Instituto de Ingeniería del Agua y Medio Ambiente

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Probable Maximum Flood estimation using upper bounded statistical models and its effect on high return period quantiles

By:

Félix Francés and Blanca Botero

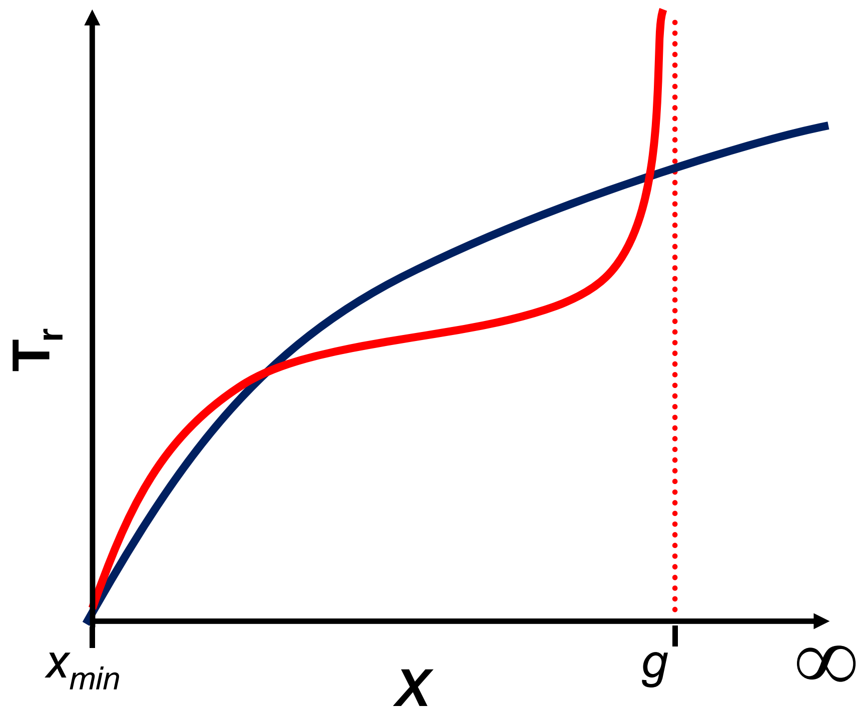
3rd International Week on Risk Analysis, Dam Safety, Dam Security, and Critical Infrastructure Management

Valencia, Spain, October 18-19, 2011

3IWRDD

Introduction

- The PMF is the biggest flood physically possible at a specific catchment (Smith and Ward, 1998)
  - It has a physical meaning and provides an upper limit for the decision maker
  - It will change the cdf behaviour at medium and high return periods



Introduction

- High return period quantile estimation main drawback: lack of available information about large events in a relatively short data series **=> increase amount of information**
  - ➢ One possibility is to included palaeoflood and/or historic information. From now, Non-Systematic information:
    - **Information different to the systematic record at the flow gauge station**
    - No differences from the statistical point of view!

Introduction

- High return period quantile estimation main drawback: lack of available information about large events in a relatively short data series **=> increase amount of information**

- **Objective**: merging deterministic knowledge and statistical analysis to better estimate high return period quantiles in a framework of enough information

LN4 - Slade transformation

- Applied by Takara and Loebis (1996)
- **Basic distribution function Y ~ LN2**
- Slade-type transformation: $y = \ln\left(\frac{x-a}{g-x}\right)$
- Resulting pdf: $f(x) = \frac{g-a}{(x-a)(g-x)\sigma_y\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left\{\frac{y-\mu_y}{\sigma_y}\right\}^2\right]$
- Parameters:
  - g = upper bound (PMF)
  - a = lower bound, set to 0 to reduce the # of parameters
  - μ_y, σ_y = LN parameters

TDF - Elíasson transformation

- Developed by Elíasson (1997)
- **Basic distribution function: Y ~ EV1 (or Gumbel)**
- Transformation: $$Y = X - \frac{ak^{*2}}{(g - X)}$$
- Resulting cdf: $$F(x) = \exp\left[-\exp\left(\frac{-x}{a} + \frac{ak^*}{(g-x)} - b\right)\right]$$
- Parameters:
  - g = upper bound (PMF)
  - k* = transformation parameter = –0.5 for better results
  - a = scale and transformation parameter
  - b = location parameter

