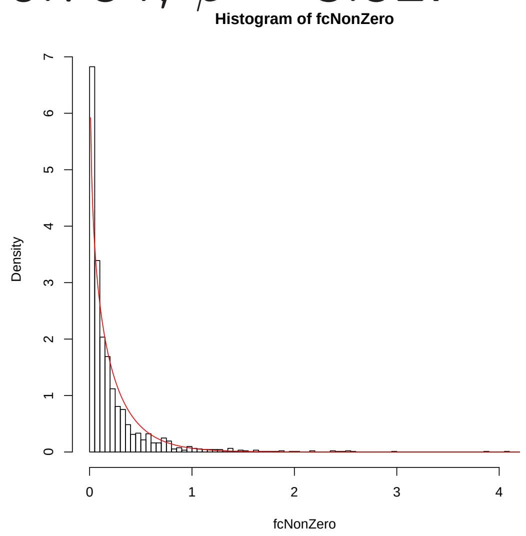
Modeling all precipitation data

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

To model precipitation, we need to account for zeroes.

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

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



All precipitation model estimate

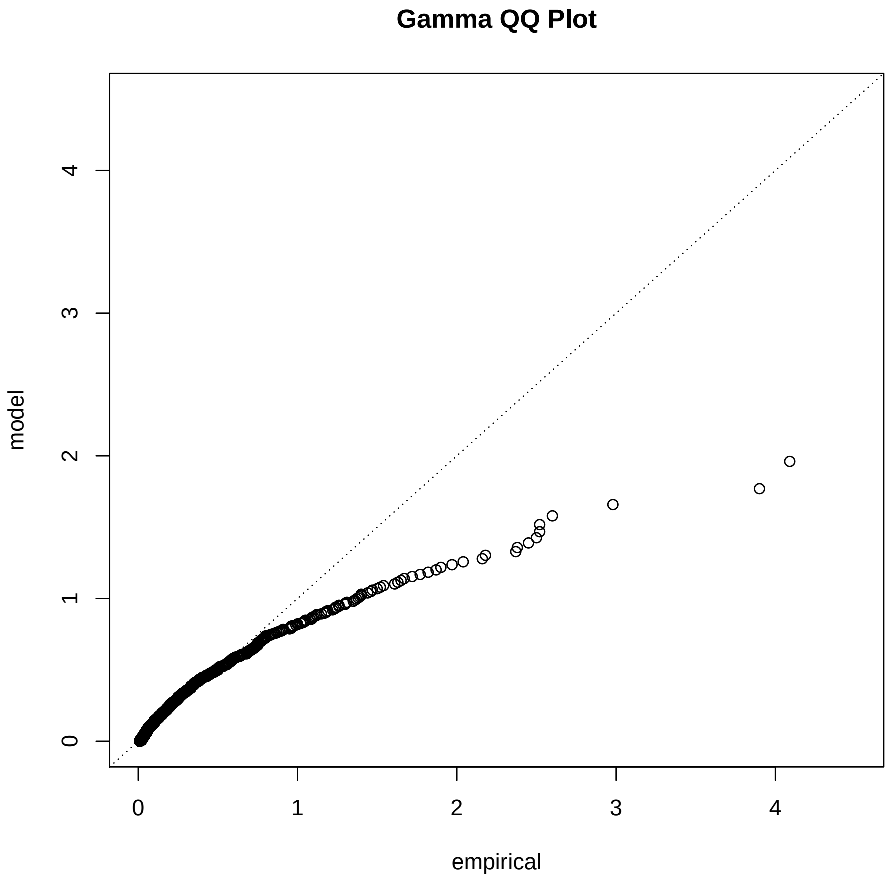
$$\begin{aligned} P(X_t > 6.18) &= P(X_t > 6.18 \mid X_t > 0)P(X_t > 0) \\ &= (1 - F_X(6.18))(0.218) \\ &= 1.47 * 10^{-10}(0.218) = 3.20 * 10^{-11} \end{aligned}$$

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

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

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

All precipitation model



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

Modeling annual maxima

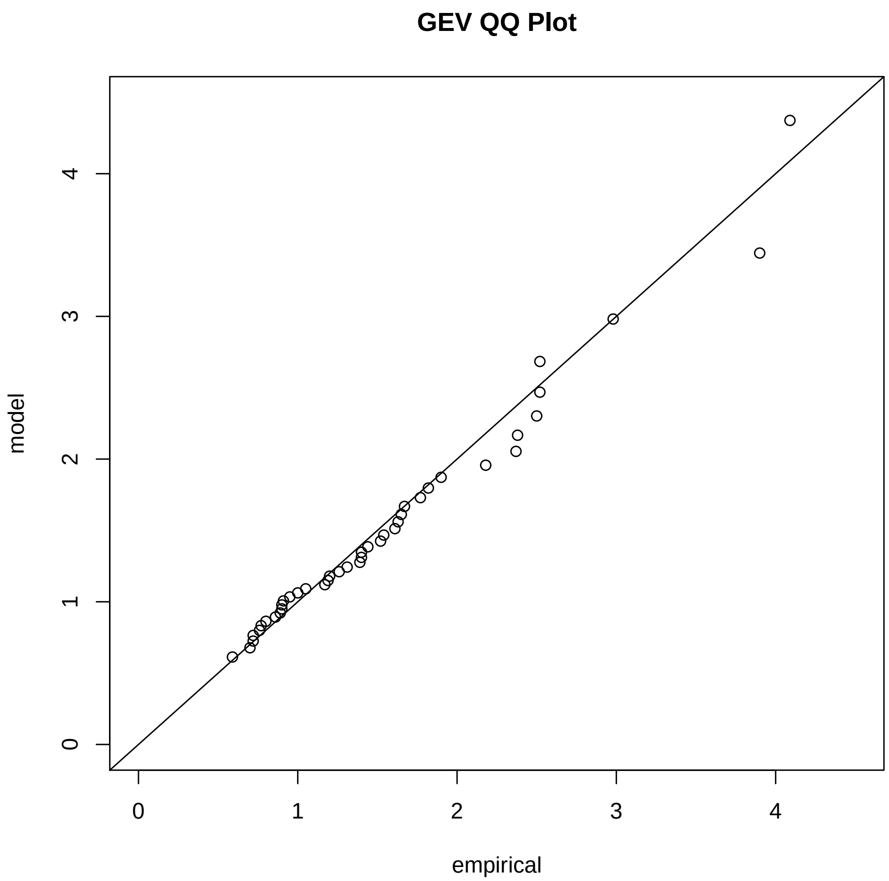
Let $M_n = \max_{t=1,\ldots,n}(X_t)$. Assume $M_n \sim \text{GEV}(\mu, \sigma, \xi)$.
(We will discuss why the GEV is the right distribution later.)

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

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

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

Modeling annual maxima



Note: Plot shows only annual maxima.

Why use only 'extreme' observations?

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

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

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

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

"Let the tails speak for themselves."

Why is the GEV the right distribution?

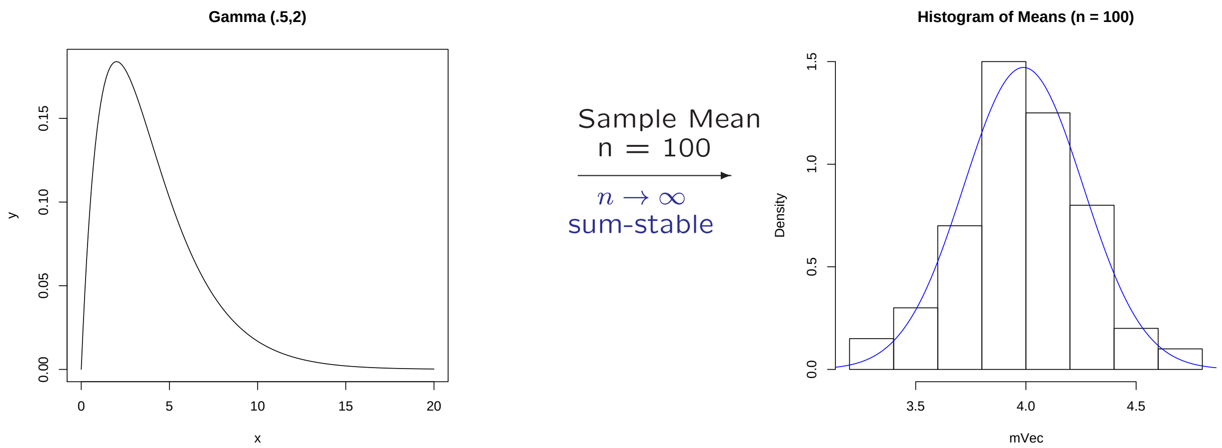
Answer: In a minute.

