

Elicitation pour l'analyse
bayésienne des événements
hydrologiques extrêmes

Eric Parent, Jacques Bernier

Eric Parent, Jacques Bernier

Laboratoire de gestion du risque en
sciences de l'eau (GRESE)

Ecole Nationale du Génie Rural des
Eaux et des Forêts (ENGREF)

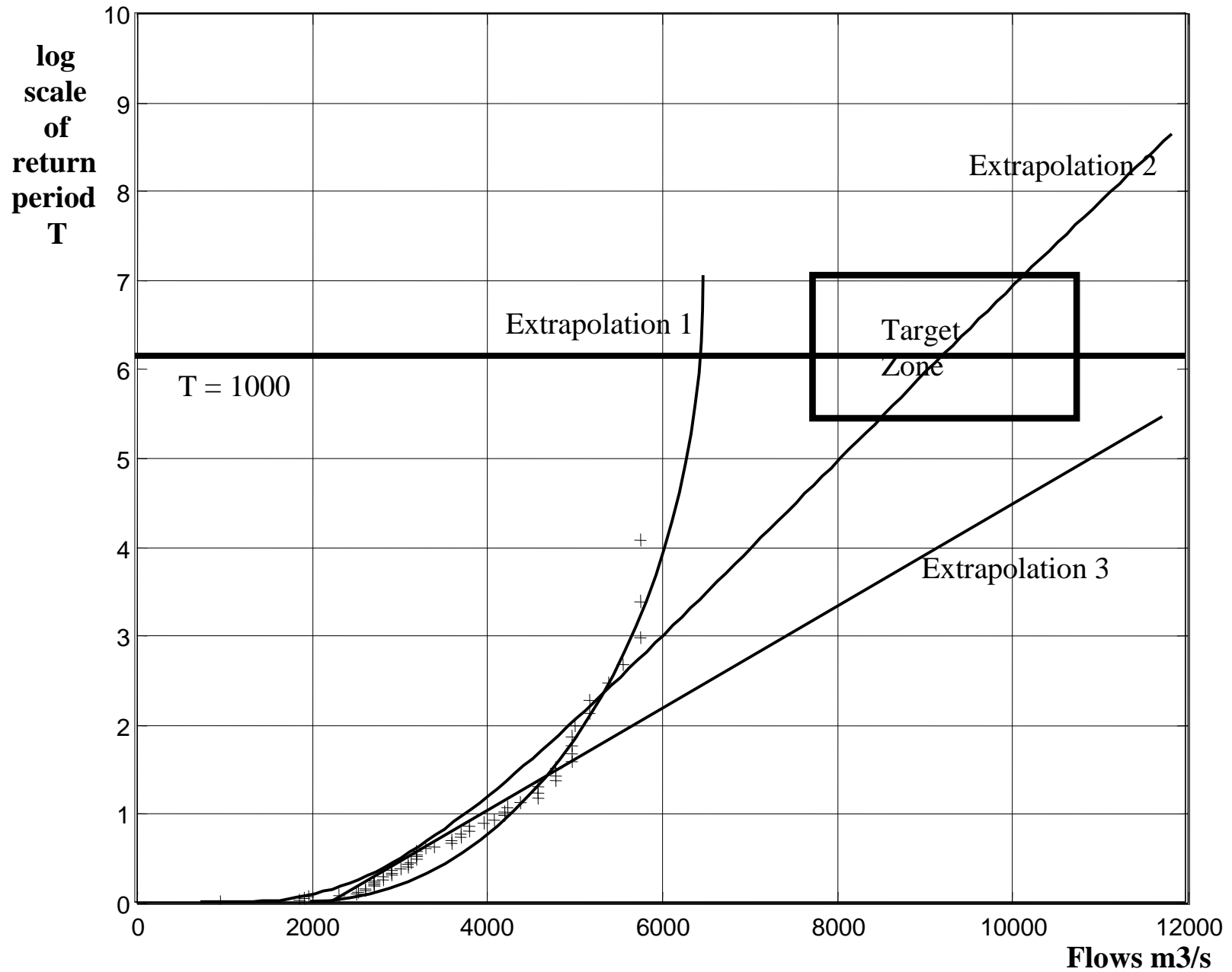
19 av du Maine 75732 PARIS Cedex 15

parent@engref.fr

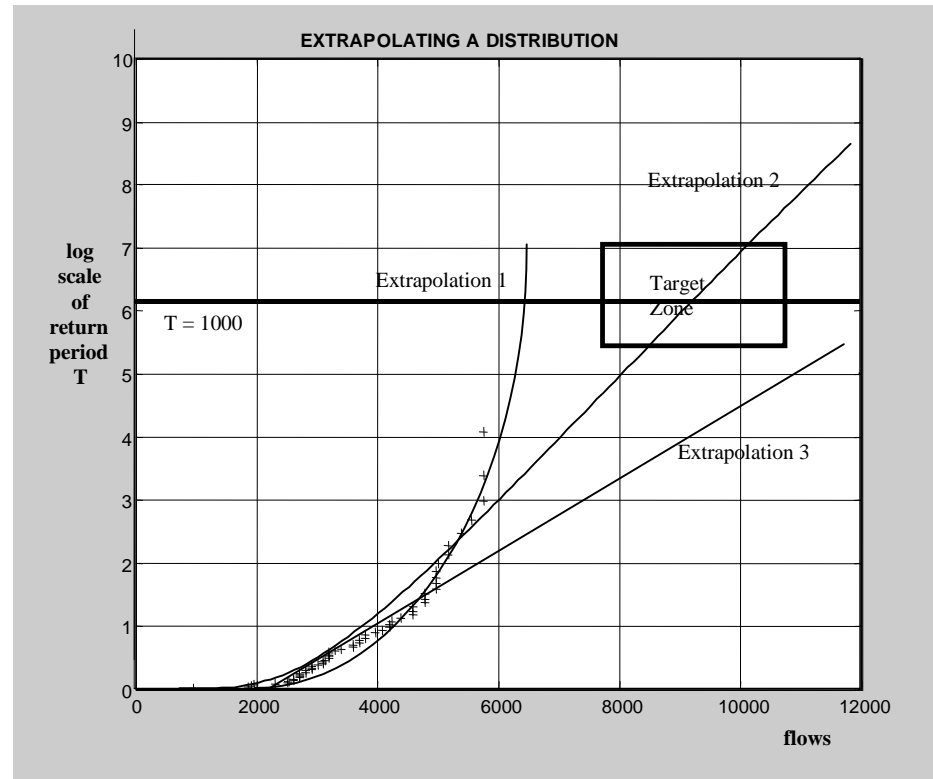
jacques.bernier2@wanadoo.fr

La Garonne en colère





Extrapolation des distributions d'événements rares



- Incertitudes d'interprétation des probabilités
- Incertitudes de mesure des grandeurs en jeu
- Incertitudes de modélisation
- Incertitudes d'échantillonnage

