Mixture of Heavy-Tailed distributions for Bivariate Precipitation Data

Julie Carreau⋆ & **Philippe Naveau**∗ & **Malaak Kallache**

julie.carreau@univ-montp2.fr philippe.naveau@lsce.ipsl.fr mk@climpact.com

⋆ HydroSciences Montpellier, FRANCE
∗ Laboratoire des Sciences du Climat et de l'Environnement (LSCE-IPSL), FRANCE
† Climpact, FRANCE

Var Flood June 15th 2010

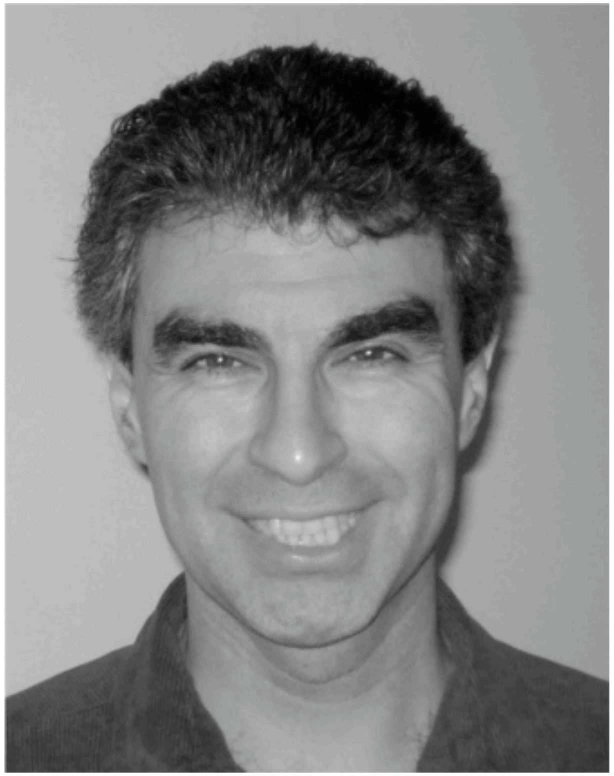
- warm sea
- mountainous landscape
- warm air from Africa

400 mm of rain in 24h $\approx$ 5 months of precipitation

no similar event since 1827

Background

Machine learning



Yoshua Bengio

Extreme-Value Theory

Vilfredo Pareto 1848-1923

Bridge the gap between non-parametric and extreme-value models

Outline

1. Precipitation data

2. Outline of the bivariate density model:

mixture of bivariate distributions with a heavy tail along 1D projections

3. Hybrid Pareto Distribution

4. Bivariate Hybrid Pareto Distribution

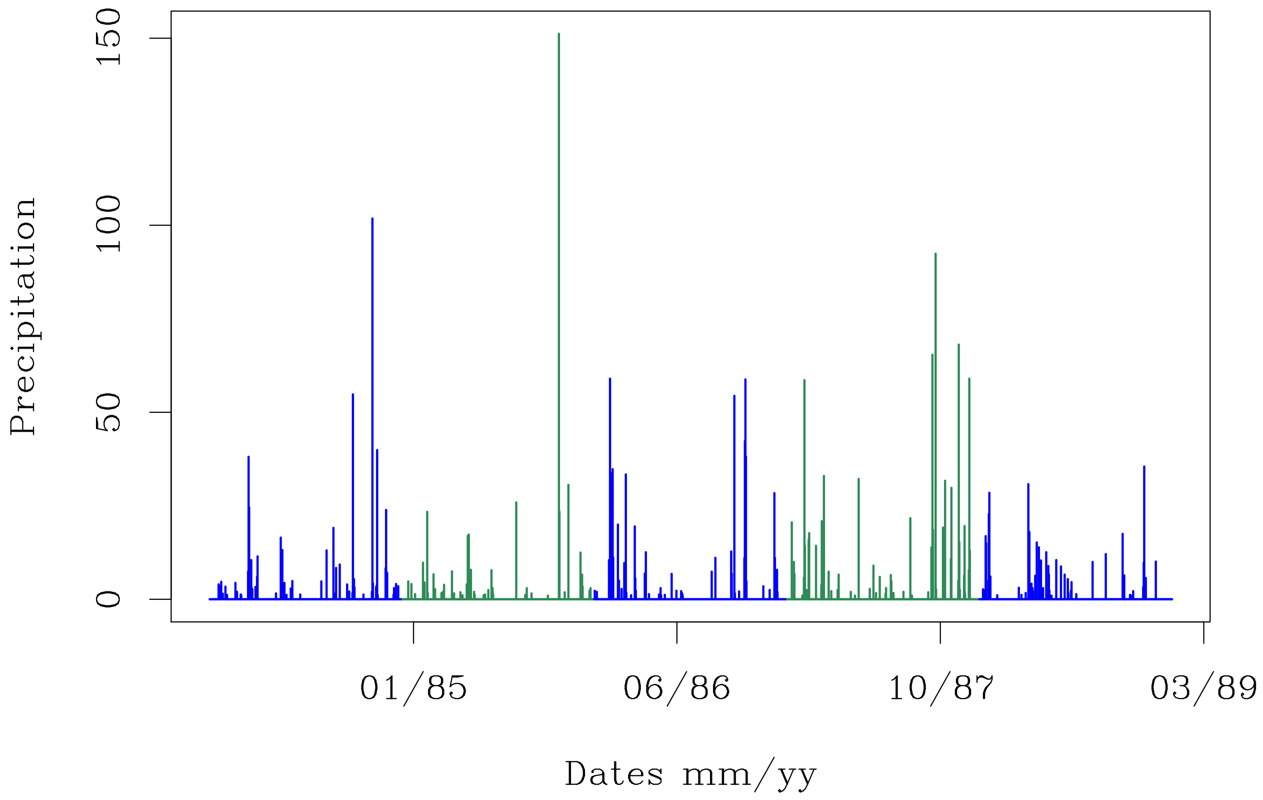
5. Mixture learning and initialization

6. Synthetic examples and precipitation data

7. Future work

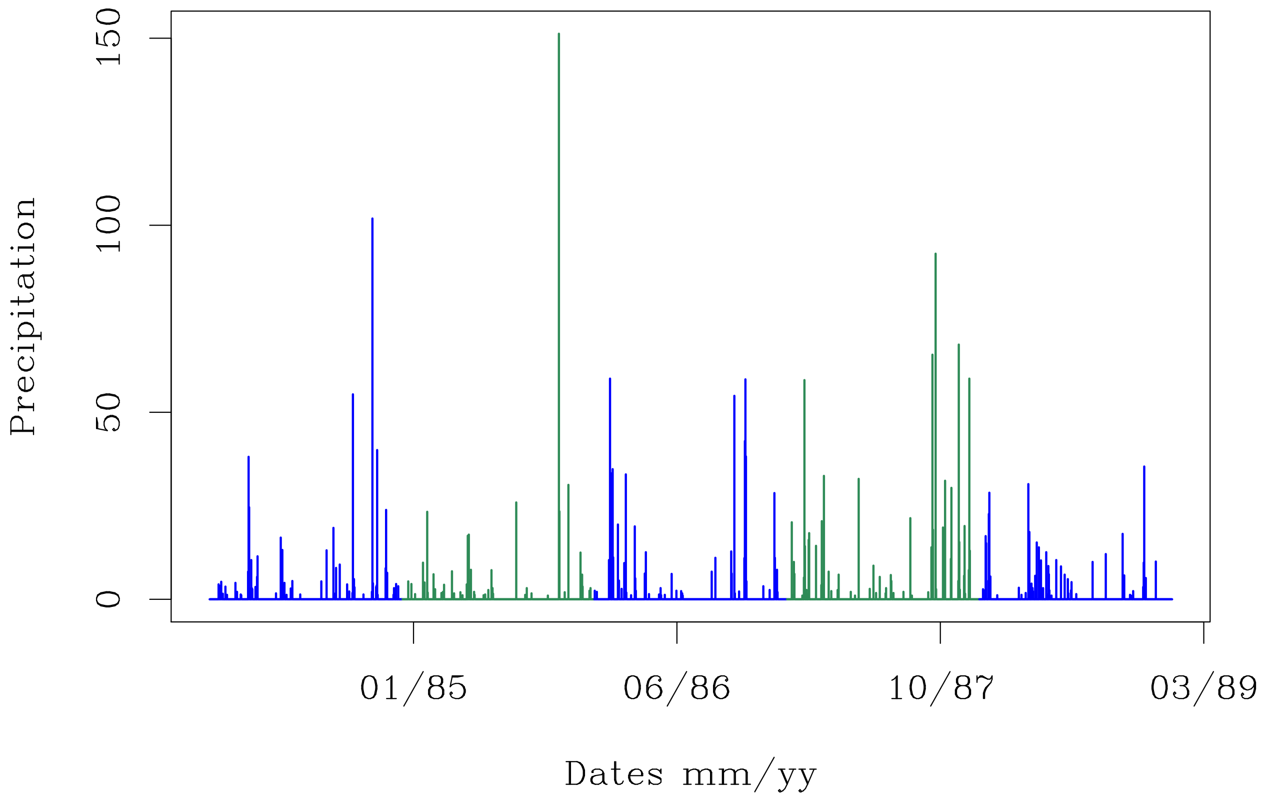
Precipitation Data

- Intermittency
- Temporal and spatial dependence
- Inter-annual / intra-annual variability
- extreme values



Precipitation Data

- Intermittency
- Temporal and **spatial dependence**
- Inter-annual / intra-annual variability
- **extreme values**



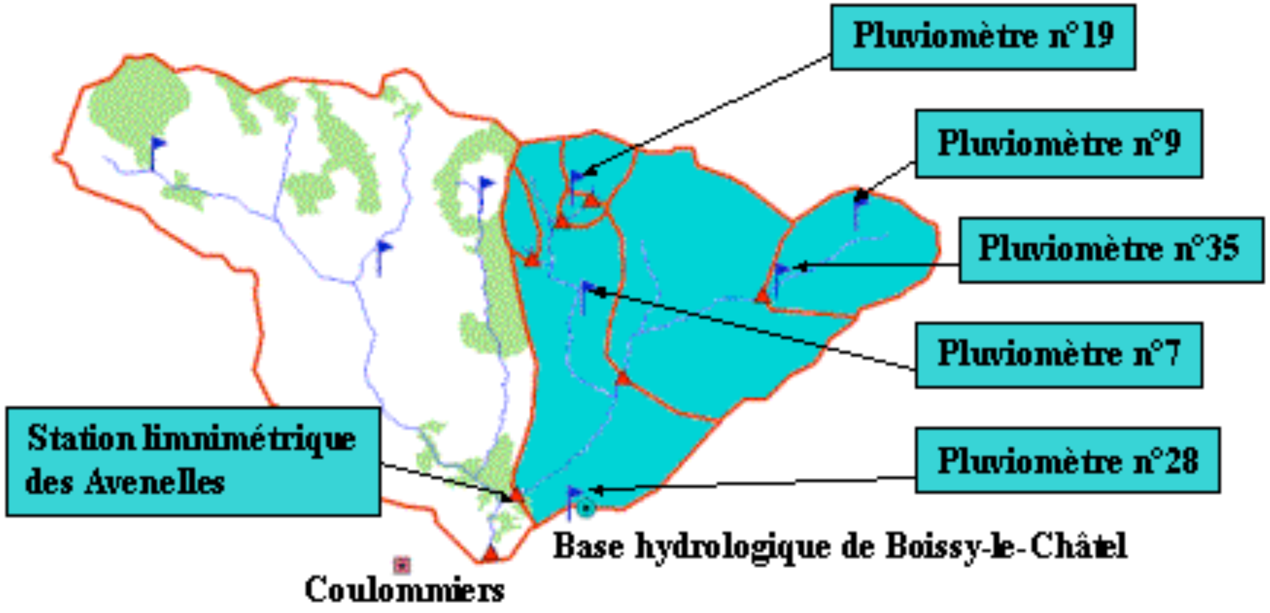
Motivation

- Simulate runoff from hydrological models:
  spatial rain field over the basin

- Evaluate the impact of climate change :
  model precipitation given large scale atmospheric variables given by GCMs

