Assessing the impacts of the choice of spatial dependence structure for flood-risk rainfall

Julie Carreau

julie.carreau@univ-montp2.fr

Jean-Noël Bacro & Christophe Bouvier & Luc Neppel & Gwladys Toulemonde

IRD, HydroSciences Montpellier, France

Avignon, September 17th

Flash Flood

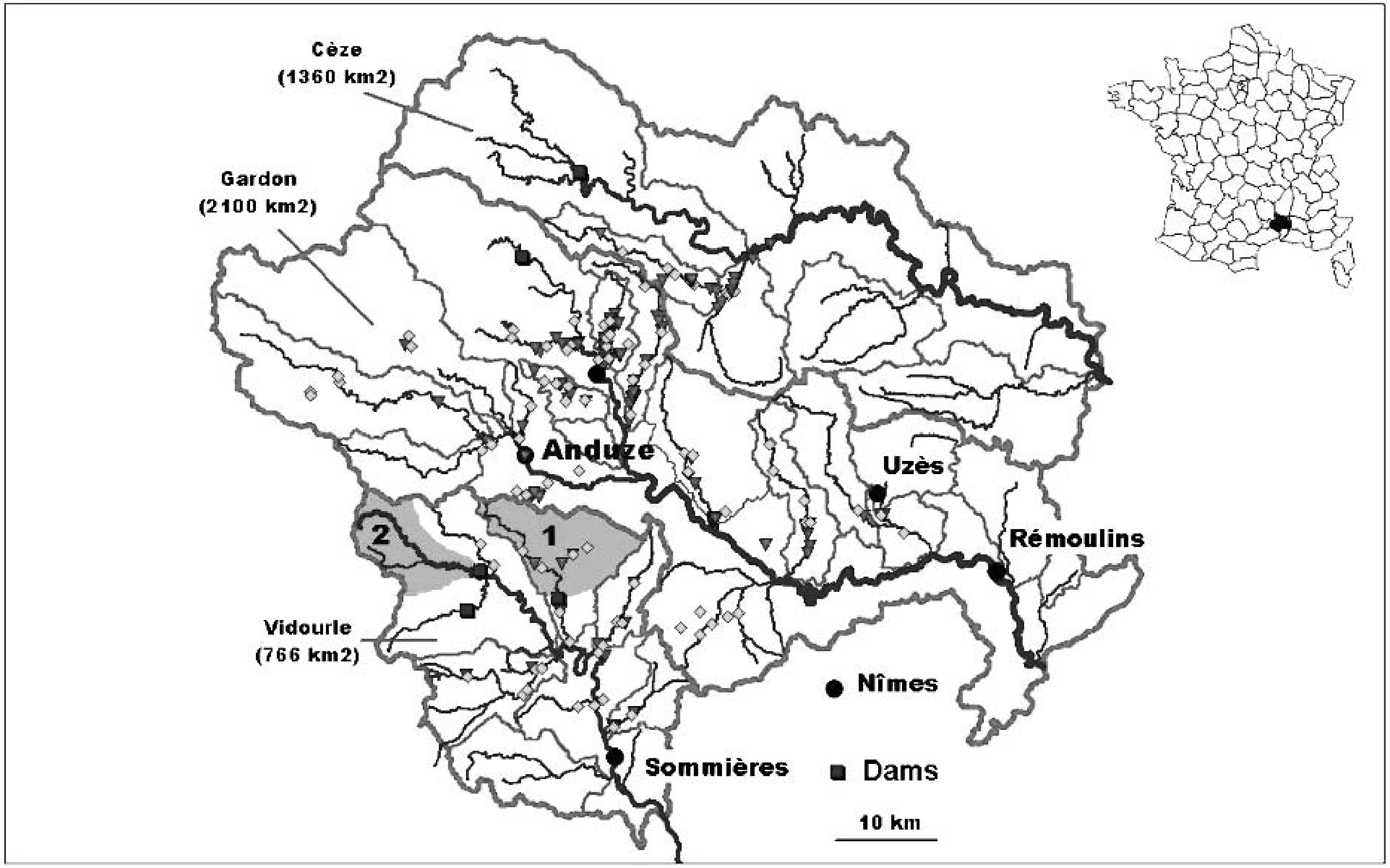
Flood attenuation dam at “La Rouvière”, Gard Dept., France

Regular day - May 2011

September 9th 2002

- 687 mm of rainfall in 24h : 30 % - 50 % of the annual rainfall
- Runoff of 830 m^3/s : control runoff 1 m^3/s

Anduze Area



From Delrieu et al (2004) "The Catastrophic Flash-Flood Event of 8-9 September 2002 in the Gard Region, France: A First Case Study for the Cévennes-Vivarais Mediterranean Hydrometeorological Observatory", Journal of Hydrometeorology

Flood Risk Measures = Return levels

☞ High level quantiles of river runoff termed **return levels**

For example, the 100-year return level is the river runoff Q which is expected to be exceeded on average once every $T = 100$ years.

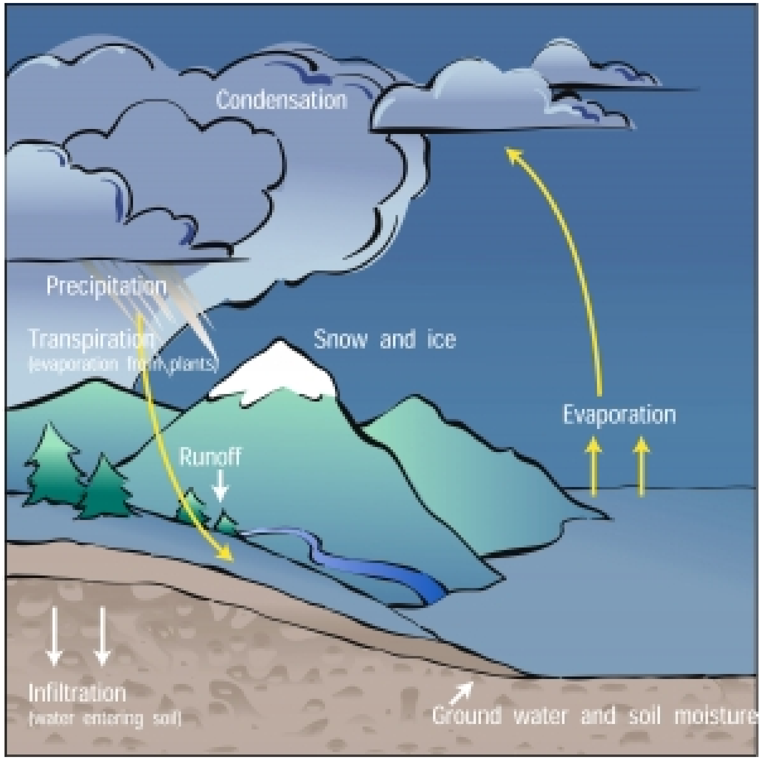
$$P(Q > R_T) = \frac{1}{T}$$

Runoff is difficult to measure : deduced from the water level

For small catchments, return levels are often estimated from **surrogate runoff** simulated from a **rainfall-runoff model**

Limnimeter

Rainfall-Runoff Models



Hydrological cycle ©UCAR

Seek to reproduce the hydrological cycle

Conceptual models : simplified modelling with few parameters (2-10)