Contexte

- Crue de projet = extrapolation au delà de l'échantillon
- Pas de stabilisation asymptotique rapide type TCL pour les estimations d'événements rares
- Beaucoup d'incertitudes = toute information est bonne à prendre



Cérons, face à Cadillac,
sous les eaux - mars 1930

Le journal La Petite Gironde du 7 mars 1930 décrit l'Etat des communes sinistrées:

“ A Cadillac, la plus grosse inondation du siècle: Jeudi matin, les eaux ont atteint 11 m . 77, à 10 heures, soit 30 centimètres de plus qu'en 1875. Le sol de la halle est recouvert par les eaux. La place du Châteaux, sur laquelle notre dépositaire avait installé un magasin de vente en plein vent ressemble à s'y méprendre au Marché neuf ou à la place Mériadeck, tant les inondés ont transporté de meubles, ustensiles et objets de toutes sortes. Le courant est d'une telle violence que, seule, la vedette des Ponts et Chaussées traverse le fleuve dont la berge gauche se situe maintenant à dix-sept kilomètres à l'ouest. La route de Saint Macaire est submergée au point que les eaux ont pénétré dans la chapelle des aliénés. Dans cet établissement aux quartiers très élevés, tous les pensionnaires sont en lieu sûr et largement ravitaillés. La décrue commencée à midi aujourd'hui semble devoir être lente ”. (...).

Records historiques de la Garonne à Mas d 'Agenais

DATE	HAUTEUR (m)	DEBIT ESTIME (m3/s)
Avril 1770	10.34	7000 à 7400
Sept. 1772		6300
Mars 1783		7000 à 7200
Mai 1827		6500
Mai 1835		6400
Janv. 1843		6500
Juin 1855	9.96	7000
Mai 1856	9.62	6200
Juin 1856	9.88	6600
Juin 1875	10.56	7000 à 7500 (peut-être 8000)
Janv. 1879	9.62	6300
Fev. 1879	10.02	7000
Mai 1918	9.51	6000
Mars 1927	9.97	6700
Mars 1930	10.72	7000 à 7500 (peut-être 8000)
Mars 1935	9.95	6700
Fev. 1952	10.26	6000 à 7000
Janv. 1955	9.32	5200 à 5700