Why is the GEV the right distribution?

Answer: In a minute.

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

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



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

Why is GEV the right distribution?

A: Three-types Theorem. (Limit theorem for maxima.)

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

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

(Form of the max-stable distributions.)

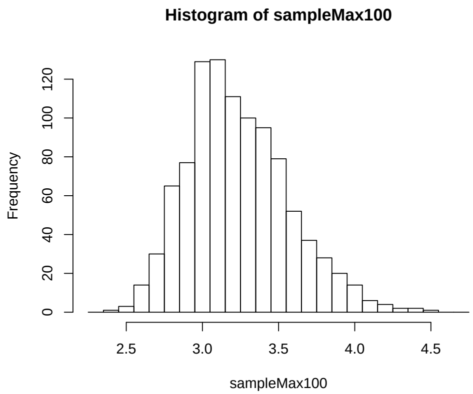
ξ determines the tail behavior.

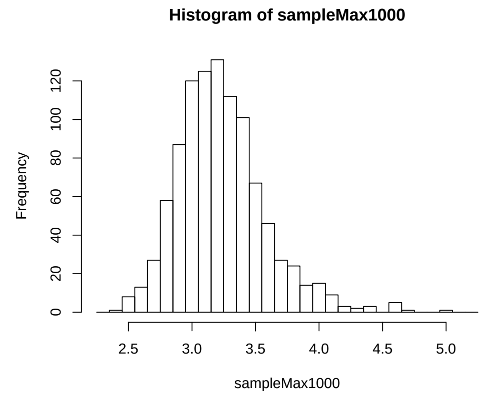
- $\xi \leq 0$: Weibull (or reverse Fréchet) case (bounded tail)
- $\xi = 0$: Gumbel case (light tail), interpreted as limit
- $\xi > 0$: Fréchet case (heavy tail, power function decay)

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

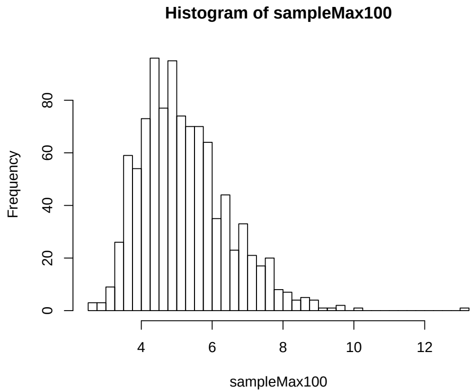
Distributions of sample maxima

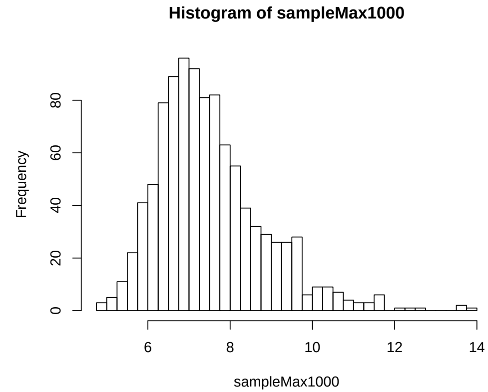
Normal





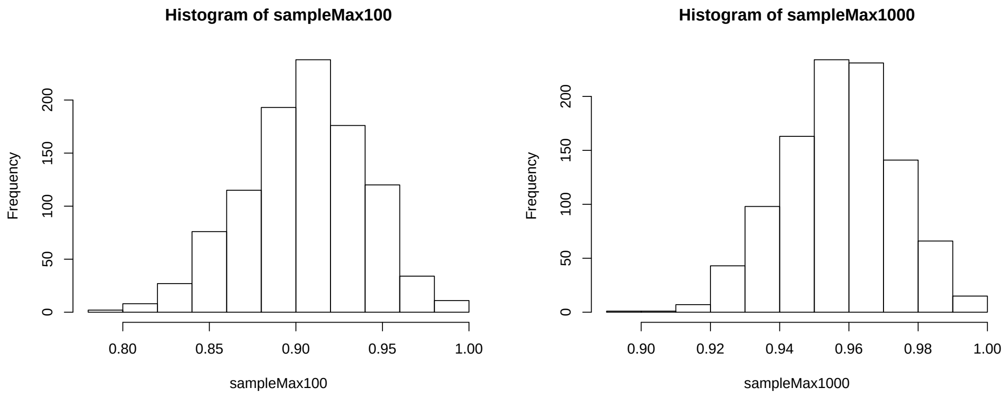
Exponential



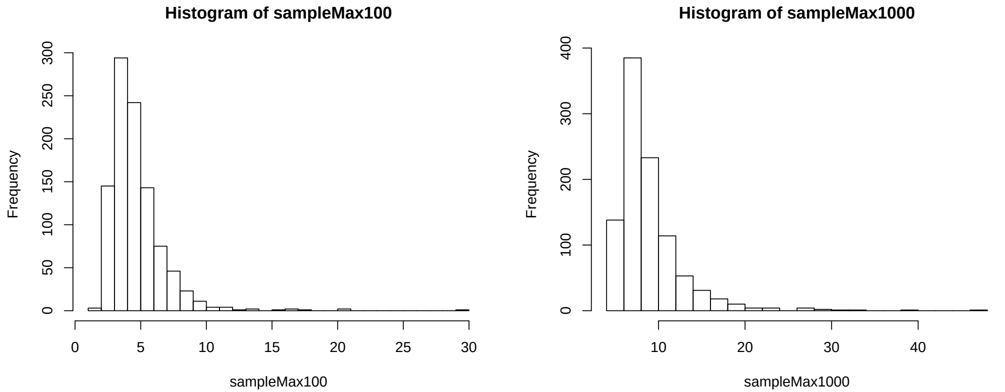


Distributions of sample maxima

Beta



Student's t (4 df)



Statistical Practice

Assume n is fixed and large enough so that:

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

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

Fort Collins Data:

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

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

95% confidence interval for 100-year return level: (3.5, 18.8)

Outline

1. Univariate
   - Analyzing a subset of extreme data.
   - Distributions for extremes.
2. Multivariate
   - What do we mean by tail dependence?
   - How multivariate extremes models capture tail dependence.
3. Spatial Processes
   - What is a maximal process?
   - Max-stable process models.

What is a Multivariate Extreme?