Main input : **rainfall**

Exploit **spatial** information of rainfall

Rainfall series are often too short :
stochastic rainfall generators provide surrogate rainfall series

Stochastic Rainfall Generators

Rainfall intermittency

Meta-Gaussian models :

Gaussian dependence structure with non-Gaussian marginals

$Z = \psi(Y) \quad Y \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma), \quad \psi(\cdot)$ is non-decreasing monotonic

Z has an atom at 0 (no-rain) and a continuous part (rain)

Vischel T. et al. (2009) " Conditional simulations schemes of rain fields and their applications to rainfall-runoff modeling studies in the Sahel" J. of Hydrology 375
Bouvier C. et al. (2003) "Generating rainfall fields using principal components decomposition of the covariance matrix: a case study in Mexico City" J. of Hydrology 278
Guillot G. (1999) "Approximation of Sahelian rainfall fields with meta-Gaussian random functions" Sto. Env. Res. & Risk Ass. 13

Indicator function :

Either with probit (truncated Gaussian) or logistic (binomial) models

Barancourt C. & Creutin J. D. (1992) "A method for delineating and estimating rainfall fields" WRR 28
Kleiber W. et al. (2011) "Daily spatiotemporal precipitation simulation using latent and transformed Gaussian processes" WRR 48
Hughes J. P. et al. (1999) "A non-homogeneous hidden Markov model for precipitation occurrence" Appl. Statist. 48 Part 1

Rainfall Inhomogeneity

Rainfall patterns : based on rainfall features and/or on atmospheric circulation

☞ mixture models

Thompson C. S. et al. (2007) "Fitting a multisite daily rainfall model to New Zealand data" J. of Hydrology 340
Leblois E. & Creutin J. D. (2013) "Space-time simulation of intermittent rainfall with prescribed advection field : adaptation of the turning band method" WRR 49
Bellone E. et al. (2000) "A hidden Markov model for downscaling synoptic atmospheric patterns to precipitation amounts" Climate Research 15
Garavaglia F. et al. (2010) "Introducing a rainfall compound distribution model based on weather patterns sub-sampling" HESS 14

Parameters vary with covariates

☞ conditional distributions

Kleiber W. et al. (2011) "Daily spatiotemporal precipitation simulation using latent and transformed Gaussian processes" WRR 48
Chandler R. E. & Wheater H. S. (2002) "Analysis of rainfall variability using generalized linear models: A case study from the west of Ireland" WRR 38

Flood-Risk Rainfall : rainfall which might lead to flooding

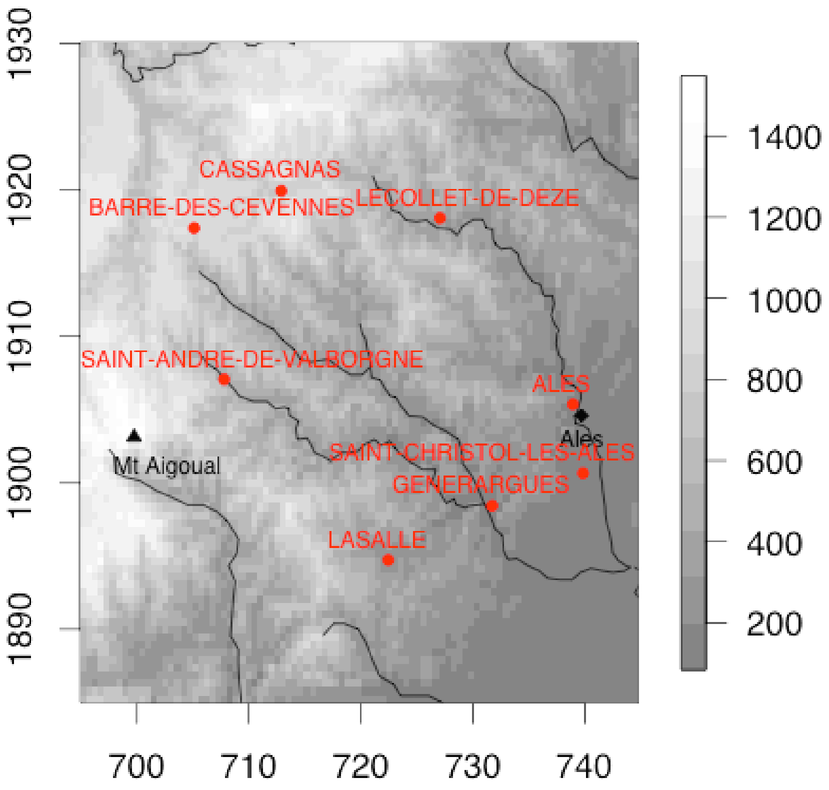
A single rainfall pattern : significant spatial average

- potentially no need to deal with intermittency
- no mixture or conditional modelling

Objectives and Data

☞Assess the impacts of the choice of **multivariate density models** as stochastic generator of rainfall which might lead to flooding

Evaluation criteria in terms of change in **rainfall** and **runoff** return levels