**Il importe d'utiliser
toute source d'information permettant
de réduire l'incertitude d'extrapolation**

**- Informations systématiques pertinentes,
modèles AMS → modèles POT**

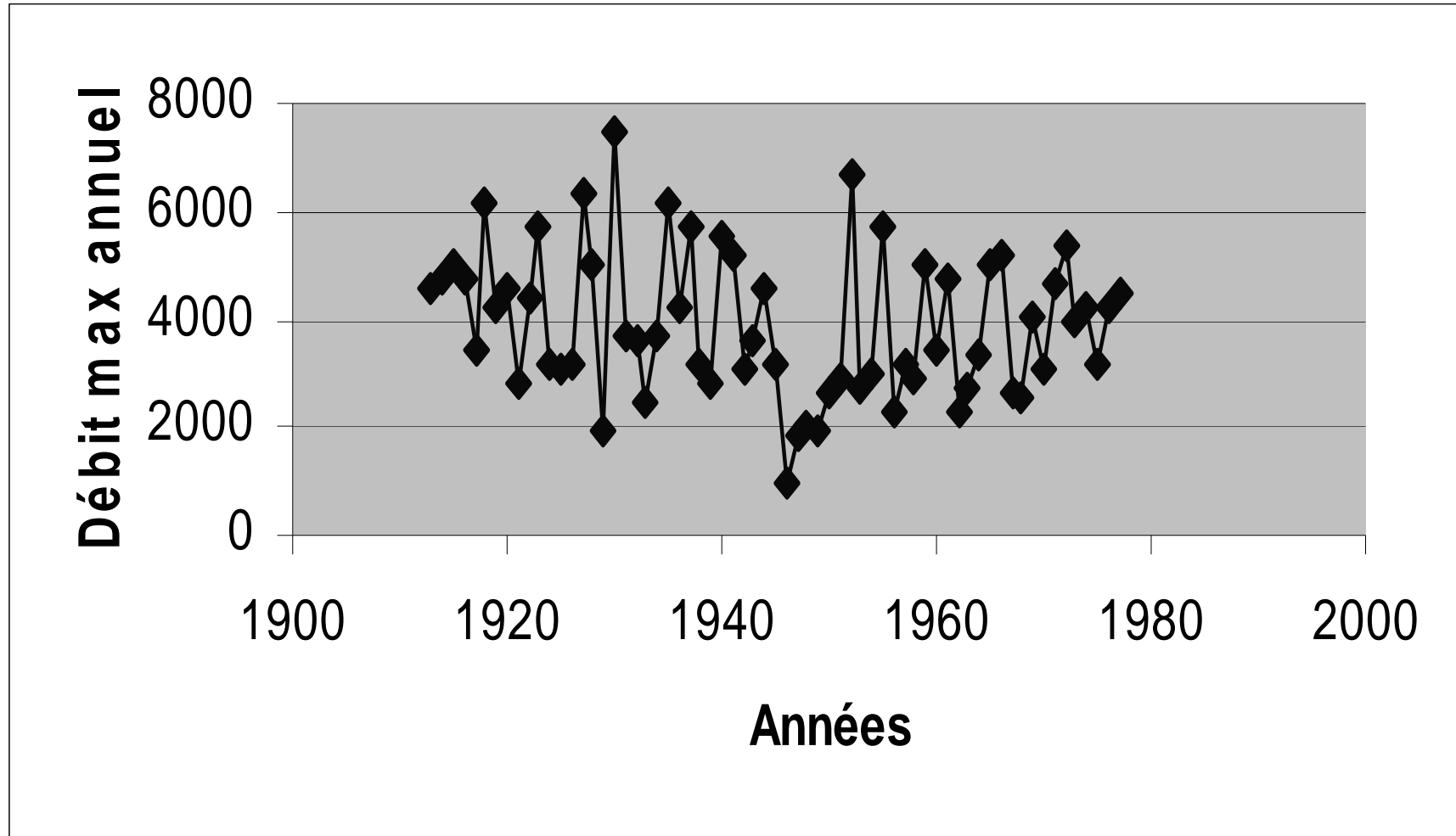
- Informations régionales

- Informations historiques

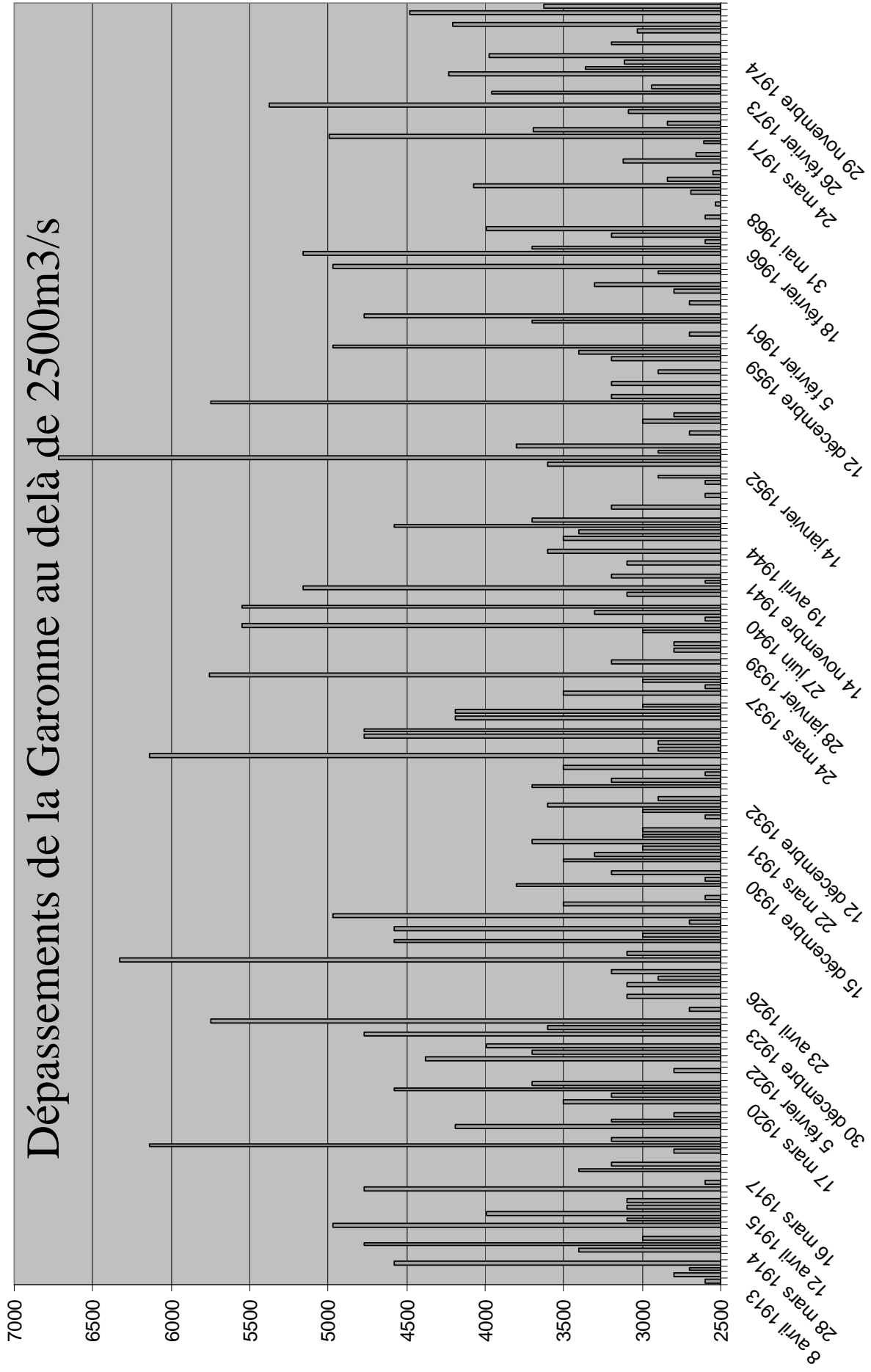
- Connaissances a priori des experts

**L'approche BAYESIENNE permet la prise en compte
d'information qualitative grâce au prior**

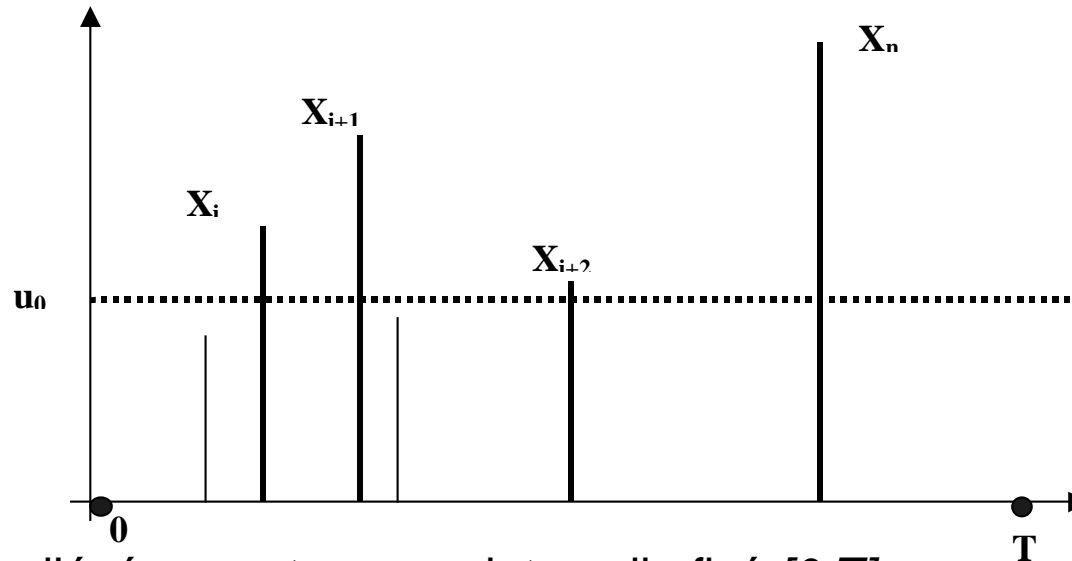
Crues Maximales Annuelles de la Garonne



DEBIT q(i) (m3/s)



Trajectoire de processus ponctuel marqué



N_T est le nombre d'événements sur un intervalle fixé $[0, T]$, ce nombre est distribué selon la « loi de Poisson » :

$$\text{Pr ob.}(N_T = n | \mu, T) = e^{-\mu T} \frac{(\mu T)^n}{n!}$$

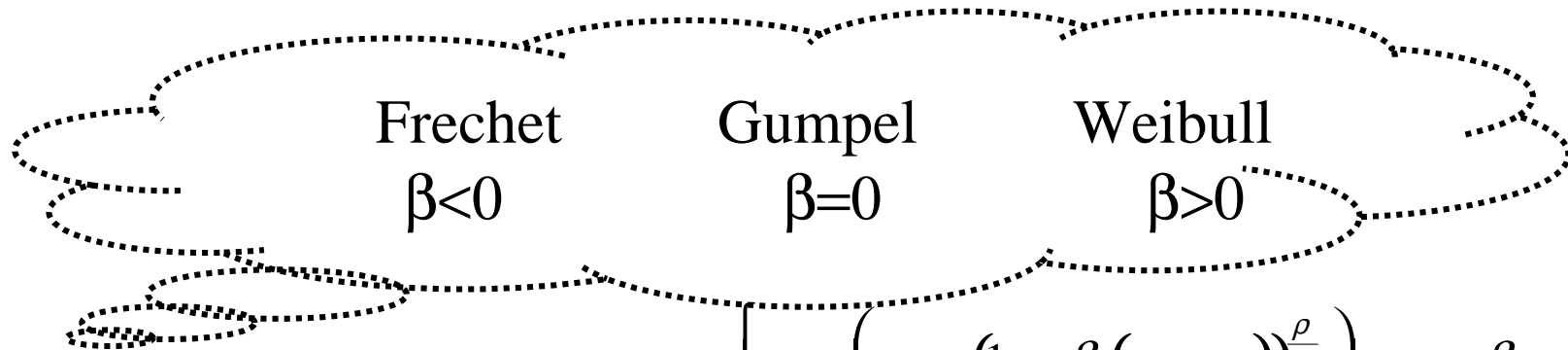
les X_i successifs sont indépendants et leurs dépassements d'un seuil u_0 sont distribués selon la « loi de Pareto » à 2 paramètres ρ et β

pour u_0 fixé, les trois paramètres μ , β et ρ caractérisent entièrement ce modèle

Modèle **P.O.T**

$$[X \leq x | \rho, \beta, \mu, x > v] = G(x | \rho, \beta, \mu, v) = \begin{cases} 1 - (1 - \beta(x - v))^{\frac{\rho}{\beta}} & \text{si } \beta \neq 0 \\ 1 - \exp(-\rho(x - v)) & \text{si } \beta = 0 \end{cases}$$

β et ρ coefficients d'échelle de Pareto,
 μ l'espérance de la loi de Poisson



$$\Pr(\max_{\text{annuel}} (X_i) \leq x | x \geq v) = \begin{cases} \exp\left(-\mu(1 - \beta(x - v))^{\frac{\rho}{\beta}}\right) & \text{pour } \beta \neq 0 \\ \exp(-\mu \exp(-\rho(x - \mu))) & \text{pour } \beta = 0 \end{cases}$$