EV4 model

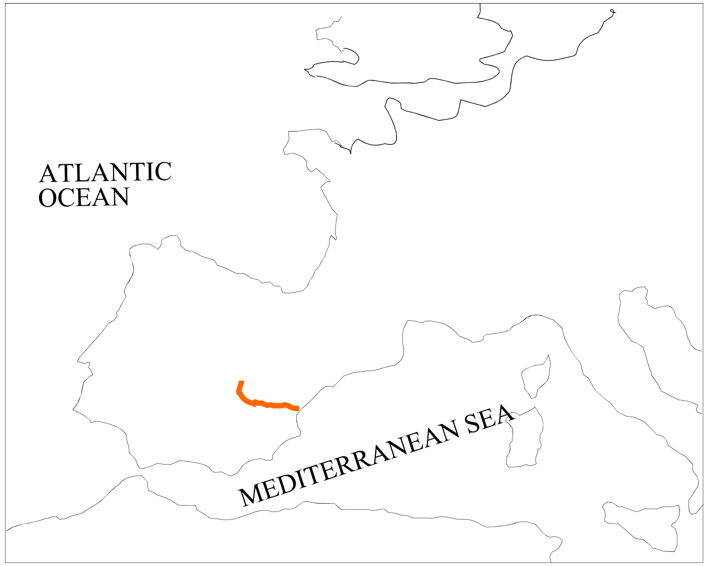
- **It was derived from a GEV** (Takara and Tosa, 1999)

- The resulting cdf is: $$F(x) = \exp\left[-\left\{ \frac{g-x}{\nu(x-a)} \right\}^{k'} \right]$$

- Parameters:
  - a = lower bound (set to 0 to reduce # of parameters)
  - g = upper bound (PMF)
  - v = scale parameter
  - k' = shape parameter

The Jucar River case study

- Relative large basin: 22,000 km^2

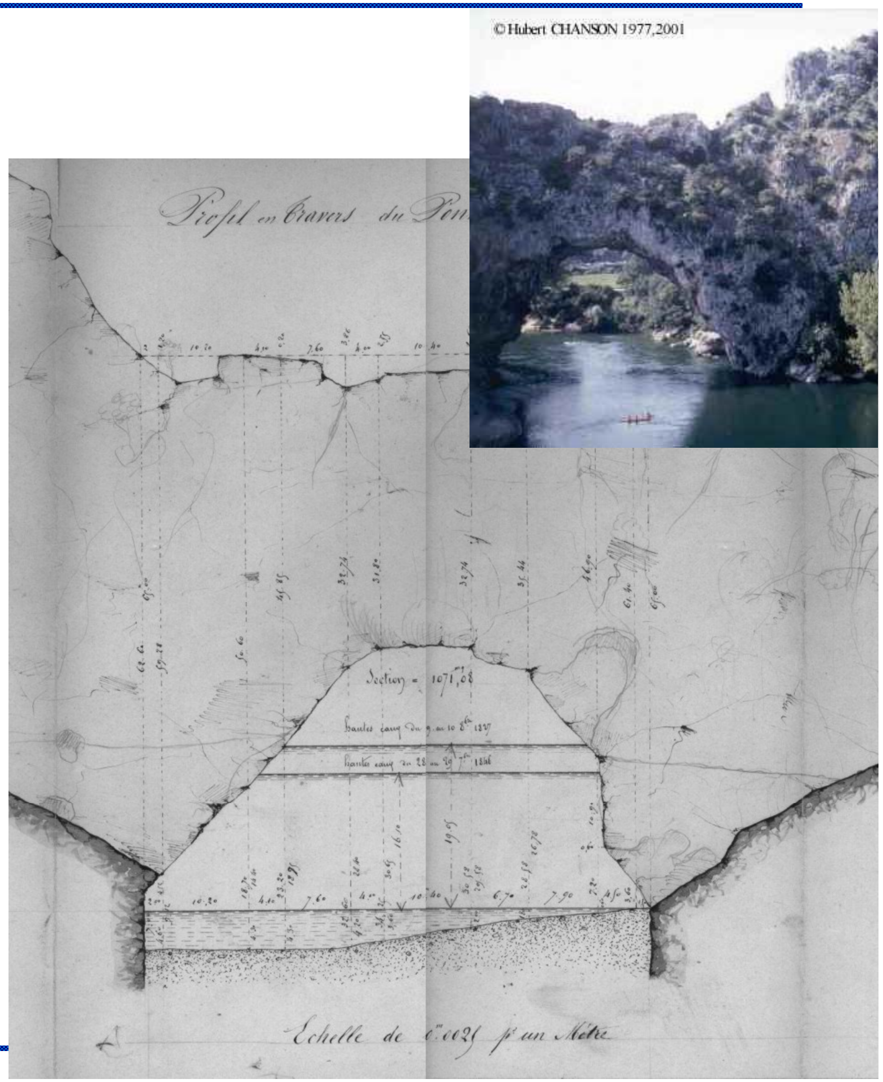


- Mediterranean torrential regime
  - Mean flow = 36 m^3/s
  - Mean flood = 713 m^3/s
  - Q_{20}= 2,000 m^3/s
  - Coefficient of variation = 2.74
  - Skewness coefficient = 5.26