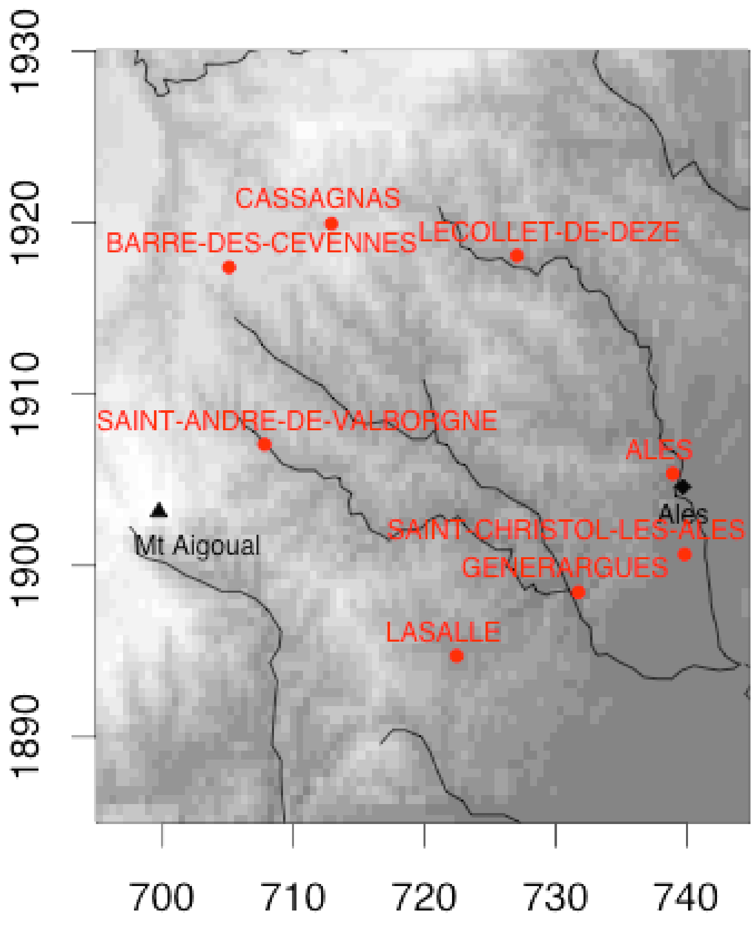
Eight daily raingauge stations

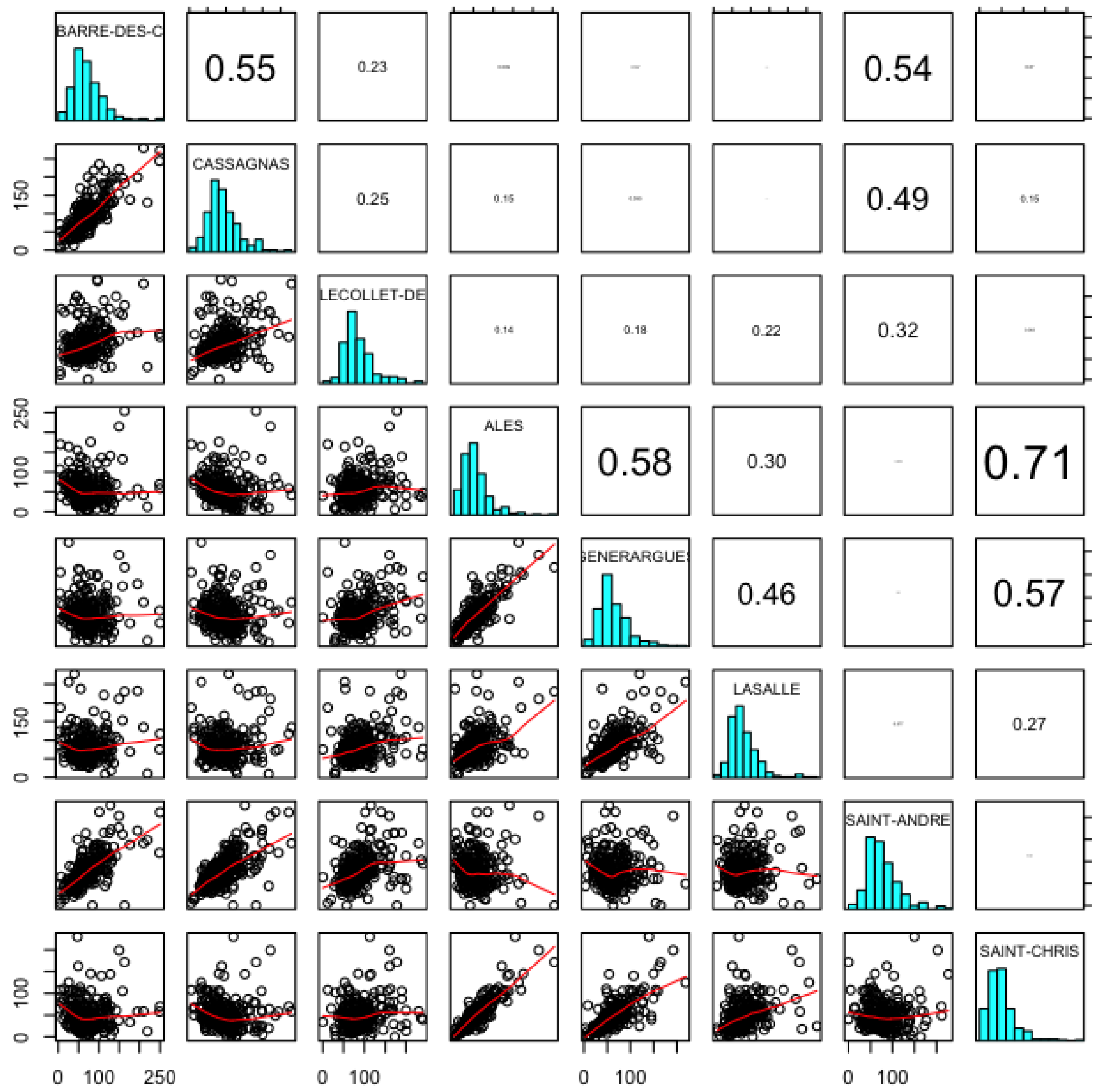
Period : 01/01/1958 - 12/31/2000
43 years or 15,706 days

Flood-risk rainfall :
daily spatial average above 50 mm
265 days less than 2%
only 4 zero obs. ↑ 0.2 mm

Assumption : no temporal dependence

Pairwise Exploratory Analysis : Kendall Correlation





☞ Both correlated and uncorrelated pairs

Pairwise Exploratory Analysis : χ-plot

Bivariate random vector (X, Y)

