Efficient Use of Local and Regional Information for Flood Frequency Analysis

Mathieu Ribatet

CEMAGREF Lyon, Unité de Recherche Hydrologie-Hydraulique, FRANCE

INRS-ETE, Université de Québec, Chaire HydroStat, CANADA

Directeurs de Thèse: T. Ouarda and J.-M. Grésillon

December 10th 2008

The Aim of an Extreme Value Analysis

"To know or predict some statistical properties of events that haven't been seen yet (or only a few)..."

Ouvèze River. Vaison-la-Romaine, France. September 1992.

- EVA tries to model the unpredictable nature of a random process
  like flood peaks, volumes, ...
- Such analysis are of great importance
  safety structure designs, failing rates estimations, ...

The Aim of an Extreme Value Analysis

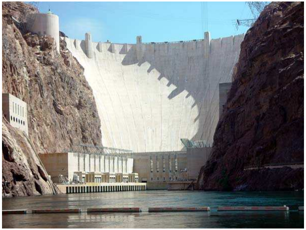
"To know or predict some statistical properties of events that haven't been seen yet (or only a few). . . "

Ouvèze River. Vaison-la-Romaine, France. September 1992.

- EVA tries to model the unpredictable nature of a random process
  like flood peaks, volumes, . . .
- Such analysis are of great importance
  safety structure designs, failing rates estimations, . . .

The Aim of an Extreme Value Analysis

"To know or predict some statistical properties of events that haven't been seen yet (or only a few)..."



The Hoover dam (Colorado)

The Taichai river (Taiwan)

- EVA tries to model the unpredictable nature of a random process
  like flood peaks, volumes, ...
- Such analysis are of great importance
  safety structure designs, failing rates estimations, ...

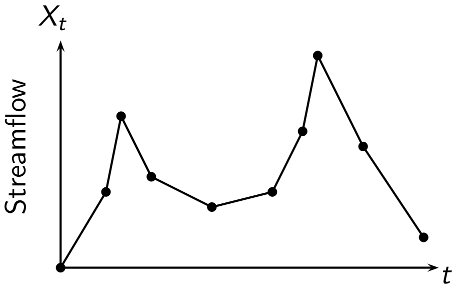
Figure: The Seine river at Paris, France (January 1910).

Question

How unpredictable is an event such as the one depicted above?

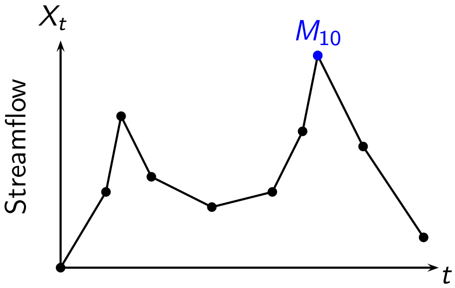
- We need a theoretical framework to answer this question
- This is the aim of the Extreme Value Theory (EVT)

AM/POT Asymptotic Distributions



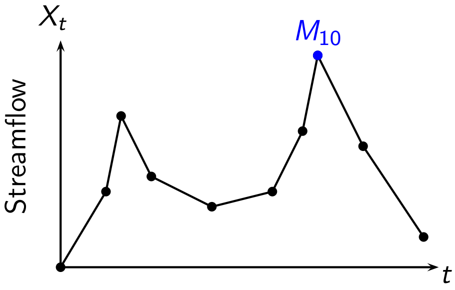
- Suppose our process is governed by a random variable X_t
- Define $Y_j = \{X_t : X_t > u, t = 1, \ldots, n\}$
- Under mild conditions, the following result holds:

AM/POT Asymptotic Distributions



- Suppose our process is governed by a random variable X_t
- Define $M_n = \max\{X_t : t = 1, \ldots, n\}$, $n \geq 1$;
- Under mild conditions, the following result holds:

AM/POT Asymptotic Distributions

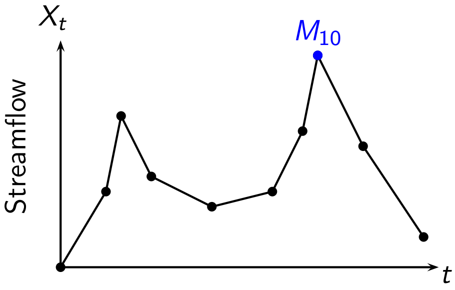


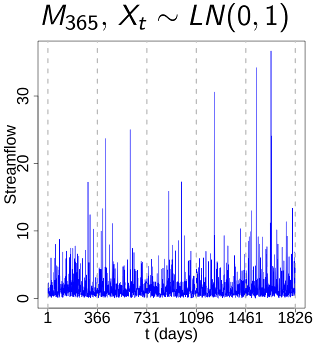
- Suppose our process is governed by a random variable X_t
- Define $M_n = \max\{X_t : t = 1, \ldots, n\}$, $n \geq 1$;
- Under mild conditions, the following result holds:

$$\Pr\left[(M_n - b_n)/a_n \leq x\right] \to G(x), \qquad n \to +\infty$$

where $G(x) = \exp\left[-\left(1 + \xi \frac{x-\mu}{\sigma}\right)_+^{-1/\xi}\right]$ is the Generalized Extreme Value distribution (GEV)

AM/POT Asymptotic Distributions



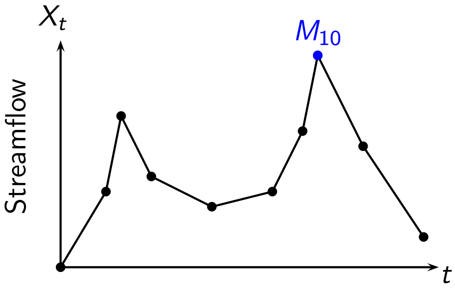


- Suppose our process is governed by a random variable X_t
- Define $M_n = \max\{X_t : t = 1, \ldots, n\}$, $n \geq 1$;
- Under mild conditions, the following result holds:

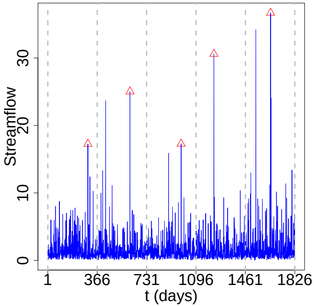
$$\Pr\left[(M_n - b_n)/a_n \leq x\right] \rightarrow G(x), \qquad n \rightarrow +\infty$$

where $G(x) = \exp\left[-\left(1 + \xi\frac{x-\mu}{\sigma}\right)_+^{-1/\xi}\right]$ is the Generalized Extreme Value distribution (GEV)

AM/POT Asymptotic Distributions



M_{365}, $X_t \sim LN(0,1)$

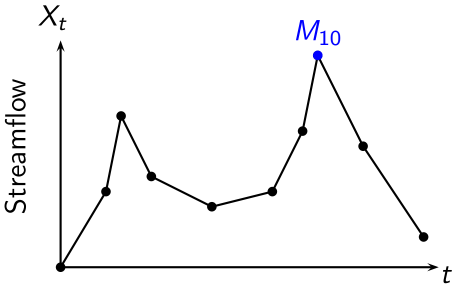


- Suppose our process is governed by a random variable X_t
- Define $M_n = \max\{X_t : t = 1, \ldots, n\}$, $n \geq 1$;
- Under mild conditions, the following result holds:

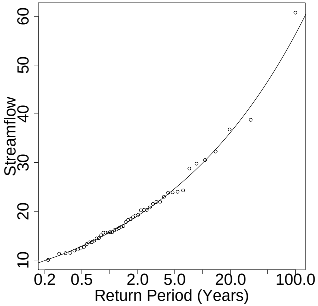
$$\Pr\left[(M_n - b_n)/a_n \leq x\right] \to G(x), \qquad n \to +\infty$$

where $G(x) = \exp\left[-\left(1 + \xi\frac{x-\mu}{\sigma}\right)_+^{-1/\xi}\right]$ is the Generalized Extreme Value distribution (GEV)

AM/POT Asymptotic Distributions



$M_{365},\ X_t \sim LN(0,1)$

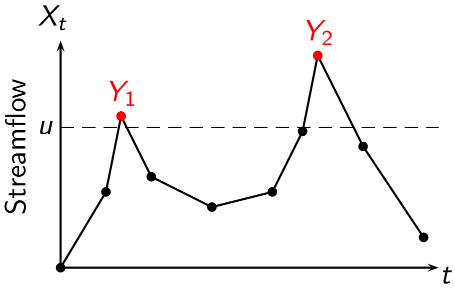


- Suppose our process is governed by a random variable X_t
- Define $M_n = \max\{X_t : t = 1, \ldots, n\}$, $n \geq 1$;
- Under mild conditions, the following result holds:

$$\Pr\left[(M_n - b_n)/a_n \leq x\right] \to G(x), \qquad n \to +\infty$$

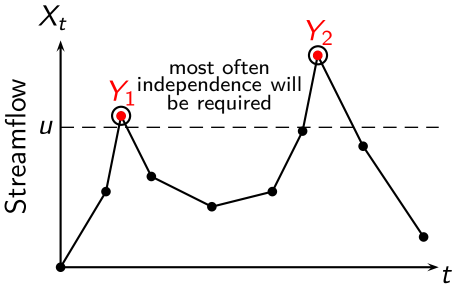
where $G(x) = \exp\left[-\left(1 + \xi \frac{x-\mu}{\sigma}\right)_+^{-1/\xi}\right]$ is the Generalized Extreme Value distribution (GEV)

AM/POT Asymptotic Distributions



- Suppose our process is governed by a random variable X_t
- Define $Y_j = \{X_t : X_t > u, t = 1, \ldots, n\}$
- Under mild conditions, the following result holds:

AM/POT Asymptotic Distributions

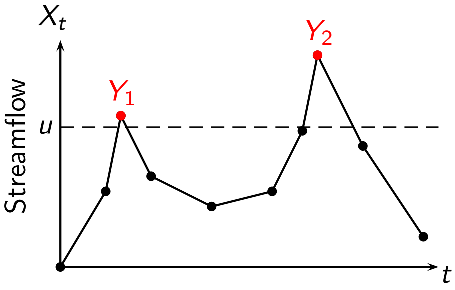


- Suppose our process is governed by a random variable X_t
- Define $Y_j = \{X_t : X_t > u, t = 1, \ldots, n\}$
- Under mild conditions, the following result holds:

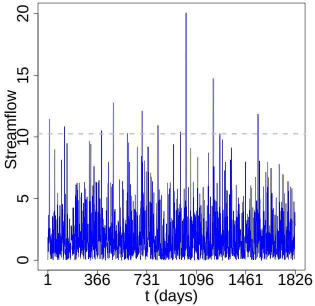
$$\Pr\left[X_t \leq x | X_t > u\right] \rightarrow H(x), \quad u \rightarrow u^+ = \inf\{x : \Pr[X_t < x] = 1\}$$

where $H(x) = 1 - \left(1 + \xi\frac{x-u}{\sigma}\right)_+^{-1/\xi}$ is the Generalized Pareto Distribution (GPD)

AM/POT Asymptotic Distributions



$X_t \sim \chi^2_2$, $u > 10.4$

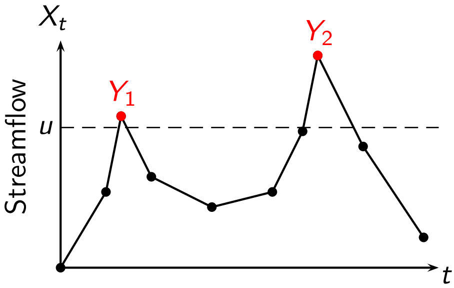


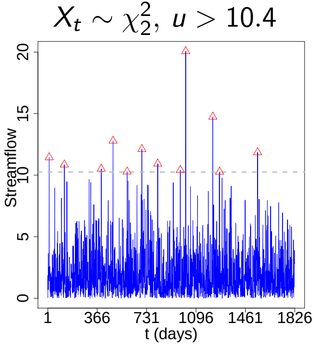
- Suppose our process is governed by a random variable X_t
- Define $Y_j = \{X_t : X_t > u, t = 1, \dots, n\}$
- Under mild conditions, the following result holds:

$$\Pr\left[X_t \leq x | X_t > u\right] \to H(x), \quad u \to u^+ = \inf\{x : \Pr[X_t < x] = 1\}$$

where $H(x) = 1 - \left(1 + \xi\frac{x-u}{\sigma}\right)_+^{-1/\xi}$ is the Generalized Pareto Distribution (GPD)

AM/POT Asymptotic Distributions



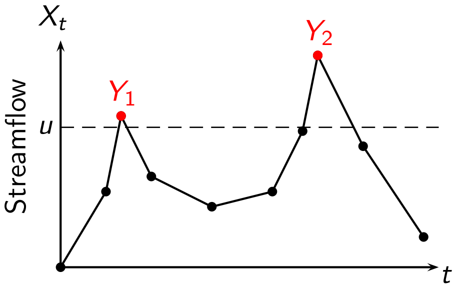


- Suppose our process is governed by a random variable X_t
- Define $Y_j = \{X_t : X_t > u, t = 1, \ldots, n\}$
- Under mild conditions, the following result holds:

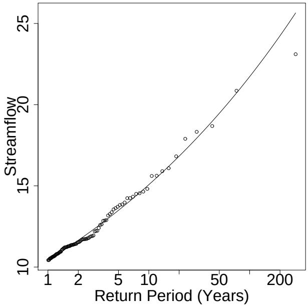
$$\Pr\left[X_t \leq x | X_t > u\right] \to H(x), \quad u \to u^+ = \inf\{x : \Pr[X_t < x] = 1\}$$

where $H(x) = 1 - \left(1 + \xi\frac{x-u}{\sigma}\right)_+^{-1/\xi}$ is the Generalized Pareto Distribution (GPD)

AM/POT Asymptotic Distributions



$X_t \sim \chi^2_2,\ u > 10.4$



- Suppose our process is governed by a random variable X_t
- Define $Y_j = \{X_t : X_t > u, t = 1, \ldots, n\}$
- Under mild conditions, the following result holds:

$$\Pr\left[X_t \leq x | X_t > u\right] \to H(x), \quad u \to u^+ = \inf\{x : \Pr[X_t < x] = 1\}$$

where $H(x) = 1 - \left(1 + \xi \frac{x-u}{\sigma}\right)_+^{-1/\xi}$ is the Generalized Pareto Distribution (GPD)

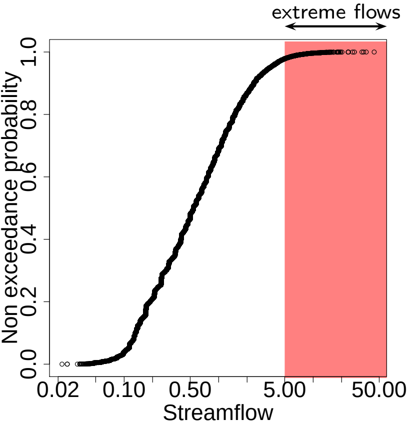
The GPD parameters

The GPD needs 3 parameters to be estimated:

location u defines which levels are extreme?

scale σ controls how "smallest extremes" increase?

shape ξ controls how "largest extremes" evolve?



Link with conventional distributions

- The Gumbel/Exponential cases correspond to $\xi = 0$
- And thus affect the "largest extremes" behavior

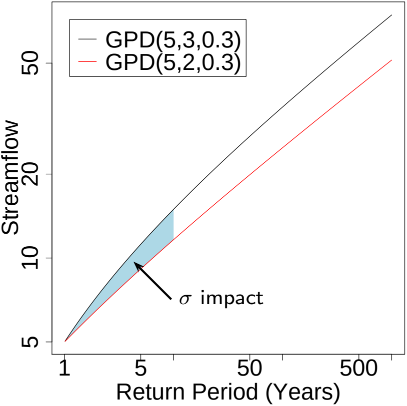
The GPD parameters

The GPD needs 3 parameters to be estimated:

location u defines which levels are extreme?

scale σ controls how "smallest extremes" increase?

shape ξ controls how "largest extremes" evolve?



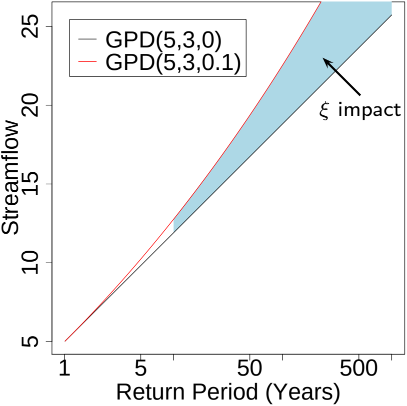
Link with conventional distributions

- The Gumbel/Exponential cases correspond to $\xi = 0$
- And thus affect the "largest extremes" behavior

The GPD parameters

The GPD needs 3 parameters to be estimated:

- location u defines which levels are extreme?
- scale σ controls how "smallest extremes" increase?
- shape ξ controls how "largest extremes" evolve?



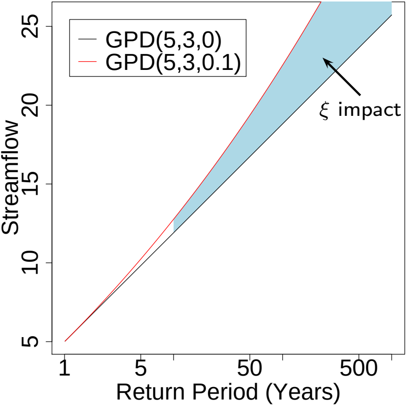
Link with conventional distributions

- The Gumbel/Exponential cases correspond to $\xi = 0$
- And thus affect the "largest extremes" behavior

The GPD parameters

The GPD needs 3 parameters to be estimated:

- location u defines which levels are extreme?
- scale σ controls how "smallest extremes" increase?
- shape ξ controls how "largest extremes" evolve?



Link with conventional distributions

- The Gumbel/Exponential cases correspond to $\xi = 0$
- And thus affect the "largest extremes" behavior

From Theory to Practice

- Using the EVT, we do not need to know X_t's distribution
  (we did not suppose X_t were LN or χ^2 distributed)
- We only need to get closer to the asymptotic hypothesis
- However, as we focus on the tail area,
- Estimations may be strongly affected

From Theory to Practice

- Using the EVT, we do not need to know X_t's distribution
  (we did not suppose X_t were LN or χ^2 distributed)
- We only need to get closer to the asymptotic hypothesis
- However, as we focus on the tail area,
- Estimations may be strongly affected

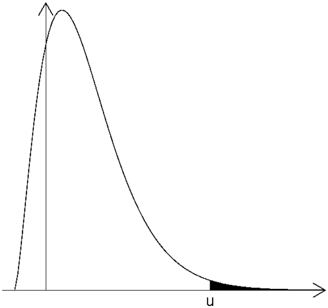


Figure: A probability density function and a right-tail area.

From Theory to Practice

- Using the EVT, we do not need to know X_t's distribution
  (we did not suppose X_t were LN or χ^2 distributed)
- We only need to get closer to the asymptotic hypothesis
- However, as we focus on the tail area, we will deal with limited sample sizes
- Estimations may be strongly affected

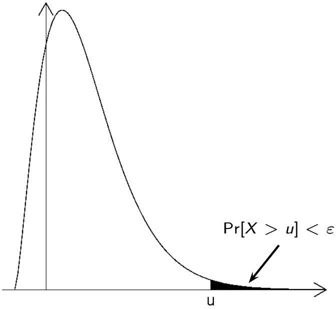
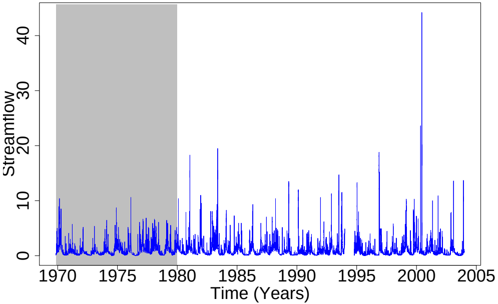


Figure: A probability density function and a right-tail area.

From Theory to Practice

- Using the EVT, we do not need to know X_t's distribution
  (we did not suppose X_t were LN or χ^2 distributed)
- We only need to get closer to the asymptotic hypothesis
- However, as we focus on the tail area, we will deal with limited sample sizes
- Estimations may be strongly affected



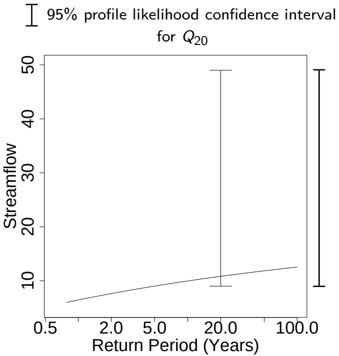


Figure: An Illustration with the Ardières River at Beaujeu

From Theory to Practice

- Using the EVT, we do not need to know X_t's distribution
  (we did not suppose X_t were LN or χ^2 distributed)
- We only need to get closer to the asymptotic hypothesis
- However, as we focus on the tail area, we will deal with limited sample sizes
- Estimations may be strongly affected

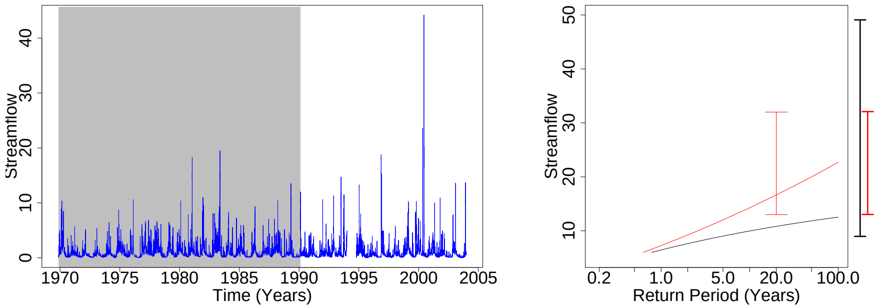


Figure: An Illustration with the Ardières River at Beaujeu

From Theory to Practice

- Using the EVT, we do not need to know X_t's distribution
  (we did not suppose X_t were LN or χ^2 distributed)
- We only need to get closer to the asymptotic hypothesis
- However, as we focus on the tail area, we will deal with limited sample sizes
- Estimations may be strongly affected

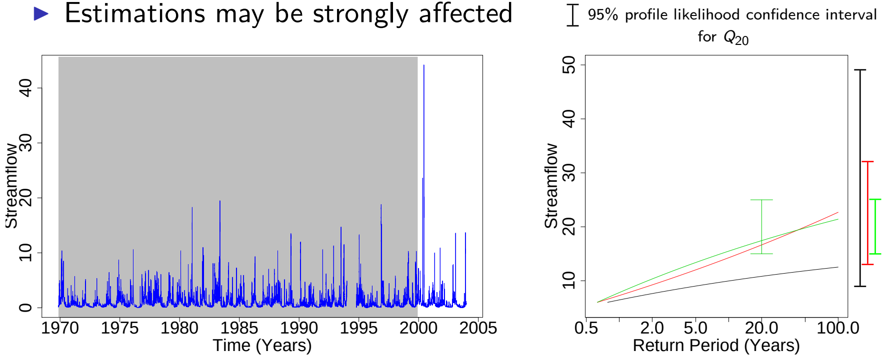
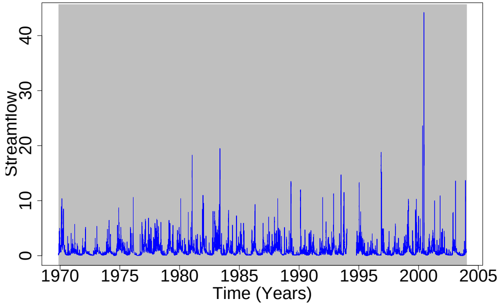


Figure: An Illustration with the Ardières River at Beaujeu

From Theory to Practice

- Using the EVT, we do not need to know X_t's distribution
  (we did not suppose X_t were LN or χ^2 distributed)
- We only need to get closer to the asymptotic hypothesis
- However, as we focus on the tail area, we will deal with limited sample sizes
- Estimations may be strongly affected



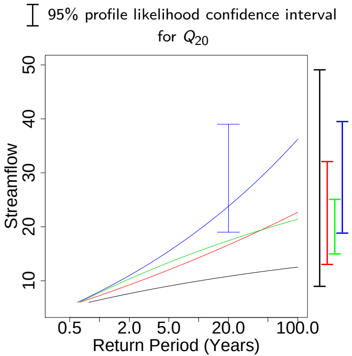


Figure: An Illustration with the Ardières River at Beaujeu

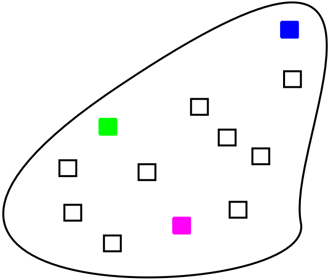
- EVT states which distribution has to be fitted to the data
- But estimations may be still unreliable even with relatively large record lengths
- There is a need to provide models which are robust enough for partially gaged stations (less than 10 years record length)

Objectives

- Provide statistical models which use the EVT potency
- Prove their accuracies on partially gaged stations
- Use efficiently all the data by:
  - using regional information: data from other gaged stations
  - using the data from the target site in a new way: modeling all excess

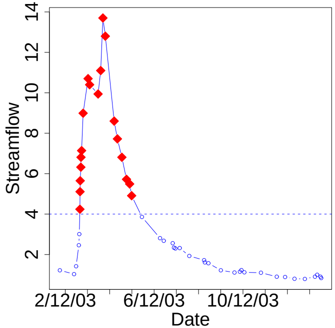
Objectives

- Provide statistical models which use the EVT potency
- Prove their accuracies on partially gaged stations
- Use efficiently all the data by:
  - using regional information: data from other gaged stations
  - using the data from the target site in a new way: modeling all excess



Objectives

- Provide statistical models which use the EVT potency
- Prove their accuracies on partially gaged stations
- Use efficiently all the data by:
    - using regional information: data from other gaged stations
    - using the data from the target site in a new way: modeling all excess



Outline

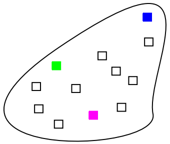
Outline

Index Flood Concept [Dalrymple, 1960]

"Within a homogeneous region, all sites have the same distribution up to a (site dependent) scale factor: the index flood."