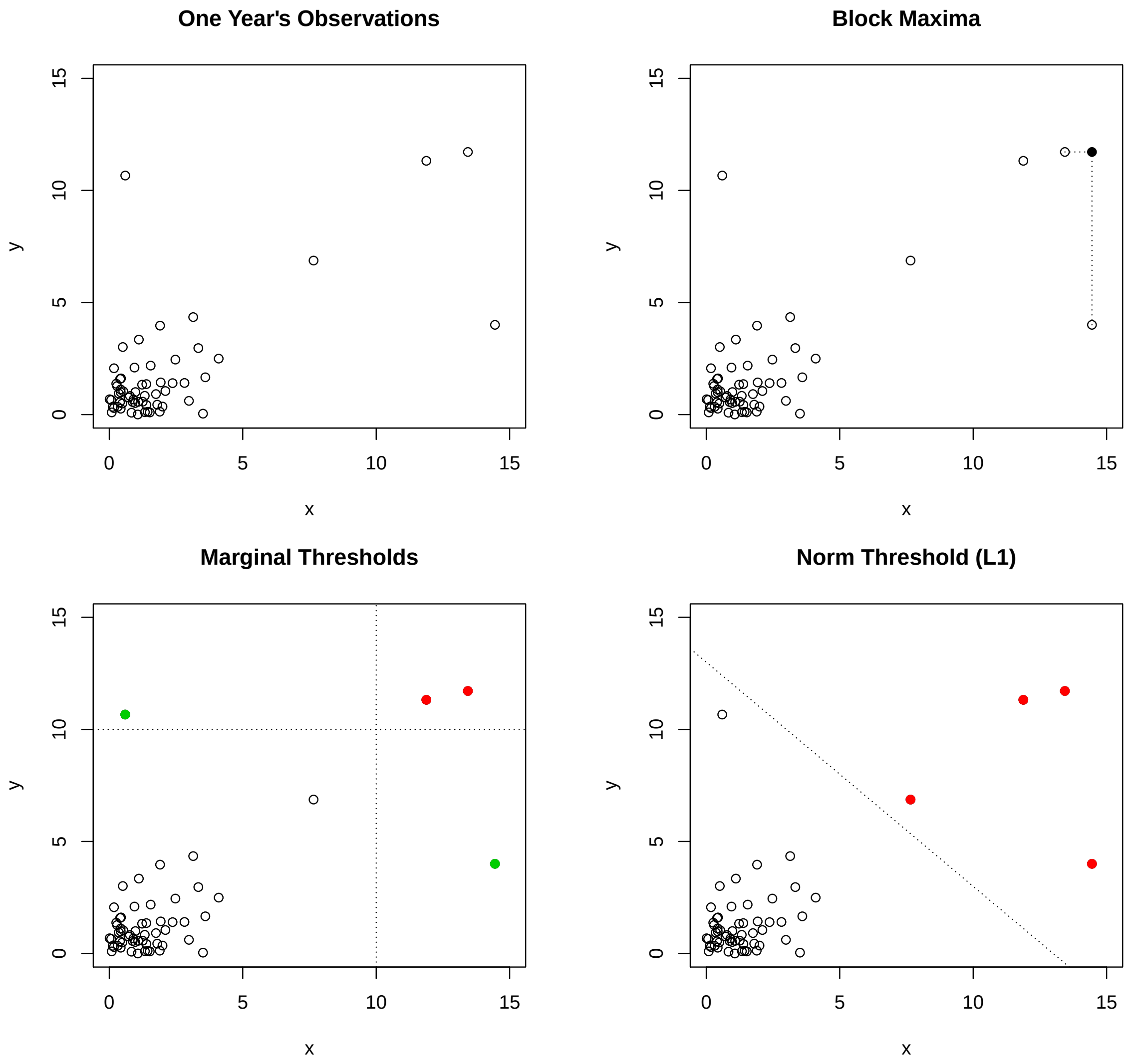
Let $\boldsymbol{Z}_m = (Z_{m,1}, \dots, Z_{m,d})^T$, $m = 1, 2, \dots$ be an iid sequence of random vectors. In an EV analysis, we extract a subset of data considered 'extreme'. How?

Block-maximum definition: Construct MV block maxima: $\boldsymbol{M}_n = (\bigvee_{m=1}^n Z_{m,1}, \dots, \bigvee_{m=1}^n Z_{m,d})$. Leads to modeling with multivariate max-stable distributions.

Marginal-exceedance definition: For each marginal $i = 1, \dots, d$, find an appropriate threshold u_i and retain data where $Z_{m,i} > u_i$. Leads to MV generalized Pareto distribution.

Norm-exceedance definition: For a given norm retain data where $\|\boldsymbol{Z}_m\| > u$. Leads to description by MV regular variation.

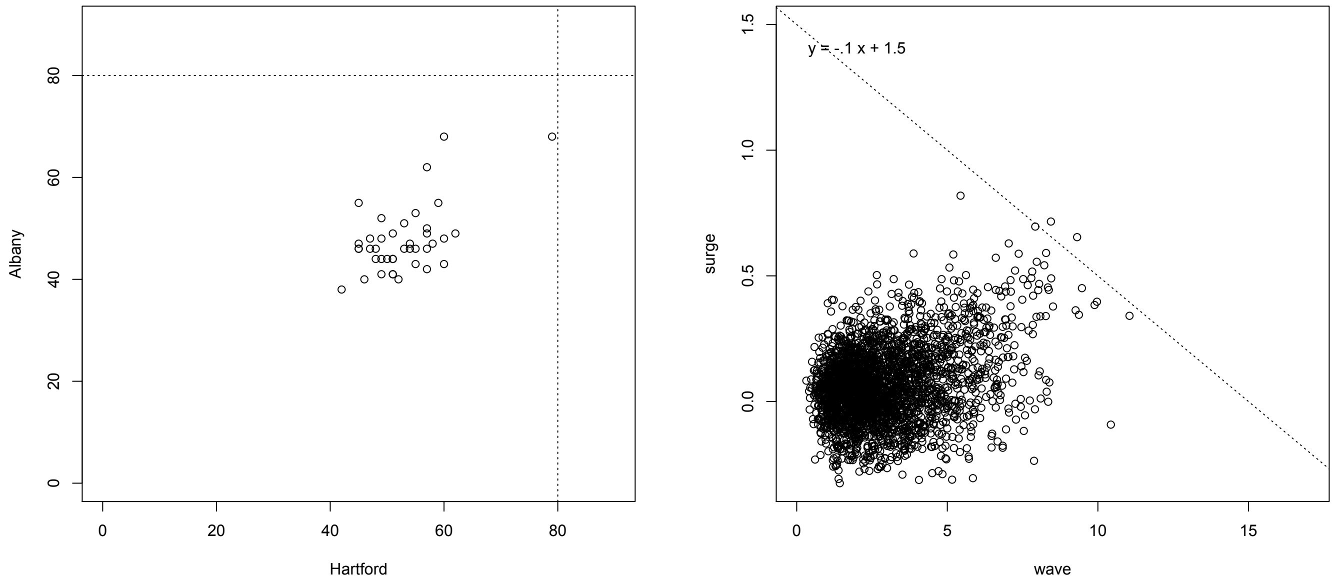
What is a Multivariate Extreme?



Goal of a MV Extreme Analysis

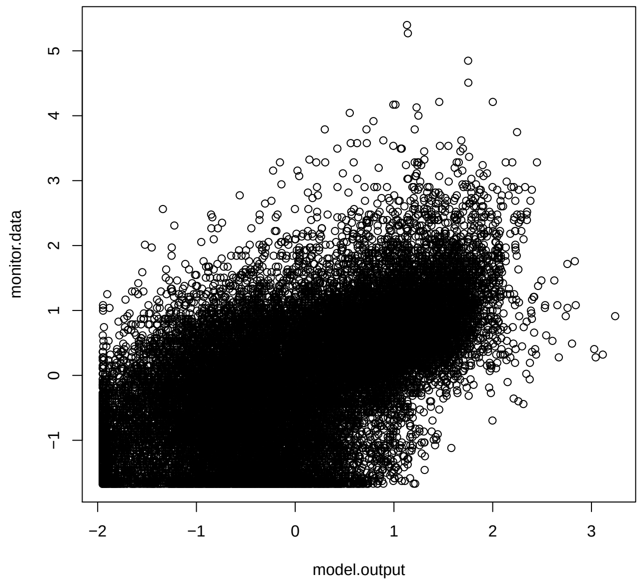
Goal: often to assess probability of falling in a risk region. Sometimes requires extrapolation.

