Multivariate quantiles with hydrological applications

Fateh Chebana, Taha B.M.J. Ouarda

INRS-ETE

Water2010, July 5-7, 2010

Outlines

- Hydrological frequency analysis (HFA)
- Multivariate quantiles in statistics
- in HFA
- Case study and illustrations
- Simulation results
- Conclusions

Hydrological frequency analysis

- Extreme hydrological events (floods, droughts, storms) have serious social and economic consequences

- The prediction of the frequency of an extreme event is of high importance

  reservoir management, construction of dams,…

- Frequency analysis of extreme events is a favourite tool in hydrology

Hydrological frequency analysis

○ HFA main aim: the study of the probability $\Pr[X \geq x_T]$ that an event x_T can be exceeded

○ The event x_T is associated to a return period T and corresponds to a ***quantile*** of exceedence probability

$$p = 1/T = \Pr\left(X \geq x_T\right) = 1 - F\left(x_T; \underline{\theta}\right) \Rightarrow x_T = F^{-1}(1 - 1/T; \underline{\theta})$$

$\underline{\theta}$: vecter of parameters to be estimated

F : distribution to be fitted

T : return period (10, 50, 100 years)

A return period : the mean passed time between the occurrences of 2 events $[X \geq x_T]$

Multivariate Freq. Anal.

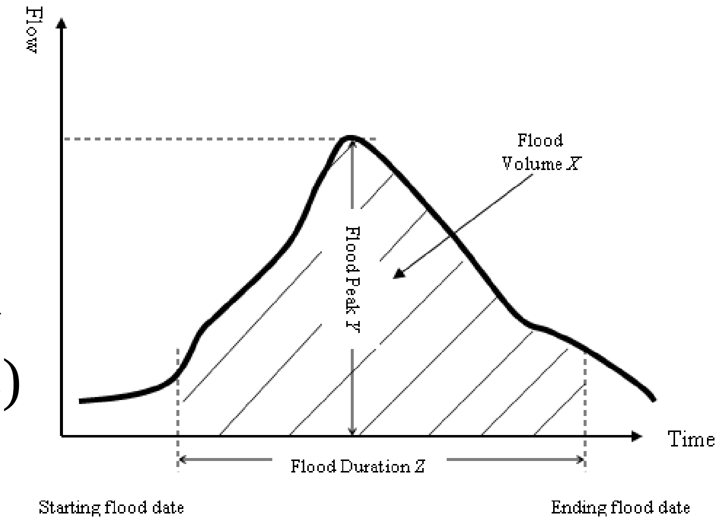
Floods are described by several correlated variables: volume, peak, duration
(e.g. Ashkar, 1980; Ouarda et al., 2000; Shiau, 2003; De Michele et al., 2005; Chebana and Ouarda, 2008)

In multivariate FA, 3 main elements are treated:

i) explaining the usefulness and importance of considering the multivariate framework

ii) modeling extreme events (fitting copula and marginal distributions, estimating parameters)

iii) defining bivariate return periods



However, the notion of multivariate quantile is not appropriately treated

Multivariate quantiles in statistics

- as inversions of mappings: for X~ F

  $$t \to -G_F(t) = E\left\{(X-t)/\|X-t\|\right\}$$ from R^d to R^d

- based on norm minimization:

  the value of θ that minimizes e.g. $E\left\{|X-\theta| + (2p-1)(X-\theta)\right\}$

- based on depth functions:

  by extending the notion of "order statistics"

- based on gradients (empirical):

  the value of θ that minimizes e.g. $D_1(\theta) = \sum_i \|X_i - \theta\|_1$

- the generalized quantile processes (<u>real-valued</u>):

  $$U(p) = \inf\left\{\lambda(c); c \in C : P(c) \ge p\right\}$$

- quadrant-based:

  $$Q_{X,Y}(p,\varepsilon) = \left\{(x,y) \in R^2 : F_\varepsilon(x,y) = p\right\}$$

  where $F_{\varepsilon-+}(x,y) = \Pr\left\{X \le x,\ Y \ge y\right\}$

Multivariate quantiles in hydrology

- The main problem in extending quantiles is the **interpretation** of the obtained values