$\xi = -\frac{\beta}{\rho}$, paramètre traditionnel, sans dimension.

Modèle **G.E.V**

Quantiles

$$q(p, \theta) = v + \frac{1}{\beta} \left[1 - \left(\frac{\log(p)}{\mu} \right)^{\frac{\beta}{\rho}} \right]$$

Queues plus ou moins lourdes...

- $\beta = 0$ (*Gumbel*)

$$\Delta_{\text{quantiles}}(\text{centennal} \rightarrow \text{millenal}) = \Delta_{\text{quantiles}}(\text{décennal} \rightarrow \text{centennal})$$

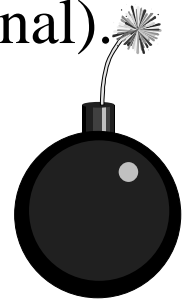
Normale, lognormale, exponentielle, Gamma.....

- $\beta > 0$ ou $\xi < 0$ (*Weibull*)

domaine de variation borné supérieurement.

- $\beta < 0$ ou $\xi > 0$ (*Fréchet*)

$$\Delta_{\text{quantiles}}(\text{centennal} \rightarrow \text{millenal}) > \Delta_{\text{quantiles}}(\text{décennal} \rightarrow \text{centennal}).$$



Vraisemblance... famille presque exponentielle!

$$S_n(x, \beta) = \frac{1}{\beta} \sum_{i=1}^n \text{Log}(1 - \beta(x_i - v))$$

$$[X, n | \mu, \rho, \beta, v, T] = \left[\frac{(\mu T)^n \exp(-\mu T)}{n!} \right] \left[\rho^n \exp((\rho - \beta) S_n(X, \beta)) \right]$$

Prior partiellement conjugué

... utilisation du caractère

presque exponentiel

$$[\theta | \pi_0(\cdot), \gamma(\cdot), \varphi(\cdot), \nu, \lambda] = \frac{\lambda^\nu}{\Gamma(\nu)} \mu^{\nu-1} \exp(-\lambda\mu) \frac{\varphi^\gamma}{\Gamma(\gamma)} \rho^{\gamma-1} \exp(-\varphi\rho) \pi_0(\beta)$$

- La loi marginale a priori pour β , $\pi_0(\beta)$ est quelconque.
- μ et (β, ρ) sont indépendants.

