**Mathieu VRAC**

Laboratoire des Sciences du Climat et de l' Environnement-IPSL
mathieu.vrac@lsce.ipsl.fr

# *Extreme Value Theory: Applications in Downscaling of precipitation*

16th Itzykson meeting on Extremes and Records
Orme des Merisiers, 14-17, 2011

# Outline

- Introduction:
  - Why do we need downscaling ?
  - What is downscaling in practice ?
  - When ? How ?
- One statistical DS approach
  - Classical “modelling” of precipitation
  - Extension to model the whole range of precipitation
- How/What about “bivariate” extreme values?
- Conclusions (some)
  - Two other SDMs including extremes

# Introduction

- 30% of the world economic activities are affected by meteorological conditions (source: IPCC)

- **IPCC** scenarios of climate change have a **coarse spatial resolution !!** Not adapted to social and economic scales of impact studies

  - Social and economic impacts: agriculture, water resources, hydrology, air pollution, human health, etc.

- How will climate change interact with existing environmental features at a regional scale ?

- Downscaling: To derive sub-grid scale (regional or local) weather or climate using General Circulation Models (GCMs) outputs or reanalysis data (e.g. NCEP)

# Why to model local precipitation?


- Key-parameter in meteorology and climatology
- Highly stochastic/variable nature compared to other meteorological parameters

# How to downscale?: The basics

≈ 250 km

**Coarse atmospheric data**
Precipitation, temperature, humidity, geopotential, wind, etc.

How to use the **coarse simulations** to produce **regional/local** climate features?

**?**

Region, city, fields, point

**Local variables (e.g., precip., temp.)**
(small scale water cycle, impacts – crops, resources – etc.)

# How to downscale?: The basics

≈ 250 km

**Coarse atmospheric data**
Precipitation, temperature, humidity, geopotential, wind, etc.

**Dynamical** downscaling:

- GCMs to drive regional models (10-50km) determining atmosphere dynamics
- Requires a lot of computer time and resources => Limited applications

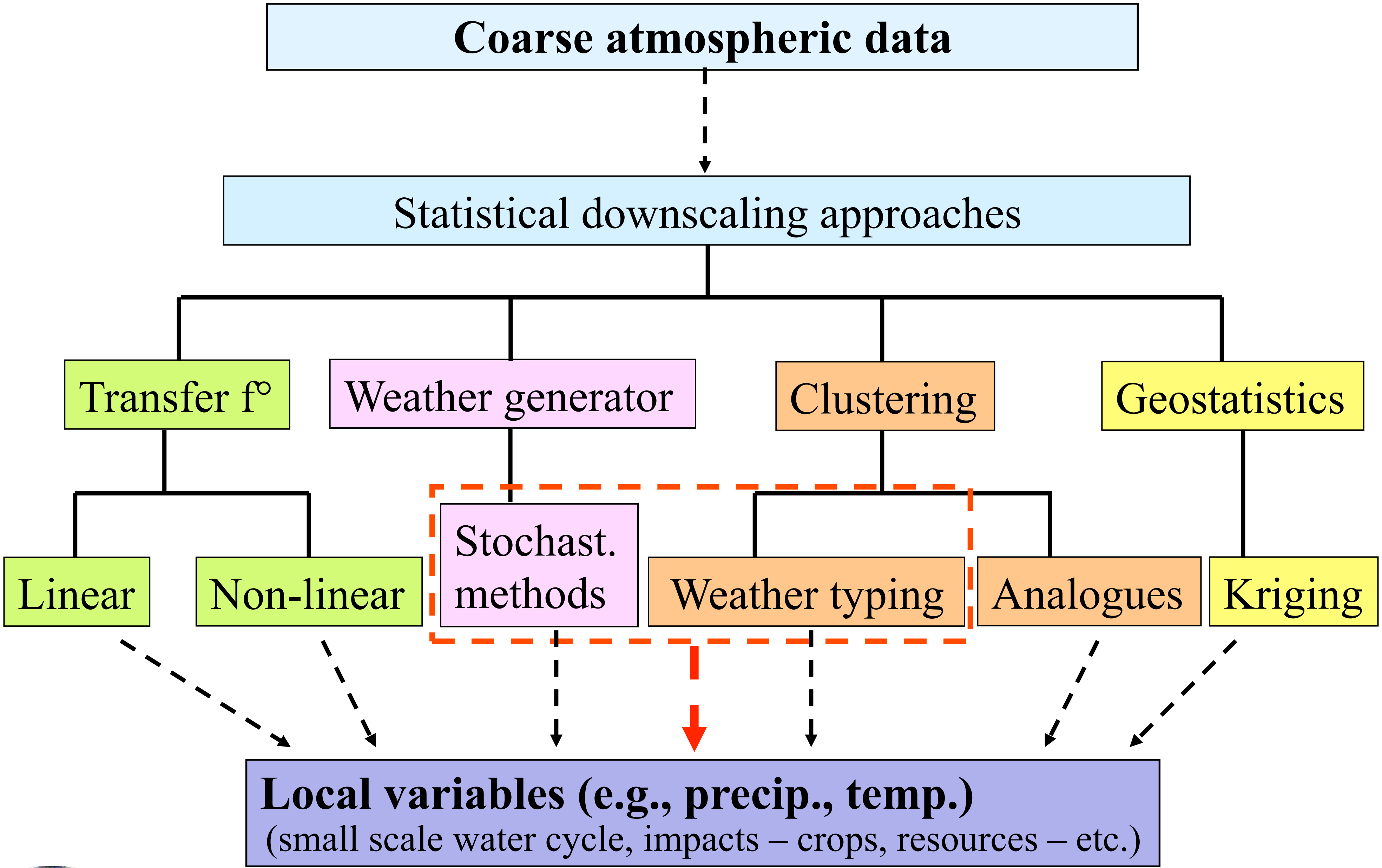
**Statistical** downscaling:

- Based on statistical relationships between large- and local-scale variables
- Low costs and rapid simulations applicable to any spatial resolution
- Uncertainties (results, propagation, etc)