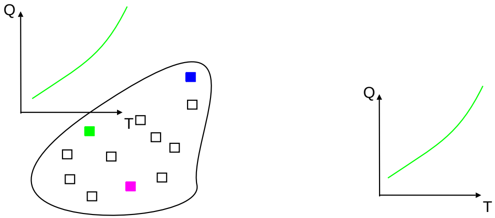
Index Flood Concept [Dalrymple, 1960]

"Within a homogeneous region, all sites have the same distribution up to a (site dependent) scale factor: the index flood."



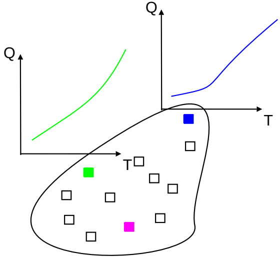
Index Flood Concept [Dalrymple, 1960]

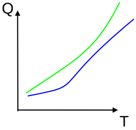
"Within a homogeneous region, all sites have the same distribution up to a (site dependent) scale factor: the index flood."



Index Flood Concept [Dalrymple, 1960]

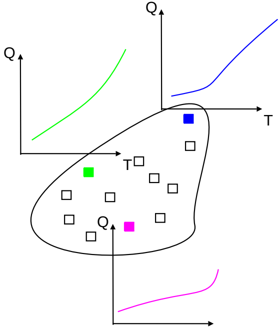
"Within a homogeneous region, all sites have the same distribution up to a (site dependent) scale factor: the index flood."

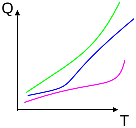




Index Flood Concept [Dalrymple, 1960]

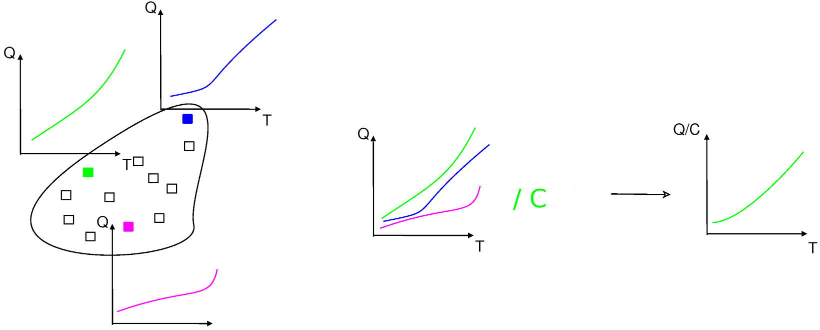
"Within a homogeneous region, all sites have the same distribution up to a (site dependent) scale factor: the index flood."





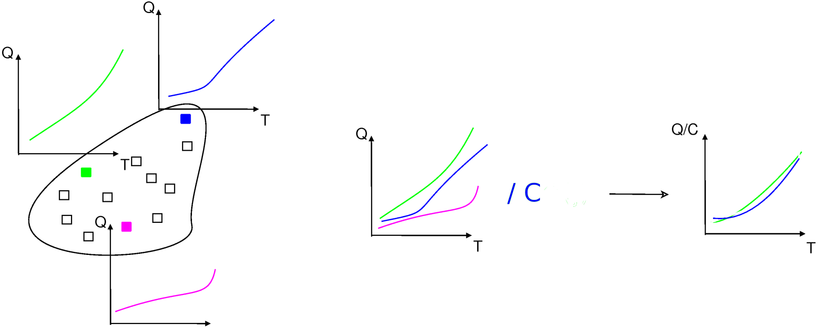
Index Flood Concept [Dalrymple, 1960]

"Within a homogeneous region, all sites have the same distribution up to a (site dependent) scale factor: the index flood."



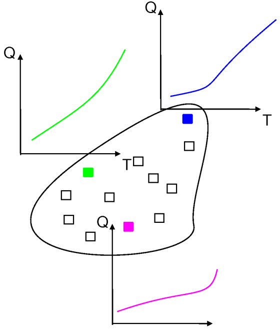
Index Flood Concept [Dalrymple, 1960]

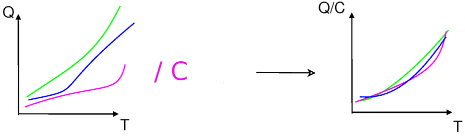
"Within a homogeneous region, all sites have the same distribution up to a (site dependent) scale factor: the index flood."



Index Flood Concept [Dalrymple, 1960]

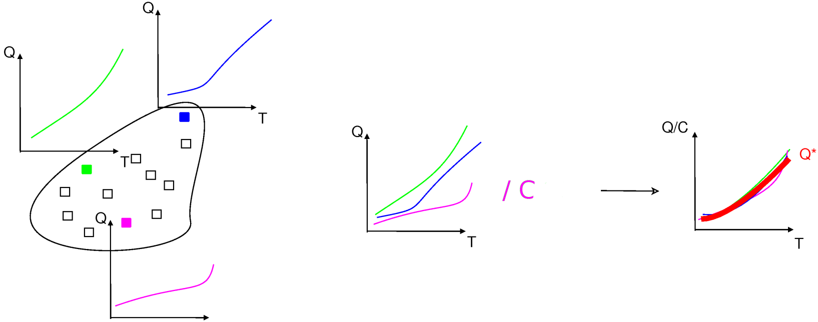
"Within a homogeneous region, all sites have the same distribution up to a (site dependent) scale factor: the index flood."





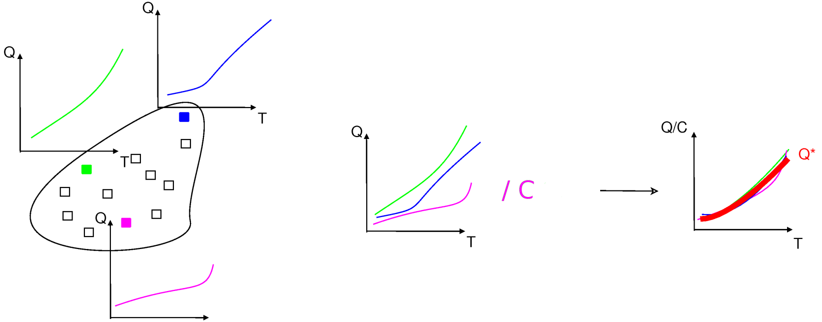
Index Flood Concept [Dalrymple, 1960]

"Within a homogeneous region, all sites have the same distribution up to a (site dependent) scale factor: the index flood."



Index Flood Concept [Dalrymple, 1960]

"Within a homogeneous region, all sites have the same distribution up to a (site dependent) scale factor: the index flood."



- Q^* is the unique distribution, the regional distribution
- Any site distribution within the region is estimated using:

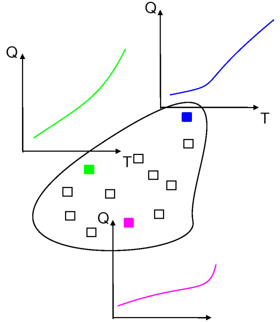
$$Q = CQ^*, \quad C \text{ : target site index flood}$$

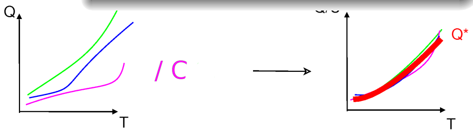
Index Flood Concept [Dalrymple, 1960]

"Within a homogeneous region, all sites have the same distribution up to a (site dependent) scale factor: the index flood."

C estimation

- Using the target site sample, e.g, sample mean, median, . . .
- Using physiographic variables - ungaged sites.





- Q^* is the unique distribution, the regional distribution
- Any site distribution within the region is estimated using:

$$Q = CQ^*, \quad C : \text{target site index flood}$$

The Index Flood Has Drawbacks

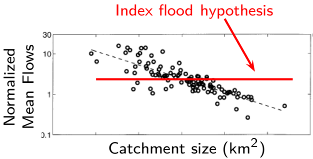
Several authors point out some problems with the IFL approach:

- No likelihood [Katz et al., 2002]
  $\Rightarrow$ What about estimate uncertainties?
- Wrong scale invariance properties [Gupta et al., 1994; Robinson and Sivapalan, 1997]
  $\Rightarrow$ Bias, artificial variance reduction
  - By definition, each observation within the pooling group has the same weight
  - Target site observation may have poor influence

The Index Flood Has Drawbacks

Several authors point out some problems with the IFL approach:

- No likelihood [Katz et al., 2002]
  ⇒ What about estimate uncertainties?
- Wrong scale invariance properties [Gupta et al., 1994; Robinson and Sivapalan, 1997]
  ⇒ Bias, artificial variance reduction



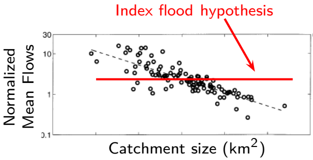
[Robinson and Sivapalan, 1997]

- By definition, each observation within the pooling group has the same weight
- Target site observation may have poor influence

The Index Flood Has Drawbacks

Several authors point out some problems with the IFL approach:

- No likelihood [Katz et al., 2002]
  ⇒ What about estimate uncertainties?
- Wrong scale invariance properties [Gupta et al., 1994; Robinson and Sivapalan, 1997]
  ⇒ Bias, artificial variance reduction



[Robinson and Sivapalan, 1997]

- By definition, each observation within the pooling group has the same weight
- Target site observation may have poor influence

Outline

Use Regional Information in Another Way

Our Goals

- The whole information must be used differently
- Especially, particular attention must be paid to the target site
- Need to relax scale invariance properties
- Need to have a rigorous statistical model

- In order to achieve these objectives, we will consider Bayesian statistics

Use Regional Information in Another Way

Our Goals

- The whole information must be used differently
- Especially, particular attention must be paid to the target site
- Need to relax scale invariance properties
- Need to have a rigorous statistical model

- In order to achieve these objectives, we will consider Bayesian statistics

Presentation of the BAY Model

The BAY model is based on the so-called Bayes' theorem:

$$\pi(\theta \mid x) = \frac{\pi(\theta)\pi(x;\theta)}{\int_\Theta \pi(\theta)\pi(x;\theta)d\theta}$$

where $\pi(\theta|x)$ is the *posterior* distribution, $\pi(\theta)$ is the *prior* distribution and $\pi(x;\theta)$ the likelihood function.

- Regional info. $\Longrightarrow \pi(\theta)$
- Data $\Longrightarrow \pi(x;\theta)$
- Estimations $\Longrightarrow \pi(\theta|x)$

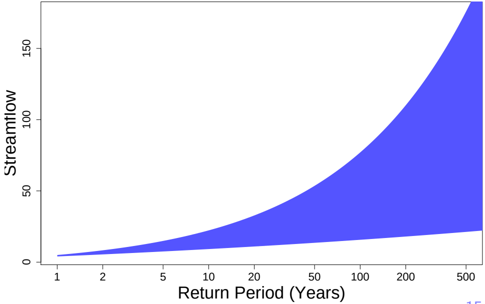
Presentation of the BAY Model

The BAY model is based on the so-called Bayes' theorem:

$$\pi(\theta \mid x) = \frac{\pi(\theta)\pi(x;\theta)}{\int_\Theta \pi(\theta)\pi(x;\theta)d\theta}$$

where $\pi(\theta|x)$ is the *posterior* distribution, $\pi(\theta)$ is the *prior* distribution and $\pi(x;\theta)$ the likelihood function.

- Regional info. $\Longrightarrow \pi(\theta)$
- Data $\Longrightarrow \pi(x;\theta)$
- Estimations $\Longrightarrow \pi(\theta|x)$



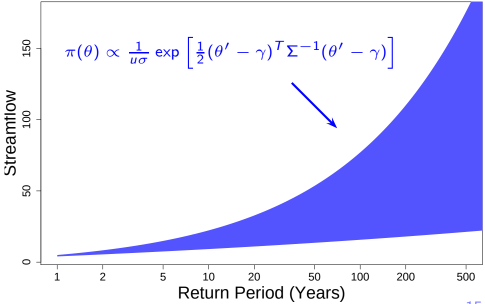
Presentation of the BAY Model

The BAY model is based on the so-called Bayes' theorem:

$$\pi(\theta \mid x) = \frac{\pi(\theta)\pi(x;\theta)}{\int_\Theta \pi(\theta)\pi(x;\theta) d\theta}$$

where $\pi(\theta|x)$ is the *posterior* distribution, $\pi(\theta)$ is the *prior* distribution and $\pi(x;\theta)$ the likelihood function.

- Regional info. $\Longrightarrow \pi(\theta)$
- Data $\Longrightarrow \pi(x;\theta)$
- Estimations $\Longrightarrow \pi(\theta|x)$



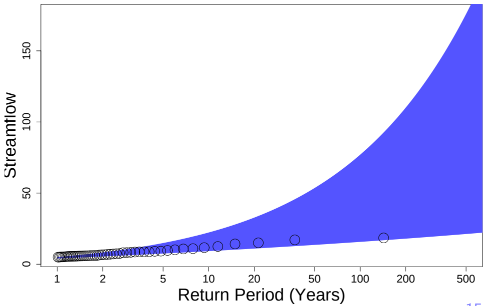
Presentation of the BAY Model

The BAY model is based on the so-called Bayes' theorem:

$$\pi(\theta \mid x) = \frac{\pi(\theta)\pi(x;\theta)}{\int_\Theta \pi(\theta)\pi(x;\theta)d\theta}$$

where $\pi(\theta|x)$ is the *posterior* distribution, $\pi(\theta)$ is the *prior* distribution and $\pi(x;\theta)$ the likelihood function.