- Dimension of dams and agricultural practice

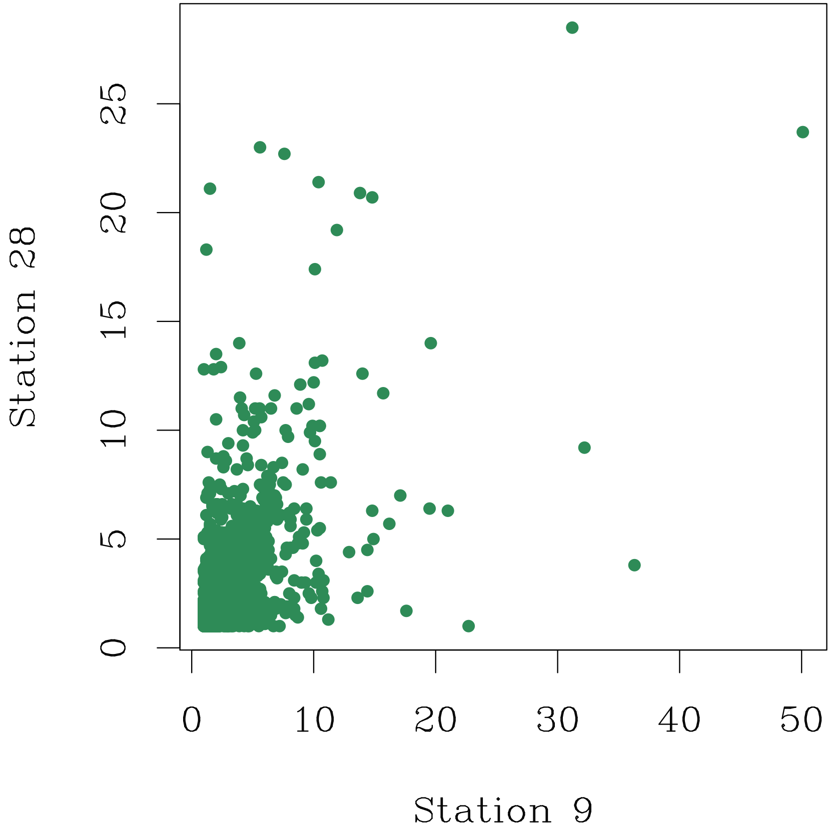
Avenelles Basin

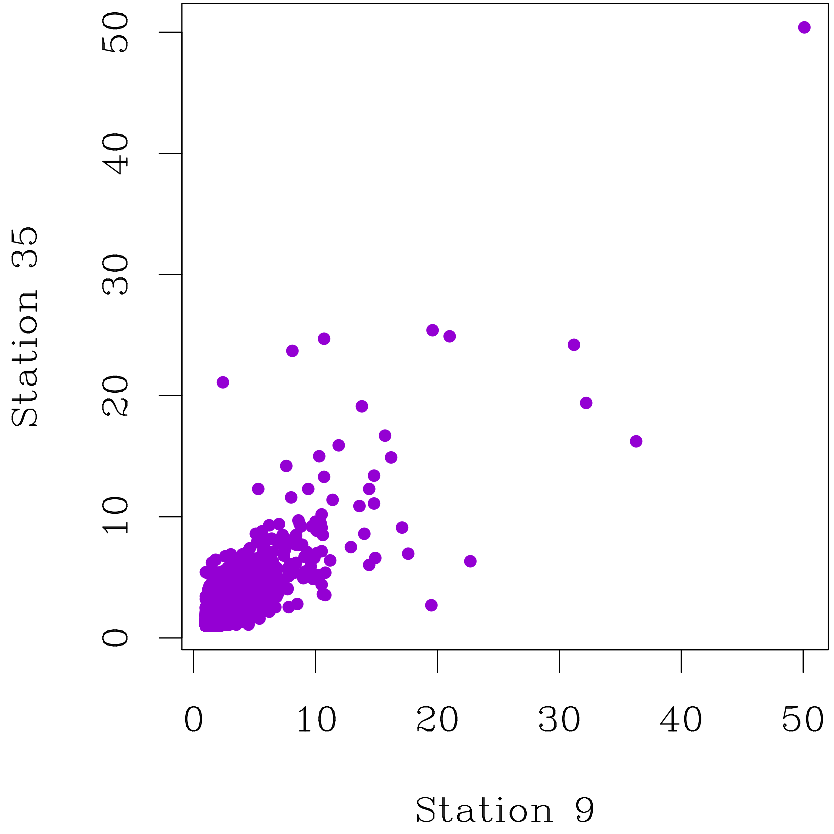


✷ **Positive hourly precipitation** (> 1 mm) for three stations of the Orgeval Basin, near Paris

✷ **Data span 1972 to 2002** for about **3000** positive observations

Precipitation





✷ **Spatially apart/close stations** show a more wide spread/narrow pattern

✷ Dependence in the **central part** might differ from dependence in the **extremes**

Bivariate Mixture Model

Two key aspects of the precipitation data

✷ **Model the dependence structure of the extreme observations**

Usually described either by the **spectral measure** or the **copula function**

Bivariate Mixture Model

Two key aspects of the precipitation data

✷ **Model the dependence structure of the extreme observations**

Usually described either by the **spectral measure** or the **copula function**

✷ **Full density estimation: central and extremal areas**

Perform estimation of the **margins** and of the **dependence structure** at once

Extremal Dependence Structure

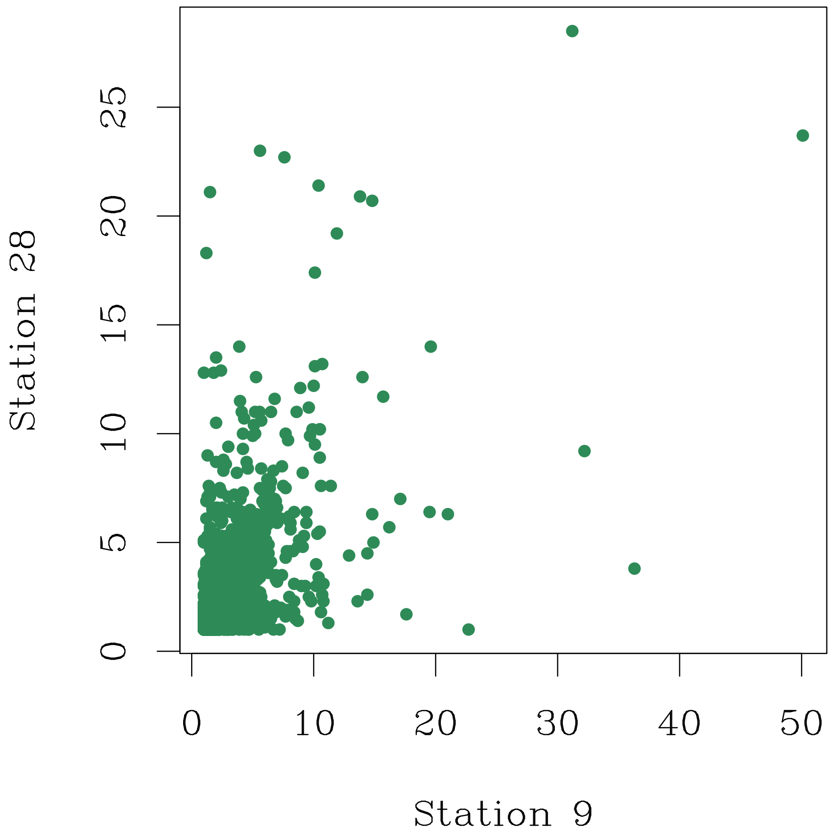
✷ **Pseudo-polar coordinates:** angle ω and radius r

✷ **Decomposition of the density for large values:**
product of the **angular spread** times the **radial distance**

✷ **Angular spread** can be described by the **spectral measure**

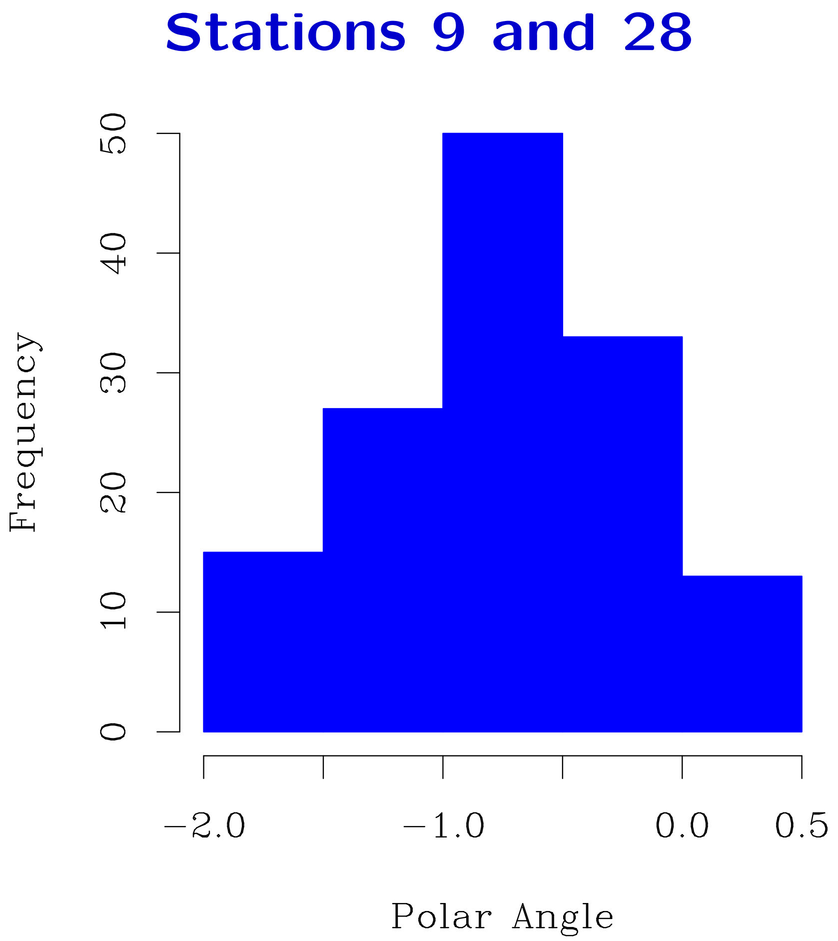
✷ **Radial distance** depends on the **margins**

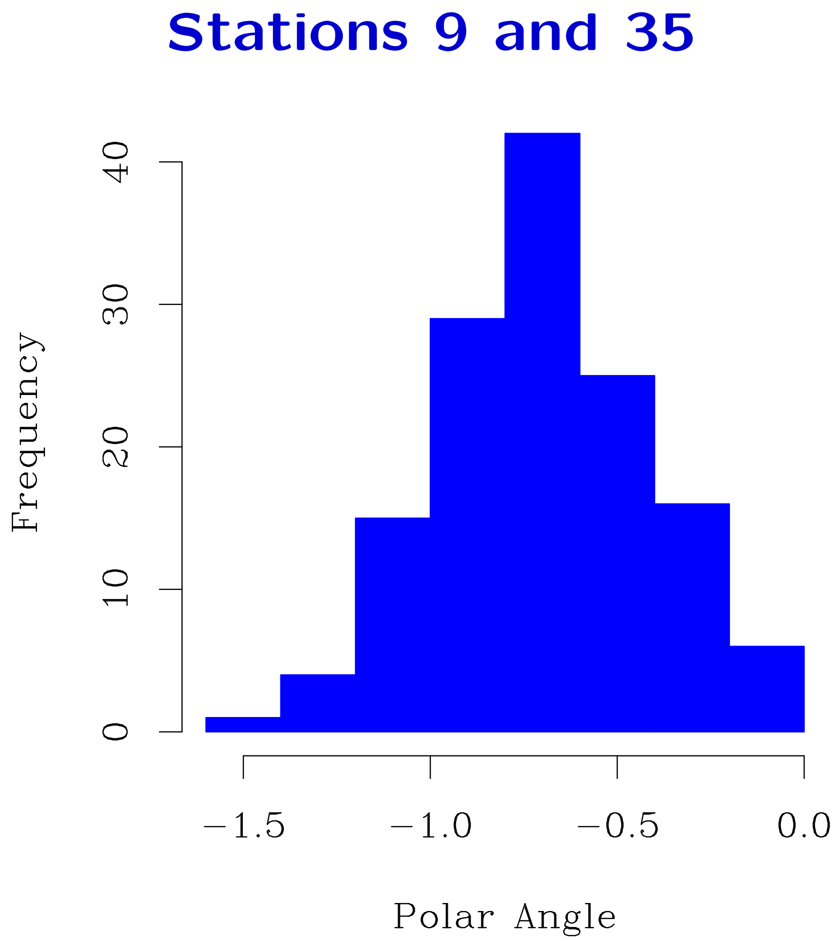
Estimate of Spectral Measure



1. Set R, a large radius
2. Extremes are outside circle of radius R
3. Map extremes on the unit circle
4. Distribution of the angles

Examples of Spectral Measure





Building Blocks

Block 1: bivariate Gaussian

convenient computationally but not adequate for heavy tails

Building Blocks

Block 1: bivariate Gaussian

convenient computationally but not adequate for heavy tails

Block 2: projection pursuit

introduce a heavy tail in a 1D projection defined by angle θ

Building Blocks

Block 1: bivariate Gaussian

convenient computationally but not adequate for heavy tails

Block 2: projection pursuit

introduce a heavy tail in a 1D projection defined by angle θ

Block 3: hybrid Pareto distribution

smooth extension of the **generalized Pareto distribution** on the whole real axis

Building Blocks

Block 1: bivariate Gaussian

convenient computationally but not adequate for heavy tails

Block 2: projection pursuit

introduce a heavy tail in a 1D projection defined by angle θ

Block 3: hybrid Pareto distribution

smooth extension of the **generalized Pareto distribution** on the whole real axis

Blocks 1 + 2 + 3: Bivariate hybrid Pareto distribution

a bivariate Gaussian with **heavy tail** in a direction determined by an angle θ

Building Blocks

Block 1: bivariate Gaussian

convenient computationally but not adequate for heavy tails

Block 2: projection pursuit

introduce a heavy tail in a 1D projection defined by angle θ

Block 3: hybrid Pareto distribution

smooth extension of the **generalized Pareto distribution** on the whole real axis

Blocks 1 + 2 + 3: Bivariate hybrid Pareto distribution

a bivariate Gaussian with **heavy tail** in a direction determined by an angle θ

Block 4: Mixture of Bivariate hybrid Pareto distribution

a discrete number of direction with heavy tails $\Rightarrow$ mixture size

Extreme Value Theory

Goal : methods to analyze and characterize extreme events

Challenges :

- few observations are extreme
- estimate a risk which was never observed

Extreme Value Theory

Goal : methods to analyze and characterize extreme events

Challenges :

- few observations are extreme
- estimate a risk which was never observed

Extremes are defined as :