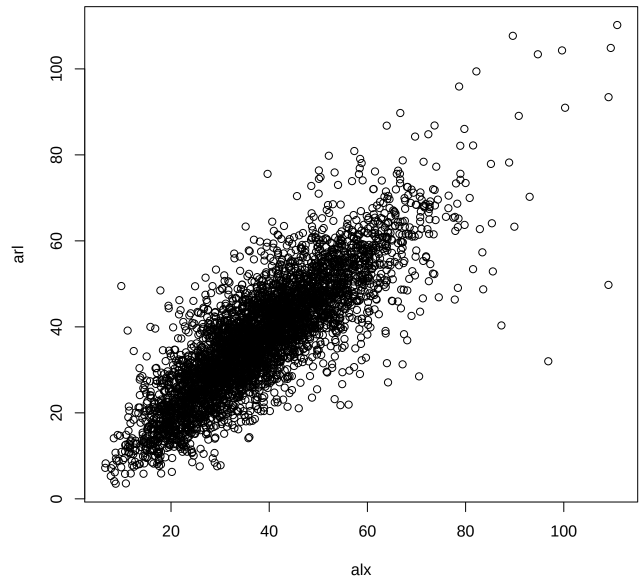
Keep in mind: A basic tenet of an extreme value analysis is to only use data considered to be extreme.



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

Tail Dependence



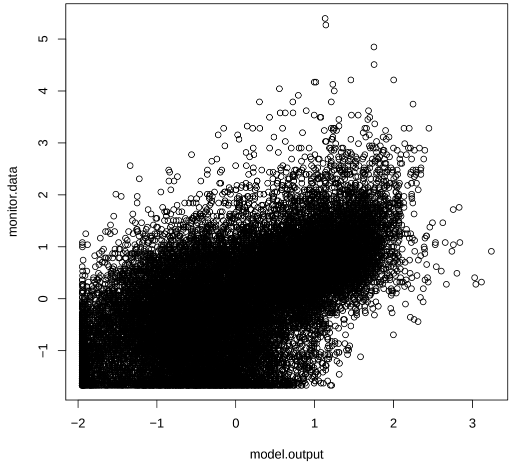


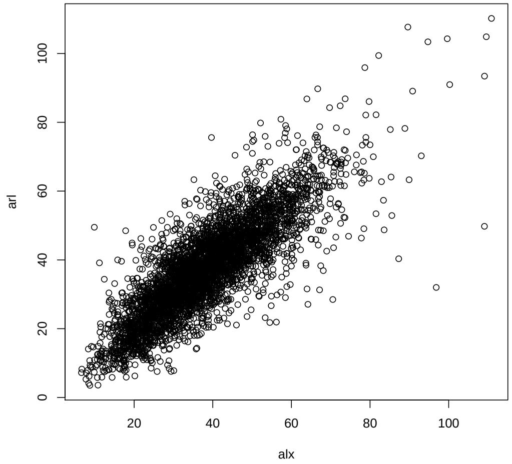
A central aim of multivariate extremes is trying to find an appropriate structure to describe *tail dependence*.

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

NOT Tail Dependence: Correlation

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





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

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

A Start: Asymptotic Dependence/Independence

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

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

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

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

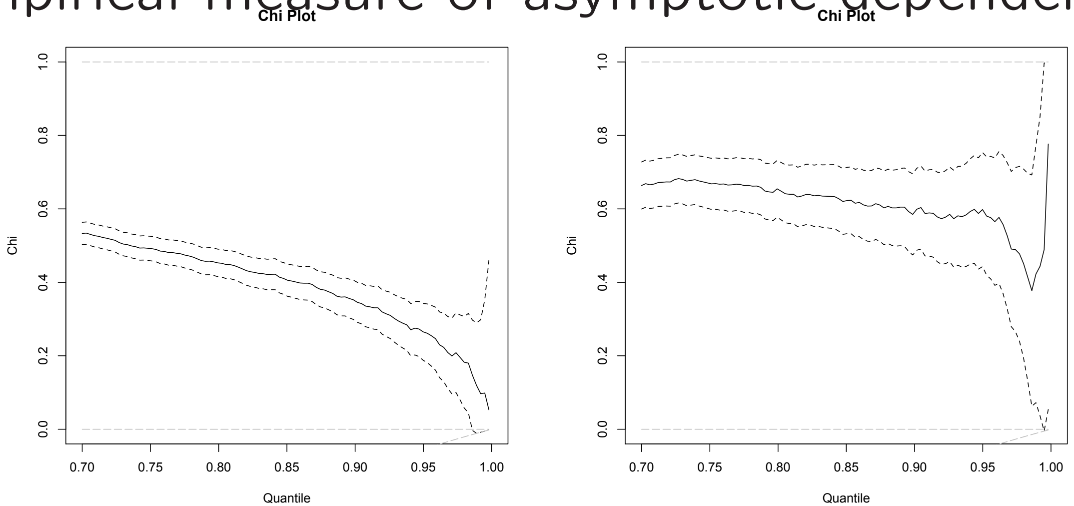
To talk about tail dependence, we need to know something about what it means to be in the tail of each component:

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

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

Tail Dependence of Examples

$\hat{\chi}$ is an empirical measure of asymptotic dependence.



Notes:

- asymptotic dependence implies a special (and strong) type of dependence.
- need dependence structures which can exhibit asymptotic dependence (few do).
- Gaussian dependence with $\rho < 1$ is asymptotically independent.

Multivariate Extremes and Marginal Distributions

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

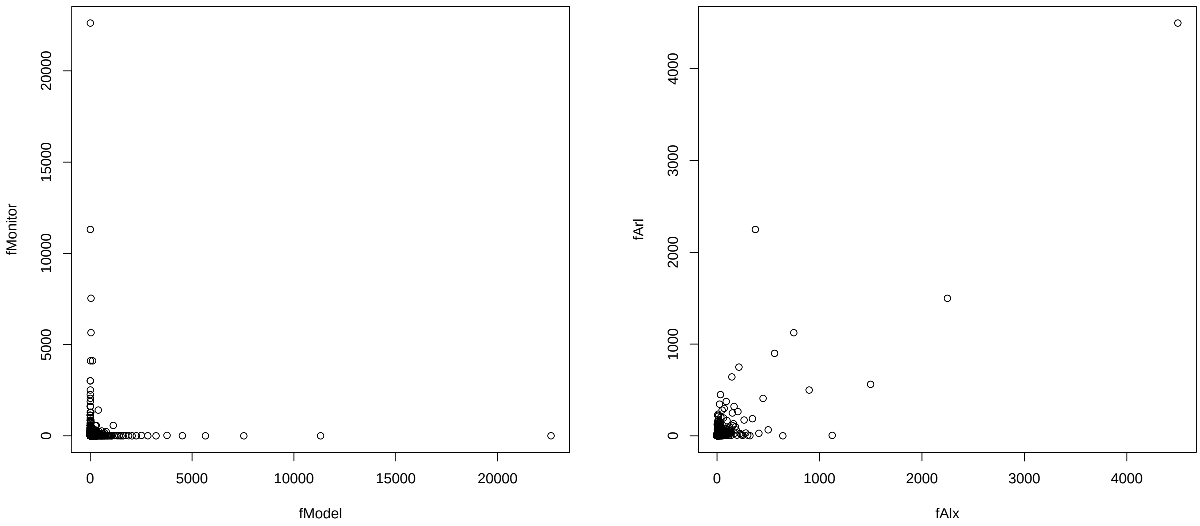
Theory: MV distributions used in extremes are described by first assuming a common marginal distribution, often unit Fréchet ($P(Z \leq z) = \exp(-z^{-1})$). Also (in theory) the marginal distribution *doesn't really matter* when describing dependence because of "domain of attraction" results.

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

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

Sounds copula-like, but with different marginals and models. Models need to accommodate tail dependence.

Marginal-transformed Example Data



- Transforming to heavy tailed marginals focuses attention on large observations.
- Asymptotic independence $\rightarrow$ large observations on axes, asymptotic dependence $\rightarrow$ large observations in interior.
- In the heavy-tailed case, there is a probabilistic framework which allows one to model tail dependence.