- We selected the quadrant-based approach:
  - simple
  - intuitive
  - does not require any symmetry assumption
  - the bivariate distribution (copula and margins) appears in its evaluation
  - probability-based (convenient for risk evaluation) rather than analytical or geometrical
  - Interpretable in freq. analysis context

Multivariate quantiles in hydrology

- The quantile curve is composed of two parts (practical reasons):
  - the naïve part (tail)
  - the proper part (central)

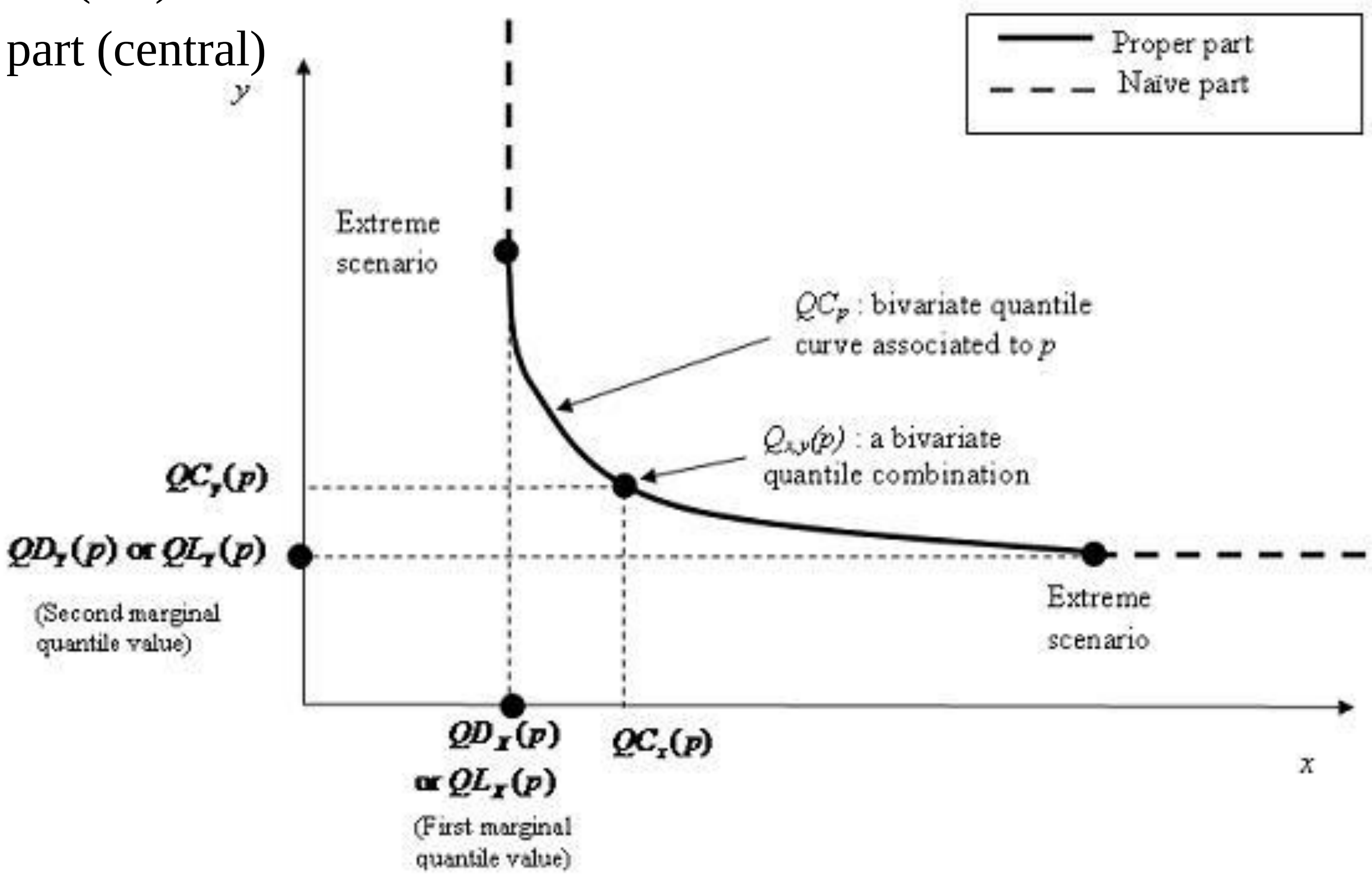


Illustration: biv. & univ. quantiles corresponding to the ***non-exceedence*** event