- Regional info. $\Longrightarrow \pi(\theta)$
- Data $\Longrightarrow \pi(x;\theta)$
- Estimations $\Longrightarrow \pi(\theta|x)$



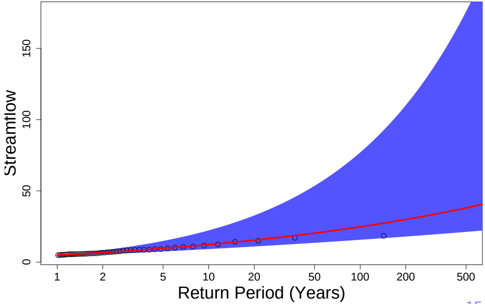
Presentation of the BAY Model

The BAY model is based on the so-called Bayes' theorem:

$$\pi(\theta \mid x) = \frac{\pi(\theta)\pi(x;\theta)}{\int_{\Theta} \pi(\theta)\pi(x;\theta)d\theta}$$

where $\pi(\theta|x)$ is the *posterior* distribution, $\pi(\theta)$ is the *prior* distribution and $\pi(x;\theta)$ the likelihood function.

- Regional info. $\Longrightarrow \pi(\theta)$
- Data $\Longrightarrow \pi(x;\theta)$
- Estimations $\Longrightarrow \pi(\theta|x)$



BAY Model Asset

- IFL supposes that there is a unique regional distribution
- Thus, a unique 3-uplet GPD parameters (u, σ, ξ)
- BAY allows the GPD parameters to lie within a specific range

Figure: Scale invariance properties relaxation.

- These specific ranges are defined using the regional consistency of the GPD parameters

BAY Model Asset

- IFL supposes that there is a unique regional distribution
- Thus, a unique 3-uplet GPD parameters (u, σ, ξ)
- BAY allows the GPD parameters to lie within a specific range

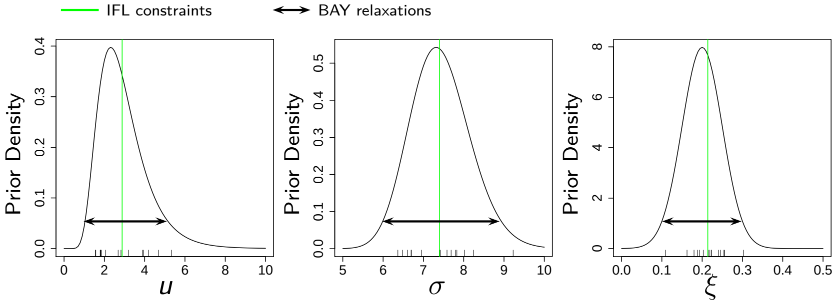


Figure: Scale invariance properties relaxation.

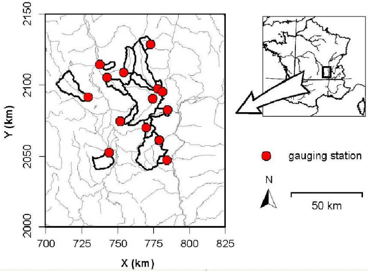
- These specific ranges are defined using the regional consistency of the GPD parameters

Outline

The Case Study

Characteristics of the Pooling Group

- 14 homogeneous stations
  (DIREN Rhône Alpes)
- Record length: 22 to 37 years
- Drainage areas: 32 to 792 km^2
- 3 particular target sites (37 years)



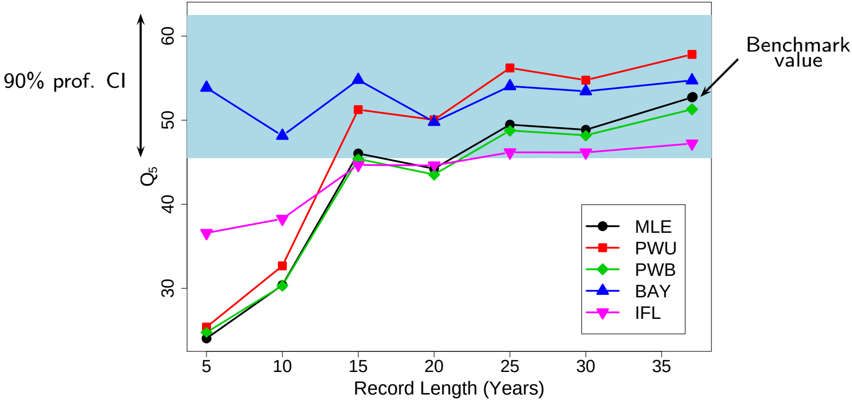
Outline

Performance for Partially Gaged Stations Assessment

Methodology

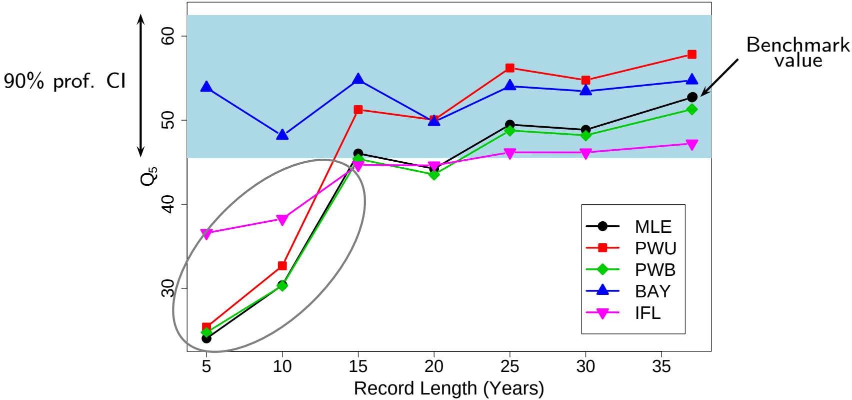
- Benchmark Values: MLE on the whole time series
- Estimators: MLE, PWU, PWB, IFL and BAY
- Inferences on several "sub-time series"
  (5, 10, 15, 20, 25, 30 and 37 years)
- Quantiles of interest: Q_2, Q_5, Q_{10} and Q_{20}

Results



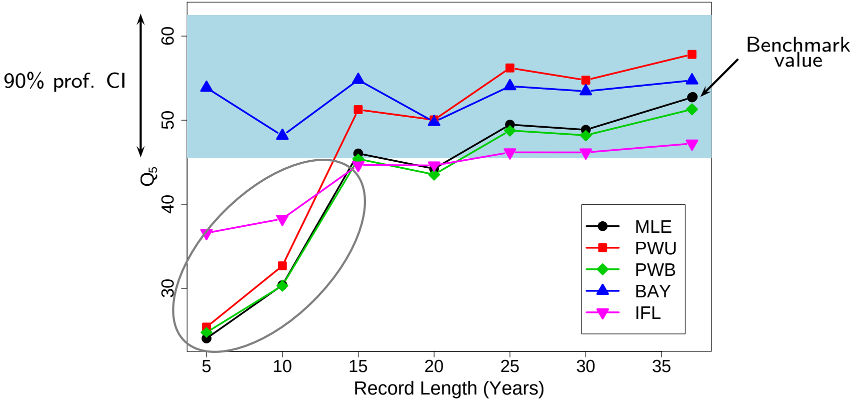
- For record length < 15 years, systematic underestimation for the local and IFL approaches:
  - local estimators: “extreme extremes” have not been “seen” yet
  - IFL estimator: underestimation of the target-site scale factor
- Bayesian approach is the most robust
  (Always in the 90% profile confidence intervals)

Results



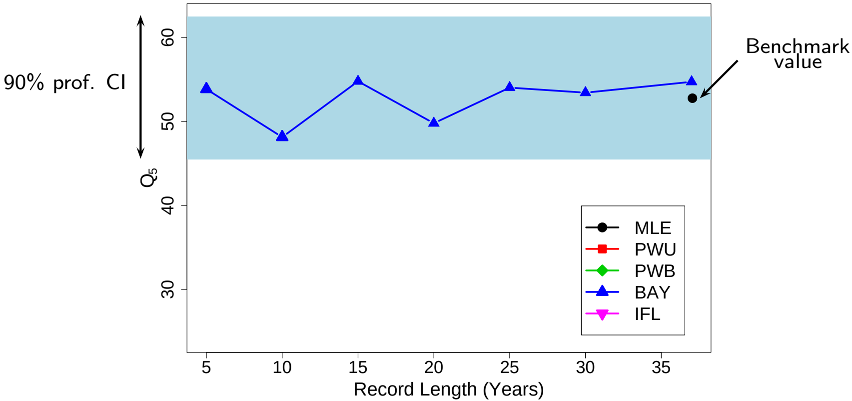
- For record length < 15 years, systematic underestimation for the local and IFL approaches:
  - local estimators: "extreme extremes" have not been "seen" yet
  - IFL estimator: underestimation of the target-site scale factor
- Bayesian approach is the most robust
  (Always in the 90% profile confidence intervals)

Results



- For record length < 15 years, systematic underestimation for the local and IFL approaches:
  - local estimators: "extreme extremes" have not been "seen" yet
  - IFL estimator: underestimation of the target-site scale factor
- Bayesian approach is the most robust
  (Always in the 90% profile confidence intervals)

Results



- For record length < 15 years, systematic underestimation for the local and IFL approaches:
  - local estimators: "extreme extremes" have not been "seen" yet
  - IFL estimator: underestimation of the target-site scale factor
- Bayesian approach is the most robust
  (Always in the 90% profile confidence intervals)

Outline

Assessment of the Homogeneity Degree Impact

Methodology

- We now consider 4 distinct pooling groups:

  He^+ Definitively heterogeneous region
  He Probably heterogeneous region
  Ho Probably homogeneous region
  Ho^+ Definitively homogeneous region

- Inferences are performed on several “sub-time series”
  (5, 10, 15, 20, 25 and 30 years)
- Studied estimators: MLE, IFL and BAY
- Comparison using a score approach
  (measure of both bias and variance for all estimates)
  - Score ≈ 1: good estimator
  - Score ≈ 0: poor estimator

Results

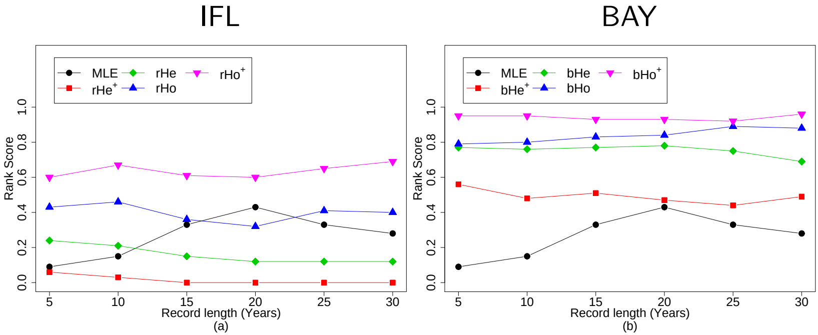


Figure: Evolution of the score as the record length increases.

- To increase homogeneity is more relevant for IFL
- Worst BAY score ≈ best IFL score
- MLE may perform better than IFL

Results

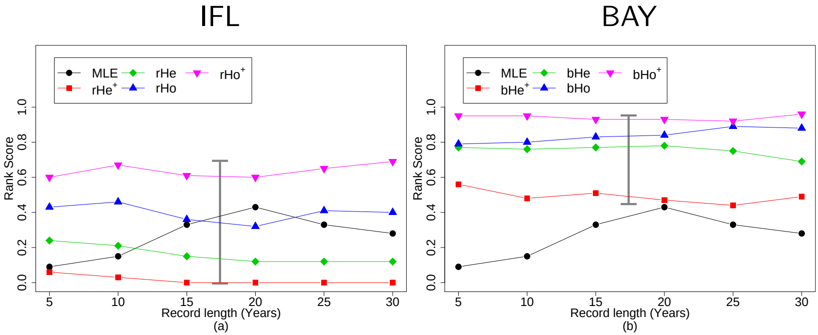


Figure: Evolution of the score as the record length increases.

- To increase homogeneity is more relevant for IFL
- Worst BAY score ≈ best IFL score
- MLE may perform better than IFL

Results

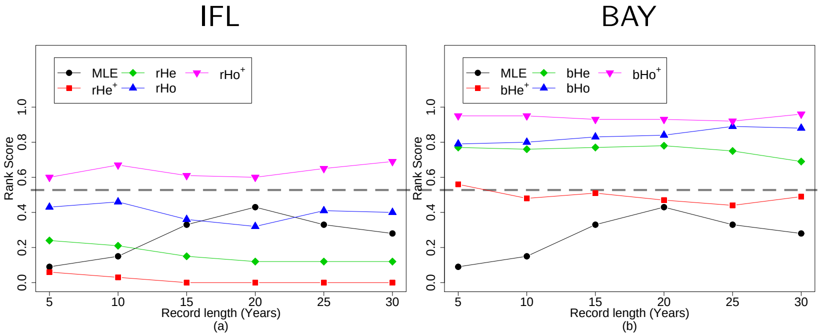


Figure: Evolution of the score as the record length increases.