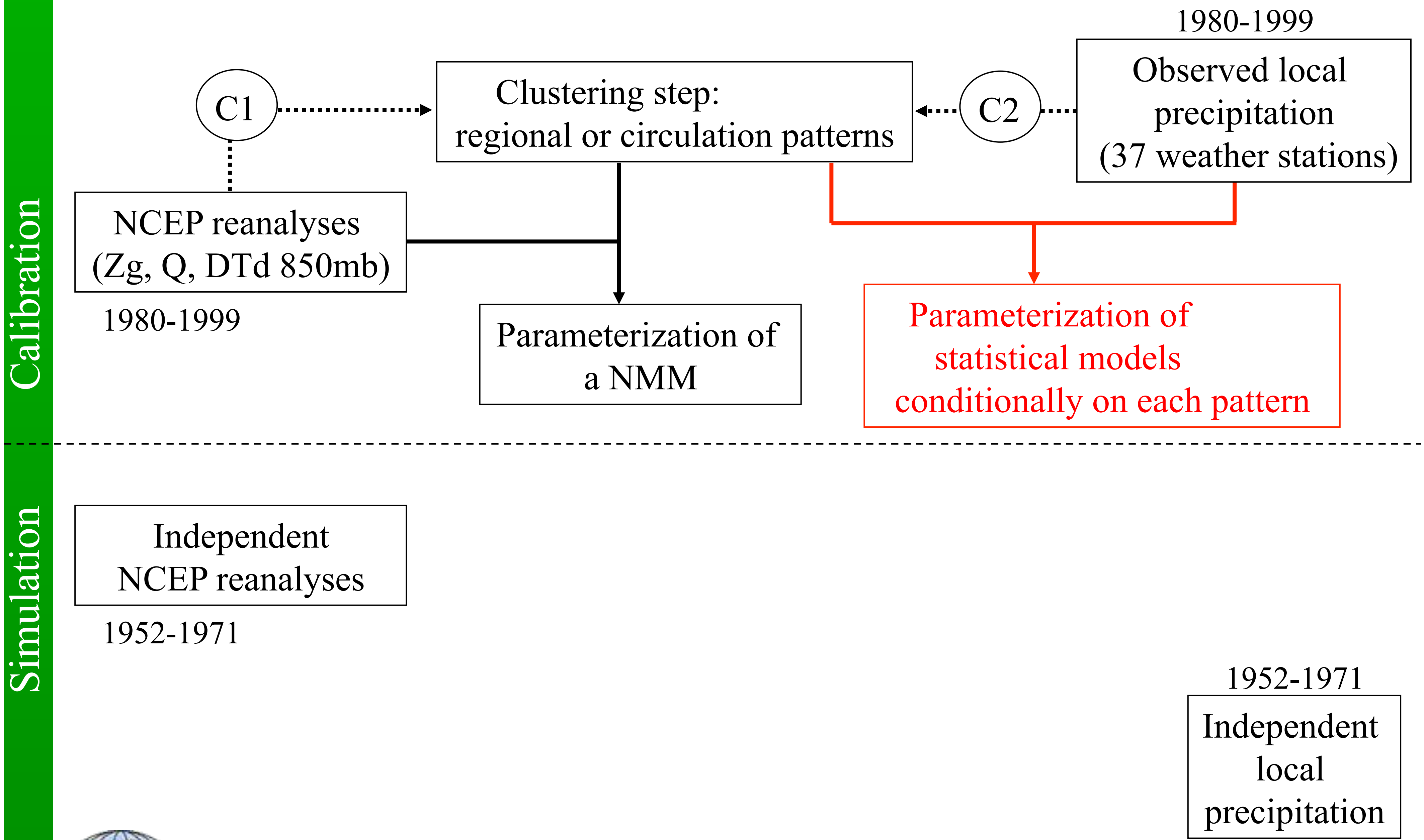
**Region, city, fields, point**

**Local variables (e.g., precip., temp.)**
(small scale water cycle, impacts – crops, resources – etc.)

# Main statistical approaches



# Scheme of the downscaling process (NSWT)



# The modelling part of NSWT

- Prior weather states ($S_t$=1,…, K)
- Fit a NMM to characterize transition probabilities:

$$P\left(S_t = j \middle| S_1^{t-1}, X_1^T\right) = P\left(S_t = j \middle| S_{t-1} = i, X_t\right) \qquad (1)$$

$$\propto \gamma_{ij}\, exp\left(-\frac{1}{2}(X_t - \mu_{ij})\Sigma^{-1}(X_t - \mu_{ij})'\right)$$

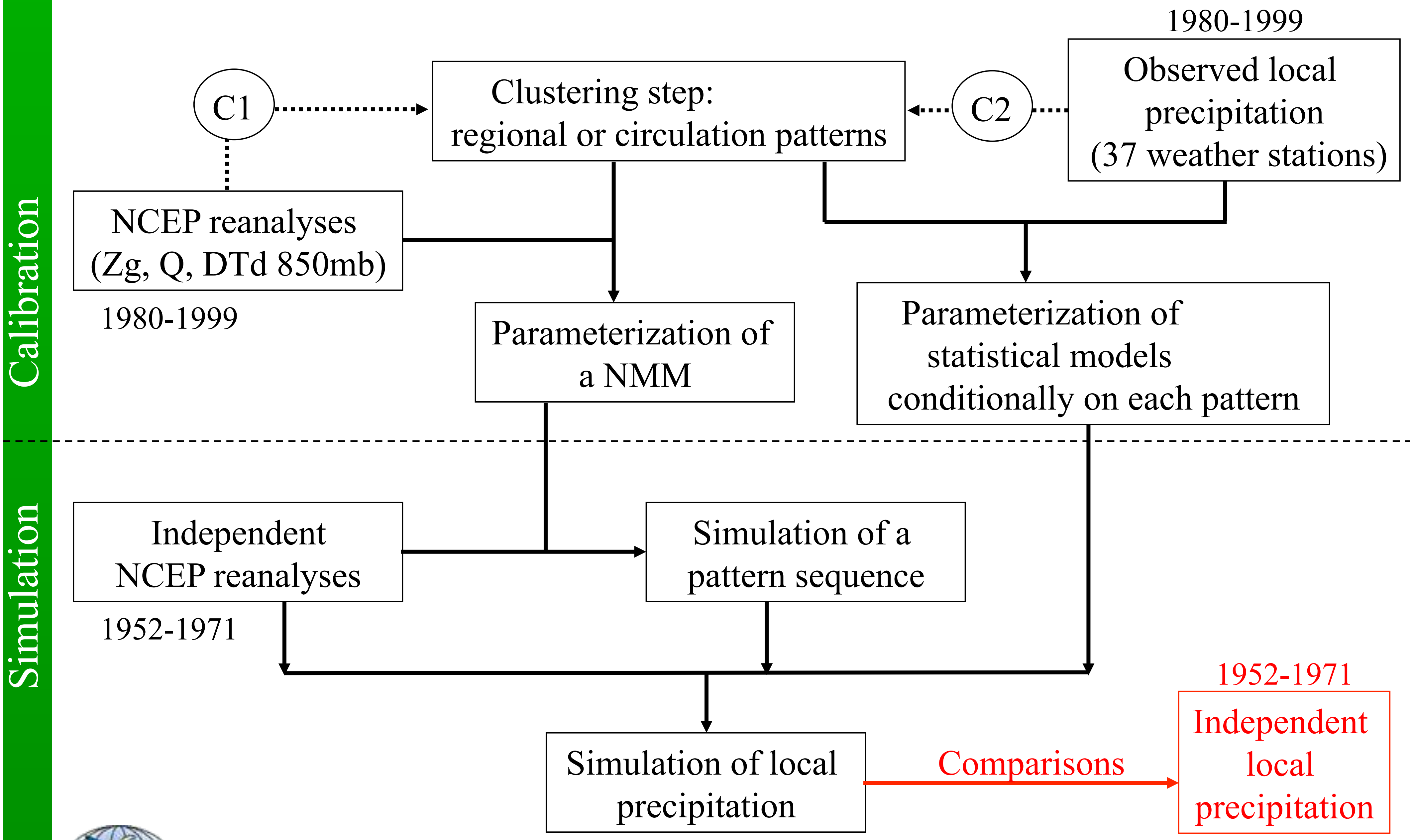
- Rainfall modeling

$$f_{R_t|S_1^T, R_1^{t-1}, X_1^T} = f_{R_t|S_t = s, X_t}(r) \qquad (2)$$

$$= \prod_{i=1}^{37}\left[\left(p_{si}(X_t) G\left(r_t^i; \alpha_{si}, \beta_{si}\right)\right)^{1_{\{r_t^i>0\}}}\left(1 - p_{si}(X_t)\right)^{1_{\{r_t^i=0\}}}\right]$$