Multivariate quantiles in hydrology

Given a bivariate sample, steps procedure are:

1. Fit a multivariate distribution to the data set (copula and margins)

2. Estimate the distribution parameters (copula and margins)

3. Specify the event of interest according to the phenomenon being studied and the specific application (e.g., $(X<x, Y<y)$)

4. Estimate the different quantile combinations $Q_{x,y}(p)$ that constitute the quantile curve for a given risk *p* in (0,1)

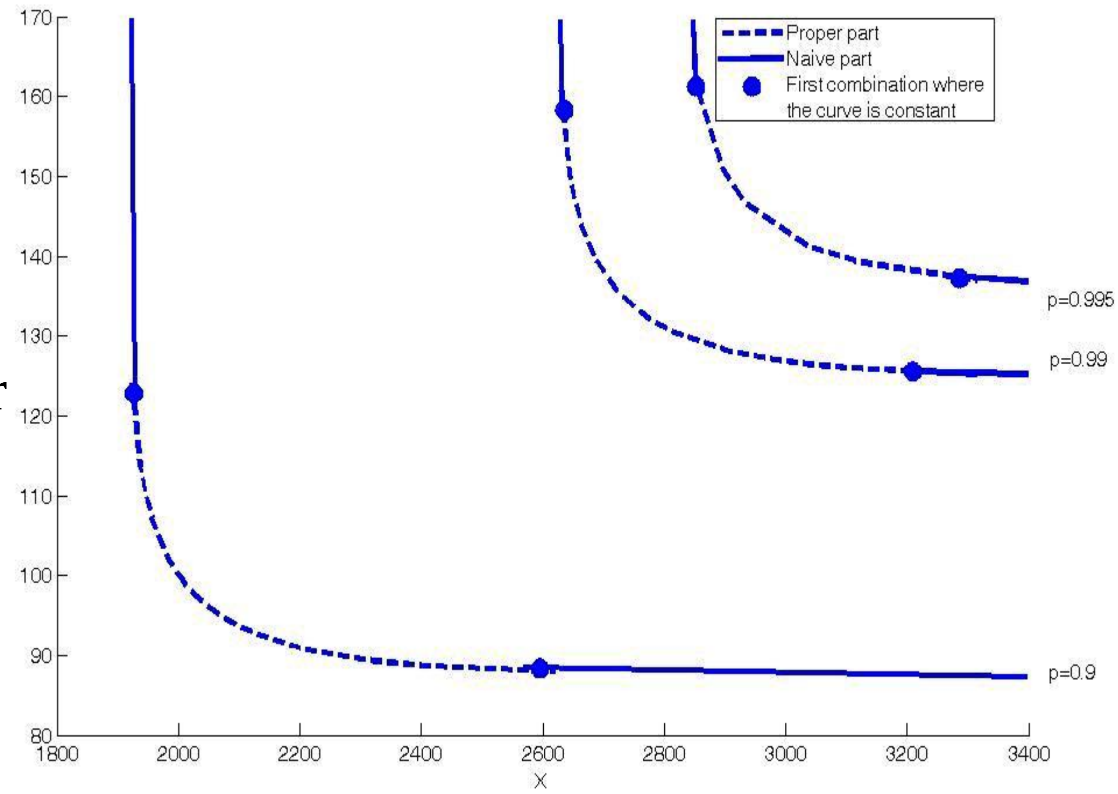
5. Select the appropriate combination(s) for the specific application

Multivariate quantile properties

- The biv. quantile gives several possible scenarios, <u>all</u> correspond to the <u>same</u> risk *p*

- The marginal (univ) quantiles correspond to the extreme scenarios

- When *p* increases, the number of scenarios decreases (the quantile curve becomes shorter)



Univ. vs. Biv.