- To increase homogeneity is more relevant for IFL
- Worst BAY score ≈ best IFL score
- MLE may perform better than IFL

Results

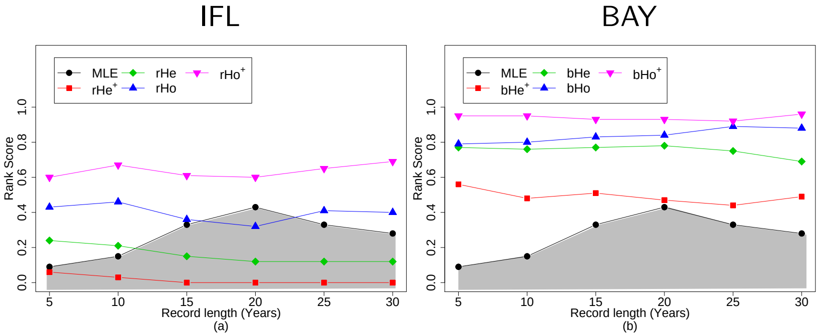


Figure: Evolution of the score as the record length increases.

- To increase homogeneity is more relevant for IFL
- Worst BAY score ≈ best IFL score
- MLE may perform better than IFL

Conclusions

- The empirical Index Flood concept was relaxed using a rigorous Bayesian model
- BAY performance was assessed on a specific French and homogeneous region (for Q_T, $T \leq 20$ years)
- Results show that:
  - BAY is more accurate than IFL (especially when the record length ≤ 10 years)
  - BAY seems more robust than IFL comparing:
    - the sample variability (presence/absence of "extreme extremes")
    - the homogeneity degree (misleading pooling-group building procedure)

Conclusions

- The empirical Index Flood concept was relaxed using a rigorous Bayesian model
- BAY performance was assessed on a specific French and homogeneous region (for Q_T, $T \leq 20$ years)
- Results show that:
  - BAY is more accurate than IFL (especially when the record length $\leq$ 10 years)
  - BAY seems more robust than IFL comparing:
    - the sample variability (presence/absence of "extreme extremes")
    - the homogeneity degree (misleading pooling-group building procedure)

Conclusions

- The empirical Index Flood concept was relaxed using a rigorous Bayesian model
- BAY performance was assessed on a specific French and homogeneous region (for Q_T, $T \leq 20$ years)
- Results show that:
  - BAY is more accurate than IFL (especially when the record length ≤ 10 years)
  - BAY seems more robust than IFL comparing:
    - the sample variability (presence/absence of "extreme extremes")
    - the homogeneity degree (misleading pooling-group building procedure)

Outline

Outline

- We proposed a regional Bayesian model which improves estimations. . . BUT . . .
- Preliminary study shows that for return periods $T \geq 50$, BAY becomes less accurate than IFL

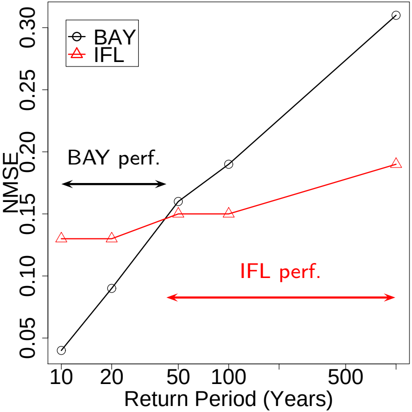


Figure: Evolution of the NMSE as the return period increases.

- We proposed a regional Bayesian model which improves estimations… BUT …
- Preliminary study shows that for return periods $T \geq 50$, BAY becomes less accurate than IFL
- Mainly because of the estimation of the GPD shape parameter, i.e., large variance

Table: *NMSE* for the GPD parameter estimates for the BAY and IFL models.

	u	σ	ξ
BAY	0.036	0.023	0.254
IFL	0.121	0.125	0.104

Idea

Suggest (but not impose) a regional shape parameter ξ_{Fix} suited for the homogeneous region under study.

- We proposed a regional Bayesian model which improves estimations. . . BUT . . .
- Preliminary study shows that for return periods $T \geq 50$, BAY becomes less accurate than IFL
- Mainly because of the estimation of the GPD shape parameter, i.e., large variance

Table: *NMSE* for the GPD parameter estimates for the BAY and IFL models.

	u	σ	ξ
BAY	0.036	0.023	0.254
IFL	0.121	0.125	0.104

Idea

Suggest (but not impose) a regional shape parameter ξ_{Fix} suited for the homogeneous region under study.

New prior distribution: Mixture

$$\pi(\theta) = \begin{cases} (1 - p_\xi)\pi_{\mathrm{in}}(\theta), & \text{for } \theta \in \Theta \backslash \Theta_0 \\ p_\xi \pi_{\xi_{\mathrm{Fix}}}(\theta), & \text{for } \theta \in \Theta_0 \end{cases}$$

$\Theta = \left\{ (u, \sigma, \xi) : u \in \mathbb{R}_*^+, \sigma \in \mathbb{R}_*^+, \xi \in \mathbb{R} \right\}$, i.e., GPD parameter space

$\Theta_0 = \left\{ (u, \sigma, \xi_{\mathrm{Fix}}) : u \in \mathbb{R}_*^+, \sigma \in \mathbb{R}_*^+ \right\}$, i.e., restricted GPD parameter space

- p_ξ controls the "belief degree" about ξ_{Fix}
- Idea: link p_ξ to the homogeneity degree of the region d_{hom}:

New prior distribution: Mixture

$$\pi(\theta) = \begin{cases} (1 - p_\xi)\pi_{\text{in}}(\theta), & \text{for } \theta \in \Theta \backslash \Theta_0 \\ p_\xi \pi_{\xi_{\text{Fix}}}(\theta), & \text{for } \theta \in \Theta_0 \end{cases}$$

$\Theta = \left\{ (u, \sigma, \xi) : u \in \mathbb{R}_*^+, \sigma \in \mathbb{R}_*^+, \xi \in \mathbb{R} \right\}$, i.e., GPD parameter space

$\Theta_0 = \left\{ (u, \sigma, \xi_{\text{Fix}}) : u \in \mathbb{R}_*^+, \sigma \in \mathbb{R}_*^+ \right\}$, i.e., restricted GPD parameter space

- p_ξ controls the “belief degree” about ξ_{Fix}
- Idea: link p_ξ to the homogeneity degree of the region d_{hom}:

New prior distribution: Mixture

$$\pi(\theta) = \begin{cases} (1 - p_\xi)\pi_{\text{in}}(\theta), & \text{for } \theta \in \Theta \backslash \Theta_0 \\ p_\xi \pi_{\xi_{\text{Fix}}}(\theta), & \text{for } \theta \in \Theta_0 \end{cases}$$

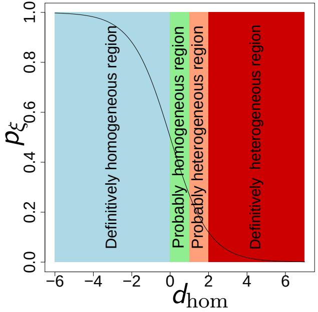
$\Theta = \left\{(u, \sigma, \xi) : u \in \mathbb{R}_*^+, \sigma \in \mathbb{R}_*^+, \xi \in \mathbb{R}\right\}$, i.e., GPD parameter space

$\Theta_0 = \left\{(u, \sigma, \xi_{\text{Fix}}) : u \in \mathbb{R}_*^+, \sigma \in \mathbb{R}_*^+\right\}$, i.e., restricted GPD parameter space

- p_ξ controls the "belief degree" about ξ_{Fix}
- Idea: link p_ξ to the homogeneity degree of the region d_{hom}:

p_ξ parametrization

$$p_\xi = \frac{\exp(-d_{\text{hom}})}{1 + \exp(-d_{\text{hom}})}$$



New prior distribution: Mixture

$$\pi(\theta) = \begin{cases} (1 - p_\xi)\pi_{\mathrm{in}}(\theta), & \text{for } \theta \in \Theta \backslash \Theta_0 \\ p_\xi \pi_{\xi_{\mathrm{Fix}}}(\theta), & \text{for } \theta \in \Theta_0 \end{cases}$$

$\Theta = \left\{(u, \sigma, \xi) : u \in \mathbb{R}_*^+, \sigma \in \mathbb{R}_*^+, \xi \in \mathbb{R}\right\}$, i.e., GPD parameter space

$\Theta_0 = \left\{(u, \sigma, \xi_{\mathrm{Fix}}) : u \in \mathbb{R}_*^+, \sigma \in \mathbb{R}_*^+\right\}$, i.e., restricted GPD parameter space

- p_ξ controls the "belief degree" about ξ_{Fix}
- Idea: link p_ξ to the homogeneity degree of the region d_{hom}:

New Prior Asset

- $\pi(\theta)$ allows to consider ξ_{Fix} with a non-null probability
- but we do not impose $\xi = \xi_{\mathrm{Fix}}$
- Only the target site data will corroborate the ξ_{Fix} relevancy (by sampling in Θ_0 within the MCMC stage.)

Outline

Impact of the Regional Configuration on Estimations

A Study on Simulated Data

Methodology

- Two regional distributions:
  - Heavy tail: Conf1–3 ($\xi > 0$)
  - near Expo.: Conf4–6 ($\xi \approx 0$)
- Several region configurations:
  - Conf1,Conf4: small regions (10 sites) with many data (450 obs.)
  - Conf2,Conf5: large regions (20 sites) with few data (450 obs.)
  - Conf3,Conf6: medium regions (15 sites) with many data (700 obs.)
- Comparison between 3 regional estimators:
  - IFL: Index Flood
  - BAY: First Model
  - REV: Reversible jump approach (i.e. mixture)
- Targeted quantiles $Q_{0.75}$, $Q_{0.95}$ and $Q_{0.995}$ are aimed (e.g. return periods $\approx$ 2, 10 and 100 years)
- Sample sizes of the target site: 10, 25, 40

Finally, 18,000 homogeneous regions were simulated

BAY/IFL Comparison

Table: Evolution of the NRMSE as the region configuration changes.

	$\xi > 0$				$\xi \approx 0$		
	$Q_{0.75}$	$Q_{0.95}$	$Q_{0.995}$		$Q_{0.75}$	$Q_{0.95}$	$Q_{0.995}$
		CONF1		Small Regions		CONF4	
BAY	0.015	0.035	0.101		0.011	0.022	0.054
IFL	0.037	0.038	0.053		0.025	0.027	0.037
		CONF2		Large Regions		CONF5	
BAY	0.015	0.063	0.326		0.012	0.037	0.144
IFL	0.035	0.037	0.049		0.029	0.031	0.039
		CONF3		Medium Regions		CONF6	
BAY	0.012	0.030	0.085		0.011	0.023	0.050
IFL	0.037	0.039	0.049		0.029	0.032	0.042

- $Q_{0.75}$: *BAY is the most accurate model* (Confirms previous results)
- $Q_{0.95}$:
  - BAY more accurate than IFL with small and medium regions
  - IFL more accurate than BAY with large regions
- $Q_{0.995}$: *IFL is the most accurate model*

BAY/IFL Comparison

Table: Evolution of the NRMSE as the region configuration changes.

	$\xi > 0$				$\xi \approx 0$		
	$Q_{0.75}$	$Q_{0.95}$	$Q_{0.995}$		$Q_{0.75}$	$Q_{0.95}$	$Q_{0.995}$
		CONF1		Small Regions		CONF4	
BAY	0.015	0.035	0.101		0.011	0.022	0.054
IFL	0.037	0.038	0.053		0.025	0.027	0.037
		CONF2		Large Regions		CONF5	
BAY	0.015	0.063	0.326		0.012	0.037	0.144
IFL	0.035	0.037	0.049		0.029	0.031	0.039
		CONF3		Medium Regions		CONF6	
BAY	0.012	0.030	0.085		0.011	0.023	0.050
IFL	0.037	0.039	0.049		0.029	0.032	0.042

- $Q_{0.75}$: *BAY is the most accurate model* (Confirms previous results)
- $Q_{0.95}$:
  - BAY more accurate than IFL with small and medium regions
  - IFL more accurate than BAY with large regions
- $Q_{0.995}$: *IFL is the most accurate model*

BAY/IFL Comparison

Table: Evolution of the NRMSE as the region configuration changes.

	$\xi > 0$				$\xi \approx 0$		
	$Q_{0.75}$	$Q_{0.95}$	$Q_{0.995}$		$Q_{0.75}$	$Q_{0.95}$	$Q_{0.995}$
		CONF1		Small Regions		CONF4	
BAY	0.015	0.035	0.101		0.011	0.022	0.054
IFL	0.037	0.038	0.053		0.025	0.027	0.037
		CONF2		Large Regions		CONF5	
BAY	0.015	0.063	0.326		0.012	0.037	0.144
IFL	0.035	0.037	0.049		0.029	0.031	0.039
		CONF3		Medium Regions		CONF6	
BAY	0.012	0.030	0.085		0.011	0.023	0.050
IFL	0.037	0.039	0.049		0.029	0.032	0.042

- $Q_{0.75}$: *BAY is the most accurate model* (Confirms previous results)
- $Q_{0.95}$:
  - BAY more accurate than IFL with small and medium regions
  - IFL more accurate than BAY with large regions
- $Q_{0.995}$: *IFL is the most accurate model*

BAY/IFL Comparison

Table: Evolution of the NRMSE as the region configuration changes.