- Very strong Convective Mesoscale Systems in Fall (Rigo and Llasat, 2007)
  - ➢ Mixed flood population (Rossi et al., 1984)

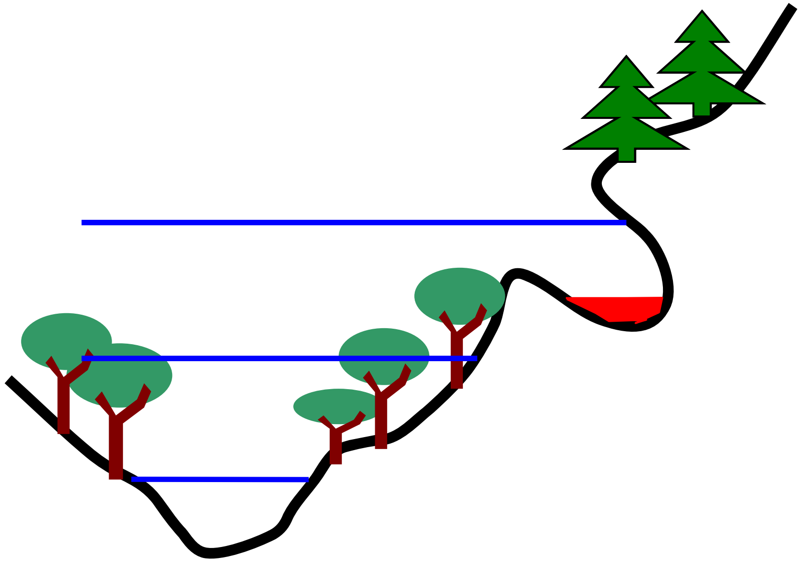
Sources of Non-Systematic Information

- **Historical information**
  - Archives: municipality records, notary notes, engineering damage reports
  - Newspapers, books, chronicles
  - Maps, photographs, plans
  - Building marks
  - Oral communications



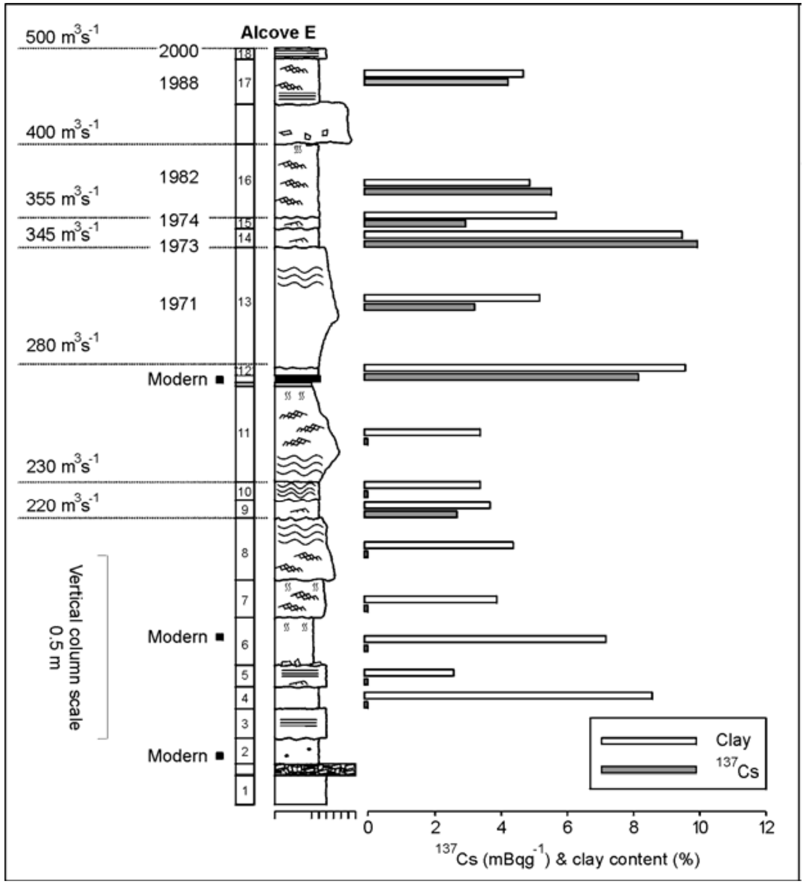
Plan of the two biggest floods in XIX century floods at Pont d’Arc (Ardèche River in France)