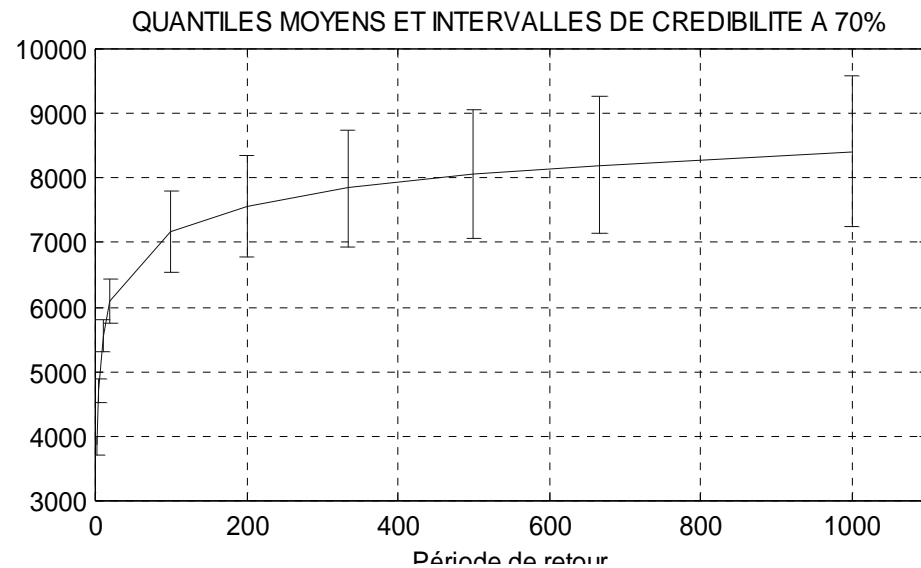
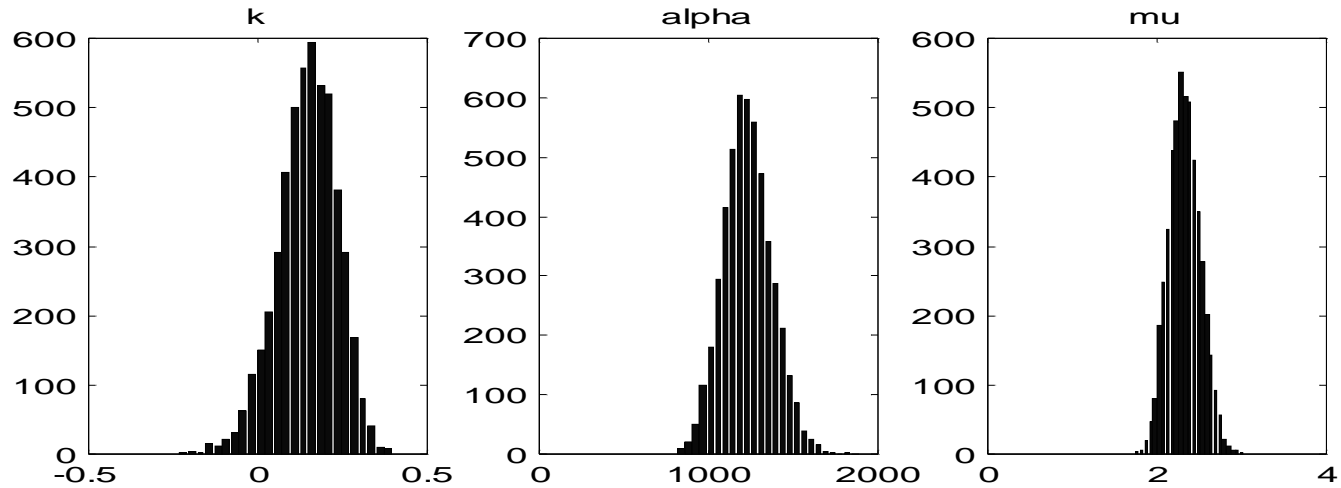
Loi a posteriori de (μ, β, ρ)

- μ demeure a posteriori indépendant de (β, ρ) , gamma $(\nu+n, \lambda+T)$.
- Conditionnellement à β , ρ est aussi gamma $(\nu+n, \varphi - S_n(x, \beta))$.

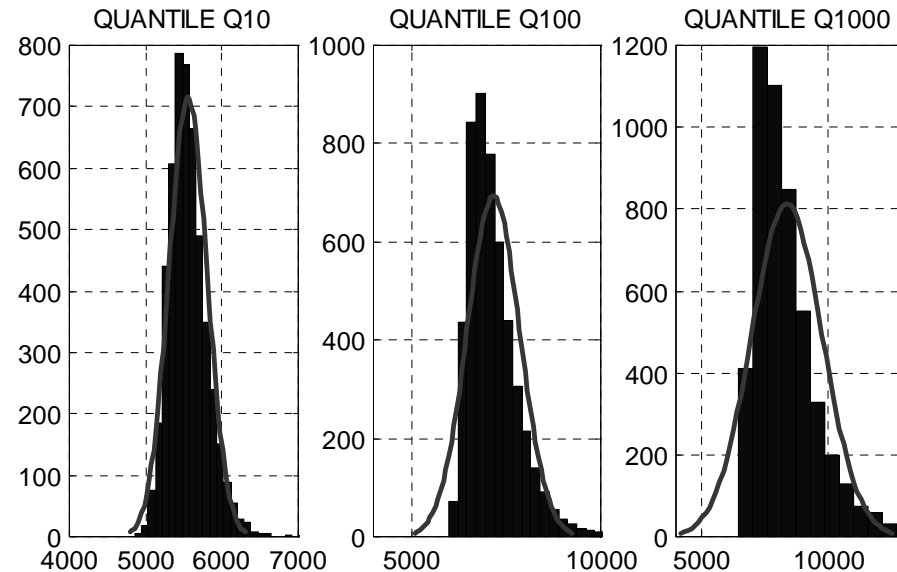
$$[\beta | x] \propto \frac{\Gamma(\gamma + n)}{\Gamma(\gamma)} \varphi^{\gamma+n} \frac{\pi_0(\beta) \exp(-\beta S_n(x, \beta))}{(\varphi - S_n(x, \beta))^{\gamma+n}}$$

Facilité pour simuler directement des tirages Monte Carlo.