- Maxima : $M_n = \max\{Z_1, \dots, Z_n\}$

Two types of approaches

- Exceedances : Z_i such that $Z_i > u$

Block Maxima Approach

In theory : $M_n = \max\{Z_1, \ldots, Z_n\}$, Z_i i.i.d. for large n

In practice : set a block size and take the maximum over each block

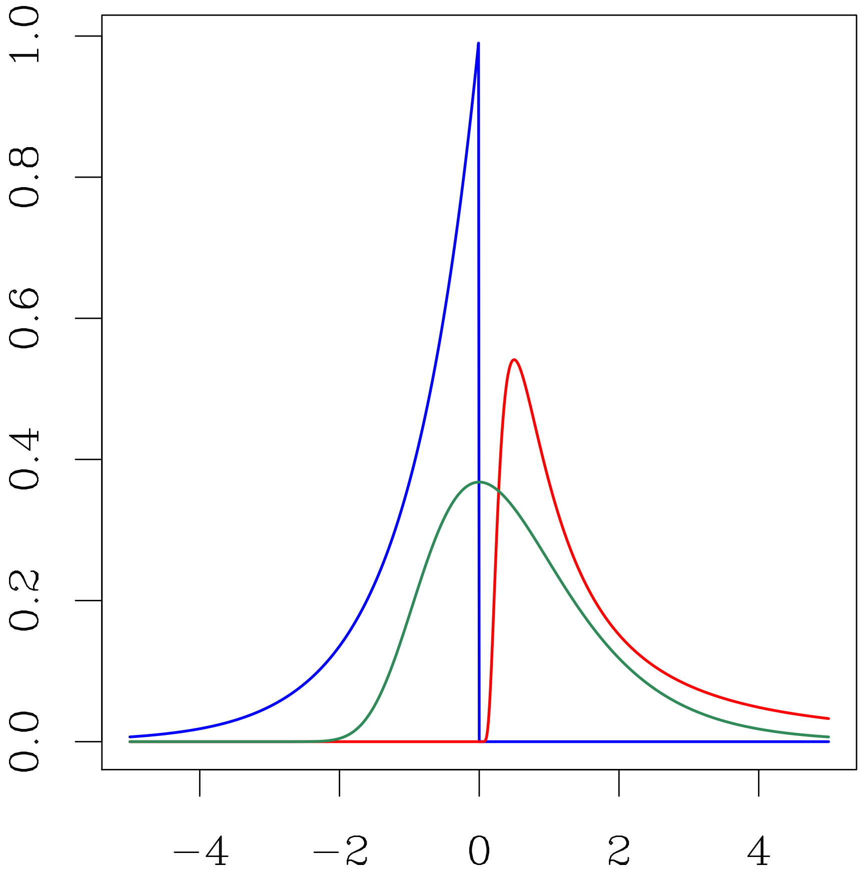
Typical example with daily data : block size = year
$\Longrightarrow$ **annual maxima**

Return level :
What is the level u that the maximum runoff would exceed once in a 100 years?

$$P(M_n > u) = 1/100$$

Extreme Value Distributions

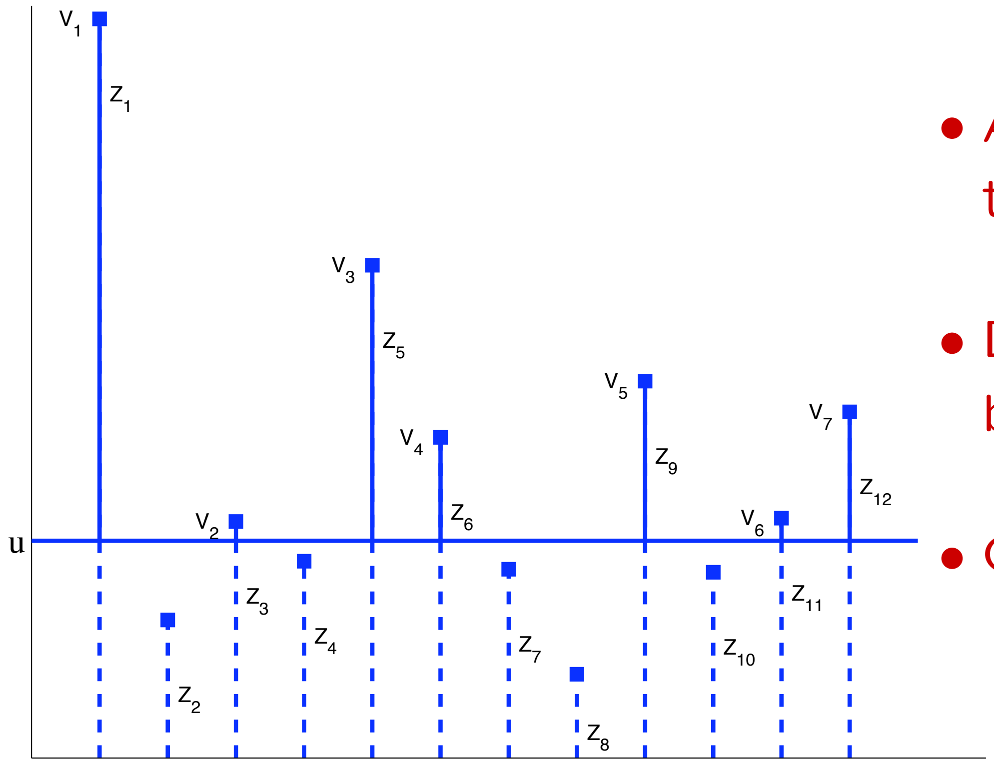
Maxima M_n, for large n, will behave either like



- Fréchet, heavy/Pareto tail: Student t, Cauchy
- Gumbel, light/exponential tail: Normal
- Weibull, finite tail: uniform

Peaks-over-Threshold

Excesses : $V_i = Z_i - u \mid Z_i > u$, u is a given threshold

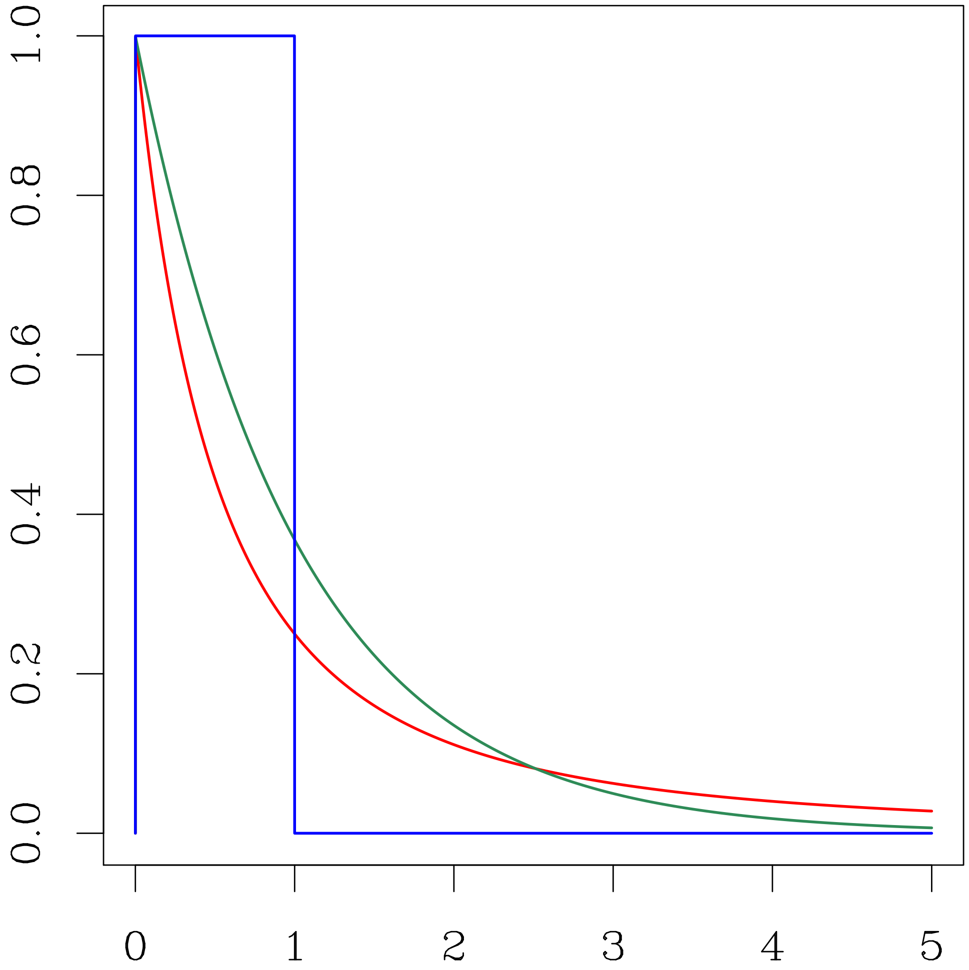


- Advantage : more data contribute to the estimation
- Difficulty : find a good threshold: bias / variance trade-off
- Questions on large events :

$$P(Z > u + \epsilon | Z > u)$$

Generalized Pareto Distribution

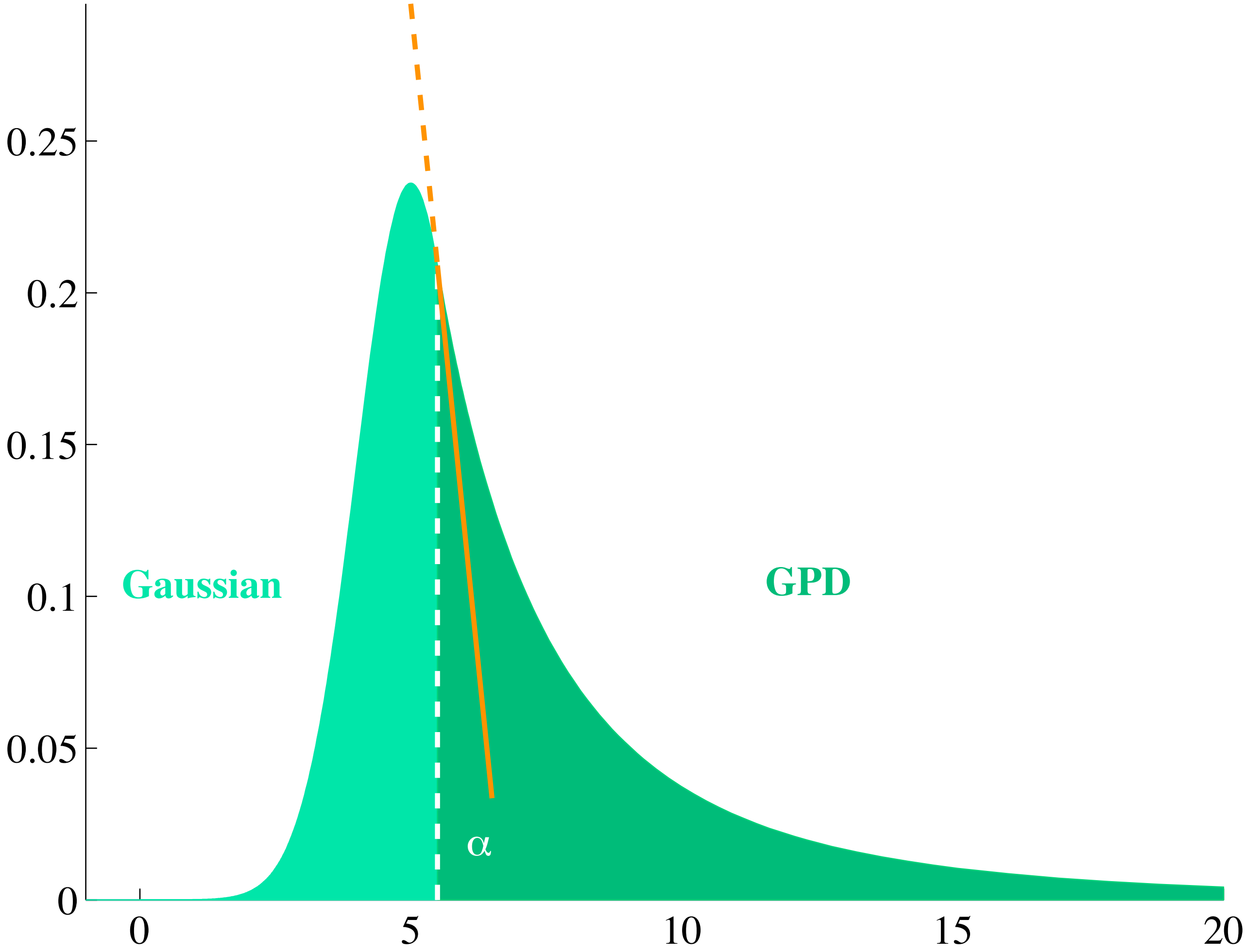
Excesses V_i, for large u will behave like a GPD with tail index ξ :



- $\xi > 0$, heavy/Pareto tail: Student t, Cauchy
- $\xi = 0$, light/exponential tail: Normal, Log-Normal, Gamma
- $\xi < 0$, finite tail: Uniform, Beta

Univariate Hybrid Pareto

Heavy-tailed distribution built by stitching together a Gaussian and a **Generalized Pareto** with continuity constraints



$h(\cdot;\xi,\mu_h,\sigma_h)$	**density**
$\xi,\ \beta$	**GPD**
μ_h,σ_h	**Gaussian**
α	**junction point**