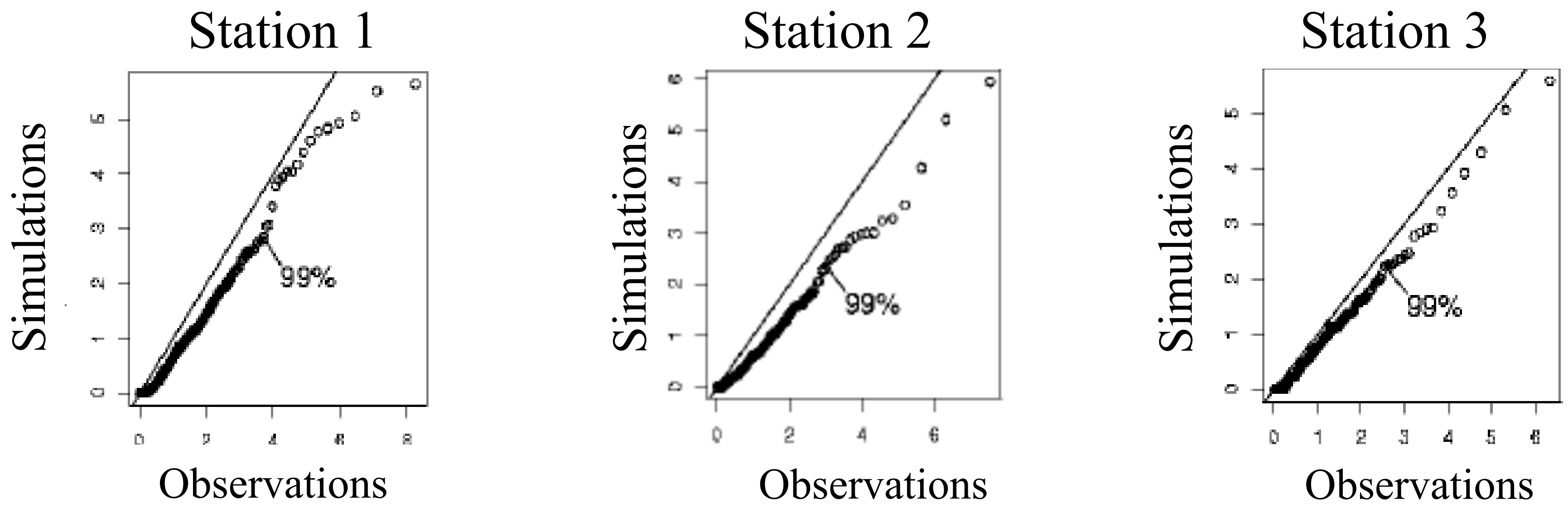
With $G$ = Gamma pdf and $p_{si}(X_t) = \dfrac{exp(X_t'\lambda_{si})}{1 + exp(X_t'\lambda_{si})}$ (3)