Extremal dependence measure

$$\chi = \lim_{u \to 1} P(F_Y(Y) > u | F_X(X) > u)$$

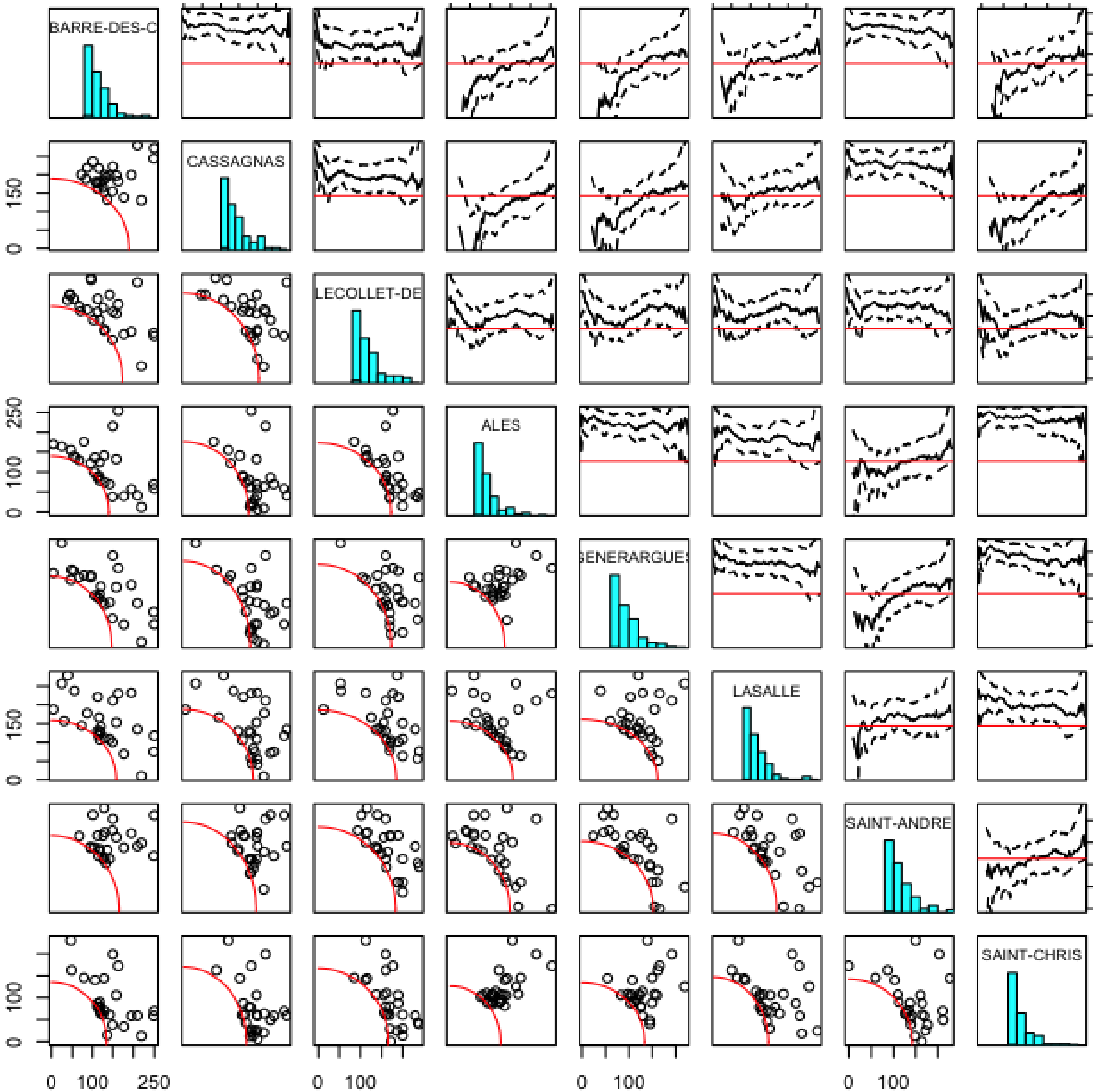
Asymptotic independence

$\chi = 0$

Asymptotic dependence

$0 < \chi \leq 1$

☞ Both A. independent and A. dependent pairs



Multivariate Density Models : Copulas

$\boldsymbol{X} = (X_1, \ldots, X_8)$ rainfall intensities such that $(X_1 + \cdots + X_8)/8 > 50\ mm$
Let $F_{X_1}(\cdot), \ldots, F_{X_8}(\cdot)$ be the marginal distribution functions

$$\underbrace{F_{\boldsymbol{X}}(x_1, \ldots, x_8)}_{\textit{joint distribution function}} = \underbrace{C(F_{X_1}(x_1), \ldots, F_{X_8}(x_8))}_{\textit{copula function}}$$

Margins

Either a Gamma or semi-parametric with GPD in the upper tail :

$$\widetilde{F}_{X_j}(x) = \begin{cases} \widehat{F}_{X_j}(x) & \text{if } x \leq u_{X_j} \\ 1 - \{1 - \widehat{F}_n(u_{X_j})\}\{1 + \xi_j (x - u_{X_j})/\sigma_j\}_+^{-1/\xi_j} & \text{if } x > u_{X_j} \end{cases}$$

where u_{X_j} is some large threshold for X_j

Copula functions

Either Gaussian (benchmark) or Student t

Multivariate Density Models : Skew Elliptical

Azzalini A. & Capitanio A. (2003) J. R. Statist. Soc.B. 65, Part 2, pp. 367-389

☞ Elliptical density multiplied by a skewing factor

Standard Skew Normal

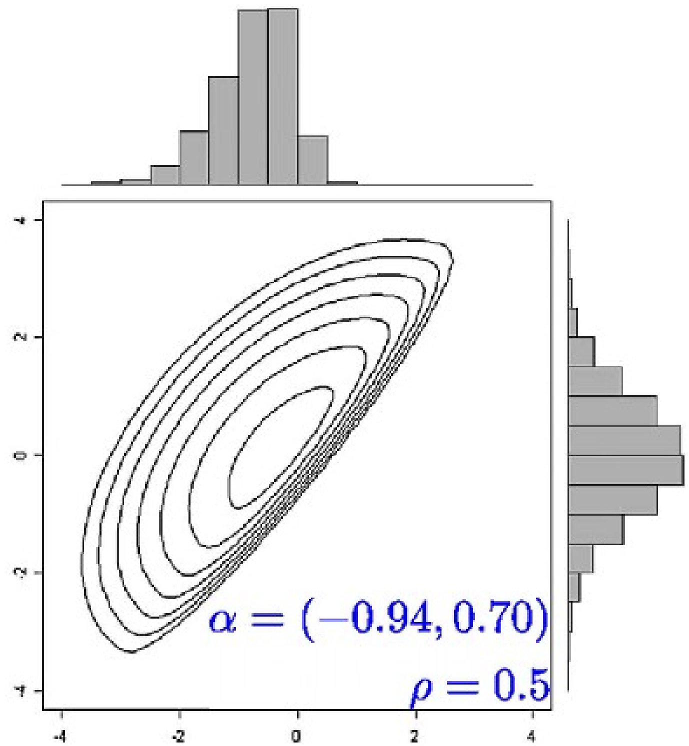
Density : $2\ \phi_d(\boldsymbol{x};\Omega)\ \Phi(\boldsymbol{\alpha}^T\boldsymbol{x}), \quad \boldsymbol{x} \in \mathbb{R}^d$

$\phi_d(\cdot;\boldsymbol{\Omega})$ d-dimensional standard Normal density with correlation matrix $\boldsymbol{\Omega}$

$\boldsymbol{\alpha} \in \mathbb{R}^d$ skewness parameter

$\boldsymbol{\alpha} = 0 \Rightarrow$ regular Normal density

Skew t defined as a transformation of the Skew Normal



Flecher et al. (2010) A stochastic weather generator for skewed data, WRR 46

Skew Elliptical Models

Multivariate Skew Normal and Skew t with fixed $\nu = 6$

Margins

Either the truncated skew distribution margins (univariate skew Normal or skew t)

Or transformed margins to standard Normal or standard t with $\nu = 6$

1. Transform to Uniform with the semi-parametric GPD

$$U = \widetilde{F}_{X_j}(X_j)$$

2. Transform to standard Normal or t with quantile functions :

$$\Phi^{(-1)}(U)$$

Multivariate Density Models : Extreme Value

$\boldsymbol{X}_i = (X_{i1}, \ldots, X_{i8})$ vector of rainfall intensities with $i = 1, \ldots, n$

Let the component-wise maxima be :

$$\boldsymbol{M}_n = (M_{n1}, \ldots, M_{n8}) \quad \text{where} \quad M_{nj} = \max_{i=1,\ldots,n} X_{ij}$$

Theorem

Assuming that X_j is standard Fréchet $\forall j$, then if

$$P\left(\boldsymbol{M}_n/n \leq \boldsymbol{x}\right) \xrightarrow{d} G(x) \quad \text{with } G \text{ non-degenerate}$$

Then G satisfies some specific constraints which define the class of multivariate extreme-value distributions.

Examples $\boldsymbol{x} = (x_1, \ldots, x_8) > 0$

Independence	$G(\boldsymbol{x}) = \exp\left\{-(x_1^{-1} + \cdots + x_8^{-1})\right\}$
Perfect dependence	$G(\boldsymbol{x}) = \exp\left\{-\max(x_1^{-1}, \ldots, x_8^{-1})\right\}$
Logistic/Gumbel	$G(\boldsymbol{x}) = \exp\left\{-(x_1^{-1/\beta} + \cdots + x_8^{-1/\beta})^{\beta}\right\}, \quad \beta \in (0,1)$

☞ **Pros**

Appropriate models for multivariate extremes defined as component-wise maxima

Also for random vectors above a high threshold : $\boldsymbol{X}|\boldsymbol{X} > \boldsymbol{u}$

☞ **Cons**

Not many models in dimension 8

Logistic/Gumbel is not flexible enough : only one parameter characterizes the dependence

Do not quite correspond to the problem expressed by hydrologists : rainfall do not have to be extreme everywhere to provoke flooding

Flood-risk rainfall problem

Let $\overline{X} = (X_1 + \cdots + X_8)/8$ be the spatial average.