	$\xi > 0$				$\xi \approx 0$		
	$Q_{0.75}$	$Q_{0.95}$	$Q_{0.995}$		$Q_{0.75}$	$Q_{0.95}$	$Q_{0.995}$
		CONF1		Small Regions		CONF4	
BAY	0.015	0.035	0.101		0.011	0.022	0.054
IFL	0.037	0.038	0.053		0.025	0.027	0.037
		CONF2		Large Regions		CONF5	
BAY	0.015	0.063	0.326		0.012	0.037	0.144
IFL	0.035	0.037	0.049		0.029	0.031	0.039
		CONF3		Medium Regions		CONF6	
BAY	0.012	0.030	0.085		0.011	0.023	0.050
IFL	0.037	0.039	0.049		0.029	0.032	0.042

- $Q_{0.75}$: *BAY is the most accurate model* (Confirms previous results)
- $Q_{0.95}$:
  - BAY more accurate than IFL with small and medium regions
  - IFL more accurate than BAY with large regions
- $Q_{0.995}$: *IFL is the most accurate model*

BAY/IFL Comparison

Table: Evolution of the NRMSE as the region configuration changes.

	$\xi > 0$				$\xi \approx 0$		
	$Q_{0.75}$	$Q_{0.95}$	$Q_{0.995}$		$Q_{0.75}$	$Q_{0.95}$	$Q_{0.995}$
		CONF1		Small Regions		CONF4	
BAY	0.015	0.035	0.101		0.011	0.022	0.054
IFL	0.037	0.038	0.053		0.025	0.027	0.037
		CONF2		Large Regions		CONF5	
BAY	0.015	0.063	0.326		0.012	0.037	0.144
IFL	0.035	0.037	0.049		0.029	0.031	0.039
		CONF3		Medium Regions		CONF6	
BAY	0.012	0.030	0.085		0.011	0.023	0.050
IFL	0.037	0.039	0.049		0.029	0.032	0.042

- $Q_{0.75}$: *BAY is the most accurate model* (Confirms previous results)
- $Q_{0.95}$: *BAY needs "clear-cut" regional information while IFL needs large regions*
  - BAY more accurate than IFL with small and medium regions
  - IFL more accurate than BAY with large regions
- $Q_{0.995}$: *IFL is the most accurate model*

BAY/IFL Comparison

Table: Evolution of the NRMSE as the region configuration changes.

	$\xi > 0$				$\xi \approx 0$		
	$Q_{0.75}$	$Q_{0.95}$	$Q_{0.995}$		$Q_{0.75}$	$Q_{0.95}$	$Q_{0.995}$
		CONF1		Small Regions		CONF4	
BAY	0.015	0.035	0.101		0.011	0.022	0.054
IFL	0.037	0.038	0.053		0.025	0.027	0.037
		CONF2		Large Regions		CONF5	
BAY	0.015	0.063	0.326		0.012	0.037	0.144
IFL	0.035	0.037	0.049		0.029	0.031	0.039
		CONF3		Medium Regions		CONF6	
BAY	0.012	0.030	0.085		0.011	0.023	0.050
IFL	0.037	0.039	0.049		0.029	0.032	0.042

- $Q_{0.75}$: *BAY is the most accurate model* (Confirms previous results)
- $Q_{0.95}$: *BAY needs "clear-cut" regional information while IFL needs large regions*
  - BAY more accurate than IFL with small and medium regions
  - IFL more accurate than BAY with large regions
- $Q_{0.995}$: *IFL is the most accurate model*

BAY/REV Comparison

Table: Evolution of the NRMSE as the region configuration changes.

	$\xi > 0$				$\xi \approx 0$		
	$Q_{0.75}$	$Q_{0.95}$	$Q_{0.995}$		$Q_{0.75}$	$Q_{0.95}$	$Q_{0.995}$
		CONF1		Small Regions		CONF4	
BAY	0.015	0.035	0.101		0.011	0.022	0.054
REV	0.014	0.026	0.047		0.011	0.018	0.033
		CONF2		Large Regions		CONF5	
BAY	0.015	0.063	0.326		0.012	0.037	0.144
REV	0.011	0.024	0.077		0.010	0.016	0.030
		CONF3		Medium Regions		CONF6	
BAY	0.012	0.030	0.085		0.011	0.023	0.050
REV	0.011	0.019	0.036		0.011	0.019	0.032

- $Q_{0.75}$: BAY and REV yield similar results
- $Q_{0.95}, Q_{0.995}$: REV is the most accurate model (smaller variance)

Impact of ξ_{Fix}

ξ_{Fix} suggestion leads to more accurate estimations for large quantiles while preserving the same accuracy for smaller quantiles

BAY/REV Comparison

Table: Evolution of the NRMSE as the region configuration changes.

	$\xi > 0$				$\xi \approx 0$		
	$Q_{0.75}$	$Q_{0.95}$	$Q_{0.995}$		$Q_{0.75}$	$Q_{0.95}$	$Q_{0.995}$
		CONF1		Small Regions		CONF4	
BAY	0.015	0.035	0.101		0.011	0.022	0.054
REV	0.014	0.026	0.047		0.011	0.018	0.033
		CONF2		Large Regions		CONF5	
BAY	0.015	0.063	0.326		0.012	0.037	0.144
REV	0.011	0.024	0.077		0.010	0.016	0.030
		CONF3		Medium Regions		CONF6	
BAY	0.012	0.030	0.085		0.011	0.023	0.050
REV	0.011	0.019	0.036		0.011	0.019	0.032

- $Q_{0.75}$: BAY and REV yield similar results
- $Q_{0.95}, Q_{0.995}$: REV is the most accurate model (smaller variance)

Impact of ξ_{Fix}

ξ_{Fix} suggestion leads to more accurate estimations for large quantiles while preserving the same accuracy for smaller quantiles

BAY/REV Comparison

Table: Evolution of the NRMSE as the region configuration changes.

	$\xi > 0$				$\xi \approx 0$		
	$Q_{0.75}$	$Q_{0.95}$	$Q_{0.995}$		$Q_{0.75}$	$Q_{0.95}$	$Q_{0.995}$
		CONF1		Small Regions		CONF4	
BAY	0.015	0.035	0.101		0.011	0.022	0.054
REV	0.014	0.026	0.047		0.011	0.018	0.033
		CONF2		Large Regions		CONF5	
BAY	0.015	0.063	0.326		0.012	0.037	0.144
REV	0.011	0.024	0.077		0.010	0.016	0.030
		CONF3		Medium Regions		CONF6	
BAY	0.012	0.030	0.085		0.011	0.023	0.050
REV	0.011	0.019	0.036		0.011	0.019	0.032

- $Q_{0.75}$: BAY and REV yield similar results
- $Q_{0.95}, Q_{0.995}$: REV is the most accurate model (smaller variance)

Impact of ξ_{Fix}

ξ_{Fix} suggestion leads to more accurate estimations for large quantiles while preserving the same accuracy for smaller quantiles

BAY/REV Comparison

Table: Evolution of the NRMSE as the region configuration changes.

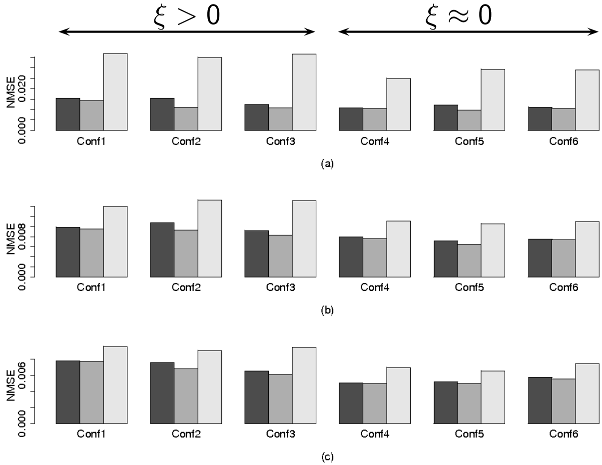
	$\xi > 0$				$\xi \approx 0$		
	$Q_{0.75}$	$Q_{0.95}$	$Q_{0.995}$		$Q_{0.75}$	$Q_{0.95}$	$Q_{0.995}$
		CONF1		Small Regions		CONF4	
BAY	0.015	0.035	0.101		0.011	0.022	0.054
REV	0.014	0.026	0.047		0.011	0.018	0.033
		CONF2		Large Regions		CONF5	
BAY	0.015	0.063	0.326		0.012	0.037	0.144
REV	0.011	0.024	0.077		0.010	0.016	0.030
		CONF3		Medium Regions		CONF6	
BAY	0.012	0.030	0.085		0.011	0.023	0.050
REV	0.011	0.019	0.036		0.011	0.019	0.032

- $Q_{0.75}$: BAY and REV yield similar results
- $Q_{0.95}, Q_{0.995}$: REV is the most accurate model (smaller variance)

Impact of ξ_{Fix}

ξ_{Fix} suggestion leads to more accurate estimations for large quantiles while preserving the same accuracy for smaller quantiles

Global Comparison



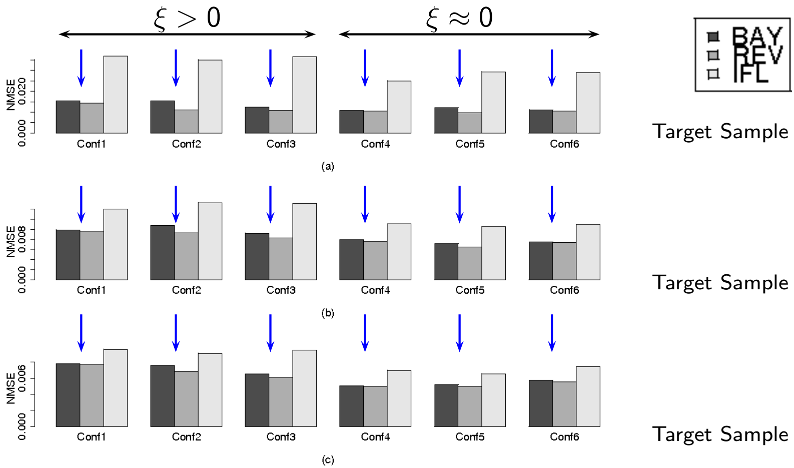
Target Sample Size: 10

Target Sample Size: 25

Target Sample Size: 40

- $Q_{0.75}$:
- $Q_{0.95}$, $Q_{0.995}$:
  - Few data at the target site: REV is the most efficient model
  - Otherwise, REV and IFL are mainly similar

Global Comparison



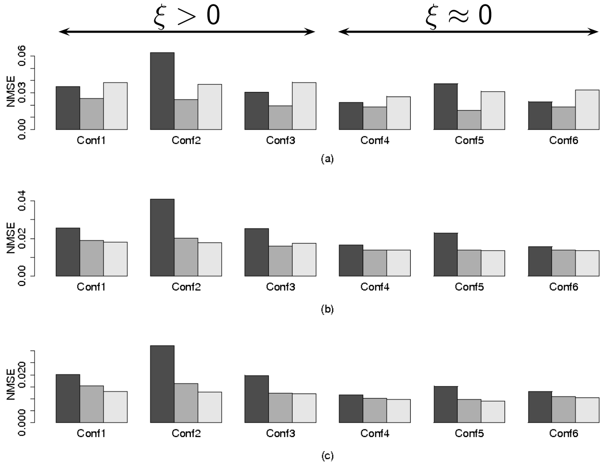
Target Sample Size: 10

Target Sample Size: 25

Target Sample Size: 40

- $Q_{0.75}$: Bayesian approaches are the most accurate
- $Q_{0.95}$, $Q_{0.995}$:
  - Few data at the target site: REV is the most efficient model
  - Otherwise, REV and IFL are mainly similar