- Univ. estimates should be used cautiously
- The combination of the univ. quantiles does not correspond to the desired risk
- They may lead to wrong conclusions
- Univ. quantiles evaluated <u>directly</u> and as extreme scenarios are very similar (<2%)

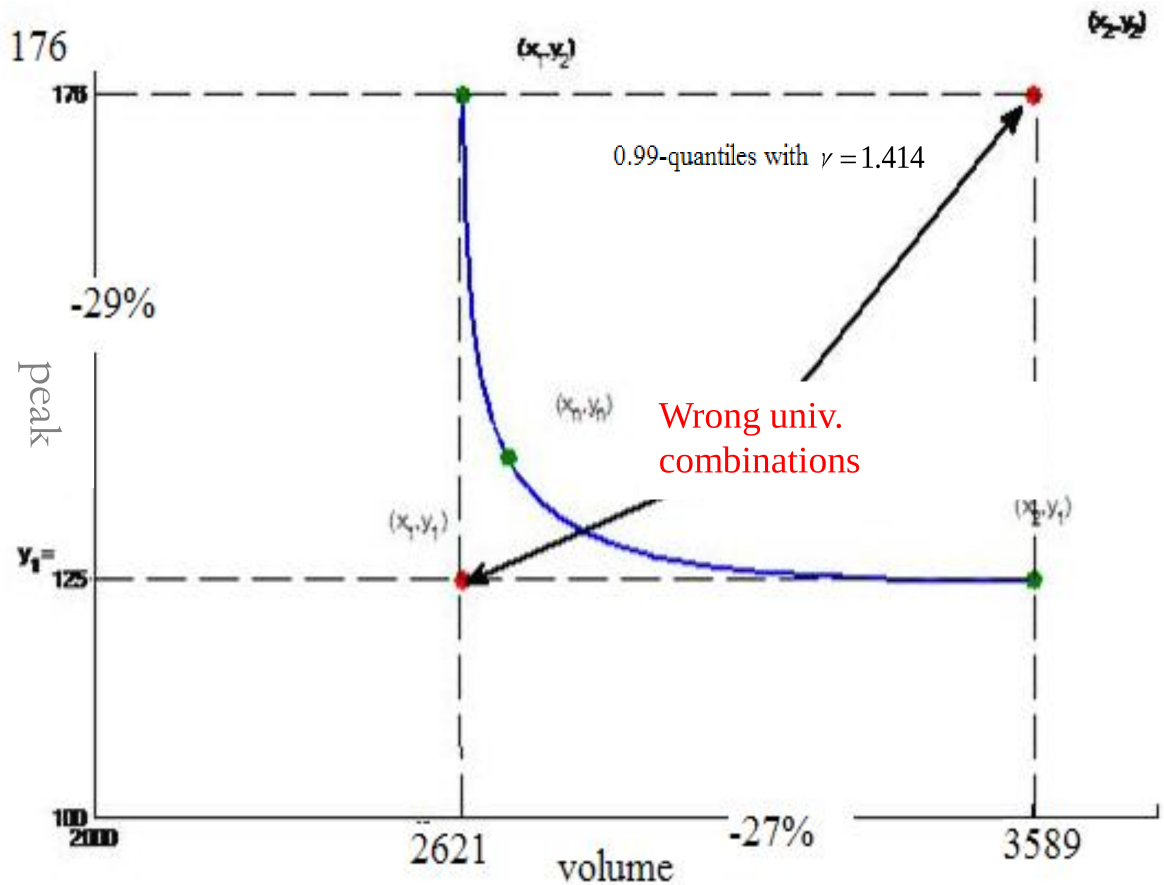


Illustration of univ. and biv. 0.99-quantile combinations

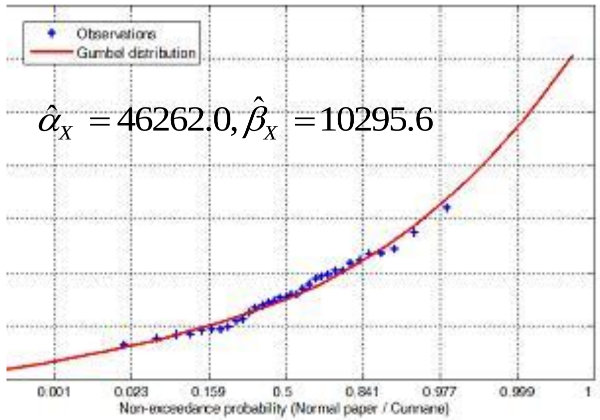
Case study : floods

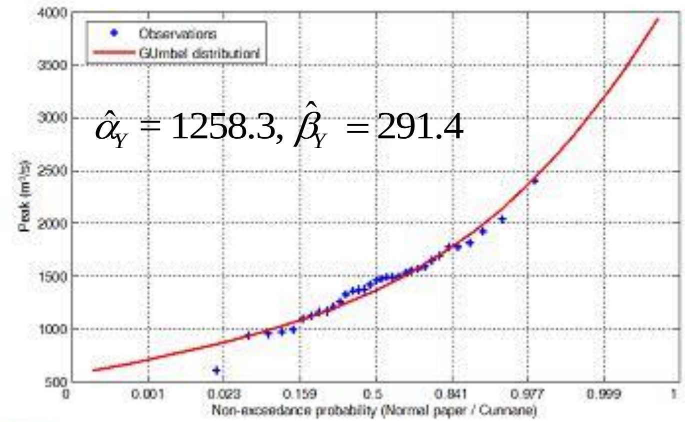
- The data set used in this case study is taken from Yue et al. (1999)
- It concerns the Ashuapmushuan basin located in the Saguenay region in the province of Québec, Canada
- The flood volume (X) and peak (Y) were extracted from a daily streamflow data set from 1963 to 1995