Then the problem is more accurately expressed as : $\boldsymbol{X}|\overline{X} > u$

Conditional Model for Multivariate Extremes

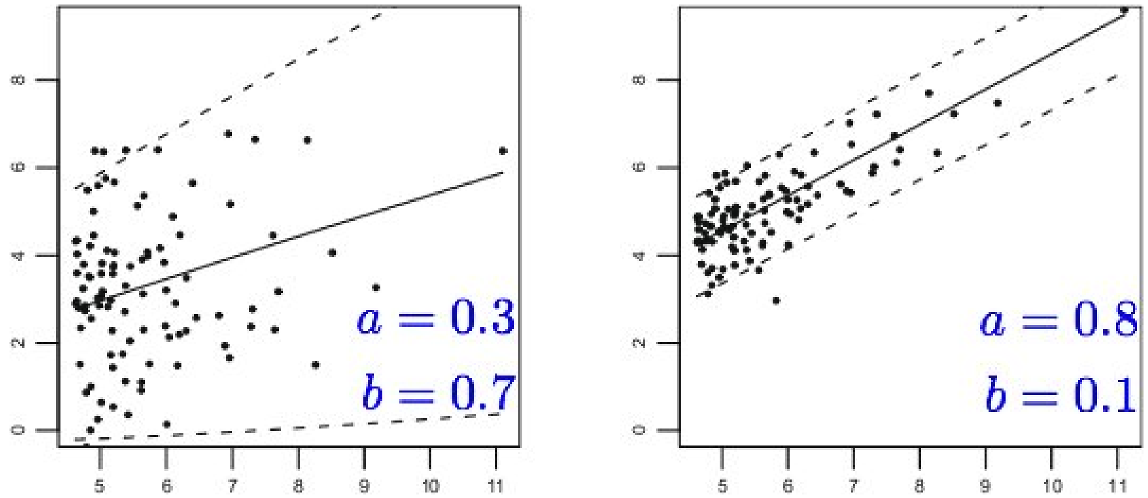
Heffernan & Tawn (2004) J. R. Statist. Soc. B 66, Part 3 pp. 497-546

$\boldsymbol{X} = (X_1, \ldots, X_8)$ rainfall intensities, $\overline{X} = (X_1 + \cdots + X_8)/8$ spatial average

For non-negatively dependent variables, the conditional distribution of $\boldsymbol{X} | \overline{X} = x$ for large x is modelled as :

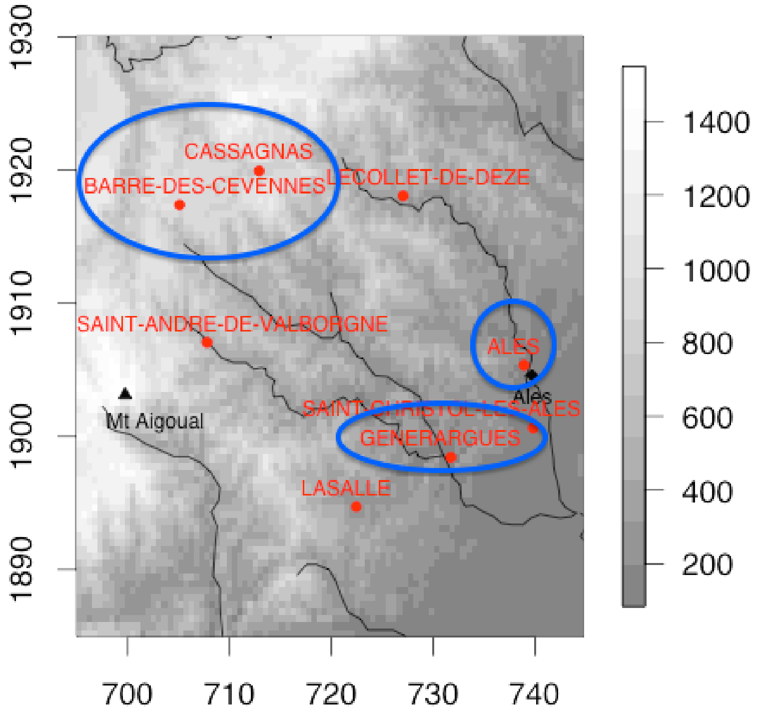
$$\boldsymbol{X} = \boldsymbol{a}\, x + x^{\boldsymbol{b}}\, \boldsymbol{Z} \qquad \text{with } \boldsymbol{Z} \in \mathbb{R}^8 \text{ and } \boldsymbol{Z} \perp \overline{X}$$

$\boldsymbol{a}$ overall strength of dependence
$\boldsymbol{b}$ how the dependence changes with $\overline{X}$
$\boldsymbol{Z}$ captures the residual dependence non-parametrically



Keef C. et al. (2009) J. of Hydrology 378

Comparative Analyses



Marginal Fit : Q-Q Plots

Either Gamma, semi-parametric GPD, Skew Normal or Skew t

Look at two stations at the top *900 m* and two in the valleys *140 m*

Spatial Average 5000 simulations

Return level curves : $\overline{X} = (X_1 + \cdots + X_8)/8$ times by each of the five models + PoT modelling

Conditional Probabilities 10000 simulations repeated 5000 times

$P(X_j > R_j(T) | X_i > R_i(T))$ where $R_j(T)$ is the T-year return level for X_j