Sources of Non-Systematic Information



- **Palaeoflood information**
  - Botanical evidences
  - Palaeolevel indicators
    - Slackwater deposits
    - Silt marks

Sources of Non-Systematic Information



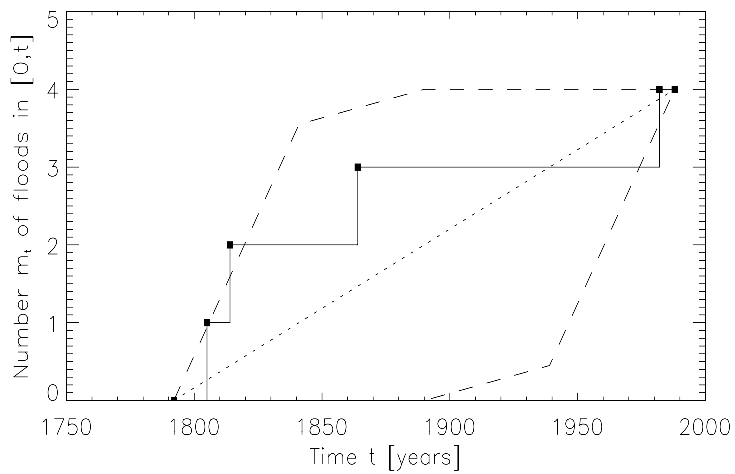
- **Palaeoflood information**
  - Botanical evidences
  - Palaeolevel indicators
    - Slackwater deposits
    - Silt marks

Sources of Non-Systematic Information

- Jucar case study with **historical information** (building marks): flooding of an ancient convent located in the floodplain
  - Threshold of inundation X_H = 6,200 m^3/s
  - Historical period: 1792 to 1945

Year	Peak Q (m^3/s)
1805	8,400
1814	6,400
1864	13,000

- Stationarity was tested and proved using the Lang Test (Lang et al., 1999)



Upper limit estimation

- **ML-PG:** *g* is fixed at the value previously calculated (G) as the best approximation for the true unknown PMF, and the other parameters are estimated by ML method:

$$\left.\begin{array}{r} g = G \\ \max L(\underline{\Theta}') \end{array}\right\}$$

- **ML-C:** the whole parameters set of the distribution function is estimated by the ML method, including *g* as another free parameter in the maximization process:

$$\max L(\underline{\Theta})$$

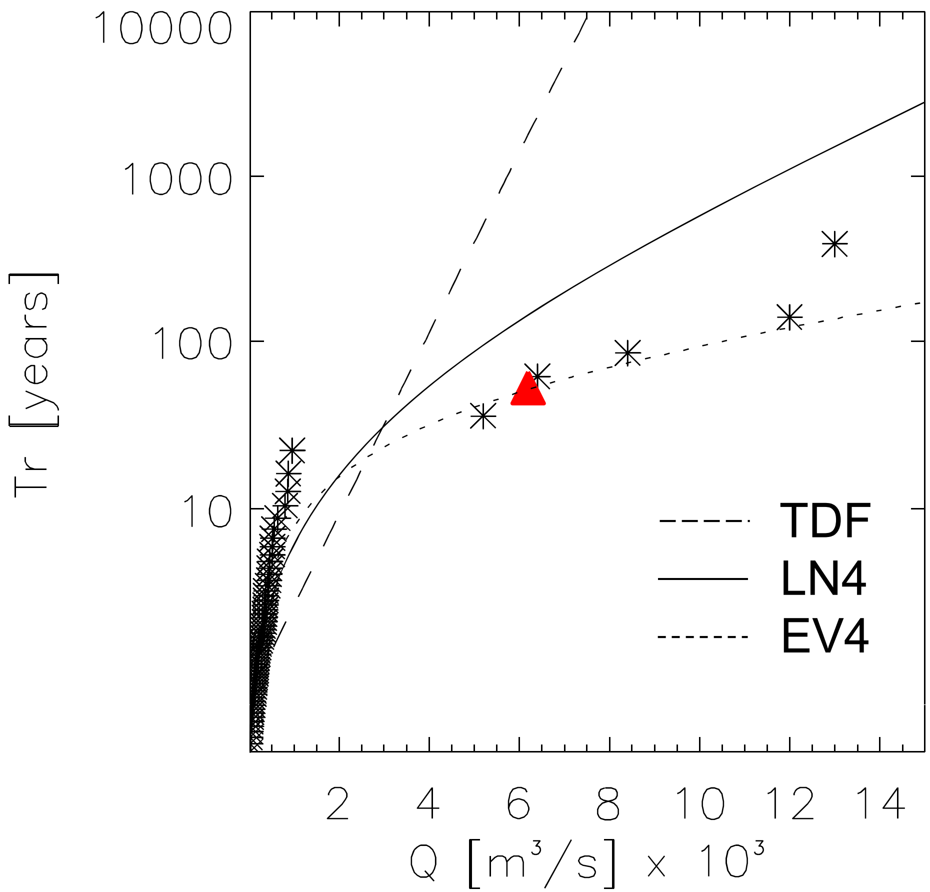
Upper limit estimation

- In some combinations of distribution + type of information, ML-C estimates *g* as the maximum observation =>