Case study: Distribution selection

$$F_X(x) = \exp\left\{-\exp\left(-\frac{x-\beta_X}{\alpha_X}\right)\right\},$$

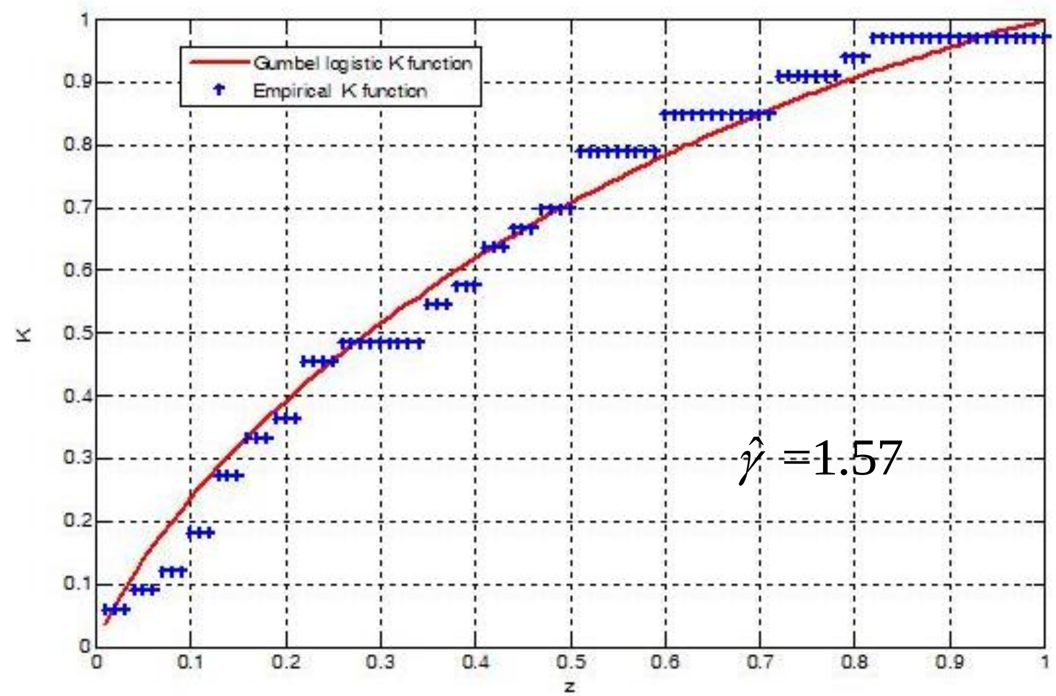