# Scheme of the downscaling process (NSWT)

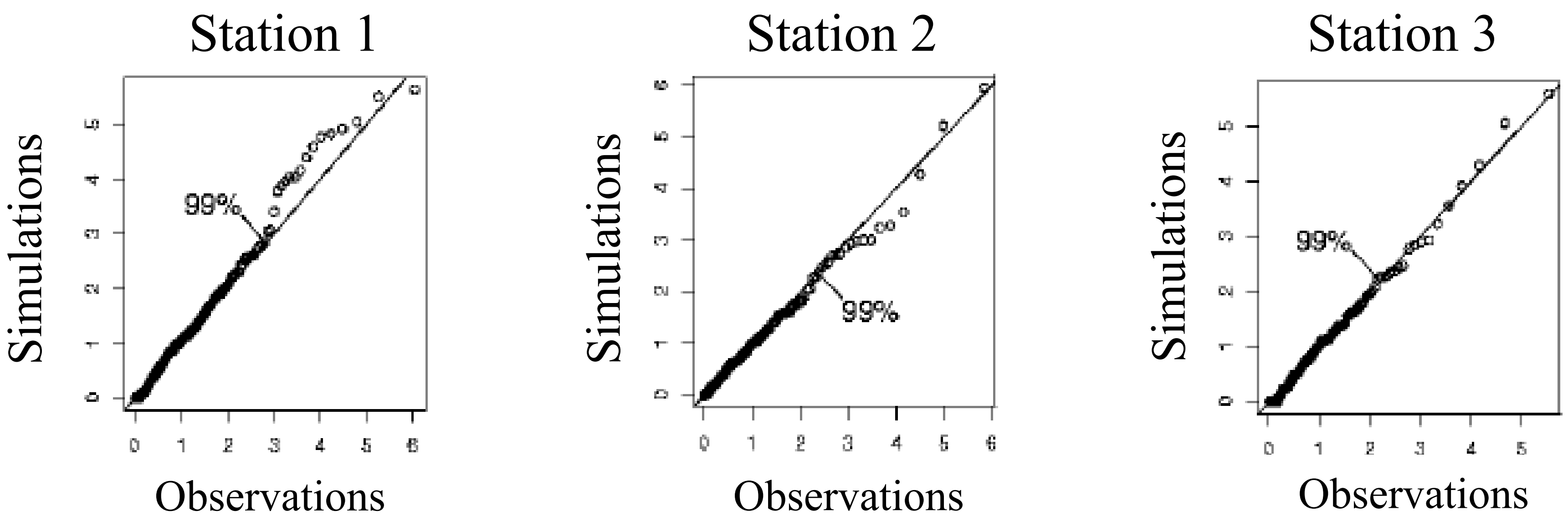


# Quantiles-Quantiles (QQ) plot

## Based on **circulation** patterns




## Based on **precipitation** patterns

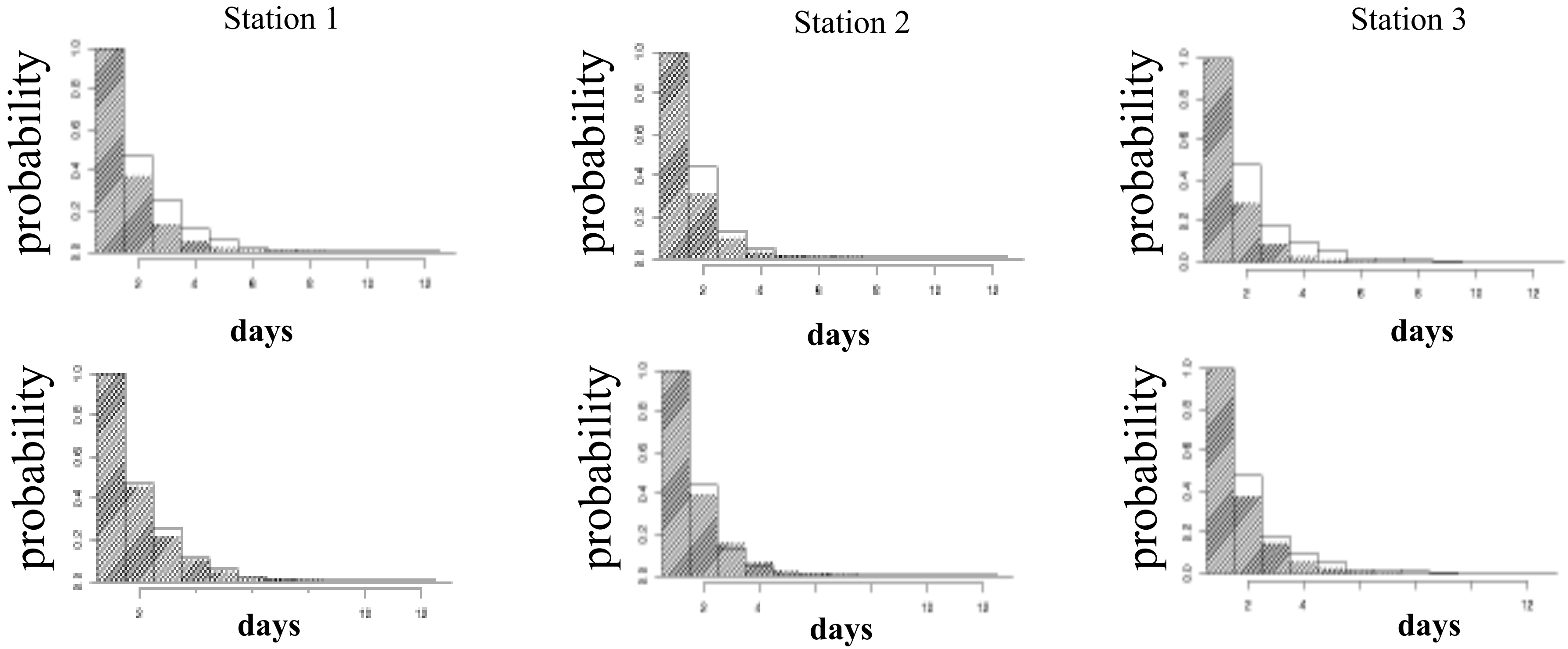


From

# Wet spells probabilities

**Circulation patterns**

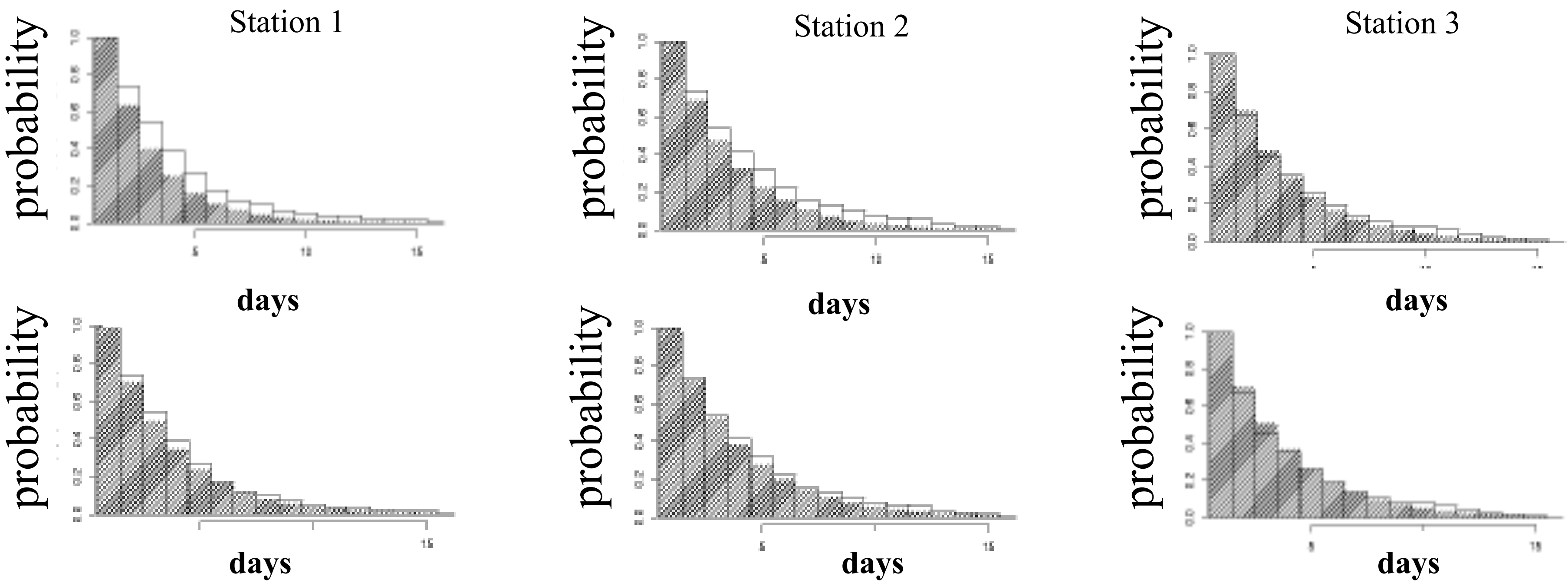
**Precipitation patterns**




# Dry spells probabilities

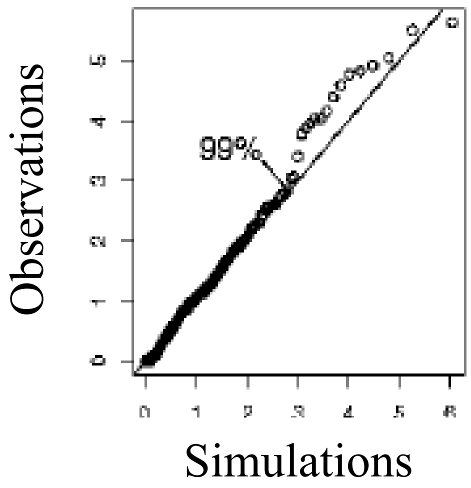
**Circulation patterns**

**Precipitation patterns**

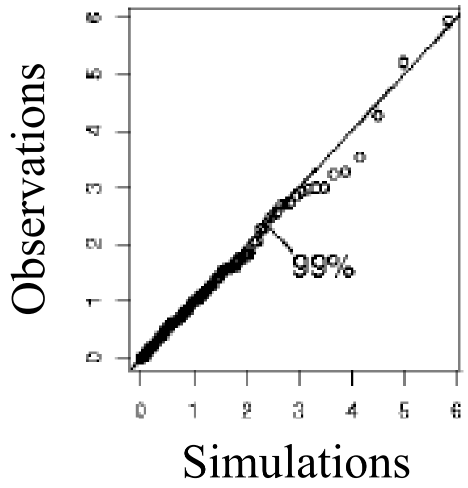


# Gamma: good… but not always enough !!

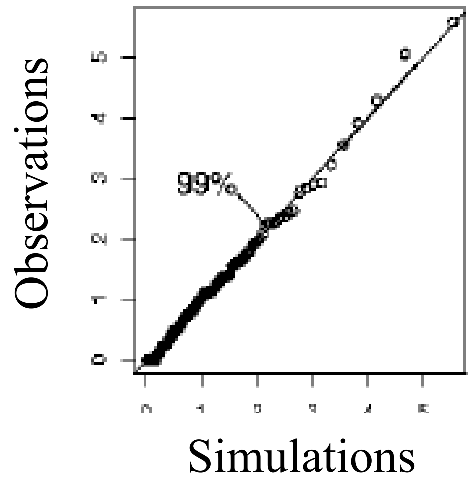
Station 1




Station 2

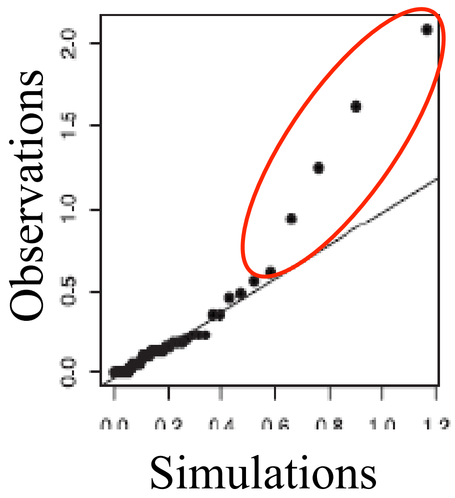



Station 3

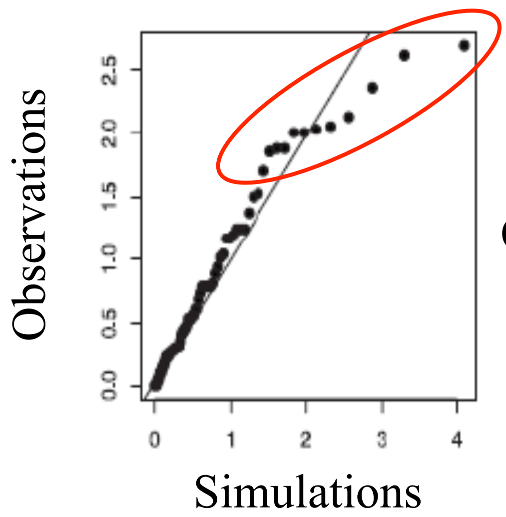



**Under-estimation** of the extremes

Station 4



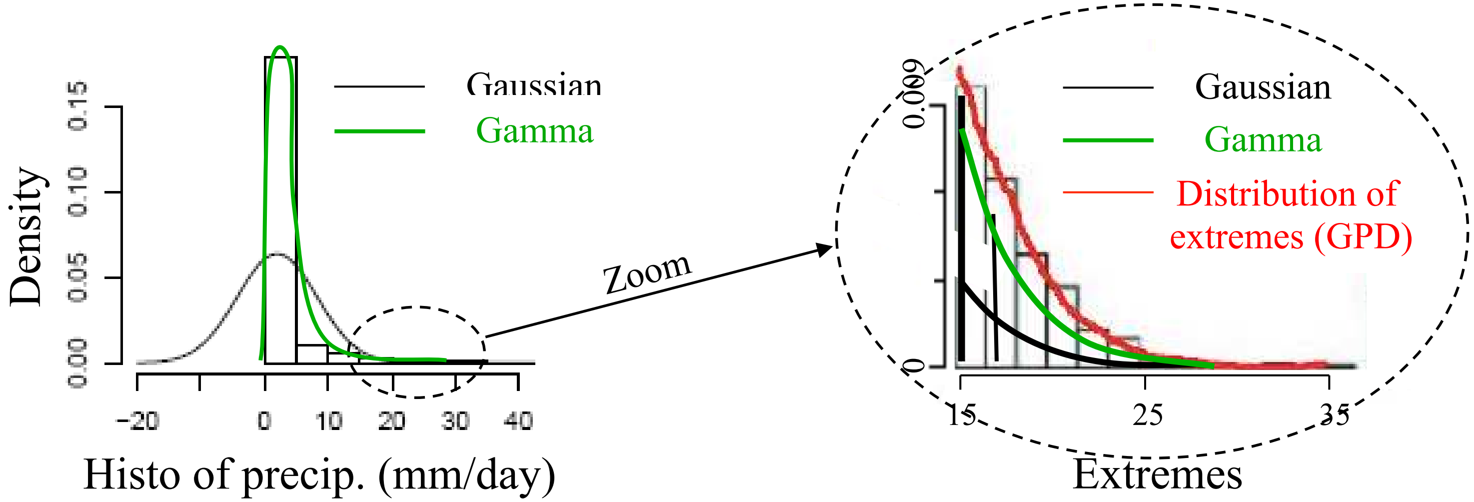

Station 5




**Over-estimation** of the extremes

# How to model local precipitation distributions?

- Classical distributions – for the "main" values