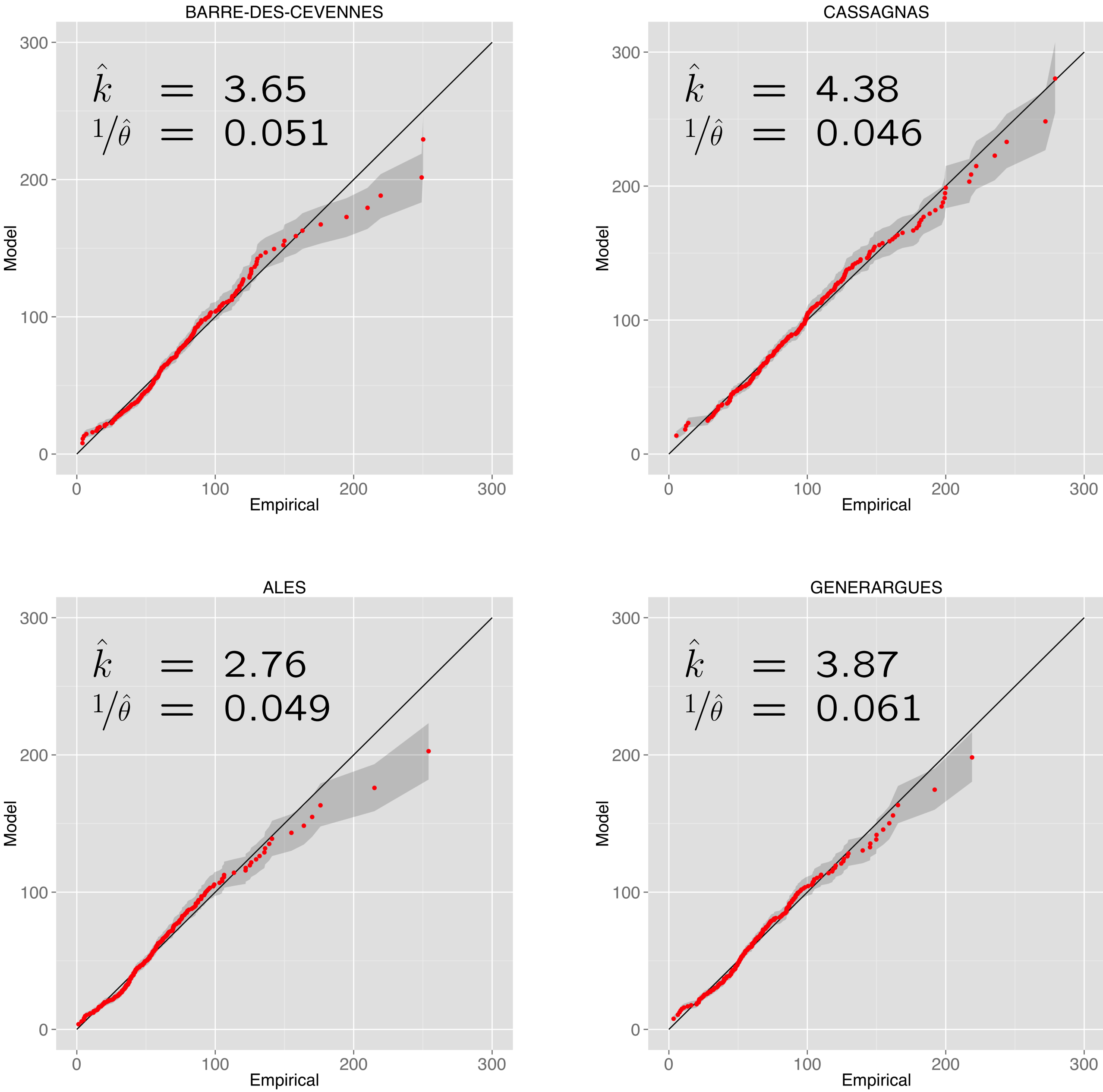
X_i is the *Barre-des-Cevennes* station

X_j is the *Cassagnas* or the *Generargues* station

Marginal Fit

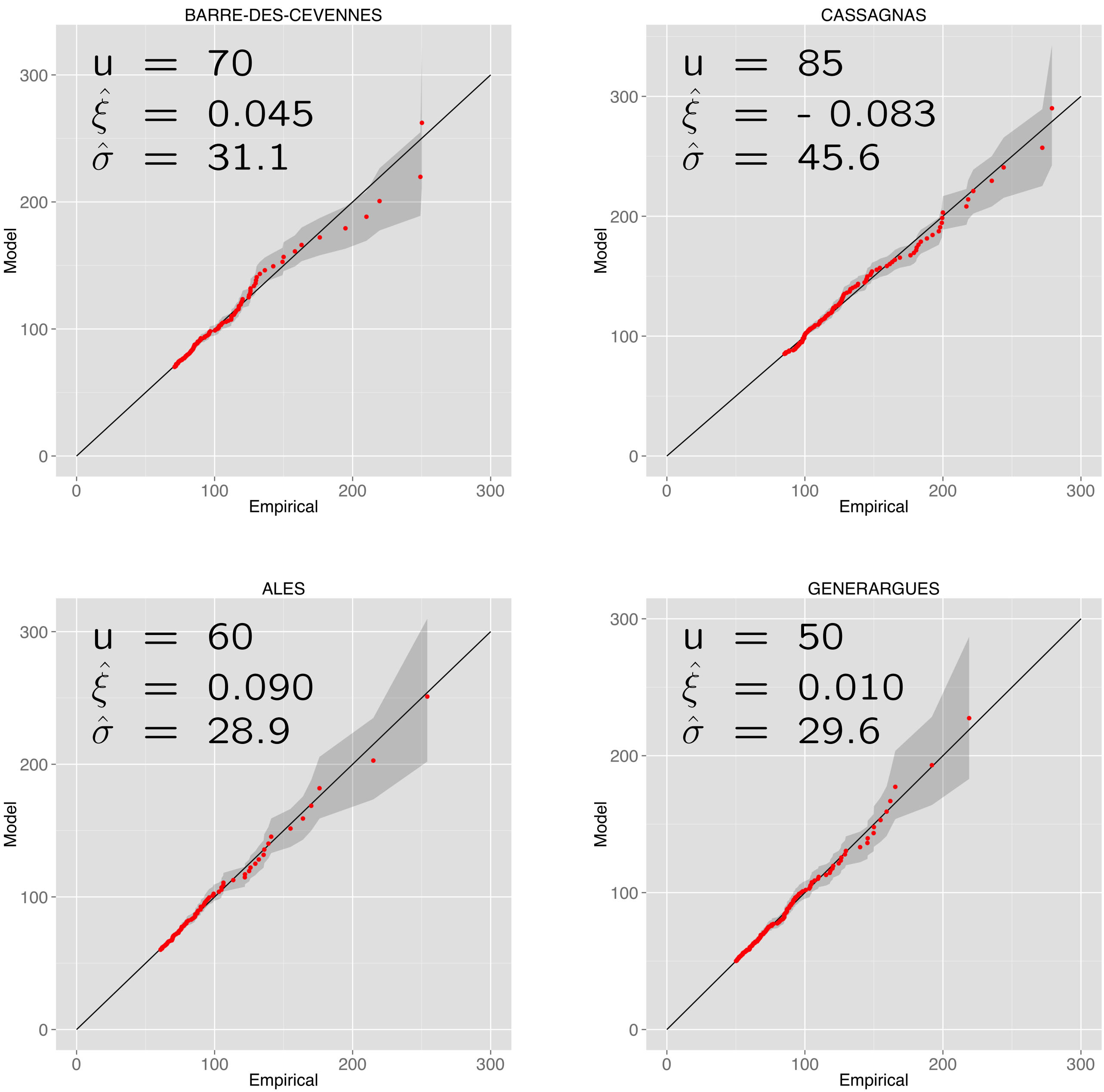
Gamma

Maximum likelihood



GPD

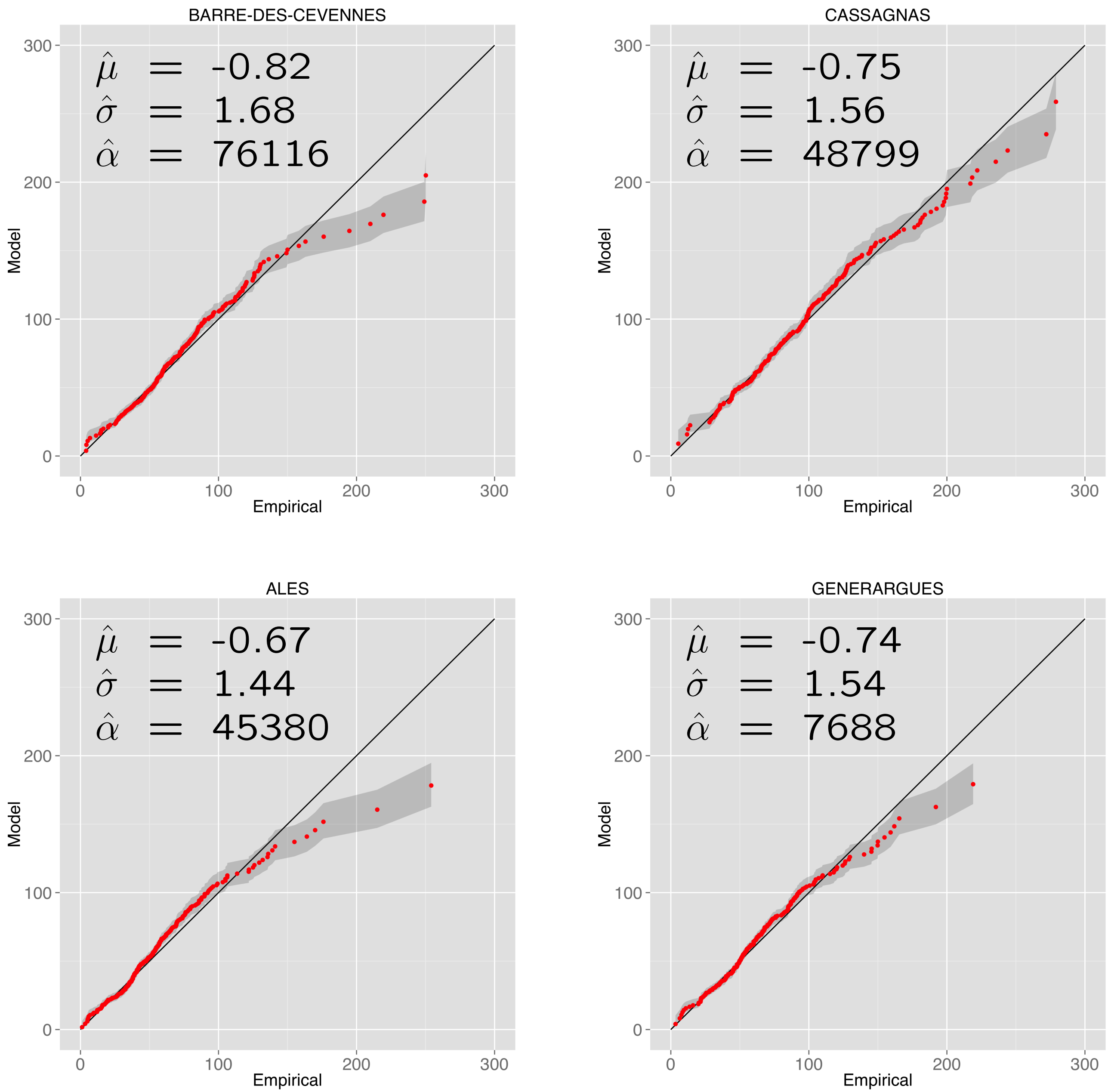
Probability weighted moments



Skew Normal

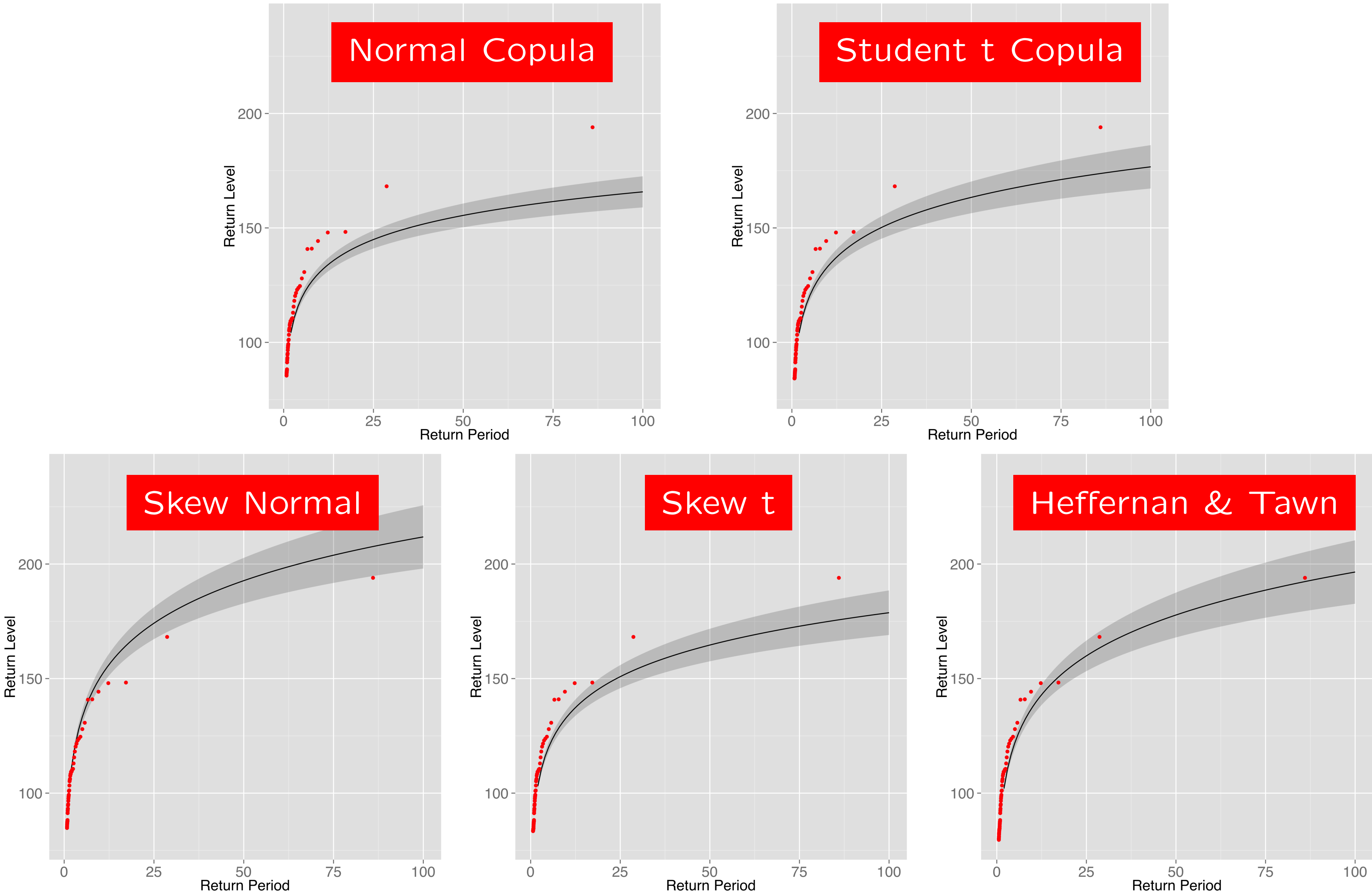
Maximum likelihood

Truncation below 0



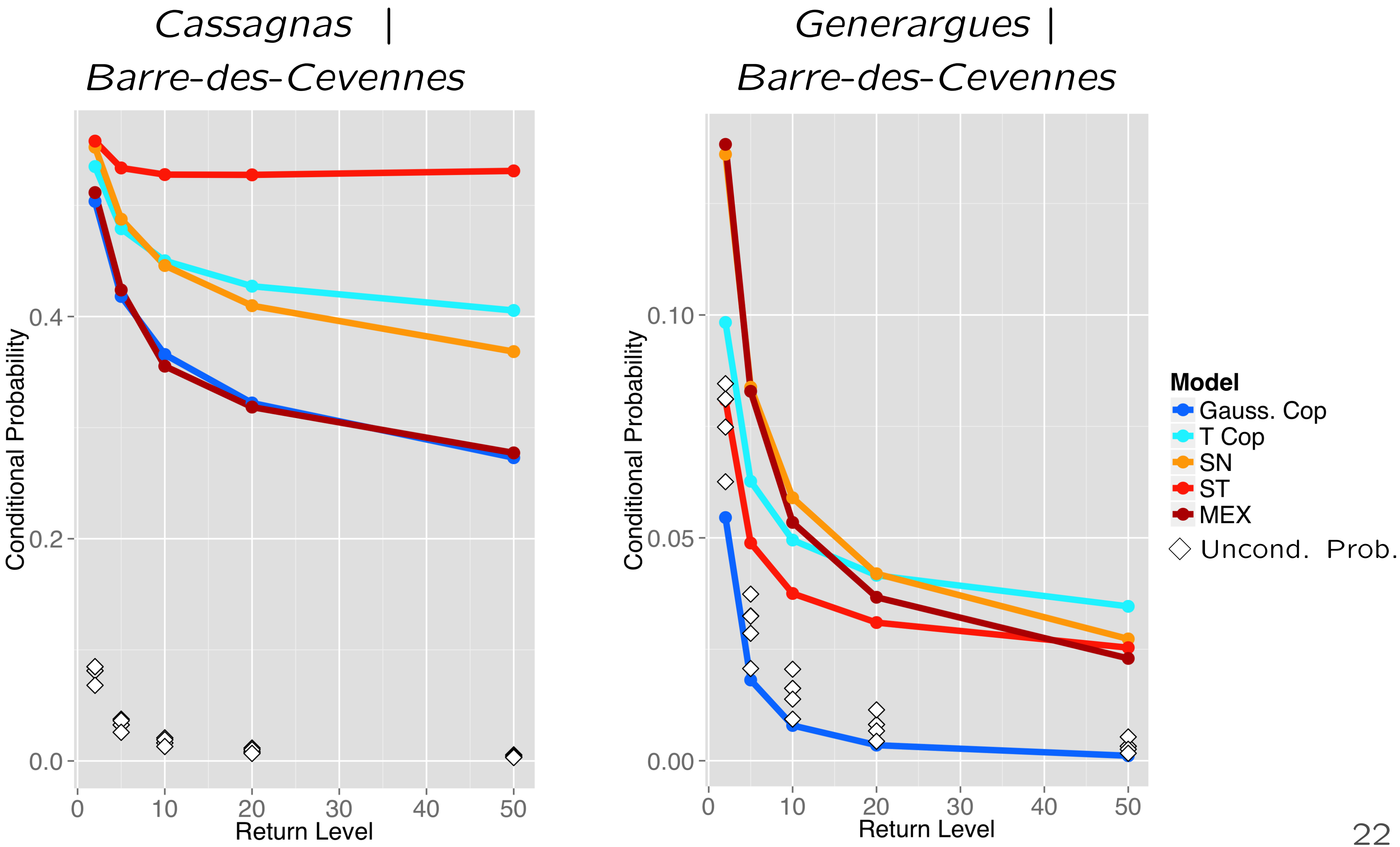
Spatial Fit : Spatial Average