x real, $\alpha_X > 0$ and β_X real



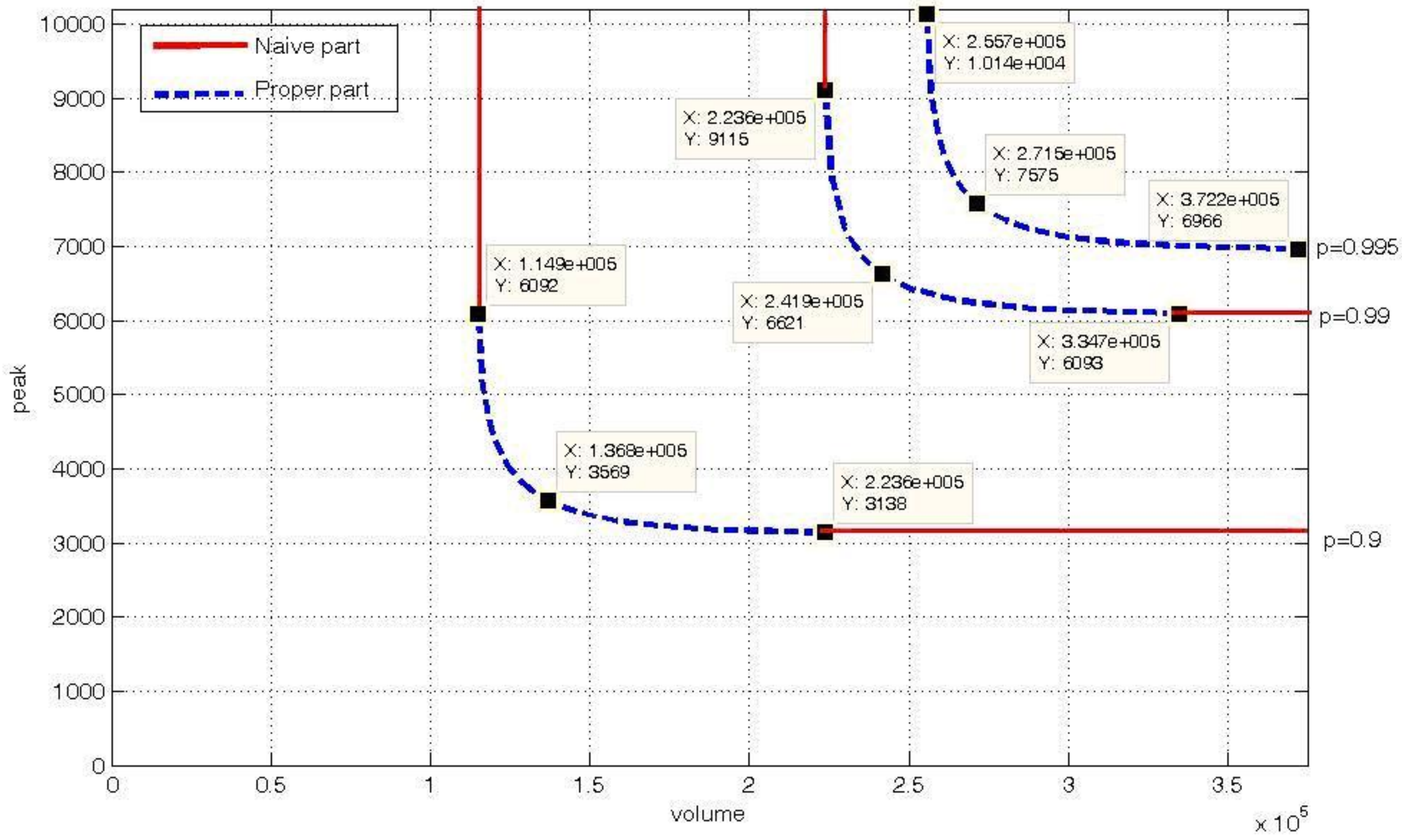


$$C_\gamma(u,v) = \exp\left\{-\left[(-\log u)^\gamma + (-\log v)^\gamma\right]^{1/\gamma}\right\},$$