  - "*Gamma*" or "*log-normal*" distributions
  - Model the most current values, i.e. extremes badly represented




- Extreme Value Distributions – for the extremes
  - "*Generalized Extreme Value*" (**GEV**) distribution for maxima
  - "*Generalized Pareto Distribution*" (**GPD**) for peaks over threshold
  - Do NOT characterize precipitation below a given threshold

# Our statistical model: Merging classical and EV distributions

- Based on Frigessi et al. (2002):

"Dynamic mixture model for unsupervised tail estimation without threshold"

$$G(r_t|\beta) = c_\beta\Big[\big(1 - w(r_t|m,\tau)\big)\underbrace{\Gamma(r_t|\gamma,\lambda)}_{\text{Gamma pdf}} + w(r_t|m,\tau)\underbrace{GPD(r_t|\xi,\sigma,u=0)}_{\text{Generalized Pareto Distribution (GPD) pdf}}\Big]$$

weight

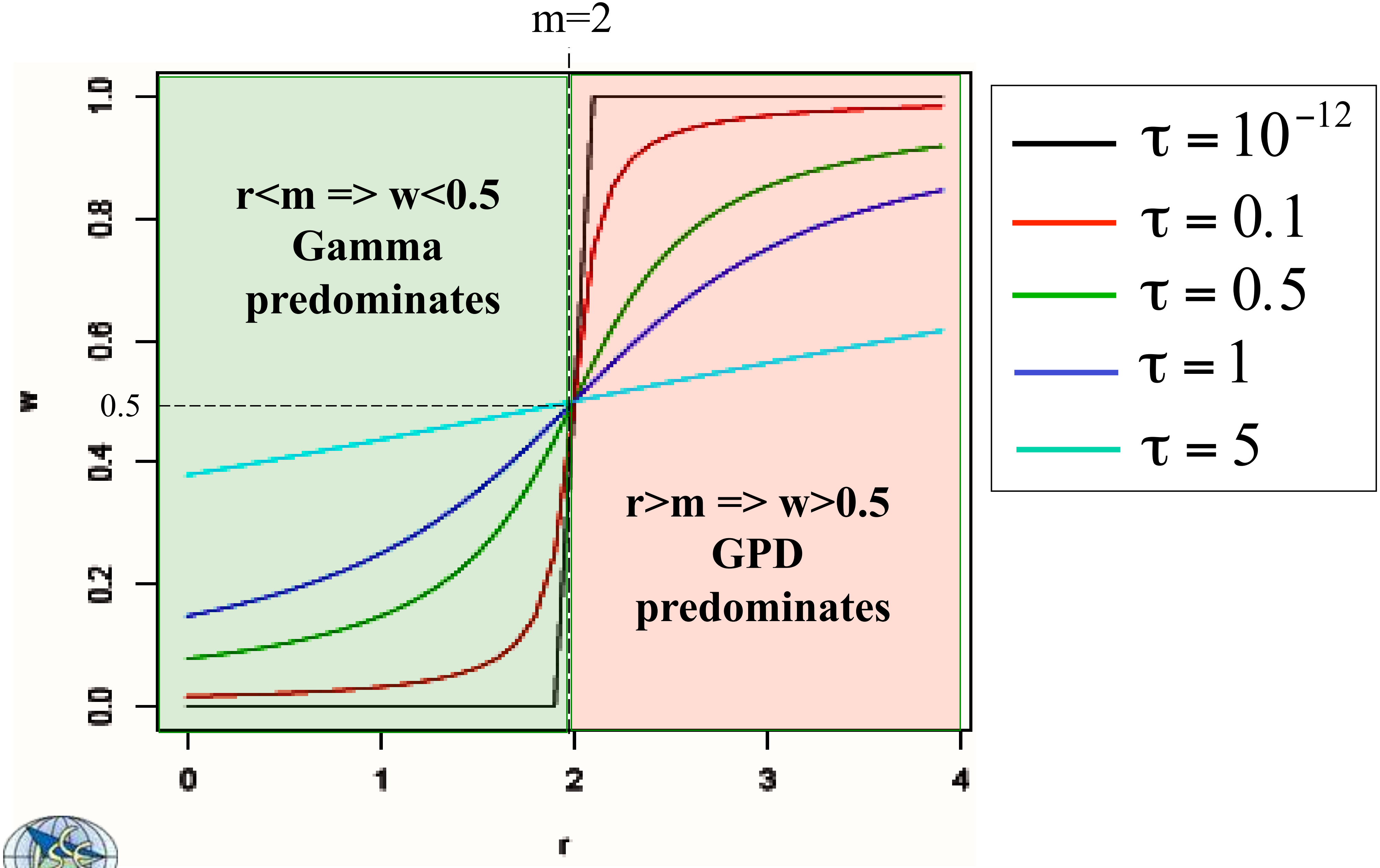
with

$$w(r_t|m,\tau) = \frac{1}{2} + \frac{1}{\pi}\arctan\left(\frac{r_t - m}{\tau}\right)$$