Global Comparison



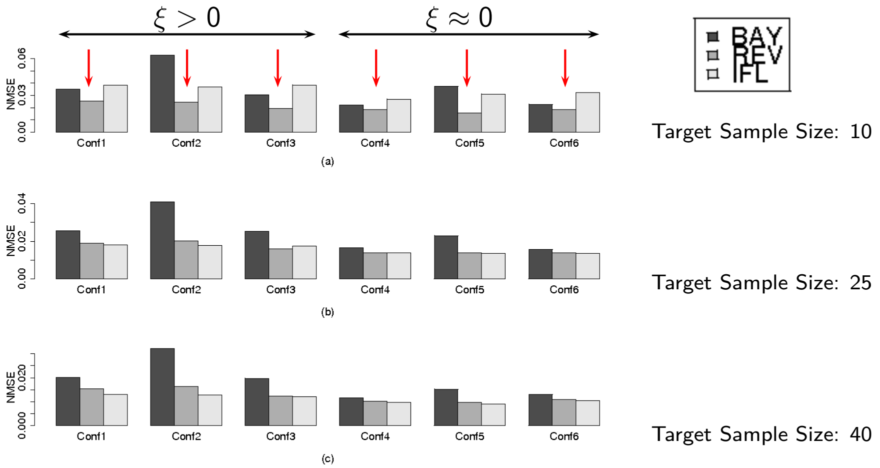
Target Sample Size: 10

Target Sample Size: 25

Target Sample Size: 40

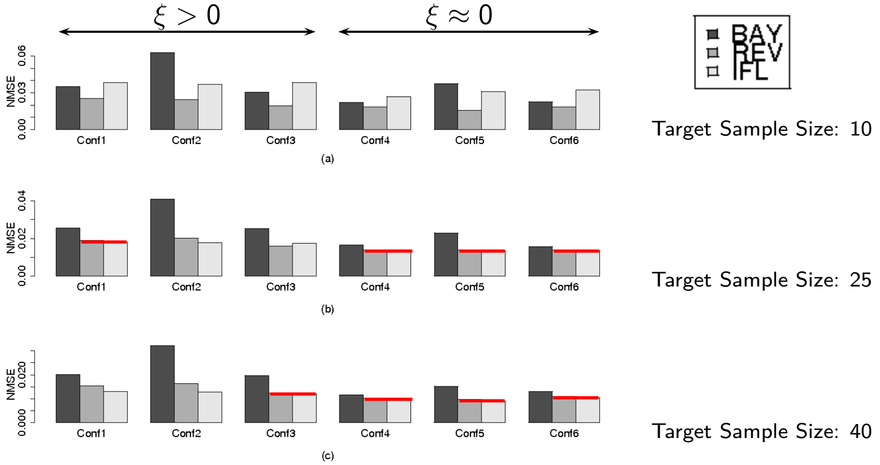
- $Q_{0.75}$: Bayesian approaches are the most accurate
- $Q_{0.95}$, $Q_{0.995}$:
  - Few data at the target site: REV is the most efficient model
  - Otherwise, REV and IFL are mainly similar

Global Comparison



- $Q_{0.75}$: Bayesian approaches are the most accurate
- $Q_{0.95}$, $Q_{0.995}$:
  - Few data at the target site: REV is the most efficient model
  - Otherwise, REV and IFL are mainly similar

Global Comparison



- $Q_{0.75}$: Bayesian approaches are the most accurate
- $Q_{0.95}$, $Q_{0.995}$:
  - Few data at the target site: REV is the most efficient model
  - Otherwise, REV and IFL are mainly similar

Conclusions

- REV generalizes BAY to derive more robust extrapolations
- A relevant regional shape parameter ξ_{Fix} was proposed
- For Q_T, $T \leq 20$ years, REV has the same performance as BAY
- For Q_T, $T > 20$ years, there are two cases:
  - Record length < 10 years: REV is the most accurate model
  - Otherwise, REV and IFL yield to the same results

Outline

Outline

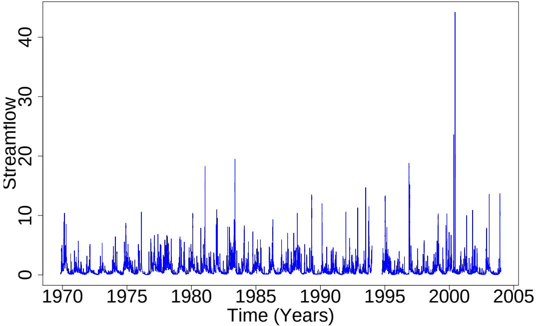
Motivations

- Previous models distinguish the target site from the other gaging stations
- Such procedures lead to more accurate estimations
  $\Longrightarrow$ Focusing on the data from the target site has a large impact on quantile estimations
- It seems to be a logical stage before looking at other kind of information (e.g. regional, historical information)

Motivations

- Previous models distinguish the target site from the other gaging stations
- Such procedures lead to more accurate estimations
  $\Longrightarrow$ Focusing on the data from the target site has a large impact on quantile estimations
- It seems to be a logical stage before looking at other kind of information (e.g. regional, historical information)

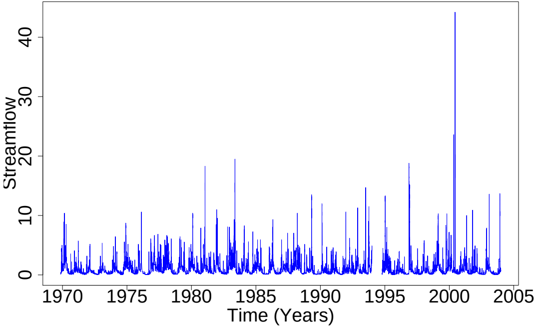
Has the information from the target site been efficiently used?



Motivations

- Previous models distinguish the target site from the other gaging stations
- Such procedures lead to more accurate estimations
  $\Longrightarrow$ Focusing on the data from the target site has a large impact on quantile estimations
- It seems to be a logical stage before looking at other kind of information (e.g. regional, historical information)

Has the information from the target site been efficiently used?

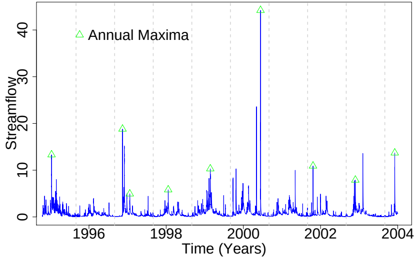


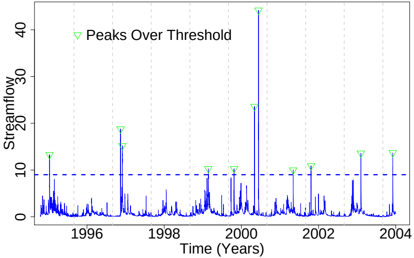
How local information is used?

From a time series to...

- Annual Maxima
- POT
- Suggestion

- Time series $\Rightarrow$ 1 obs/year
- Time series $\Rightarrow$ λ obs/year
- Time series $\Rightarrow$ all exceedances



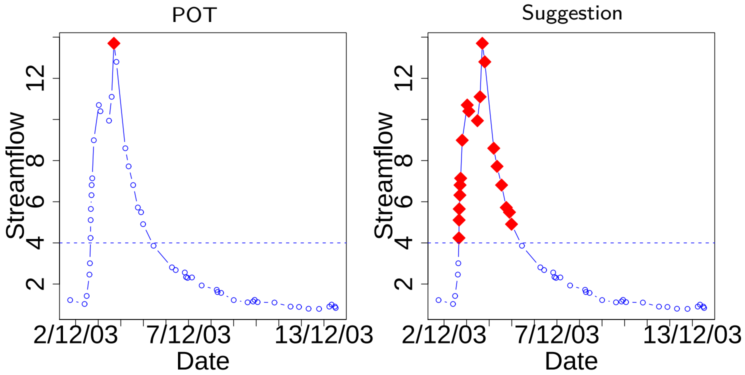


How local information is used?

From a time series to...

- Annual Maxima
- POT
- Suggestion

- Time series $\Rightarrow$ 1 obs/year
- Time series $\Rightarrow$ λ obs/year
- Time series $\Rightarrow$ all exceedances

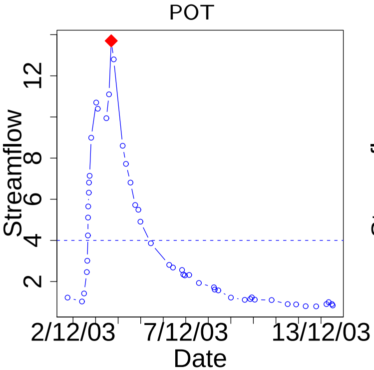


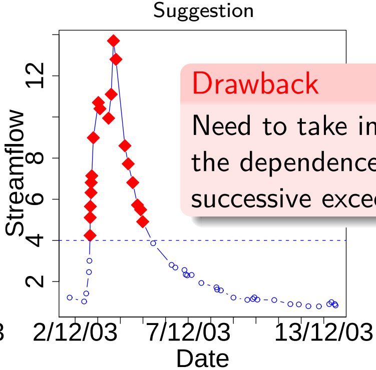
How local information is used?

From a time series to...

- Annual Maxima
- POT
- Suggestion

- Time series $\Rightarrow$ 1 obs/year
- Time series $\Rightarrow$ λ obs/year
- Time series $\Rightarrow$ all exceedances





Drawback

Need to take into account the dependence between successive exceedances

Outline

Presentation of the Model

X_t gives information about the X_{t+1} value.

Hypothesis

- First order Markov chain approach

$$L(y_1, \ldots, y_n; \phi, \psi) = \frac{\prod_{i=2}^{n} f(y_i, y_{i-1}; \phi, \psi)}{\prod_{i=2}^{n-1} f(y_i; \phi)}$$

- $f(\cdot; \phi) \sim GPD$
- $f(\cdot, \cdot; \phi, \psi) \sim MEVD$ (bivariate)
  $f(\cdot, \cdot; \phi, \psi)$ depends on a dependence structure V
- V governs "how Y_t and Y_{t+1} are related"

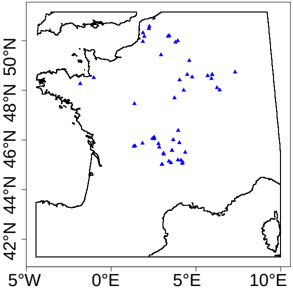
Outline

Comparison between all Markovian Models

- 50 daily discharge stations (B. Renard PhD)
- Record lengths: 39–97 years
- Catchment Areas: 13–5660 km^2
- 6 different dependence structures V:

  symmetric: *log*, *nlog*, *mix*

  asymmetric: *alog*, *anlog*, *amix*



- Mixed models are more accurate than log. and nlog. families
- The asymmetric models seem more efficient than symmetric ones BUT this is not such a surprise

Comparison between all Markovian Models

- 50 daily discharge stations (B. Renard PhD)
- Record lengths: 39–97 years
- Catchment Areas: 13–5660 km^2
- 6 different dependence structures V:

  symmetric: *log*, *nlog*, *mix*

  asymmetric: *alog*, *anlog*, *amix*

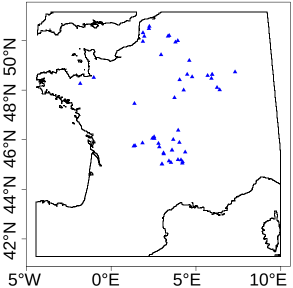


Table: Evolution of the Q_{50} *NMSE* as the record length increases.

	5 years	10 years	15 years
log	0.66	0.42	0.32
nlog	0.24	0.15	0.12
mix	0.14	0.08	0.06
alog	0.41	0.24	0.17
anlog	0.21	0.12	0.09
amix	0.12	0.06	0.05

- Mixed models are more accurate than log. and nlog. families
- The asymmetric models seem more efficient than symmetric ones BUT this is not such a surprise. . .

Comparison between all Markovian Models

- 50 daily discharge stations (B. Renard PhD)
- Record lengths: 39–97 years
- Catchment Areas: 13–5660 km^2
- 6 different dependence structures V:

  symmetric: *log*, *nlog*, *mix*

  asymmetric: *alog*, *anlog*, *amix*

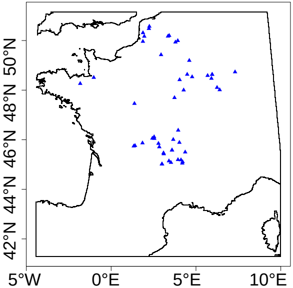
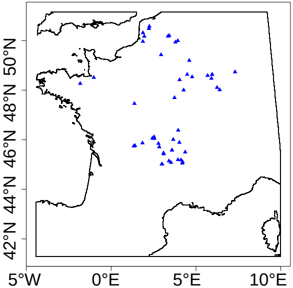


Table: Evolution of the Q_{50} *NMSE* as the record length increases.

	5 years	10 years	15 years
log	0.66	0.42	0.32
nlog	0.24	0.15	0.12
mix	0.14	0.08	0.06
alog	0.41	0.24	0.17
anlog	0.21	0.12	0.09
amix	0.12	0.06	0.05

- Mixed models are more accurate than log. and nlog. families
- The asymmetric models seem more efficient than symmetric ones BUT this is not such a surprise. . .

Comparison between all Markovian Models



- 50 daily discharge stations (B. Renard PhD)
- Record lengths: 39–97 years
- Catchment Areas: 13–5660 km^2
- 6 different dependence structures V:

  symmetric: *log*, *nlog*, *mix*

  asymmetric: *alog*, *anlog*, *amix*

Table: Evolution of the Q_{50} *NMSE* as the record length increases.

	5 years	10 years	15 years
log	0.66	0.42	0.32
nlog	0.24	0.15	0.12
mix	0.14	0.08	0.06
alog	0.41	0.24	0.17
anlog	0.21	0.12	0.09
amix	0.12	0.06	0.05

- Mixed models are more accurate than log. and nlog. families
- The asymmetric models seem more efficient than symmetric ones BUT this is not such a surprise. . .