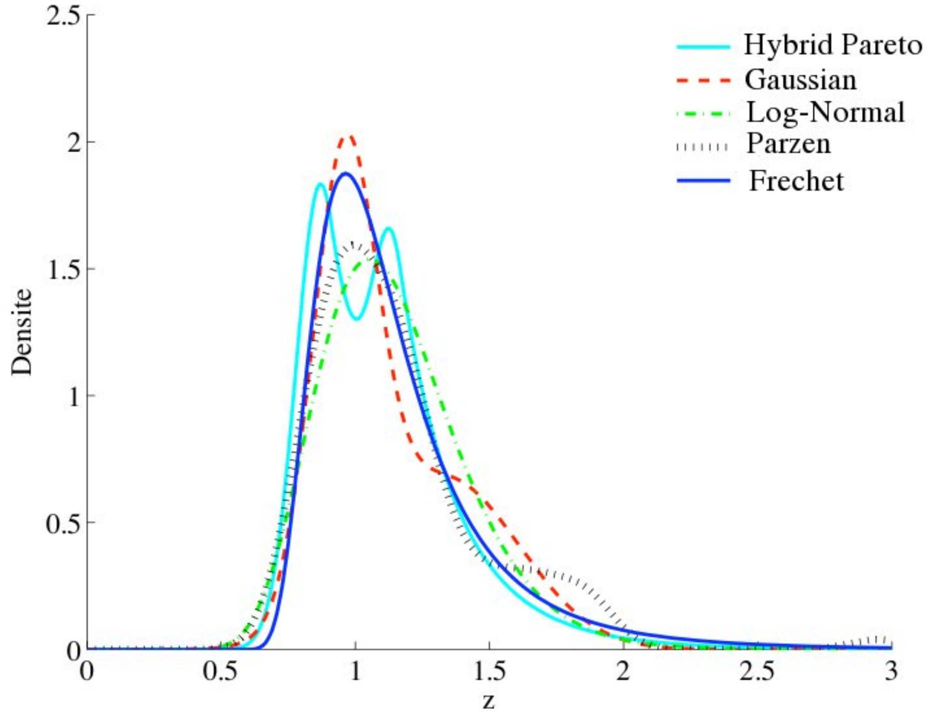
Modelling with the hybrid Pareto

✷ **Mixture of hybrid Paretos : univariate heavy tailed data**

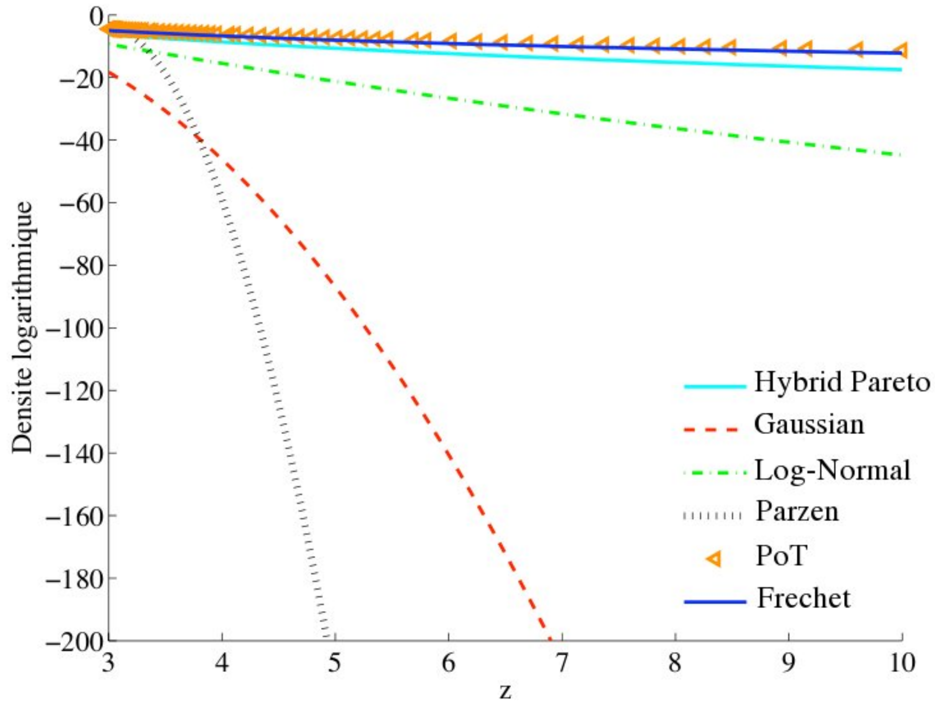
- **non parametric** in the central part: mixture of Gaussians
- **parametric** in the upper tail: combination of GPDs

Comparing Mixture Components

100 points from a Frechet distribution with $\xi = 0.2$



Central part



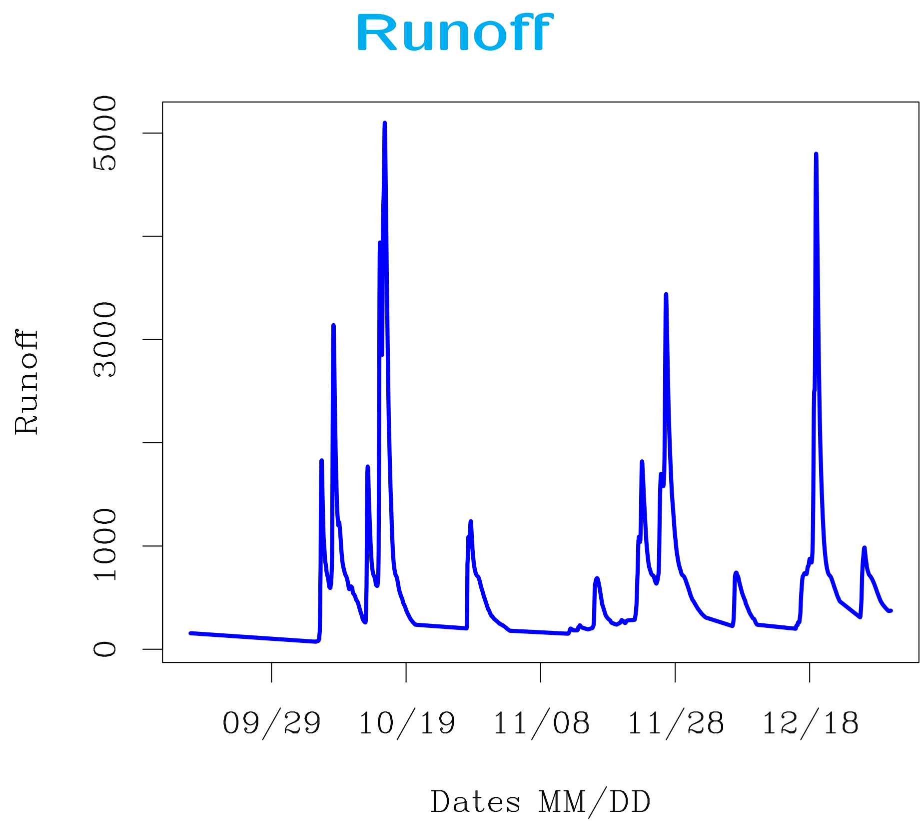
Upper tail

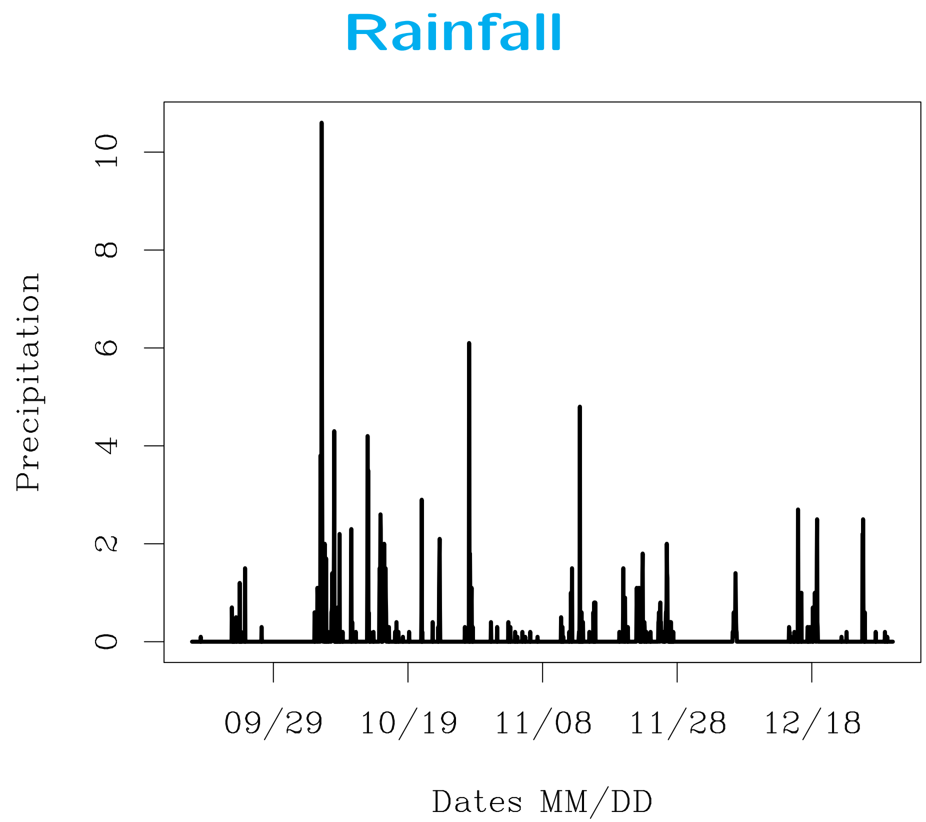
Modelling with the hybrid Pareto

* **Mixture of hybrid Paretos : univariate heavy tailed data**
  - **non parametric** in the central part: mixture of Gaussians
  - **parametric** in the upper tail: combination of GPDs

* **Conditional mixture: parameters are functions of the input**
  - **environmental application :** rainfall-runoff modelling
  - **downscaling:** precipitation given large scale variables

Rainfall-Runoff process





Projection Pursuit

1. **Find a 1D projection which is interesting**
   interestingness could mean **heavy tails**

Projection Pursuit

1. **Find a 1D projection which is interesting**
   interestingness could mean **heavy tails**

2. **Remove interestingness using a rotation trick**

   - **Rotate** to align the 1D projection with the x-axis
   - **Gaussianize** by modelling with a univariate density along the x-axis
   - **Rotate** back

Projection Pursuit

1. **Find a 1D projection which is interesting**

   interestingness could mean **heavy tails**

2. **Remove interestingness using a rotation trick**

   Rotate, Gaussianize, Rotate back

3. **Iterate steps 1 and 2**

   Stop when no more interesting 1D projection can be found

Projection Pursuit

1. **Find a 1D projection which is interesting**

   interestingness could mean **heavy tails**

2. **Remove interestingness using a rotation trick**

   Rotate, Gaussianize, Rotate back

3. **Iterate steps 1 and 2**

   Stop when no more interesting 1D projection can be found

4. **Model resulting density with a multivariate Gaussian**

   Combine the multivariate Gaussian with the 1D densities

Bivariate Hybrid Pareto Construction

Last steps of Projection Pursuit and reverse

4'. **Start from a bivariate standard Gaussian :**
assume no **interestingness == heavy tail** is present

Bivariate Hybrid Pareto Construction

Last steps of Projection Pursuit and reverse

4'. **Start from a bivariate standard Gaussian :**
assume no **interestingness** == **heavy tail** is present

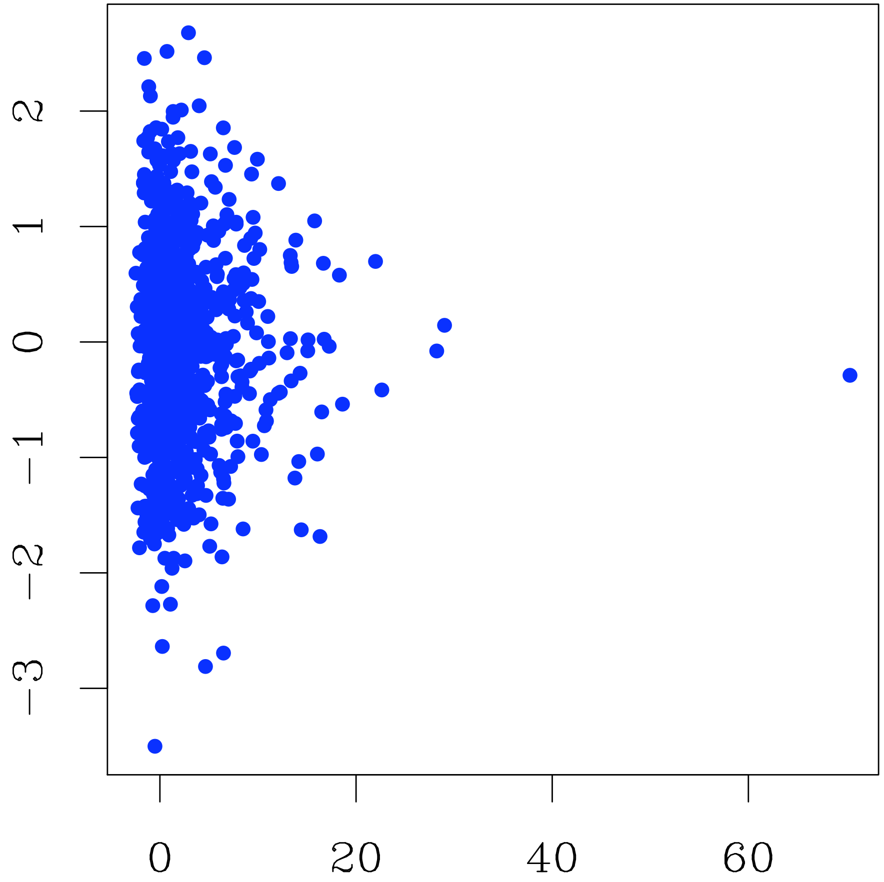
3'. **Iterate once, i.e. apply steps 1 and 2 once**
introduce interestingness in a **1D projection defined by an angle** θ

- Transform the density along the x-axis into a **heavy-tailed density**
- **Rotate back** to align heavy-tailed density with **angle** θ