- **ML-GE:** This method consists on the use of the Generic Equation to estimate *g* (Kijko, 2004) and the ML method for the rest of parameters:

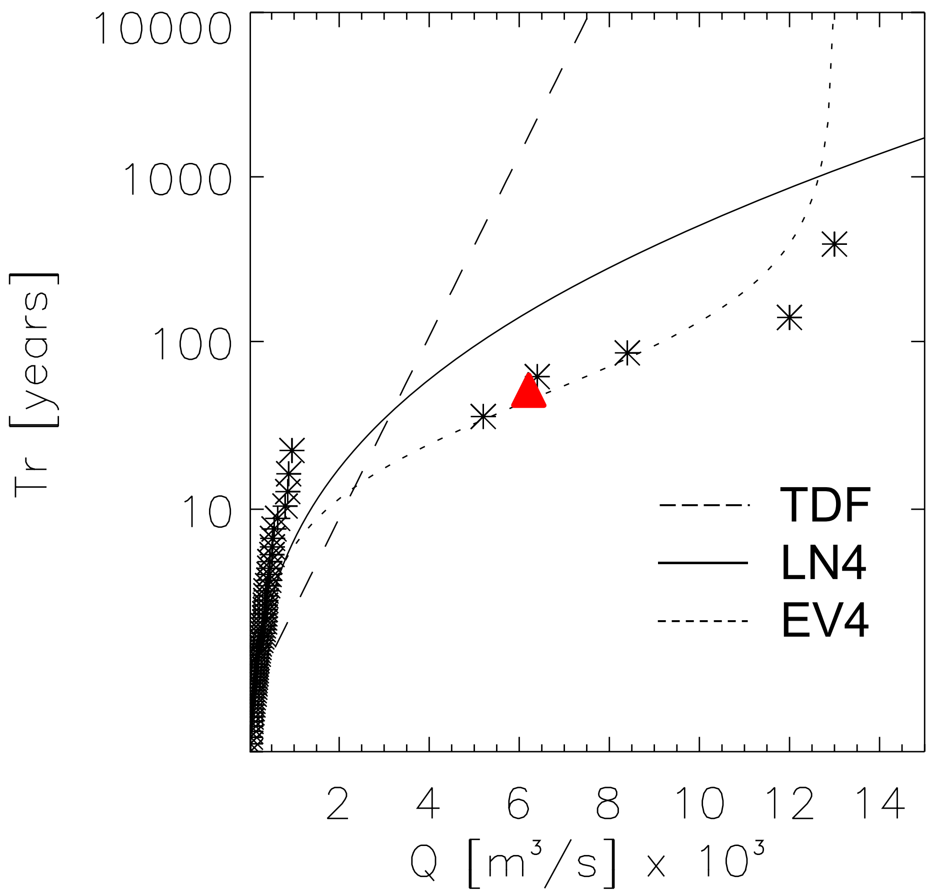
$$\left.\begin{array}{c} g = x_{\max} + \int_{-\infty}^{g} [F_X(x;\underline{\Theta}', g)]^n \, dx \\ \max L(\underline{\Theta}') \end{array}\right\}$$

ML-PG (prefixed *g*)