Comparison between all Markovian Models

- 50 daily discharge stations (B. Renard PhD)
- Record lengths: 39–97 years
- Catalogue Areas: 13–5660 km^2
- 6 different dependence structures V:

  symmetric: *log*, *nlog*, *mix*

  asymmetric: *alog*, *anlog*, *amix*

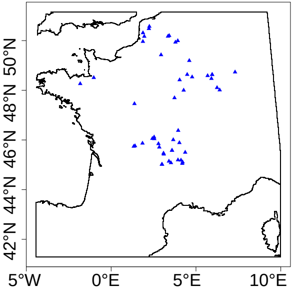


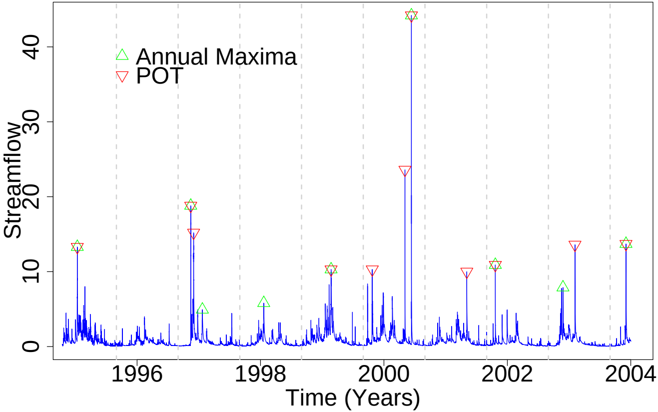
Table: Evolution of the Q_{50} *NMSE* as the record length increases.

	5 years	10 years	15 years
log	0.66	0.42	0.32
nlog	0.24	0.15	0.12
mix	0.14	0.08	0.06
alog	0.41	0.24	0.17
anlog	0.21	0.12	0.09
amix	0.12	0.06	0.05

- Mixed models are more accurate than log. and nlog. families
- The asymmetric models seem more efficient than symmetric ones BUT this is not such a surprise...

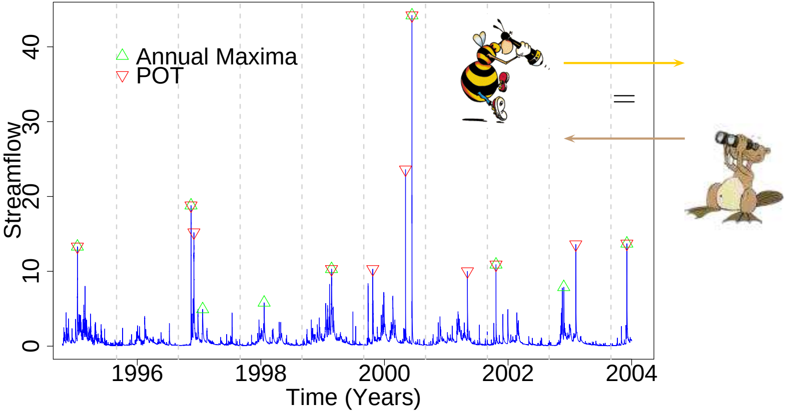
Time reversibility?

- For stationary time series and AM/POT approach,
- When modelling all exceedances, IT HAS!
  Flood hydrographs are clearly asymmetric
- Thus, Markovian models should respect this asymmetry



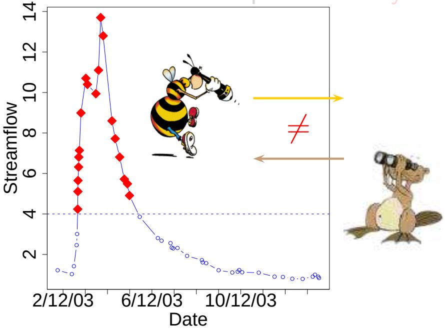
Time reversibility?

- For stationary time series and AM/POT approach, the time direction has no importance
- When modelling all exceedances, IT HAS!
  Flood hydrographs are clearly asymmetric
- Thus, Markovian models should respect this asymmetry



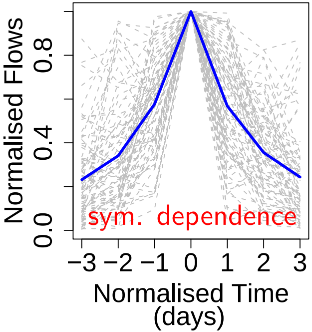
Time reversibility?

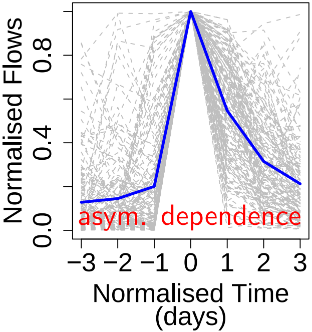
- For stationary time series and AM/POT approach, the time direction has no importance
- When modelling all exceedances, IT HAS!
  Flood hydrographs are clearly asymmetric
- Thus, Markovian models should respect this asymmetry



Time reversibility?

- For stationary time series and AM/POT approach, the time direction has no importance
- When modelling all exceedances, IT HAS!
  Flood hydrographs are clearly asymmetric
- Thus, Markovian models should respect this asymmetry





Comparison between *amix* and classical estimators

Table: Several characteristics of the *amix*, *MLE*, *PWU* and *PWB* estimators for Q_{50} estimation as the record length increases.

Model	5 years			10 years			15 years		
	NBIAS	*SD*	*NMSE*	*NBIAS*	*SD*	*NMSE*	*NBIAS*	*SD*	*NMSE*
amix	−0.06	0.33	0.12	−0.05	0.25	0.07	−0.04	0.21	0.05
MLE	−0.13	0.50	0.27	−0.14	0.36	0.14	−0.13	0.29	0.10
PWU	0.08	0.55	0.31	−0.01	0.39	0.15	−0.03	0.31	0.10
PWB	−0.07	0.45	0.21	−0.10	0.33	0.12	−0.11	0.27	0.09

- *amix* has the same biases than classical estimators
- but *amix* has the smallest variance (as more data are inferred)
- Thus, *amix* has the smallest MSE
- This is true whatever the return period is!

Comparison between *amix* and classical estimators

Table: Several characteristics of the *amix*, *MLE*, *PWU* and *PWB* estimators for Q_{50} estimation as the record length increases.

Model	5 years			10 years			15 years		
	NBIAS	*SD*	*NMSE*	*NBIAS*	*SD*	*NMSE*	*NBIAS*	*SD*	*NMSE*
amix	−0.06	0.33	0.12	−0.05	0.25	0.07	−0.04	0.21	0.05
MLE	−0.06	0.50	0.27	−0.14	0.36	0.14	−0.13	0.29	0.10
PWU	0.08	0.55	0.31	−0.01	0.39	0.15	−0.03	0.31	0.10
PWB	−0.07	0.45	0.21	−0.10	0.33	0.12	−0.11	0.27	0.09

- *amix* has the same biases than classical estimators
  $\Longrightarrow$ *amix* dependence structure seems to be suited to model two consecutive flood observations
- but *amix* has the smallest variance (as more data are inferred)
- Thus, *amix* has the smallest MSE
- This is true whatever the return period is!

Comparison between *amix* and classical estimators

Table: Several characteristics of the *amix*, *MLE*, *PWU* and *PWB* estimators for Q_{50} estimation as the record length increases.

Model	5 years			10 years			15 years		
	NBIAS	*SD*	*NMSE*	*NBIAS*	*SD*	*NMSE*	*NBIAS*	*SD*	*NMSE*
amix	−0.06	0.33	0.12	−0.05	0.25	0.07	−0.04	0.21	0.05
MLE	−0.13	0.50	0.27	−0.14	0.36	0.14	−0.13	0.29	0.10
PWU	0.08	0.55	0.31	−0.01	0.39	0.15	−0.03	0.31	0.10
PWB	−0.07	0.45	0.21	−0.10	0.33	0.12	−0.11	0.27	0.09

- *amix* has the same biases than classical estimators
- but *amix* has the smallest variance (as more data are inferred)
- Thus, *amix* has the smallest MSE
- This is true whatever the return period is!

Comparison between *amix* and classical estimators

Table: Several characteristics of the *amix*, *MLE*, *PWU* and *PWB* estimators for Q_{50} estimation as the record length increases.

Model	5 years			10 years			15 years		
	NBIAS	*SD*	*NMSE*	*NBIAS*	*SD*	*NMSE*	*NBIAS*	*SD*	*NMSE*
amix	−0.06	0.33	0.12	−0.05	0.25	0.07	−0.04	0.21	0.05
MLE	−0.13	0.50	0.27	−0.14	0.36	0.14	−0.13	0.29	0.10
PWU	0.08	0.55	0.31	−0.01	0.39	0.15	−0.03	0.31	0.10
PWB	−0.07	0.45	0.21	−0.10	0.33	0.12	−0.11	0.27	0.09

- *amix* has the same biases than classical estimators
- but *amix* has the smallest variance (as more data are inferred)
- Thus, *amix* has the smallest MSE
- This is true whatever the return period is!

Comparison between *amix* and classical estimators

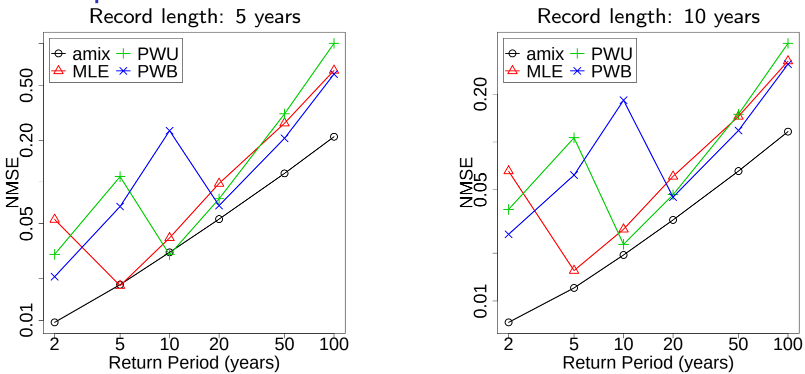


Figure: Evolution of the NMSE as the return period increases.

- *amix* has the same biases than classical estimators
- but *amix* has the smallest variance (as more data are inferred)
- Thus, *amix* has the smallest MSE
- This is true whatever the return period is!

An Attempt to Estimate Flood Hydrograph

- Flood peak were accurately estimated BUT...
- Severity of a flood event is not only defined by its flood peak
- Volume, duration and the shape of the hydrograph play a major role
- The proposed model can estimate design-flood hydrographs

Figure: Estimation of the normalized flood hydrograph.The Rance river at Guenroc (left) and the Loire river at Villerest (right).

An Attempt to Estimate Flood Hydrograph

- Flood peak were accurately estimated BUT...
- Severity of a flood event is not only defined by its flood peak
- Volume, duration and the shape of the hydrograph play a major role
- The proposed model can estimate design-flood hydrographs

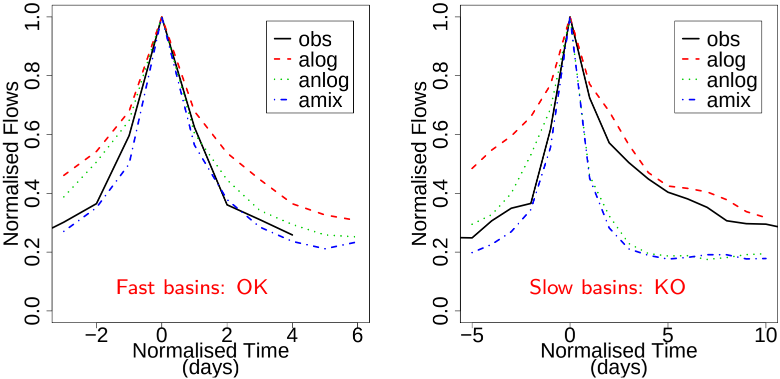


Figure: Estimation of the normalized flood hydrograph.The Rance river at Guenroc (left) and the Loire river at Villerest (right).

Conclusions

- A purely local model was proposed
- Most of the data from the target site are used
- MEVT can be applied in practice
- Our results show that it is worthy
  (especially when $T > 50$ years)
- Markovian models are attractive as they can predict flood peaks, durations, volumes...
- But there is still a lot of work to be done
  (for flood hydrograph estimation)

Outline

Regional Models

Conclusions

- The whole information is used more rationally than the IFL does (distinguishing the target site data)
- Models are based on well-established theory (Bayes)
- Bayesian models are definitively adapted to partially gaged stations
- BAY is accurate for return periods such as $T \leq 20$ years
- REV generalizes BAY
- And allows accurate estimations for $T > 20$ years

Perspectives

- Take into account the potential inter-site dependence (copula, MEVT, pairwise likelihood, ...)
- Test other parametrizations for p_ξ as a function of d_{hom} (maybe try relationships that are more clear-cut)
- Extend these models to the non-stationary case

Purely Local Model

Conclusions

- Need not to waste the target site data
- Uses all exceedances by applying a theoretically based structure for the extremal dependency
- This approach seems to be more accurate than conventional local estimators (mse divided by 3... when $T > 50$ years)
- Modeling flood volumes and duration is now available

Perspectives

- Improve the estimation of flood volume and duration (test other V choices)
- Link catchment characteristics to the V functions
- Maybe consider higher order Markov chains
- Or other "lag conditioning" values (time step adapted to the basin dynamic)


Thank you very much for your attention!