Semi-parametric GPD margins $P(\overline{X} > x_T) = 1/T$



Spatial Fit : Conditional Probabilities

Semi-parametric GPD margins



Model Selection

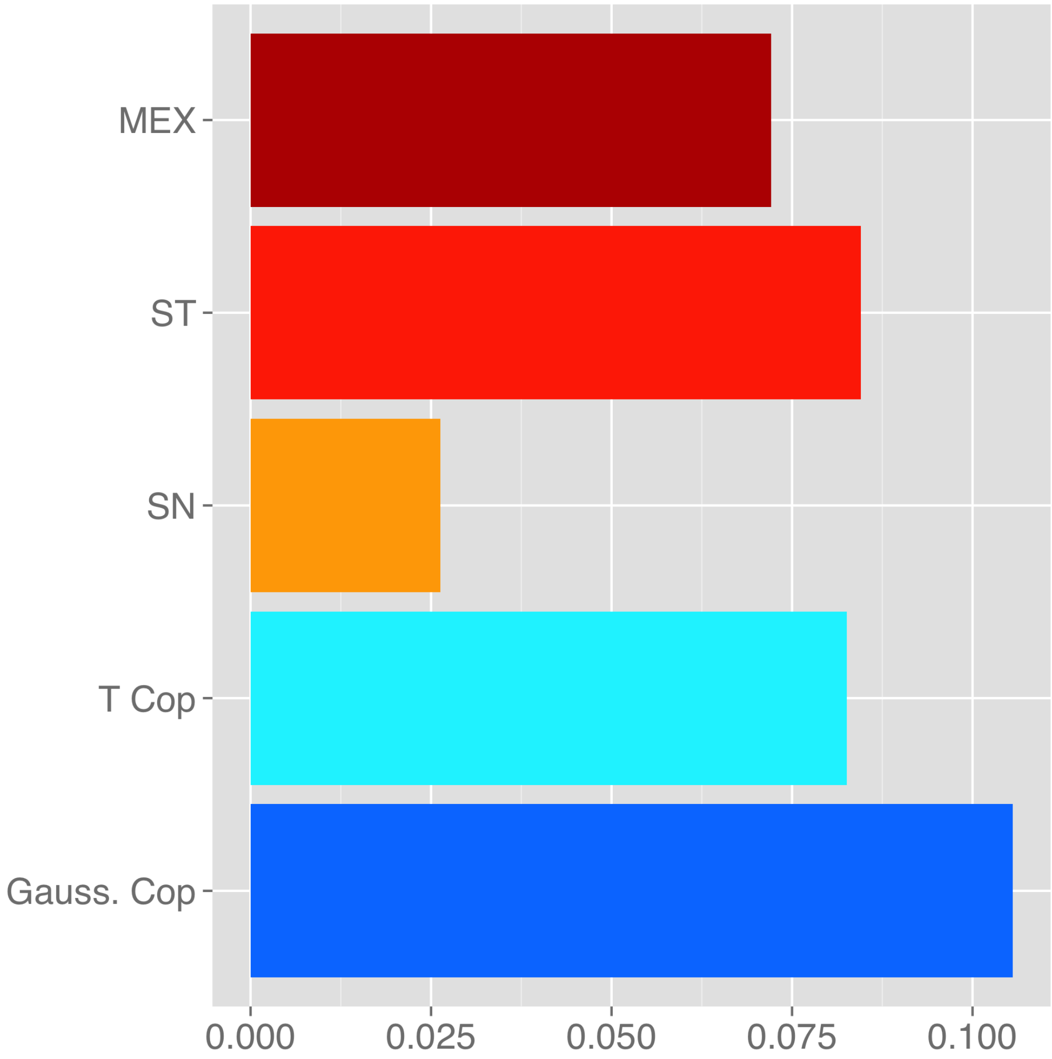
Semi-parametric GPD margins

Five dependence structures

Leave-one-out cross-validation

Cramer-von Mises statistic :

$$\sum_{j=1}^{265} \{C_n(\boldsymbol{U_j}) - C_{\theta_n}(\boldsymbol{U_j})\}^2$$



Conclusions and Perspectives

- spatial dependence structure has a strong influence on the spatial average of rainfall
- choice of margins too influences the spatial average
- a variety of multivariate models readily available in R packages
- MEV is not easily applicable because all variables are not extremes at the same time

- assess the impacts on runoff return levels
- spatial process modelling
- hourly time-step
- Anderson-Darling goodness-of-fit
- other models ?

R Librairies

- evd, extRemes, fExtremes, mex
- MASS
- copula
- sn
- ggplot2, fields

Thank you for your attention.