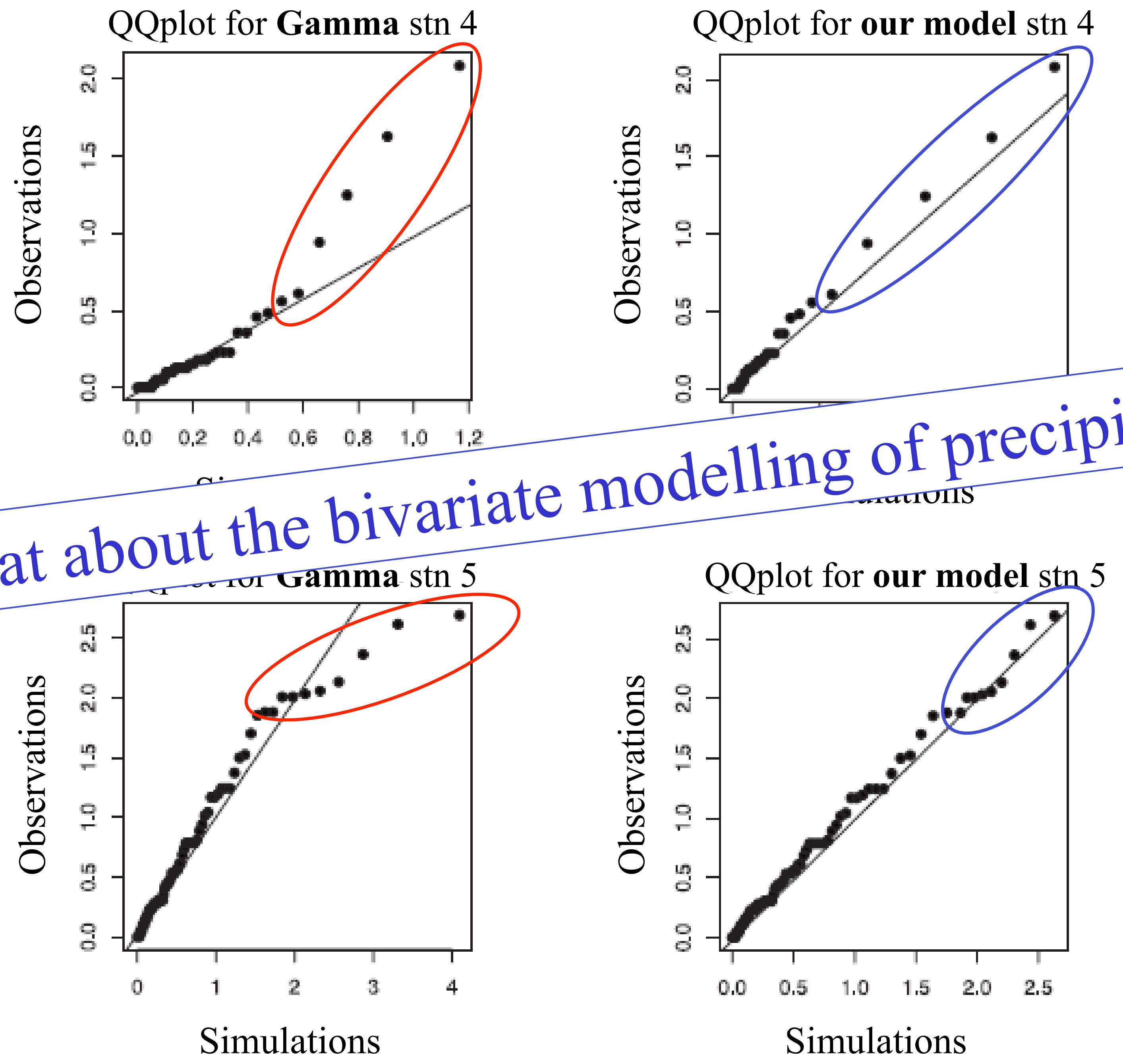
$m$: Value where transition from $\Gamma$ to GPD

$\tau$: Transition rate

$$w\left(r_t \middle| m,\tau\right)=\frac{1}{2}+\frac{1}{\pi}\arctan\left(\frac{r_t-m}{\tau}\right)$$



# Illustration on two stations




What about the bivariate modelling of precipitation?

# Why is it difficult to model local precipitation in "Mediterranean"?

- Physical/Geographical constraints:
  - Mediterranean sea effect
  - Chiseled/indented coasts
  - Mountain effect (Alpe
  - Rhone valley
  - Strong urbanization


Var region
June, 15th, 2010
400 mm in one day

- Consequences:
  - Important temporal and **spatial variability**
  - Frequent **extreme events** of precipitation

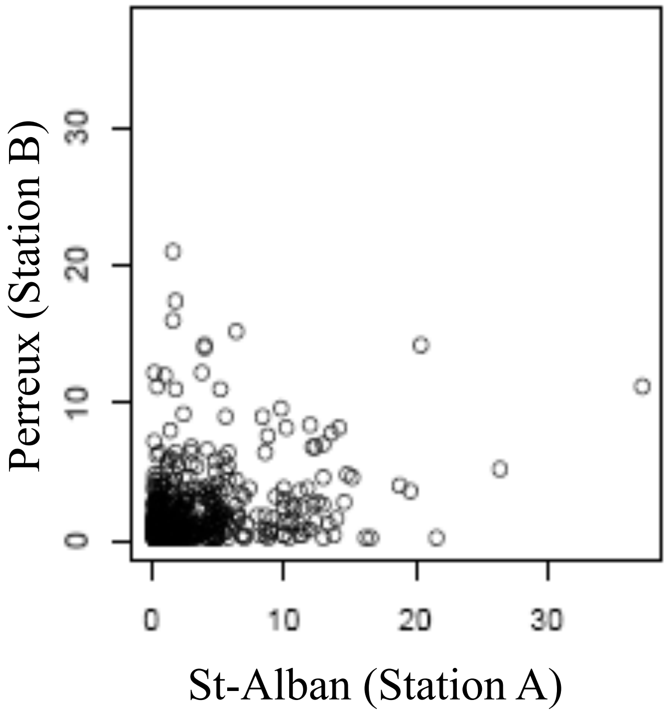
    Strong intensities (flooding) / long wet or dry spells (droughts)

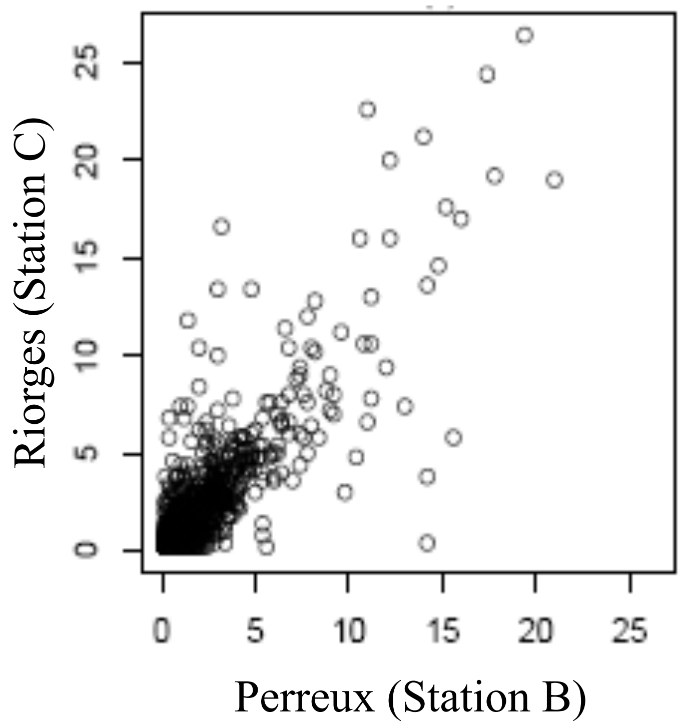
# Dependencies:

## Why is this so important for precipitation?

Taking **spatial coherence/dependencies** into account:

- Necessary for runoff modelling
- Necessary for flooding/droughts modelling (overall tail dep.)
- Necessary for impacts studies, etc.