Framework for Tail Dependence: Regular Variation and Polar Representation

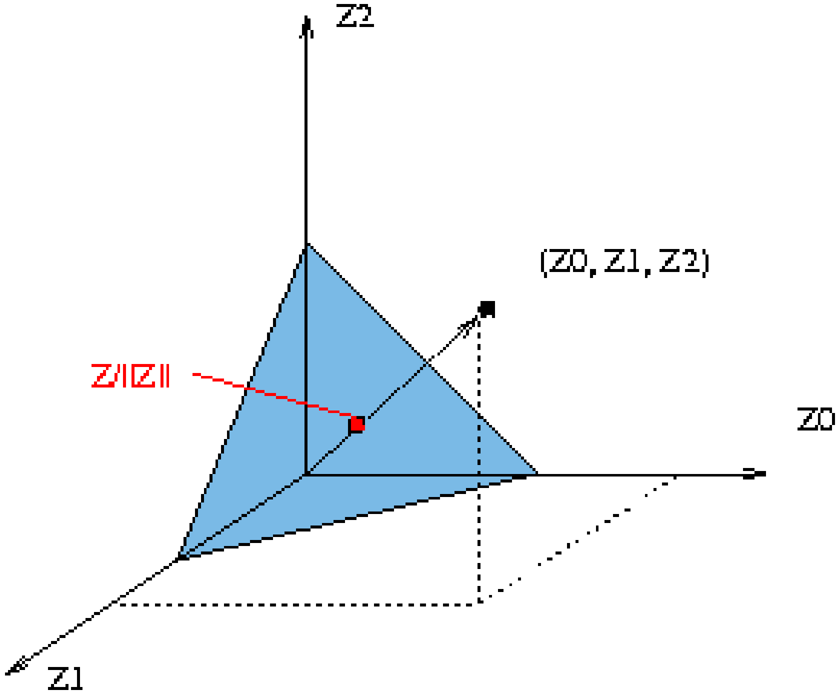
Let $R = \|\boldsymbol{Z}\|$ and $\boldsymbol{W} = \|\boldsymbol{Z}\|^{-1}\boldsymbol{Z}$. $\boldsymbol{Z}$ is regular varying if there exists a normalizing sequence $\{b_n\}$ where $P(b_n^{-1}\|\boldsymbol{Z}\| > r) \sim 1/n$, such that

$$nP\left(b_n^{-1}R > r, \boldsymbol{W} \in A\right) \xrightarrow{v} r^{-1/\xi}H(A)$$

where d is the dimension of $\boldsymbol{Z}$, and where H is some probability measure on the unit 'ball' $S_d = \{\boldsymbol{z} \in \mathbb{R}^d \mid \|\boldsymbol{z}\| = 1\}$.

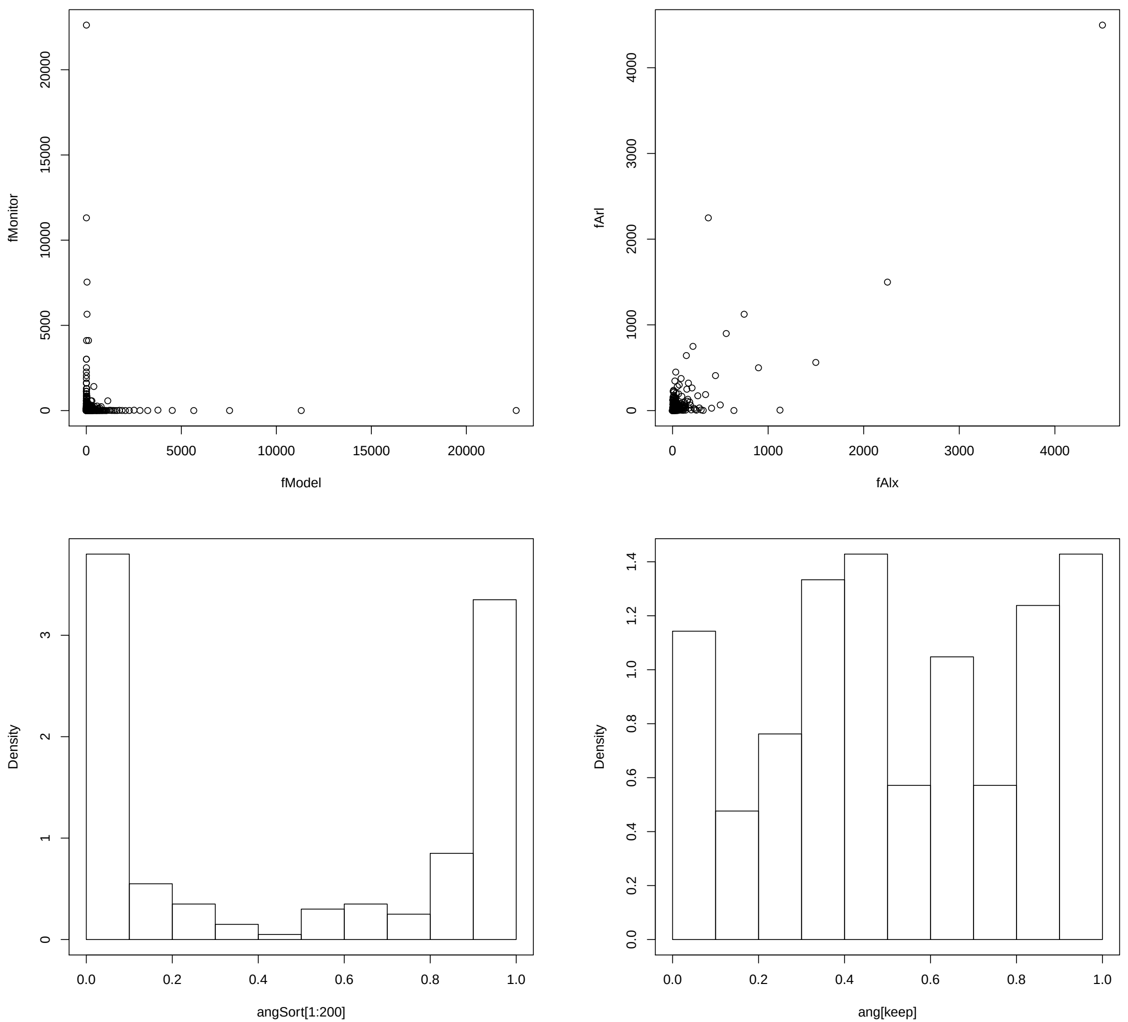
- 'regular variation' = heavy tail (described by ξ)
- LHS: "as points get big (radial component)"
- RHS: "radial and angular comps. become independent"
- angular measure H describes distribution of directions – completely describes dependence.

MV Regular Variation in a Picture



Idea: Distribution of large points described by radial component and angular component (which has a probability distribution on the unit simplex.

Transformed Data: Air Pollution Datasets



MV EVD's: Limit Dists for Component Maxima

Let $\boldsymbol{Z}_m = (Z_{m,1}, \dots, Z_{m,d})^T$, $m = 1, 2, \dots$ be an iid sequence of random vectors, with $Z_{m,i}$ in the domain of attraction of a unit Fréchet distribution.

Let $\boldsymbol{M}_n = (\bigvee_{m=1}^n Z_{m,1}, \dots, \bigvee_{m=1}^n Z_{m,d})$. If there exist normalizing sequences $\boldsymbol{a}_n$ and $\boldsymbol{b}_n$ such that $P\left(\frac{\boldsymbol{M}_n - \boldsymbol{b}_n}{\boldsymbol{a}_n} \leq \boldsymbol{z}\right) \to G(\boldsymbol{z})$ (nondegenerate) as $n \to \infty$, then

$$G(\boldsymbol{z}) = \exp(-V(\boldsymbol{z})),$$

where

$$V(\boldsymbol{z}) = d \int_{S_{d-1}} \max_{i=1,\dots,d} \left(\frac{w_i}{z_i}\right) dH(\boldsymbol{w}).$$

One can think of $V(\boldsymbol{z})$ as linking the angular measure $H(\boldsymbol{w})$ to Cartesian coordinates required by a cdf.