$\gamma \geq 1$ and $0 \leq u,v \leq 1$



Case study: **quantile curves**



Simulation study

- We generate M= 10 000 bivariate samples with sizes n = 30 and 60
- The generated samples are from a bivariate distribution composed by Gumbel margins and a Gumbel logistic copula
- The sample generation is based on the algorithm developed by Ghoudi et al. (1998).
- p = 0.9, 0.99 and 0.995.
- **Performance evaluation criteria:** Basically, the evaluation consists in the assessment of the distance between the true and estimated quantile curves.

Table 3: Relative errors (%) of univariate quantiles evaluated directly and relative errors (%) of the bivariate quantile curve for the simultaneous non-exceedence and exceedence events when *n*=30

			p=0.9		***p=0.99***		***p=0.995***	
			RB	*RRMSE*	*RB*	*RRMSE*	*RB*	*RRMSE*
Exceedence	$\gamma = 3.162$	QD_X	0.09	6.35	-0.39	9.75	-0.34	9.62
		QD_Y	0.10	8.63	-0.44	15.13	-0.06	16.38
		Biv^*	0.92	9.32	2.24	14.90	2.94	16.17
	$\gamma = 1.414$	QD_X	-0.15	6.35	-0.43	9.77	-0.58	9.65
		QD_Y	-0.21	8.71	-0.40	15.96	-0.32	18.19
		Biv^*	0.53	10.16	1.19	15.78	1.40	17.53
	$\gamma = 1$	QD_X	0.02	6.41	-0.40	9.76	-0.46	9.68
		QD_Y	0.04	8.60	0.04	16.13	-0.26	18.77
		Biv^*	0.67	9.98	0.89	15.59	0.79	17.58
Non-exceedence	$\gamma = 3.162$	QD_X	-0.02	7.09	0.04	9.52	-0.04	9.96
		QD_Y	-0.02	8.17	0.03	10.55	-0.01	10.91
		Biv^*	0.59	13.04	0.65	16.66	0.46	16.76
	$\gamma = 1.414$	QD_X	0.03	7.09	-0.04	9.36	0.11	9.93
		QD_Y	0.16	8.27	-0.02	10.43	0.20	10.88
		Biv^*	0.40	12.32	0.27	15.63	0.36	15.07
	$\gamma = 1$	QD_X	-0.04	7.16	-0.12	9.44	0.03	9.99
		QD_Y	0.06	8.26	0.15	10.63	0.02	11.04
		Biv^*	-0.09	11.09	-0.09	13.47	-0.16	12.88

*The *RB* and *RRMSE* are evaluated using respectively $RIE^{*[n]}(p)$ and $RIE^{[n]}(p)$

Simulation results

- The estimation procedure performs better for large sample sizes in all considered situations
- The univ. estimation does not take into account the dependence structure between variables and should be used cautiously.
- The relative errors of both biv. and univ. estimations are of the same order of magnitude with similar behaviours with respect to *n* and *p*
- The biv. procedure provides univ. quantile estimates that are very close to those obtained directly and also with equivalent precisions
- The main performance differences between univ. and biv. estimations are conceptual.

Conclusions

- The extension of the quantile notion to high dimensions leads to several multivariate quantile versions
- The selected version is simple, intuitive, probability-based and interpretable
- Even though, the focus was on the bivariate context, the study can be conducted in higher dimensions with the appropriate adaptations
- The univ. estimated quantiles, correctly combined, are particular cases corresponding to the extreme scenarios of the biv. quantile curve
- Depending on the available resources and the nature of the project, one or more convenient scenarios may be selected
- Aside from being more accurate and realistic, the bivariate setting offers more flexibility to designers than the univariate framework

Merci

Thank you