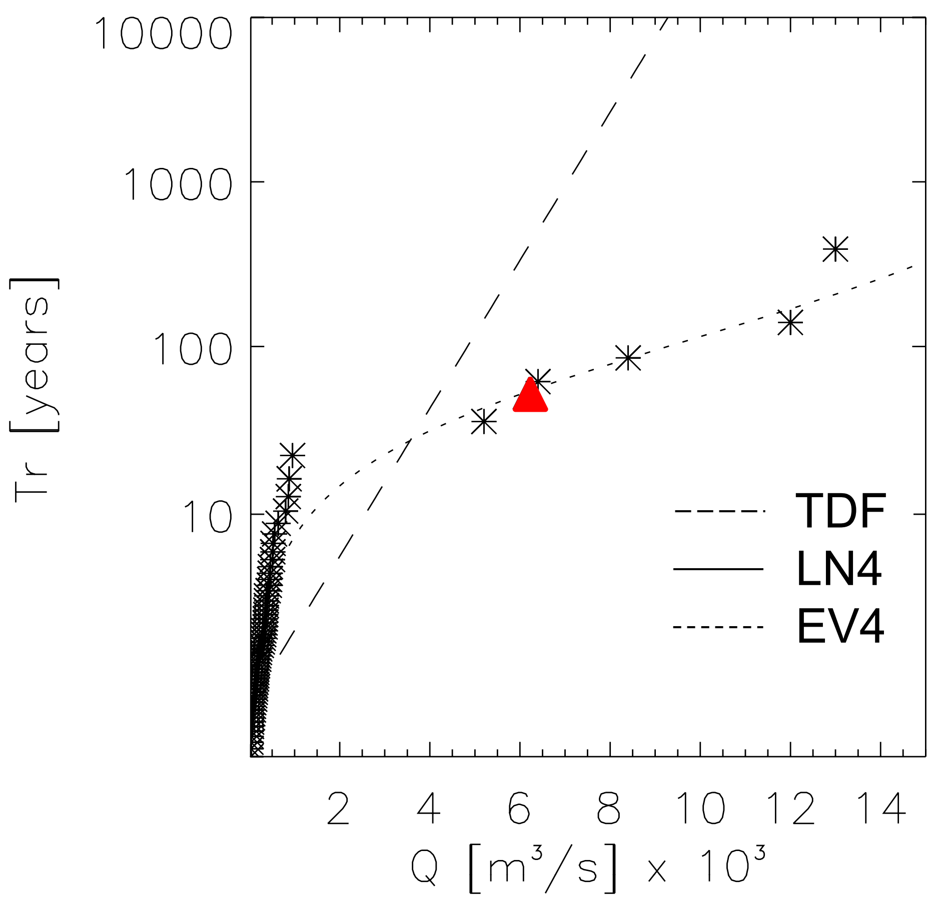
- PMF using **specific discharge** from upstream PMF study = 33,900 m^3/s => possible overestimation
- "Dog leg" effect due to mixed populations
- Clear different approach to the upper limit
  - Slower for TDF
  - Faster for EV4

ML-C (complete ML)



- PMF with EV4 and TDF ML-C = 13,000 m^3/s (maximum observation)
- PMF with LN4 ML-C = 93,300 m^3/s

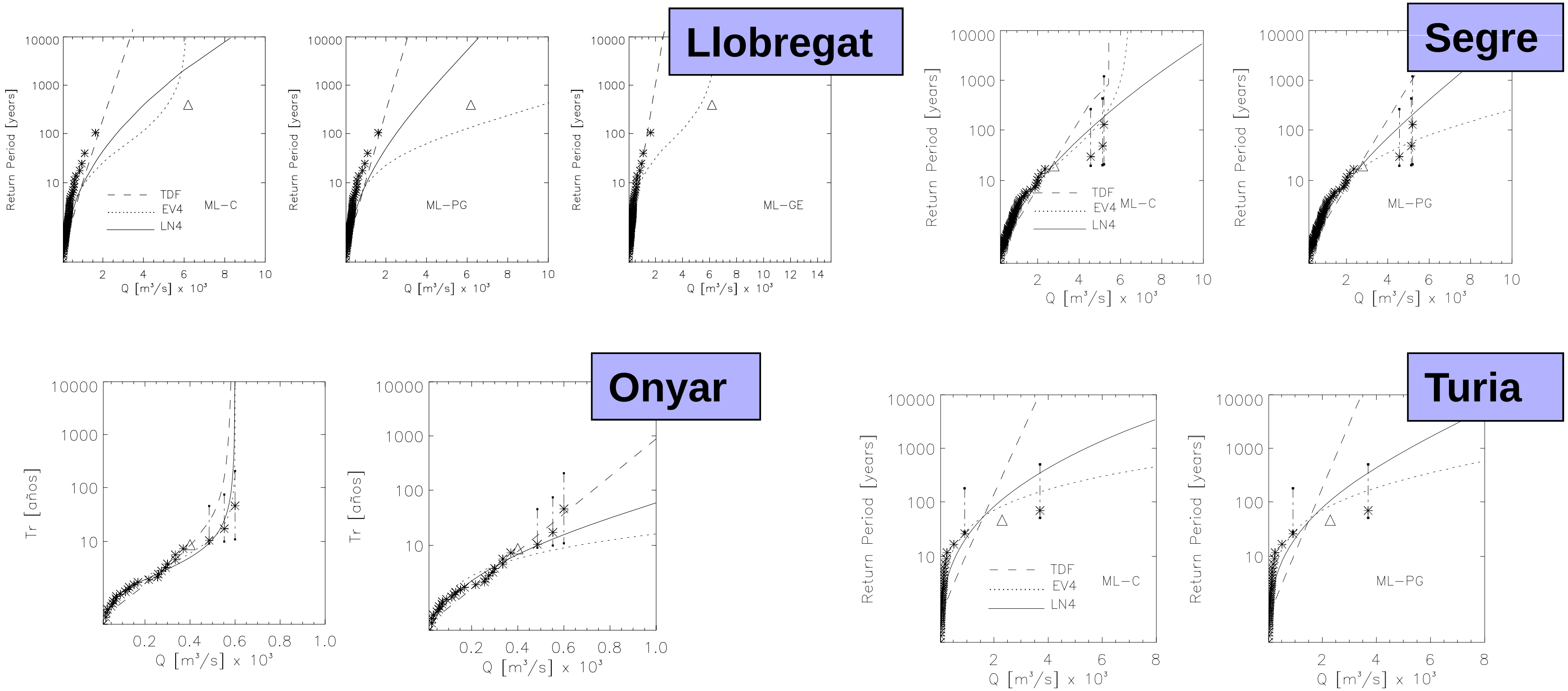
ML-GE (mixed with generic equation)