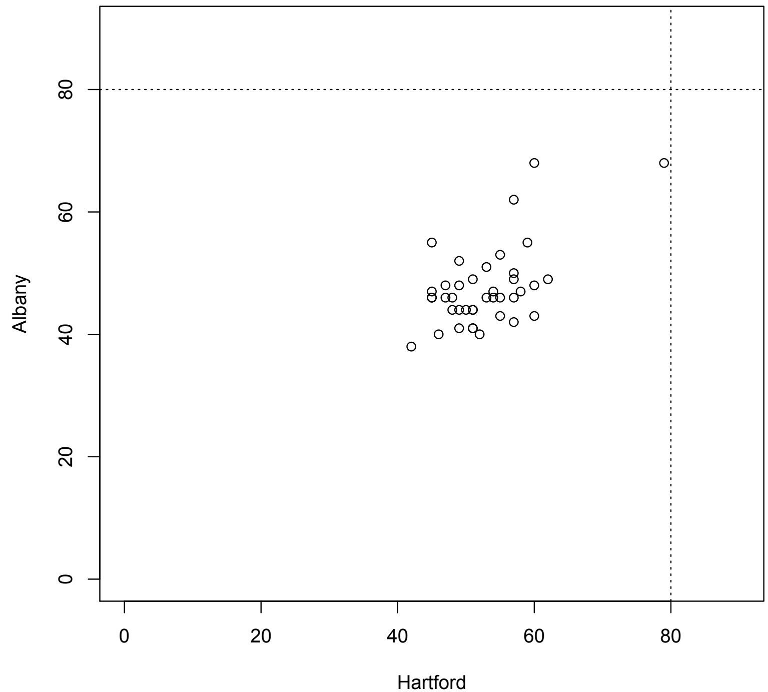
There are a number of parametric subfamilies of MVEVDs (e.g., Logistic).

Statistics: Fitting a MV EVD

Logistic Model

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

Annual max wind speed at Hartford and Albany.



Fitted Logistic Model to Wind Data

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

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

Estimation of Risk

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

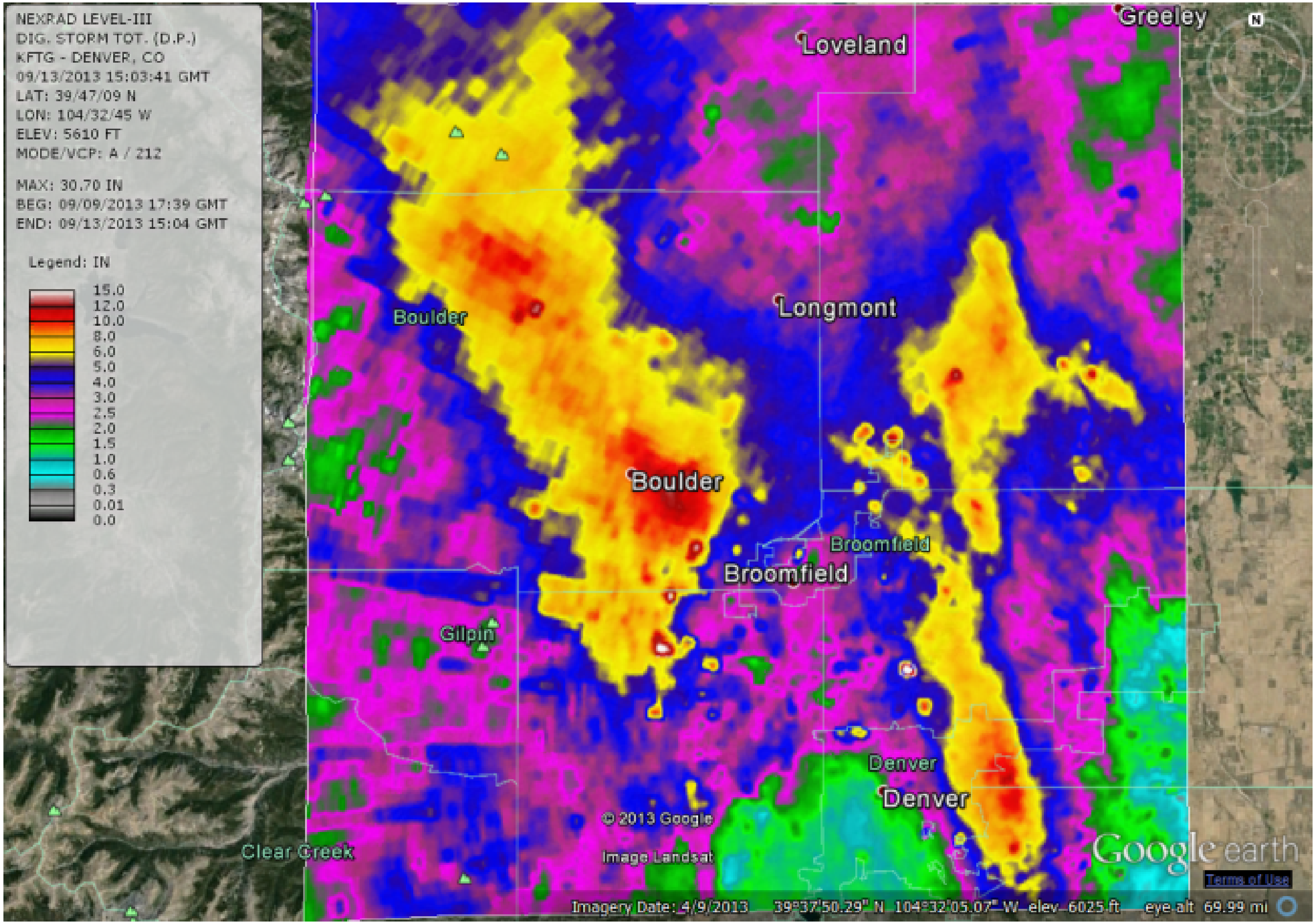
There is dependence in this data. Note the difference between the "joint" and "independent" estimates.

Outline

1. Univariate
   - Analyzing a subset of extreme data.
   - Distributions for extremes.
2. Multivariate
   - What do we mean by tail dependence?
   - How multivariate extremes models capture tail dependence.
3. Spatial Processes
   - What is a maximal process?
   - Max-stable process models.

The Need for Spatial Extremes

Colorado Precipitation September 9-13, 2013



Limit Distributions for Maxima of *Processes*

We know:

- GEV models limiting dists of univariate block maxima.
- MV EVD's are limiting dists of componentwise maxima.
- form of MV EVDs if marginal is assumed.