Bivariate Hybrid Pareto Construction

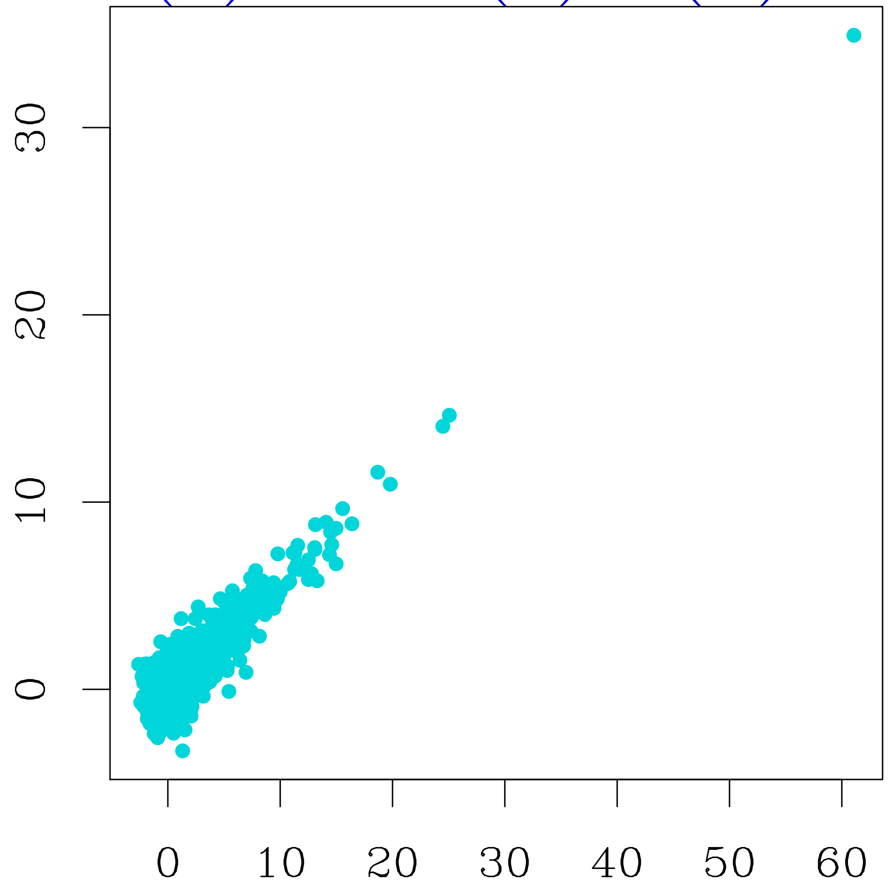
Independent random variables

$\widetilde{X} \rightarrow h(\cdot;\xi,0,\sigma_h)$ and $\widetilde{Y} \rightarrow \phi(\cdot;0,\sigma)$



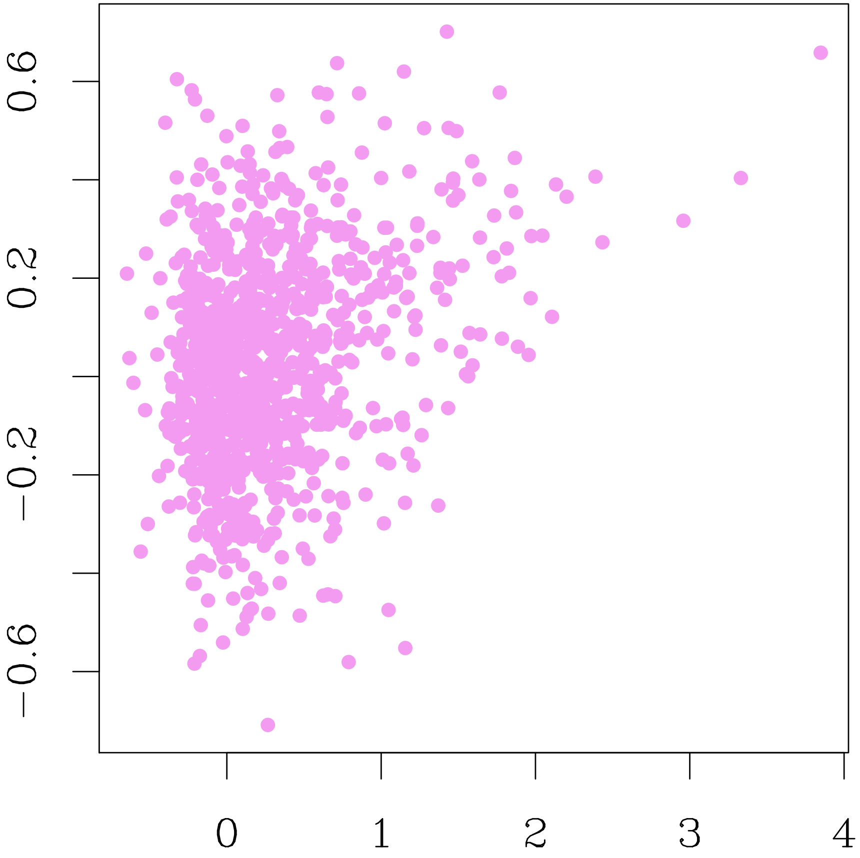
Rotation with angle $\theta \in [-\pi, \pi)$

$$\begin{pmatrix} X \\ Y \end{pmatrix} = R(\theta)^t \begin{pmatrix} \widetilde{X} \\ \widetilde{Y} \end{pmatrix} + \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

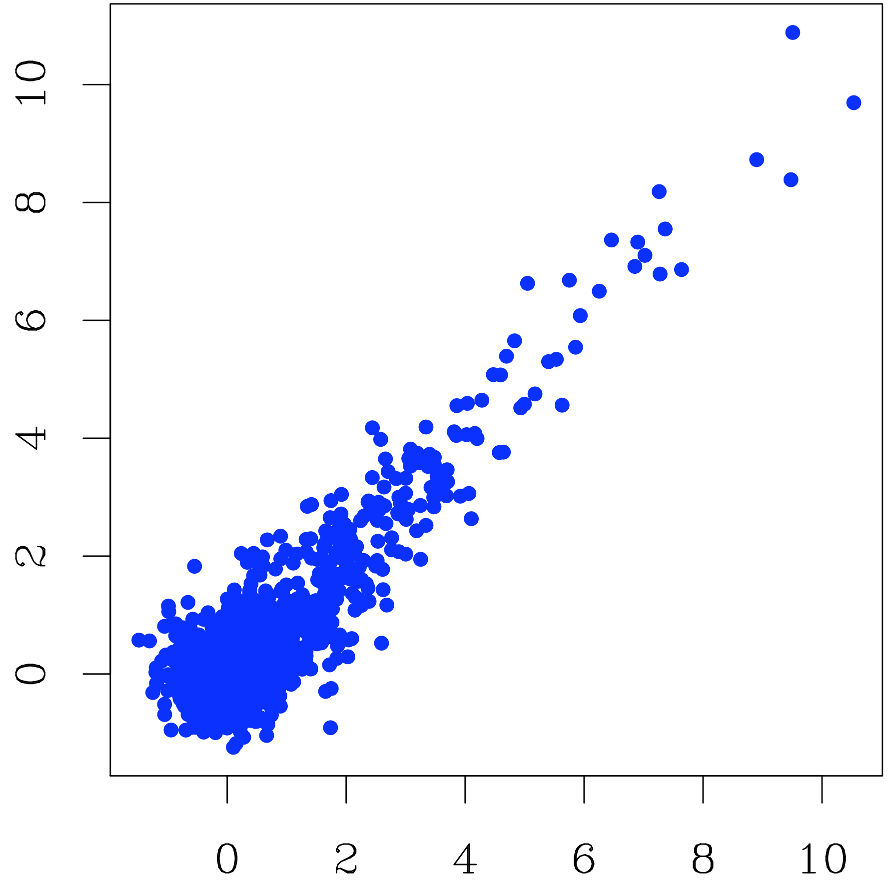


Sample from Bivariate Hybrid Paretos

$\theta = -\pi/24$ and $\xi = 0.001$



$\theta = -\pi/4$ and $\xi = 0.2$



Density of the Bivariate Hybrid Pareto

According to the PP construction :

$$h_2(x, y; \psi) = \phi_2(x, y) \frac{h(p_\theta(x, y))}{\phi(p_\theta^\perp(x, y))}$$

✷ h and ϕ are the densities of the hybrid Pareto and the standard Gaussian respectively

✷ p_θ and $p_\theta^\perp$ are the projections along the line defined by θ and the line orthogonal to it

Density of the Bivariate Hybrid Pareto

According to the PP construction :

$$h_2(x, y; \psi) = \phi_2(x, y) \frac{h(p_\theta(x, y))}{\phi(p_\theta^\perp(x, y))}$$

✷ h and ϕ are the densities of the hybrid Pareto and the standard Gaussian respectively

✷ p_θ and $p_\theta^\perp$ are the projections along the line defined by θ and the line orthogonal to it

✷ **The density of the bivariate hybrid decomposes into**

$$h_2(x, y; \psi) = \underbrace{h(p_\theta(x - \mu_1, y - \mu_2); \xi, \sigma^{(h)})}_{\text{hybrid Pareto}} \ \underbrace{\phi(p_\theta^\perp(x - \mu_1, y - \mu_2)/\sigma)}_{\text{Gaussian}}$$

✷ $\psi = (\mu_1, \mu_2, \sigma, \theta, \xi, \sigma^{(h)})$ is the bivariate hybrid Pareto parameter vector

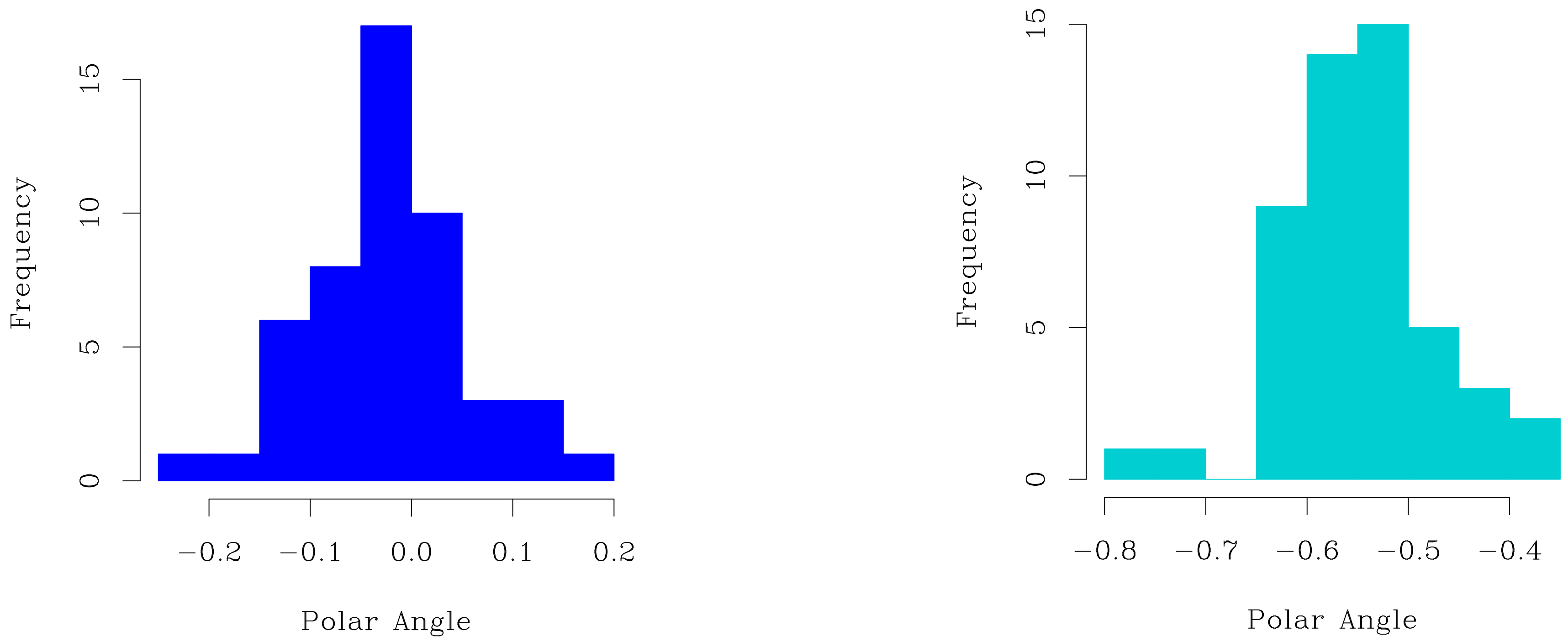
Rotation and Dependence

* **Rotation transforms the covariance matrix :** $R(\theta)\Sigma R(\theta)^t$

Rotation and Dependence

- **Rotation transforms the covariance matrix :** $R(\theta)\Sigma R(\theta)^t$
- Rotation introduces **dependence in the extremes** as well

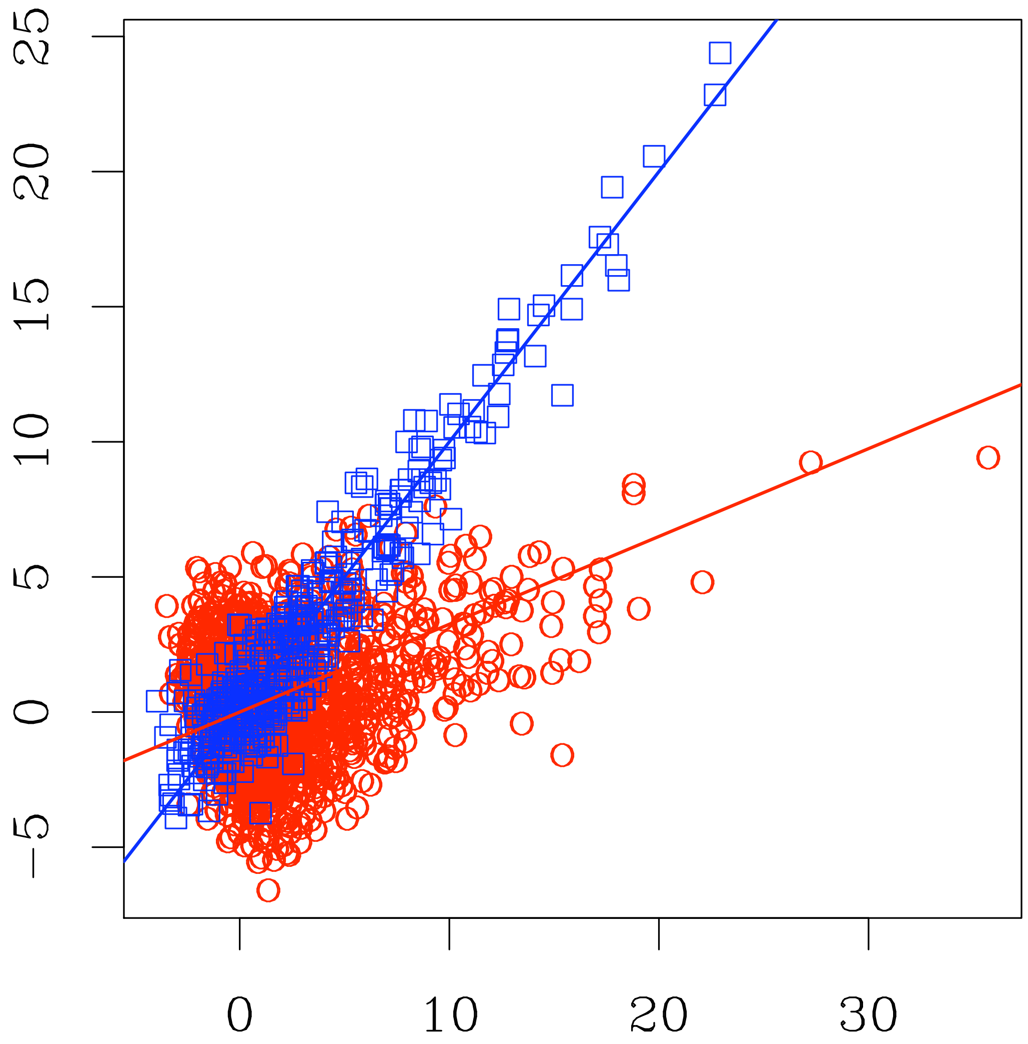
Polar angle corresponding to large radius