# Extension of our model in 2D…

- For small and medium values: bivariate Gamma (Cheriyan and Ramabhadran)
- For extreme values: Pickand's coordinates (R, Z) from ($R_1$, $R_2$)

$$R = R_1 + R_2 \quad \text{and} \quad Z = \frac{R_2}{R_1 + R_2}$$

2D-Gamma pdf (for the "core")

$$f_\Theta(r_1, r_2) = c_\Theta \left[ \left(1 - w(R|m,\tau)\right) \Gamma_{2d}(r_1, r_2|\theta) + w(R|m,\tau)\, gpd(R|\xi,\sigma)\, \beta(Z|a,b) \right]$$

Normalizing constant

Extreme intensities pdf

Dependencies pdf

# Example of tail dependencies

Independence ......... Dependence

$\beta(.|a,b)$

$f_\Theta(r_1,r_2)$

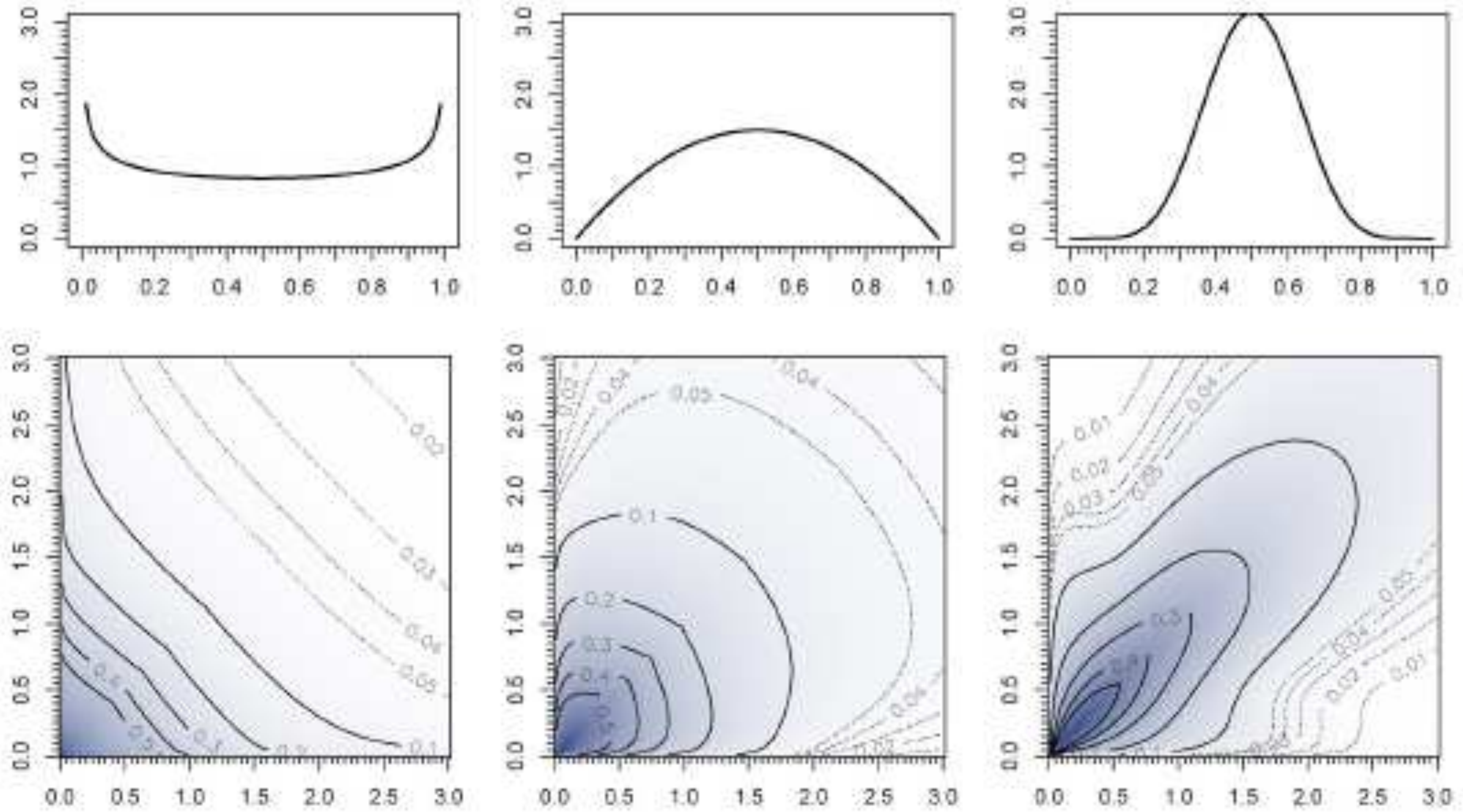
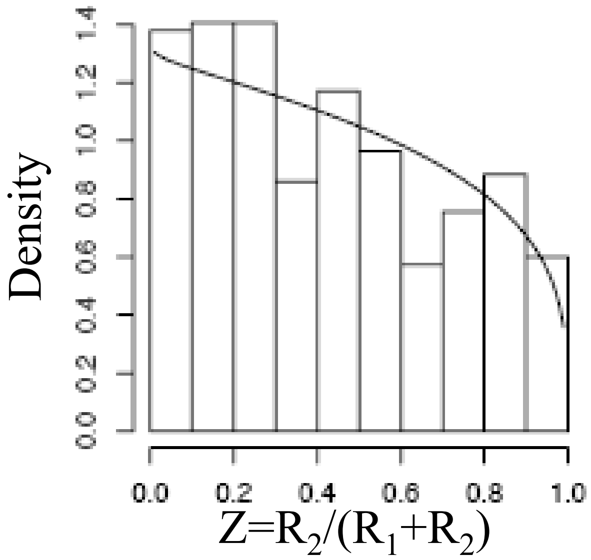


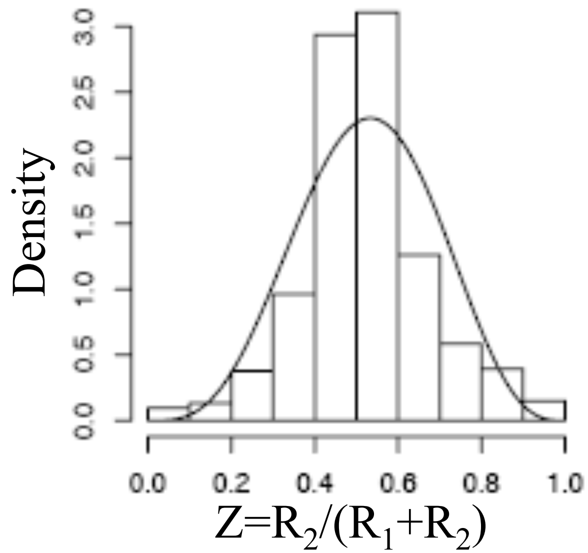

**Figure:** $f_{B(a,b)}(\omega)$ (top) and $f_{mix}(r_1, r_2)$ for three different parameter sets

# Results

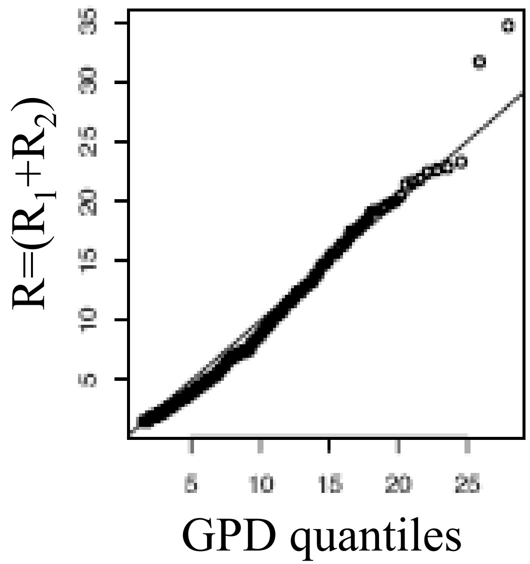
Stations A and B (distant) | Stations B and C (close)