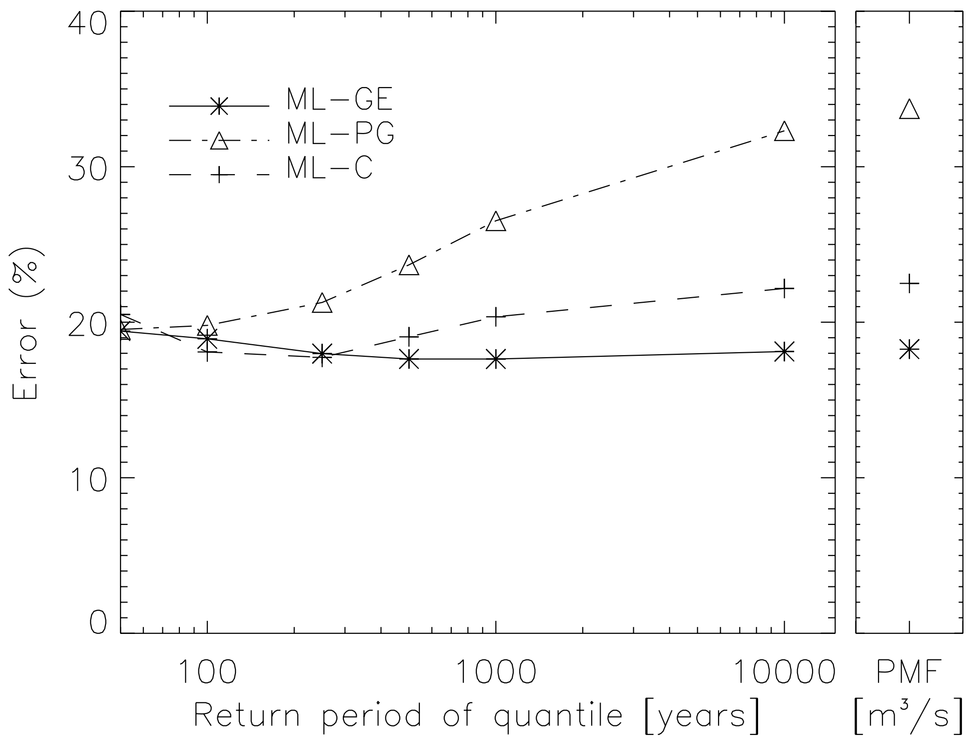
- Numerical problems with LN4
- PMF with EV4 ML-GE = 18,100 m^3/s
- PMF with TDF ML-GE = 93,100 m^3/s

Models comparison

- EV4 better performance with high skewnees coefficient, as it was pointed out by Takara and Tosa (1999)