Estimation Bayésienne sur données systématiques

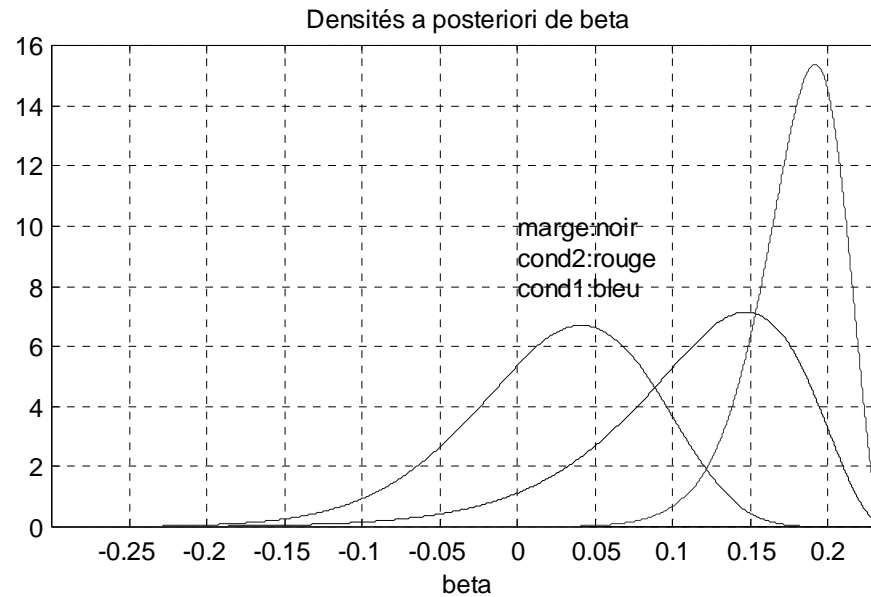
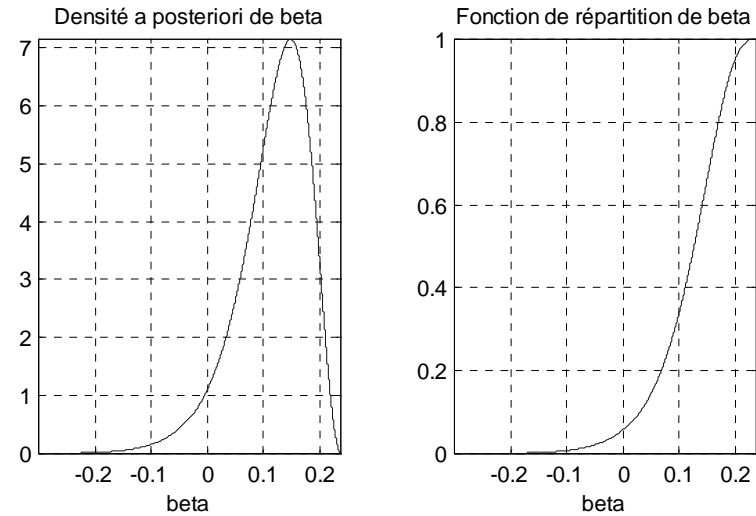


Quantiles (données systématiques)



	Moyenne	Médiane	Borne Sup 95%	Borne Inf 95%
μ	2.32	2.32	2.64	2.02
k	0.15	0.16	0.28	-0.01
alpha	1232.82	1226.71	1472.94	1011.12
Q_{10}	5551.04	5523.10	5989.21	5193.49
Q_{100}	7153.01	7004.00	8450.48	6353.65
Q_{1000}	8362.84	8017.89	10880.40	6927.07

Loi de Béta (éch. systématique)



Quelles quantités éliciter?

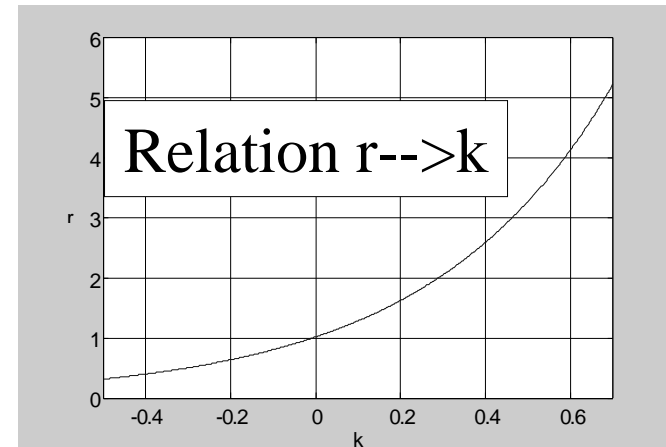
- Le nombre moyen d'événements dommageables
- Q10, Q100 domaine des variations ordinaires locales
- Q1000 appel à l'expérience trans-basins

METHODE D'ELICITATION

* Paramètres directement quantifiés

$$\mu, q_2 = Q_{100} - Q_{10}, r = (Q_{1000} - Q_{100}) / (Q_{100} - Q_{10})$$

Garonne (Agen)	Méd.	Déc9.
μ	1.7	2.1
q_2	1000 m ³ / s	1600 m ³ / s
r	2	3.5



Les divers objets a priori

- indépendances? Gamma? Hypothèses?
- **Prior élicité** (Coles et Tawn)
 $\mu, q_2 = Q_{100} - Q_{10}, q_3 = Q_{1000} - Q_{100}$
indépendants et gamma distribués
 $\mu, q_2, q_3 \text{ ----} \rightarrow \mu, q_2, r$
- **Prior modélisé** (semi conjugué naturel de POT), ajusté au prior élicité