Independence $\theta = 0$ Dependence $\theta \neq 0$

Directions of Extremes in the Mixture

$$P(||(X,Y)|| > R, S=j) = P(||(X,Y)|| > R|S=j)P(S=j) \approx \frac{\pi_j}{\gamma}\left(\frac{\xi}{\beta_j}\right)^{-1/\xi} R^{-1/\xi}$$



$$P(\theta = \theta_j) \propto \pi_j (\sigma_j^{(h)})^{1/\xi}$$

- Same tail index: $\xi_j = \xi \ \forall j$
- π_j is the mixture weight

Initialize θ_j

Angle Clustering : Find good extremal directions

a) Center and sphere the data

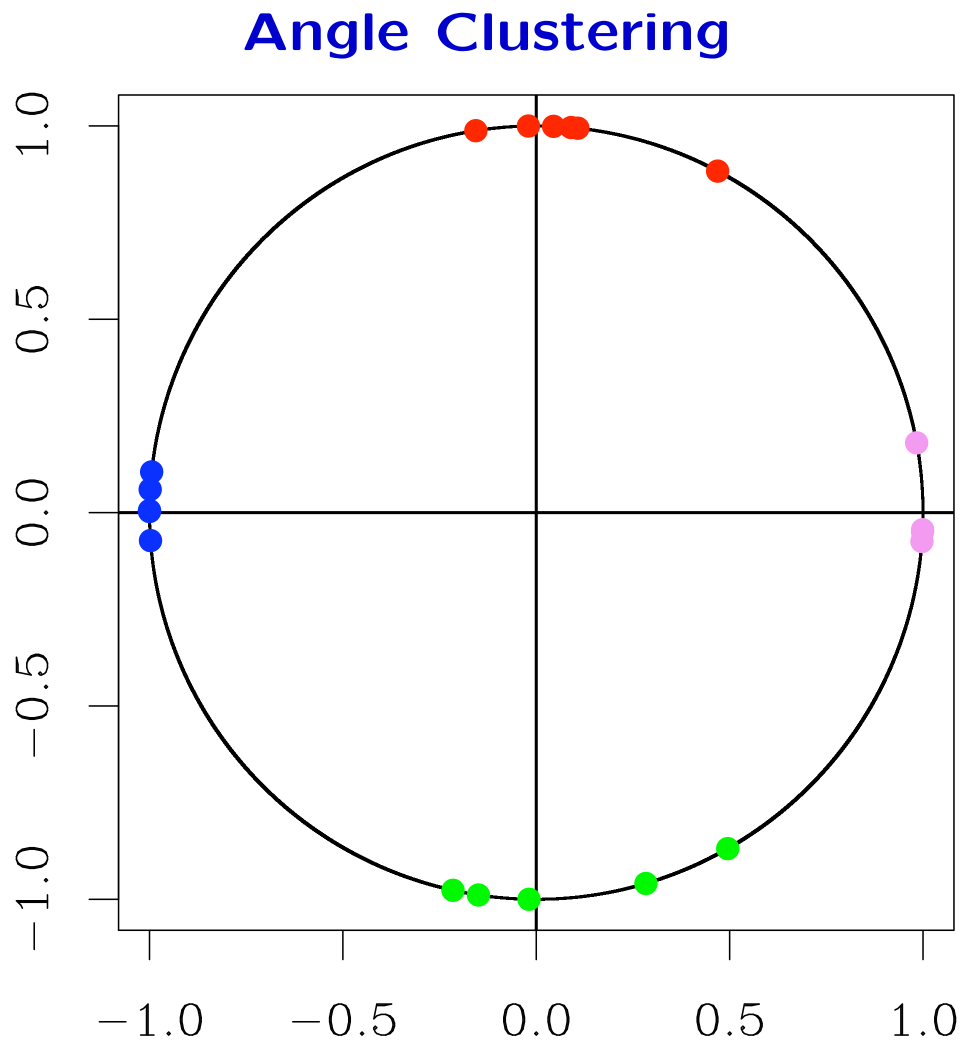
b) Transform into polar coordinates: $R \in \mathbb{R}^+$ and $\omega \in [-\pi, \pi)$

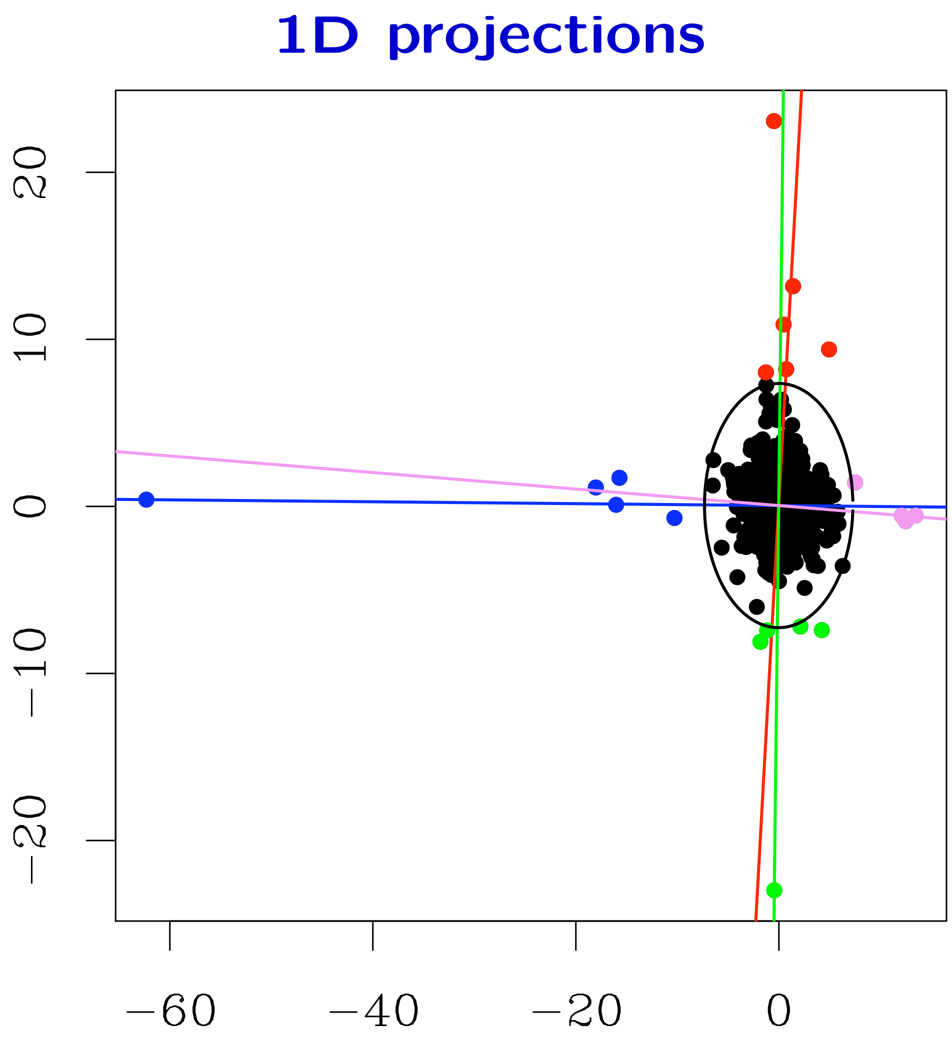
c) Consider ω such that $R > u$

d) Compute **circular distances**

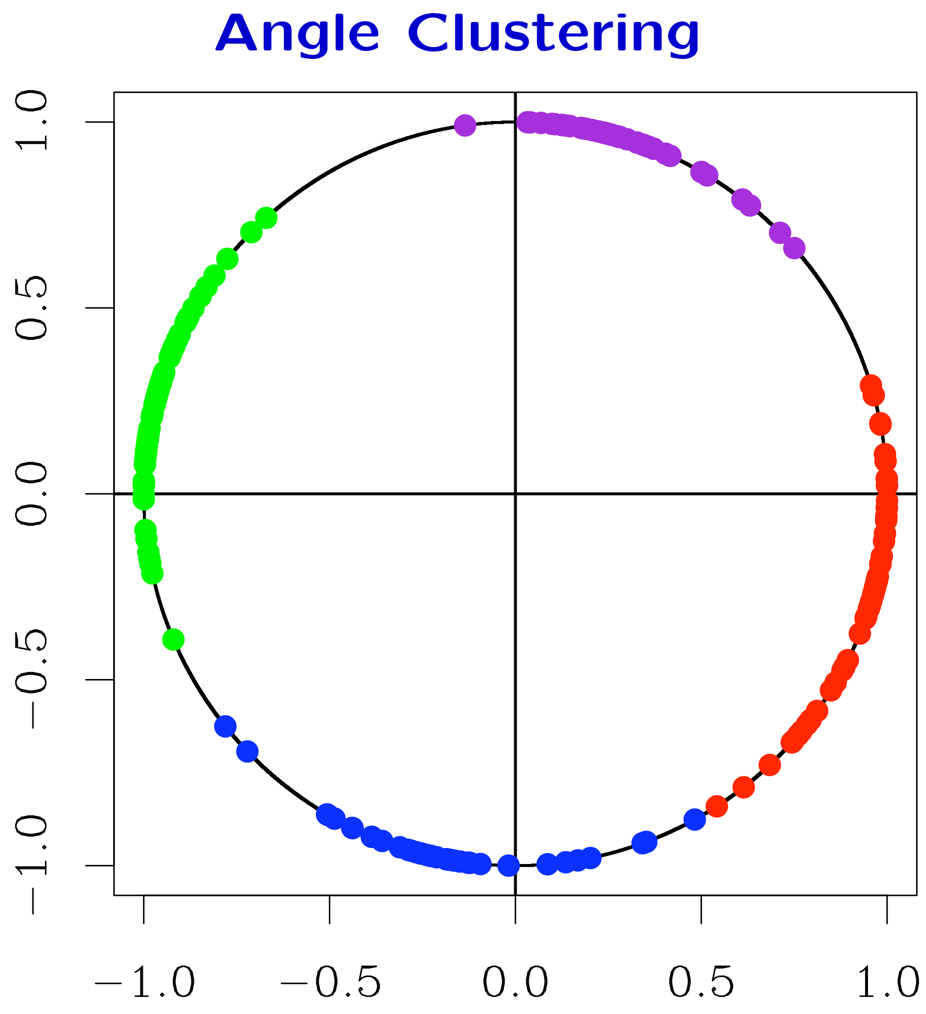
e) Perform **clustering** and cluster centers are taken as initial angles