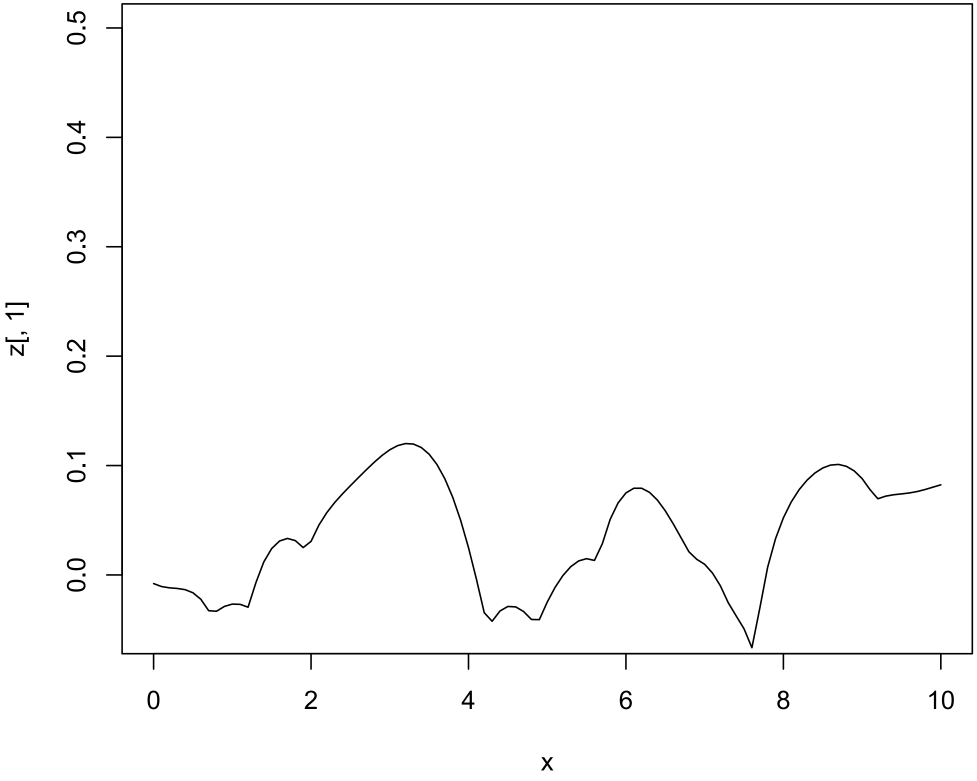
Q: What is the limiting distribution of locationwise maxima of a process?
A: A max-stable process.

Let $Z_m(\boldsymbol{s}), \boldsymbol{s} \in \mathcal{D}, m = 1, \ldots, n$ be independent copies of $Z(\boldsymbol{s})$, and let $M_n(\boldsymbol{s}) = \max_m Z_m(\boldsymbol{s})$. If there exist $a_n(\boldsymbol{s})$ and $b_n(\boldsymbol{s})$ such that

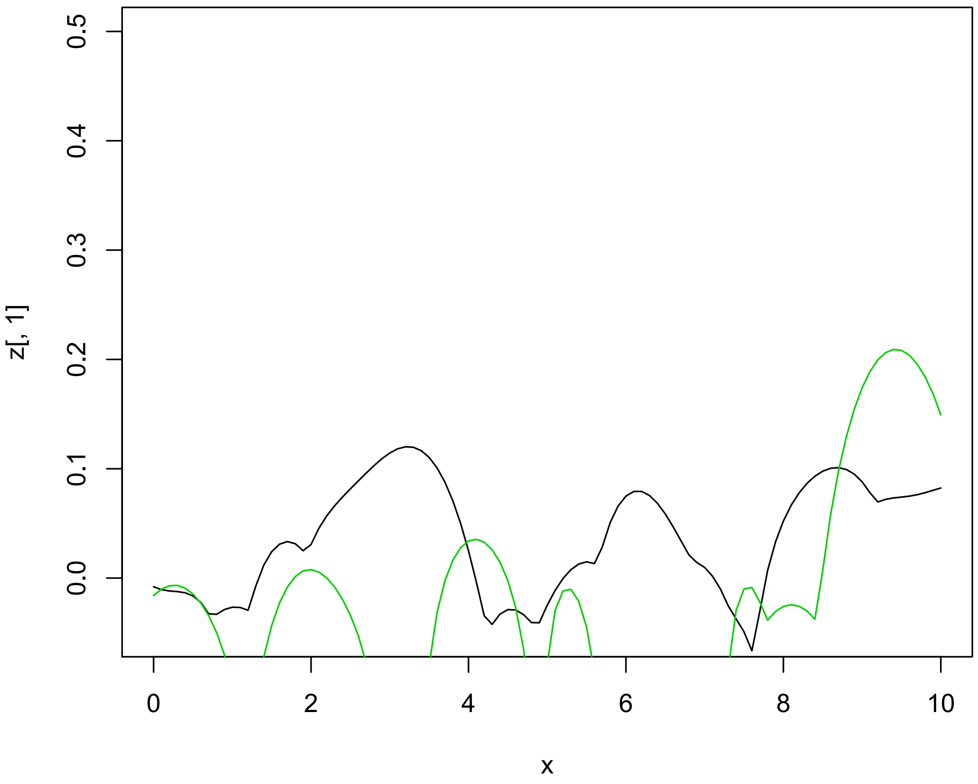
$$\frac{M_n(\boldsymbol{s}) - b_n(\boldsymbol{s})}{a_n(\boldsymbol{s})} \rightarrow Y(\boldsymbol{s}),$$

then $Y(\boldsymbol{s})$ is a max-stable process.