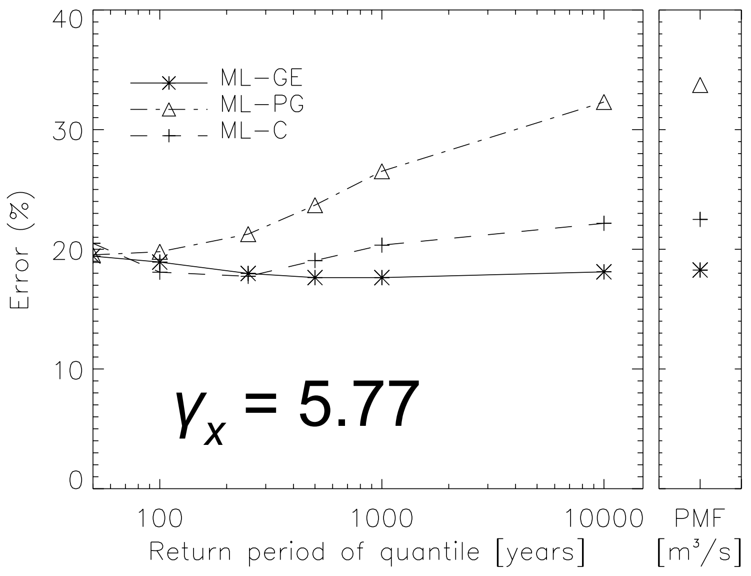
EV4 uncertainty analysis



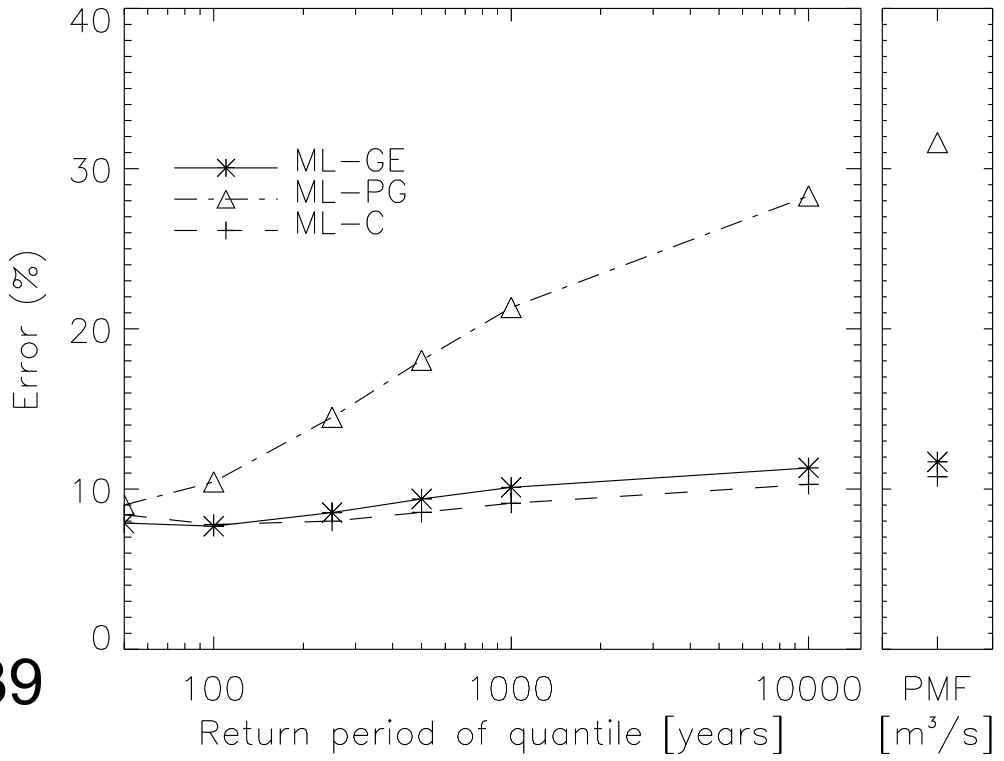
$$E(\%) = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{\theta}_i - \theta\right)^2}}{\theta}100$$

- Monte Carlo simulations with:
  - N= 50 years
  - M = 400 years
  - H = 50 years return period
  - γ_x = 5.77
  - Errors in G with
    - CV = 0.3
    - bias = +10%

EV4 uncertainty analysis



$\gamma_x = 5.77$



$\gamma_x = 2.39$

Conclusions

- Why not explore and use upper bounded distribution functions?
  - There is an upper limit
  - Upper bounded and unbounded distributions behave different at medium and high return periods

Conclusions

- Why not explore and use upper bounded distribution functions?

- The upper limit represents the PMF and either can be:
  - **Causal information expansion** (Merz and Blöschl, 2008), if it is fixed a priori

  - Within a **temporal information expansion** framework (Merz and Blöschl, 2008), considered as one more parameter to be estimated in the statistical model fitting

Remarks

Program **AFINS** available at http://lluvia.dihma.upv.es

Poster: *High return period annual maximum reservoir water level quantiles estimation using synthetic generated flood events*

Many thanks for your attention!

Prof. Félix Francés (ffrances@hma.upv.es)

Research Group in Hydrological and Environmental Modelling (GIMHA)