Elicitation phase 1

Hydrological meaning	elicited θ	Médian ($p = 0.5$)	90% quar
annual mean number of floods $> 2500m^3/s$	μ	1.7	2.1
Difference between 100 and 10 year return floods	$\tilde{q}_2 = Q_{100} - Q_{10}$	1000 m3/s	1600 m3/s
ratio of quantile differences	$r = \frac{Q_{1000} - Q_{100}}{Q_{100} - Q_{10}}$	2	3.5

Table 1 : Expert's beliefs for meaningful hydrological quantities μ, \tilde{q}_2 and r

a_μ	b_μ	a_2	b_2	a_3	b_3
34	0.05	6.50	162.08	100	20

$$z = \frac{1}{1 + \frac{b_2}{b_3} r} \quad \text{Distribuée comme beta } (a_2, a_3)$$

Table 2 : Hyper-parameters for prior gamma distributions of μ, \tilde{q}_2 and \tilde{q}_3

Elicitation: phase 2

Par simulation: $\mu, q_2, r \rightarrow \mu, \rho, \beta$

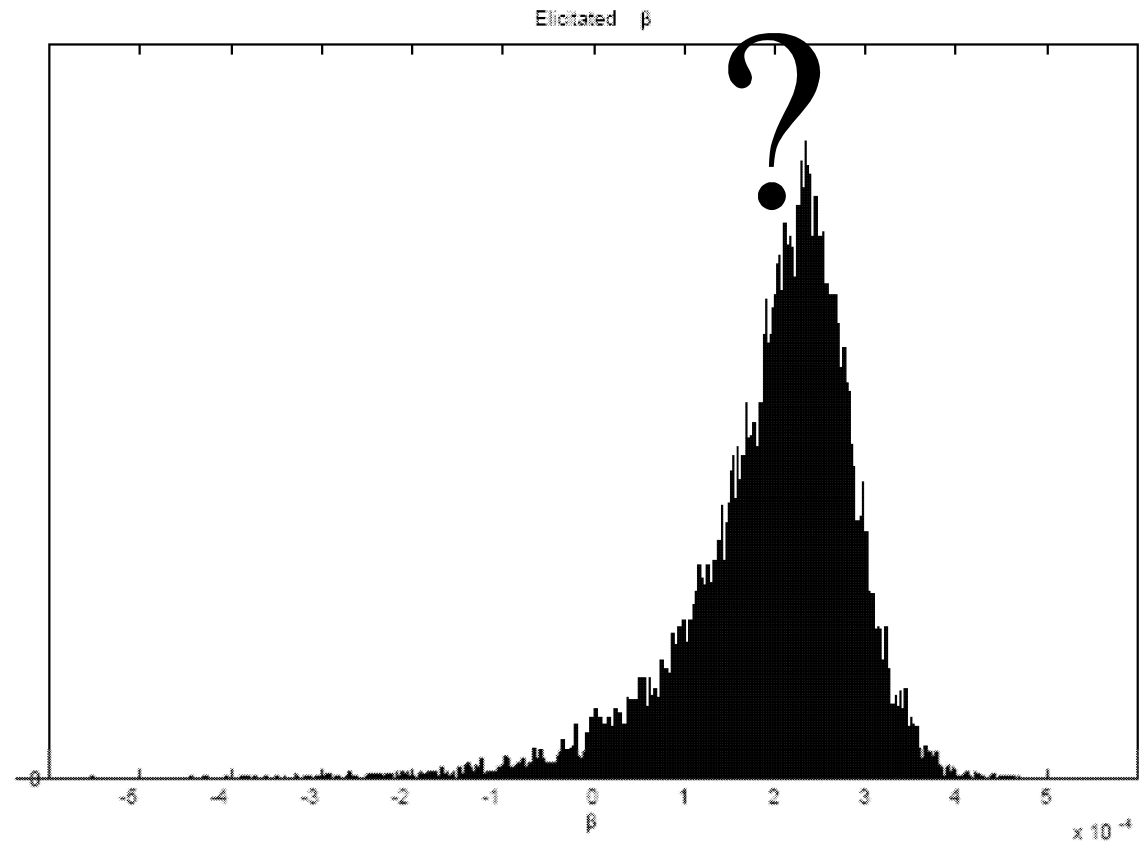
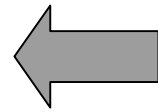
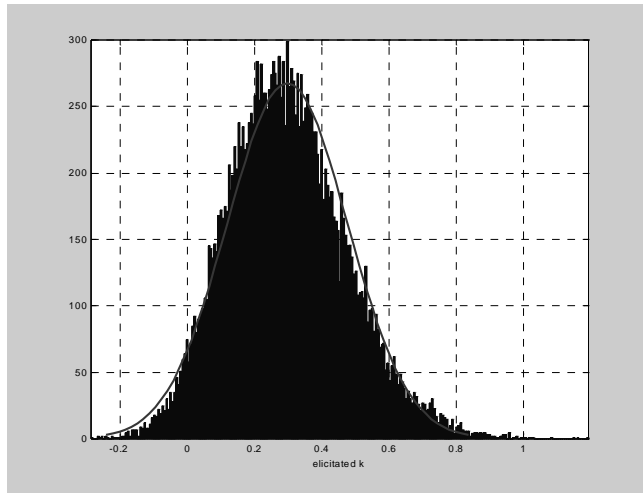
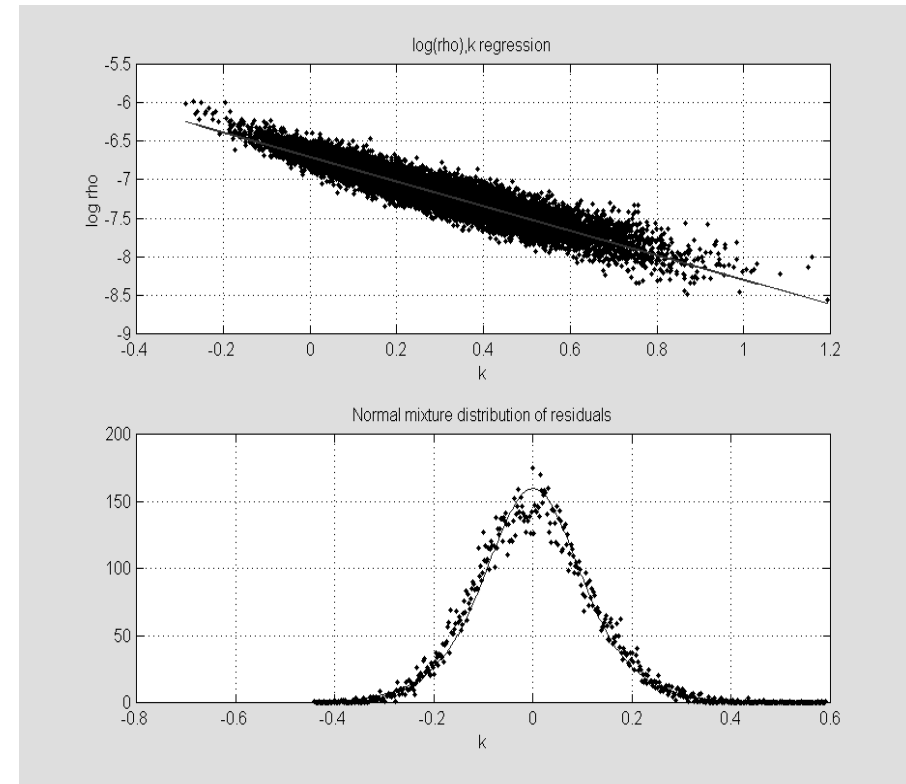