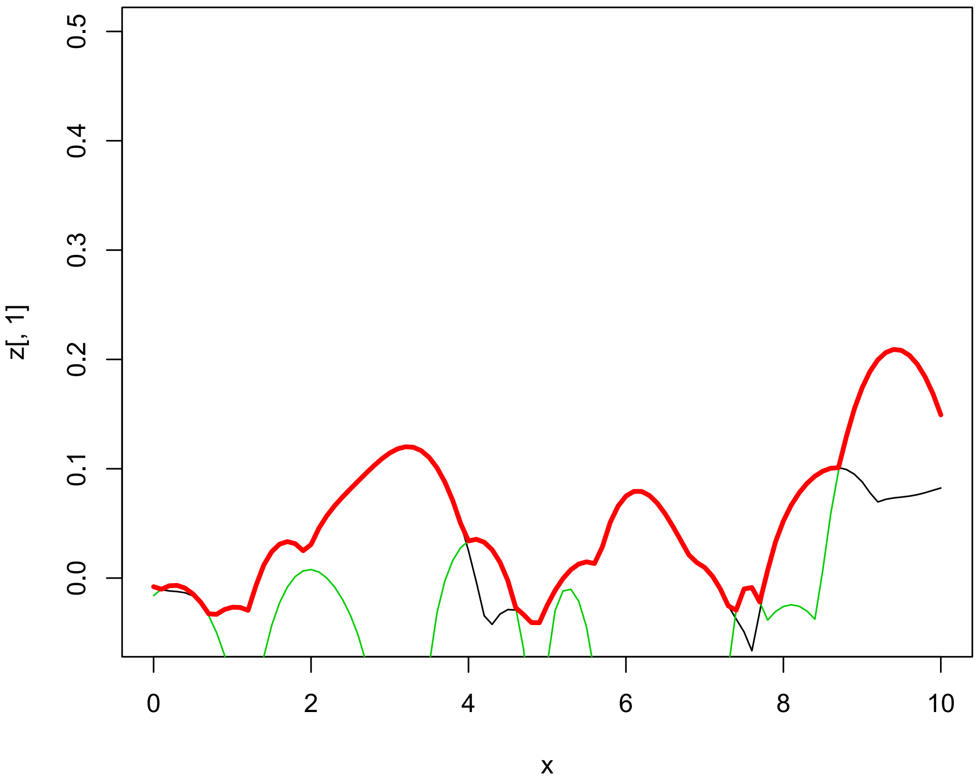
Maxima of Processes



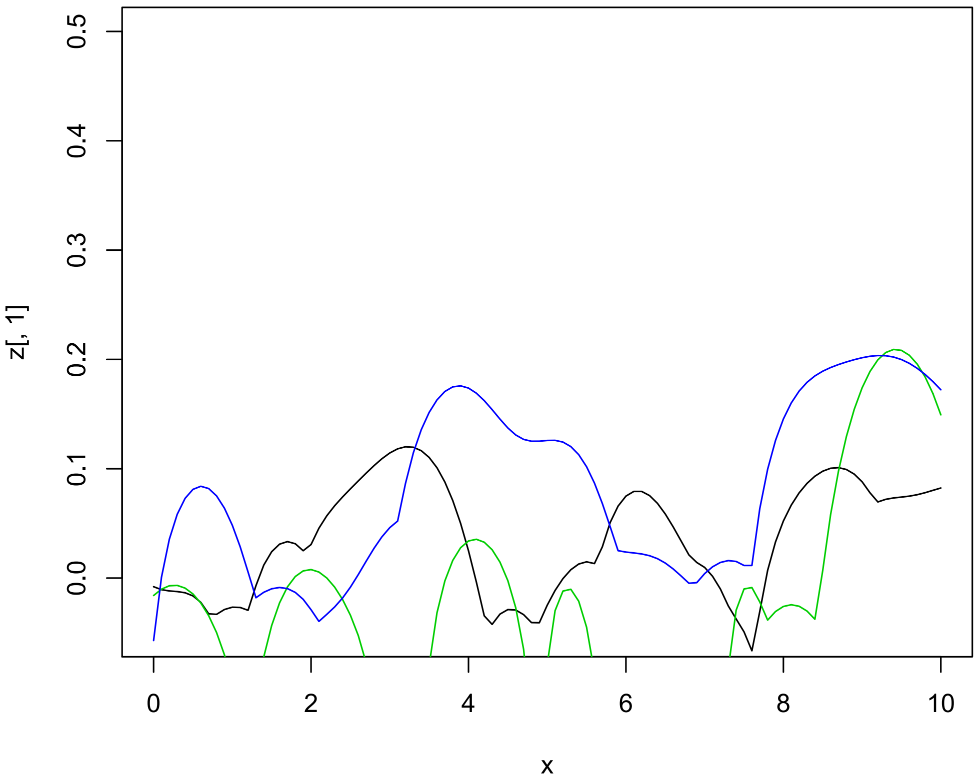
Maxima of Processes



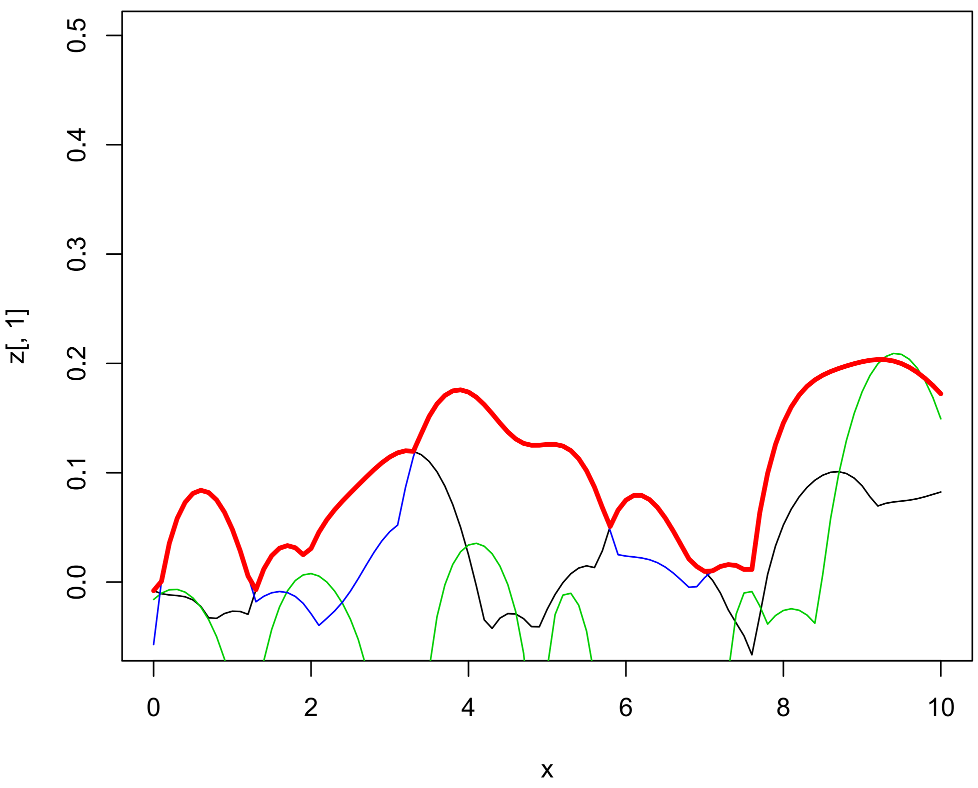
Maxima of Processes



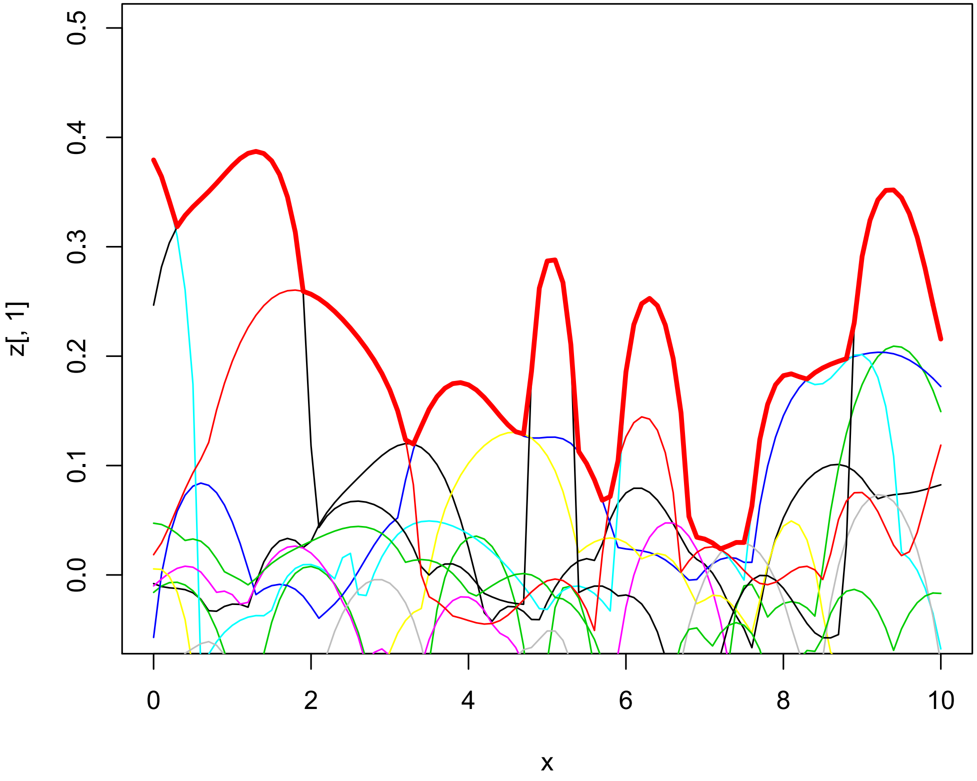
Maxima of Processes



Maxima of Processes



Maxima of Processes



A Max-stable Process Model

Brown Resnick Process:

$$\Pr[Z(x_i) \le z_1, Z(x_j) \le z_2] =$$
$$\exp\left\{ - \left[e^{-z_1} \Phi \left(\frac{\sqrt{\rho(x_i - x_j)}}{2} + \frac{z_2 - z_1}{\sqrt{\rho(x_i - x_j)}} \right) \right. \right.$$
$$\left. \left. + \, e^{-z_2} \Phi \left(\frac{\sqrt{\rho(x_i - x_j)}}{2} + \frac{z_1 - z_2}{\sqrt{\rho(x_i - x_j)}} \right) \right] \right\}$$

- Best model available for both theoretical and practical reasons.
- Above representation assumes marginals are unit Frechet.
- Model known in closed form only for two dimensions.
- One method for inference: composite likelihoods.
- Inference only uses extreme observations (i.e., annual maxima); model can capture tail dependence.

Take-Away Messages

1. Extremes goal is often to extrapolate.
2. Extremes methods 'let tail speak for itself', use a subset of extreme data.
3. Two general approaches:
   - Block maxima (GEV/MVEVD's/Max-Stable Processes)
   - Threshold exceedances (GPD/Various Methods)
4. Do not need to know underlying dist'n to model extremes.
5. Dependence for extremes
   - is not described with correlations.
   - requires structures which can allow for tail dependence.
6. Max-stable processes are
   - theoretically justified models for spatial extremes.
   - (practical standpoint) process models which can exhibit tail dependence.