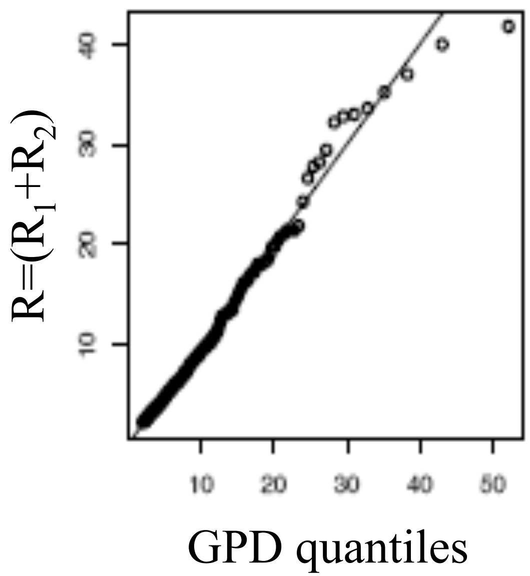
Beta fit for Z




GPD QQplots for R



# Conclusions

- NSWT: Regional. prec. patterns better than classical circulation patterns
- Our 1D-stochastic mixture model:
  - Can describe the whole precipitation range (small/medium precipitation AND extreme values)
  - No threshold selection needed due to functional weight *w(.)*
  - Shows better performance than "classical" models on simulated and real data
  - Proved successful in a downscaling context
    - M. Vrac, M. Stein, and K. Hayhoe (2007) Statistical downscaling of precipitation through nonhomogeneous stochastic weather typing, *Climate Research*, 34: 169-184, doi: 10.3354/cr00696
    - M. Vrac, and P. Naveau (2007) Stochastic downscaling of precipitation: From dry events to heavy rainfalls, *Water Resources Research*, 43, W07402, doi:10.1029/2006WR005308.
- Our 2D-extension:
  - Captures core and tail dependencies
  - Estimation and simulation processes developed
    - M. Vrac, P. Naveau, and P. Drobinski (2007). Modeling pairwise dependencies in precipitation intensities, submitted to *Non-linear Processes in Geophysics*, 14, 789-797

# Two other SDMs for/with extremes (1)

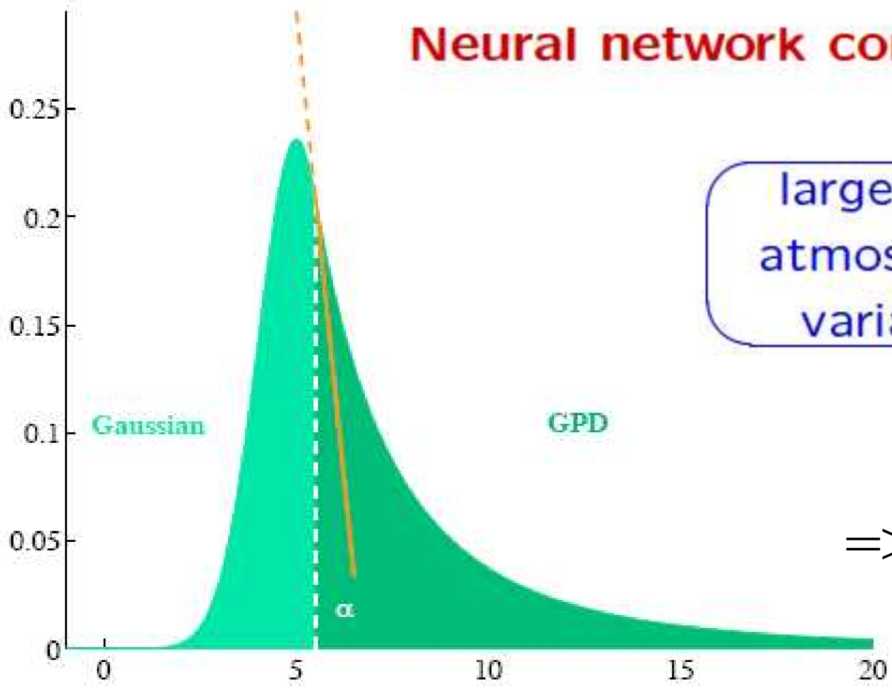
- Neural Network Conditional Mixture Models (Carreau and Vrac, 2011)
  - e.g., with "hybrid Pareto" pdf to take extremes into account

**Neural network transfer function:**

large-scale atmospheric variables ⟹ neural network ⟹ estimate of precipitation

**Neural network conditional mixture** (*weather generator*):

large-scale atmospheric variables ⟹ neural network ⟹ mixture parameters




$\alpha$ is the junction point

=> One mixture model per day

# Two other SDMs for/with extremes (2)

- Cumulative Distribution Function – Transform (CDF-t) (Michelangeli, Vrac, and Loukos, 2009, GRL):
  - $F_{Sp}$ Verifies eq.(1) by definition
  - Assumption: Eq. (2) remains valid in the futur

| | Present | Futur |
|---|---|---|
| GCM | $F_{Gp}$ ↓ $T$ | $F_{Gf}$ ↓ $T$ |
| Station | $F_{Sp}$ | ? $F_{Sf}$ ? |

$$T\left(F_{Gp}(x)\right) = F_{Sp}(x) \quad (1)$$

$$\text{Let } x = F_{Gp}^{-1}(u) \quad \text{with } u \in [0,1]$$

$$\Rightarrow \boxed{T(u) = F_{Sp}\left(F_{Gp}^{-1}(u)\right)} \quad (2)$$

$$F_{Sf}(x) = T\left(F_{Gf}(x)\right) \quad \Leftrightarrow \quad \boxed{F_{Sf}(x) = F_{Sp}\left(F_{Gp}^{-1}\left(F_{Gf}(x)\right)\right)} \quad (3)$$

  - XCDF-t: for GPD-extremes (Kallache, Vrac, Michelangeli, Naveau, 2011, JGR)

# Two other SDMs for/with extremes (2)

- XCDF-t: for GPD-extremes (Kallache, Vrac, Michelangeli, Naveau, 2011, JGR)

  F's are GPD's

$$F_{Rf}(y - c_{fac}) = 1 - \left(1 + \frac{\xi_{Rp}}{\xi_{Gp}}\frac{\sigma_{Gp}}{\sigma_{Rp}}\left[\left(1 + \frac{\xi_{Gf}}{\sigma_{Gf}} y\right)^{\xi_{Gp}/\xi_{Gf}} - 1\right]\right)^{-(1/\xi_{Rp})},$$

which is defined on $\{y : y > -c_{fac}\}$, $y = x - u_{Gf}$ and $x \sim F_{Gf}$.

Set $\xi_{Gf} = \xi_{Gp}$, then $F_{Rf}$

$$F_{Rf}(y) = 1 - \left(1 + \frac{\xi_{Rp}}{\sigma_{Gf}}\frac{\sigma_{Gp}}{\sigma_{Rp}}(y + c_{fac})\right)^{-(1/\xi_{Rp})},$$

- Inclusion of covariate information:

$$\begin{aligned} F(\cdot) &= GPD(\sigma, \xi) \\ \sigma_t &= exp(a_0 + a_1 \mathrm{cov}_1^t + \ldots + a_n \mathrm{cov}_n^t) \end{aligned}$$

  - The parameters may have different link functions,
  - The covariates vary with time, therefore the CDF transform results in one cdf $F^t_{Rf}(\cdot)$ per time step.
  - Conditional on a covariate state, the local and global variables may evolve differently in the future time period.

# Perspectives (DS & extremes)

- New SDM approaches still on development
  - ✓ e.g., “probabilistic” downscaling
- Only few SDMs are used in the IPCC AR4 (5?)
  - ✓ need more since RCMs and SDMs have both pros and cons
- Needs for multivariate model(s)
  - ✓ especially for dependence of extreme
- Spatial models: DS even at locations where no data
  - ✓ especially for extremes (maxima and excesses)
- Impacts studies
  - ✓ Properly estimate future costs/impacts for environment & society

Mathieu Vrac

mathieu.vrac@lsce.ipsl.fr

http://www.lsce.ipsl.fr/Pisp/58/mathieu.vrac.html