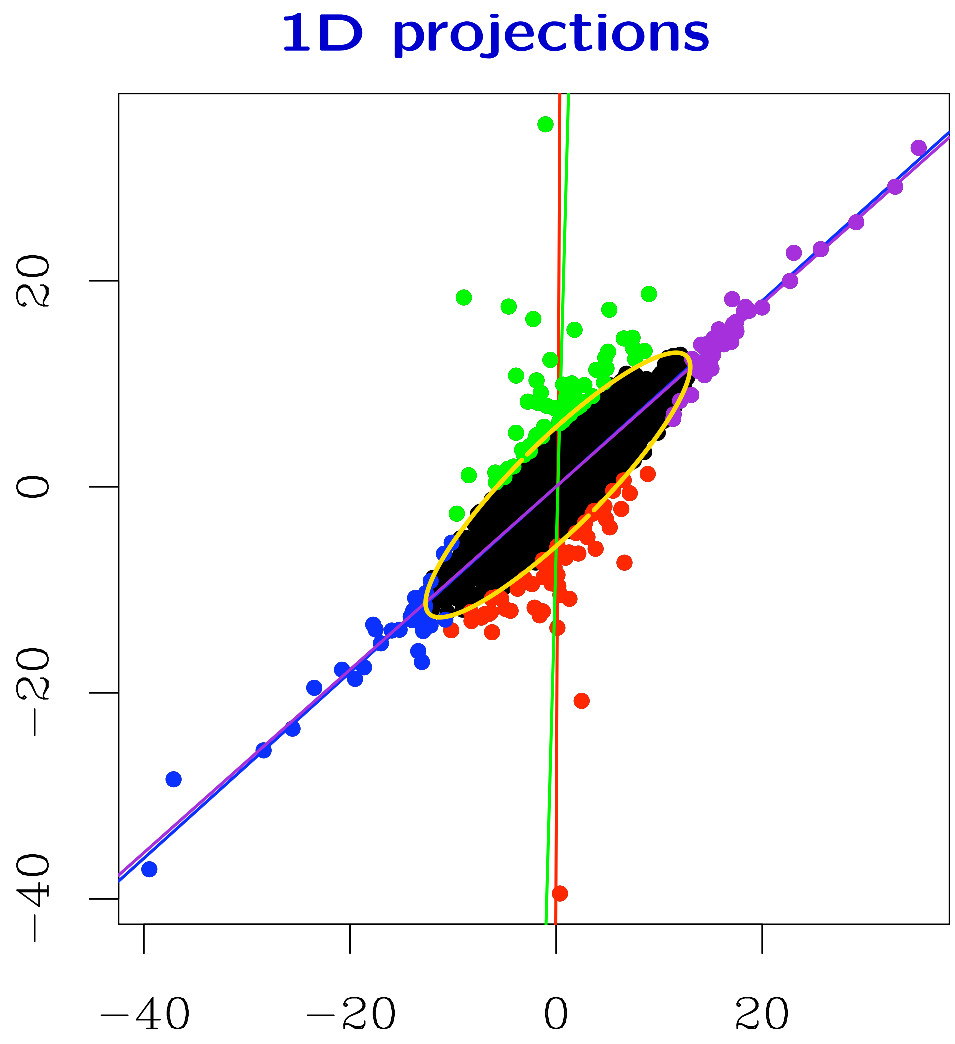
Independent Bivariate α-stable Data





AR(1) with Student t Noise





Mixture Initialization

1. Estimate the rotation angles θ_j

✷ **Clustering** of angles corresponding to large radius

Mixture Initialization

1. Estimate the rotation angles θ_j

✷ **Clustering** of angles corresponding to large radius

2. Iterative classification

✷ **Decrease the threshold** used for angle estimation

✷ **Classify new points** according to previous classification

Mixture Initialization

1. Estimate the rotation angles θ_j

✷ **Clustering** of angles corresponding to large radius

2. **Iterative classification**

✷ **Decrease the threshold** used for angle estimation

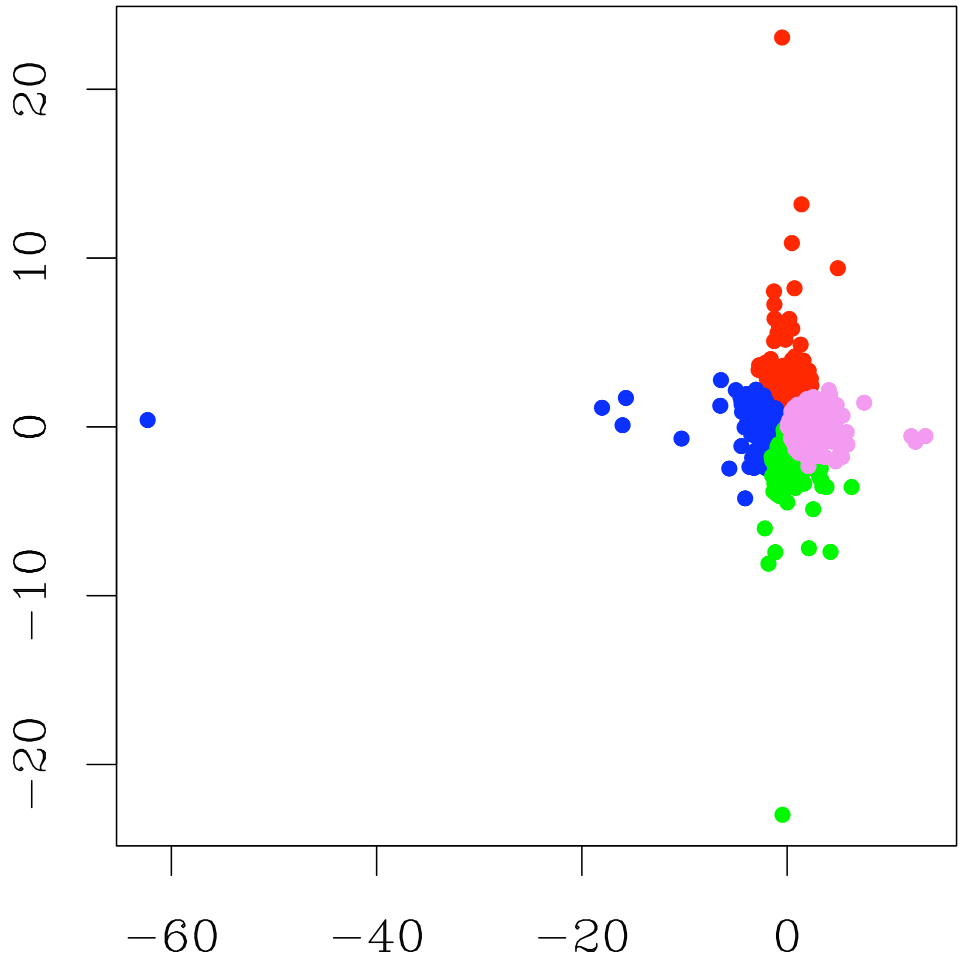
✷ **Classify new points** according to previous classification

3. Initialize one component per cluster

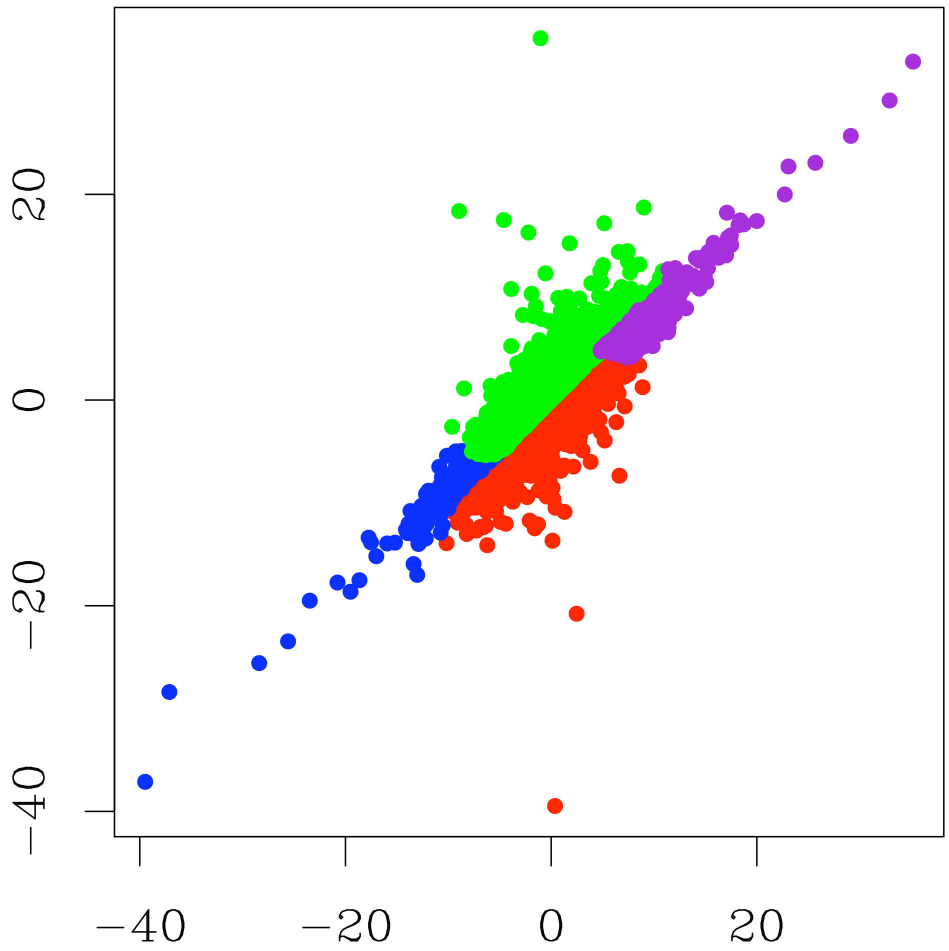
✷ **Estimate univariate density parameters** on projected data

Iterative Clustering

Independent Bivariate α-stable Data



AR(1) with Student t Noise



GEM for the Bivariate Mixture

$\sum_{j=1}^{m} \pi_j h_2(x, y; \psi_j)$, with $h_2(\cdot, \cdot; \psi_j)$ the bivariate hybrid pareto density

$\{\pi_j, \psi_j\}_{j=1:m}$ are estimated by **maximizing the log-likelihood**

GEM for the Bivariate Mixture

$\sum_{j=1}^{m} \pi_j h_2(x, y; \psi_j)$, with $h_2(\cdot, \cdot; \psi_j)$ the bivariate hybrid pareto density

$\{\pi_j, \psi_j\}_{j=1:m}$ are estimated by **maximizing the log-likelihood**

Generalized EM algorithm:

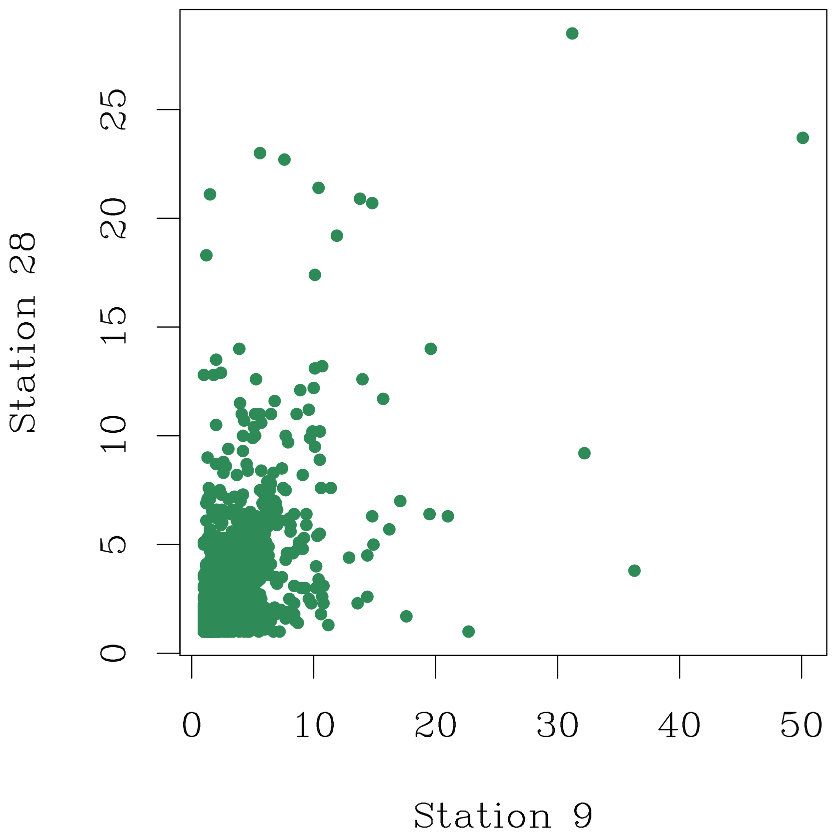
E-step: compute posteriors $\tau_{i,j}$ according to current parameters

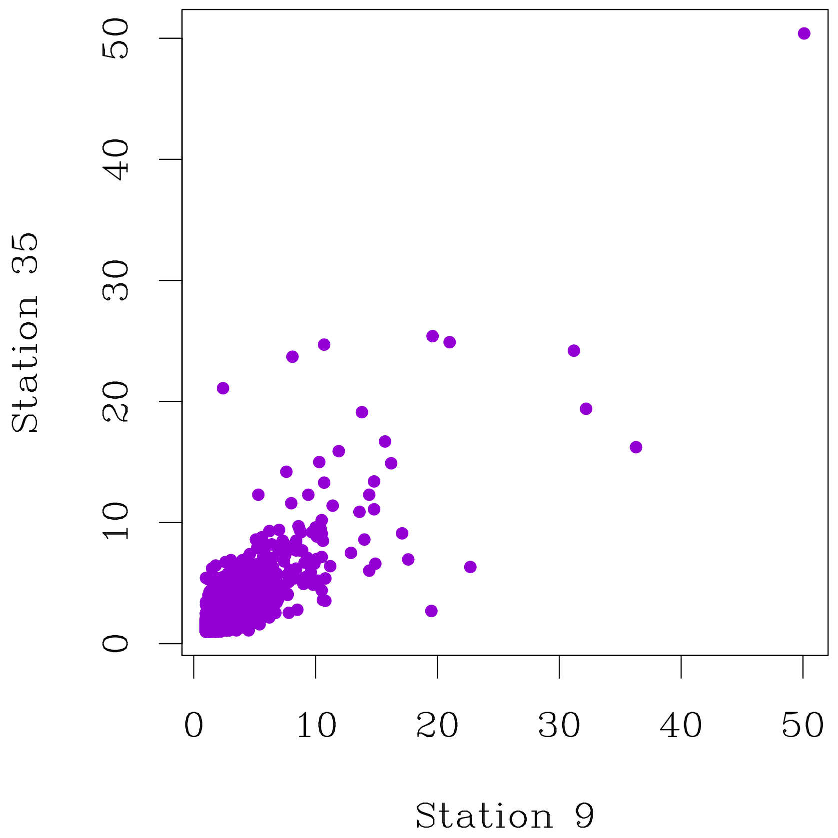
M-step:

1. Update the priors $\pi_j = 1/n \sum_{i=1}^{n} \tau_{i,j}$
2. For each j, optimize numerically w/r to ψ_j:

$$\sum_{i=1}^{n} \tau_{i,j} \log\left(h_2(x_i, y_i; \psi_j)\right)$$

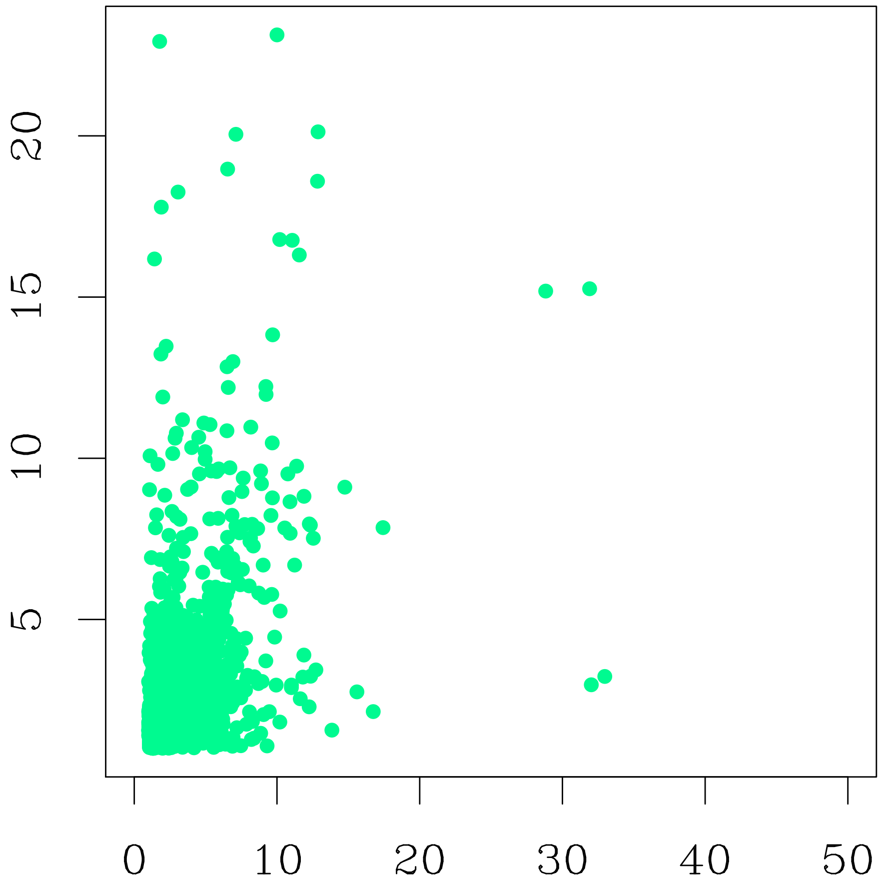
Precipitation



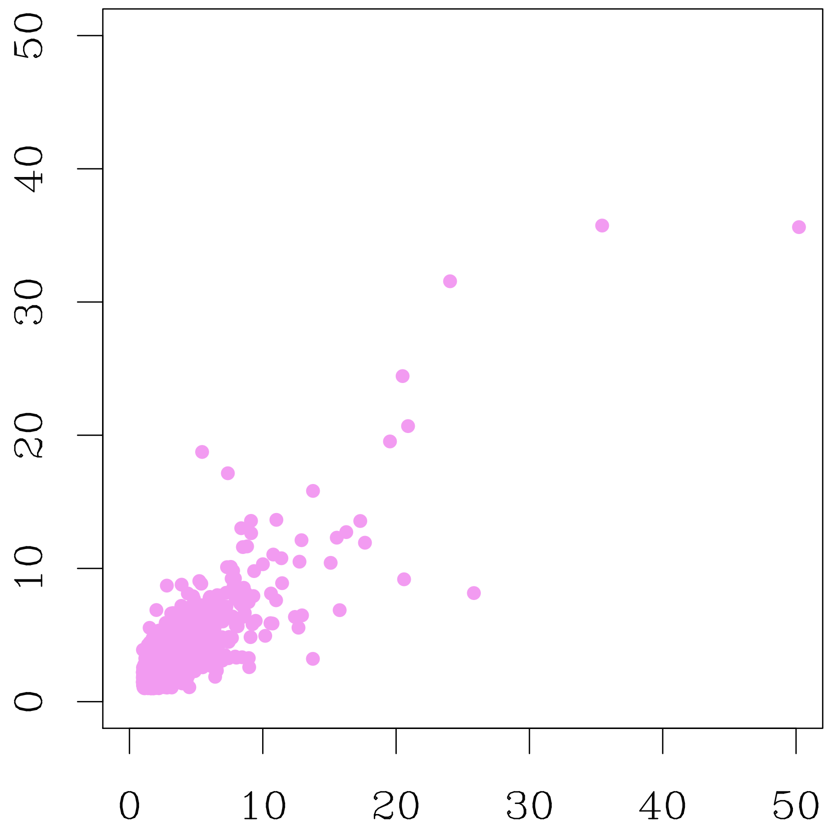


Data generated from the trained model

Trained on data from stations 9 and 28

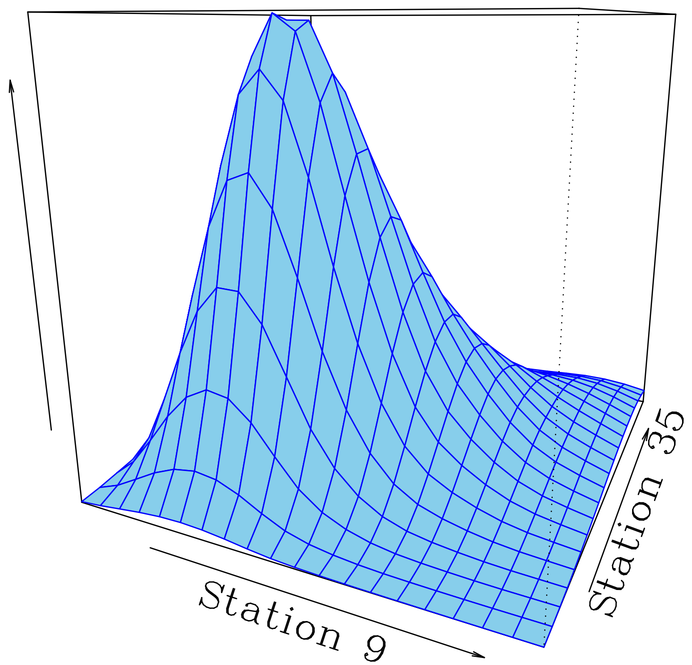


Trained on data from stations 9 and 35

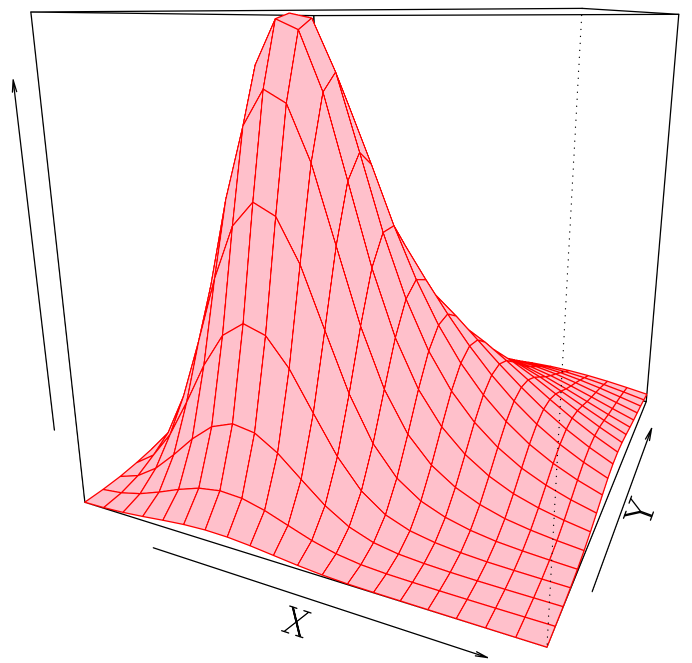


Full Density Estimate

Stations 9 and 35



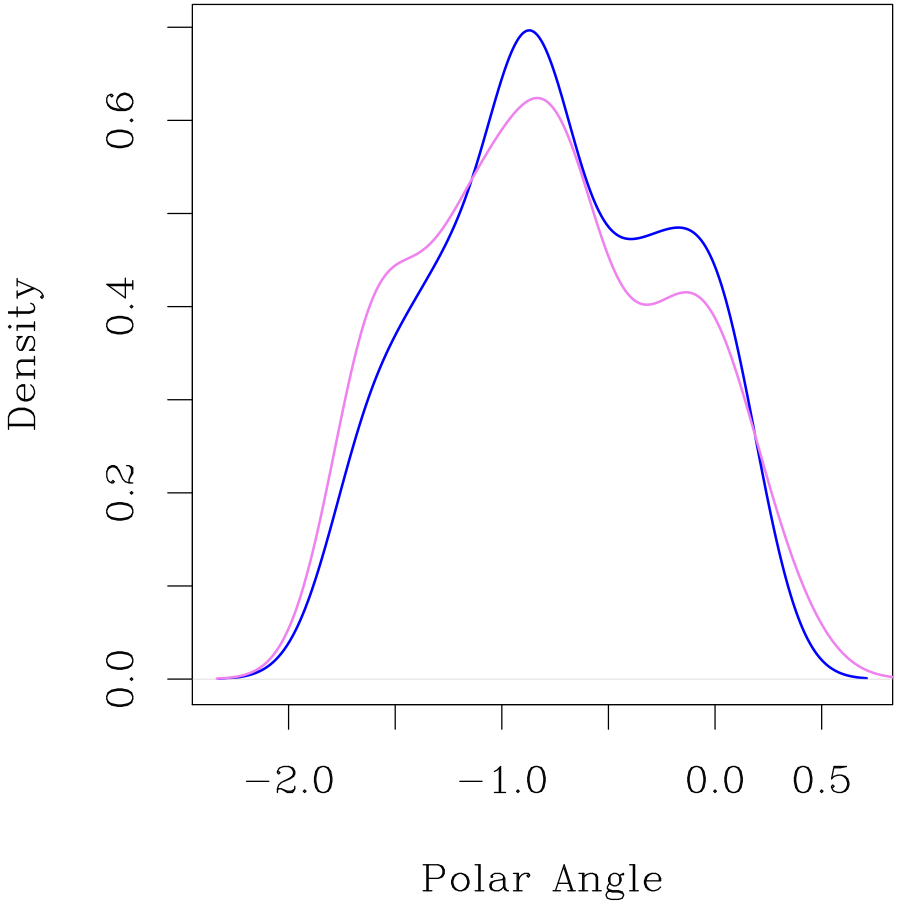
Trained mixture



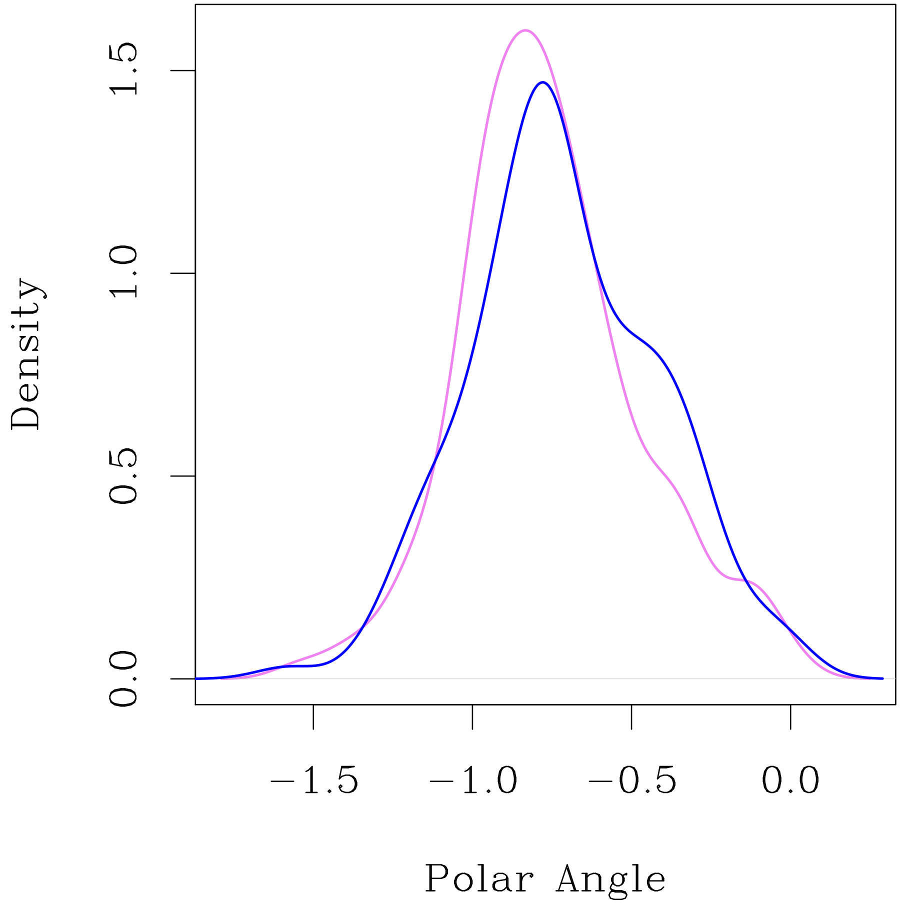
Kernel Estimates of the Angular Spread

Precipitation data versus **Trained Mixture**

Stations 9 and 28



Stations 9 and 35



Next Steps

* Introduce **multivariate Gaussian** components to model the central part

* Automated **model selection**: number of components

* Higher dimensions...

References

[1] Carreau J. and Bengio Y. (2009) **A Hybrid Pareto Model for Asymmetric Fat-tailed Data: the Univariate Case**, *Extremes*, Vol. **12**, 53-76.

[2] Friedman, J. H. (1987) **Exploratory Projection Pursuit**, *JASA*, Vol. **82**, 249-266.

[3] Chen, S. S. and Gopinath, R. A. (2001) **Gaussianization**, *NIPS*, Vol. **13**.

[4] Coles, S., Heffernan J. and Tawn J. (1999) **Dependence measures for extreme value analysis**, *Extremes*, Vol. **2**, 339-365