Figure 4: The distribution of 20000 simulated values of β is neither normal nor gamma

Représentation du prior élicité de β, ρ

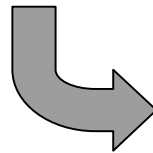


$k = \beta/\rho$ distribué comme $N(m, \sigma)$



$$\lambda = \log \rho = a + b.k + \varepsilon \text{ (k fixé)}$$

$\varepsilon = \text{normal (ou mélange normal)}$



Changement analytique
de variables $k, \rho \rightarrow \beta, \rho$

m_k	σ_k^2	δ	l_0	l_1	σ_1^2	σ_2^2
0.298	0.032	0.5	-6.7029	-1.6029	0.01989	0.00663

Table 3 : Hyper-parameters for normal distributions describing $(\kappa, \log(\lambda))$

Elicitation : phase 3 ... vers le prior semi-conjugué

$$\pi_0(\beta) = \int_{\rho} p(k, \lambda | m_k, \sigma_k, l_0, l_1, \delta, \sigma_1, \sigma_2) \left| \frac{\partial(\beta, \rho)}{\partial(k, \lambda)} \right| d\rho$$

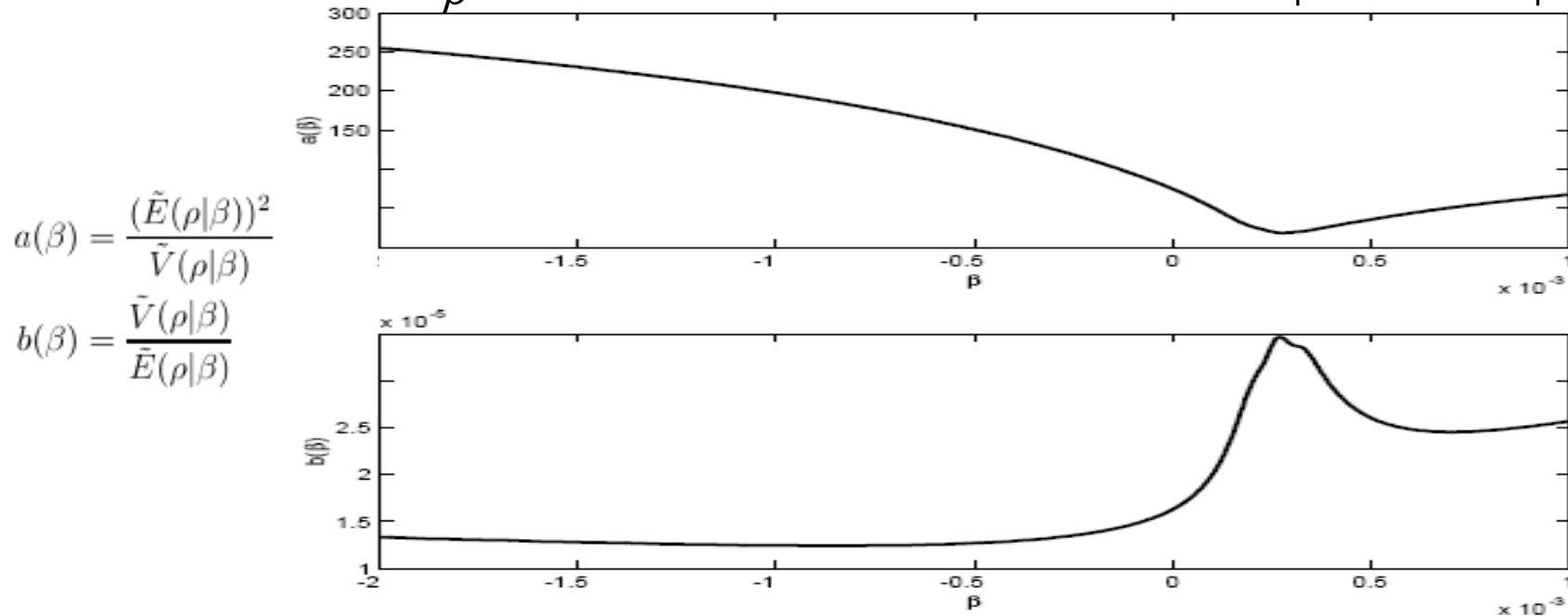
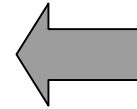
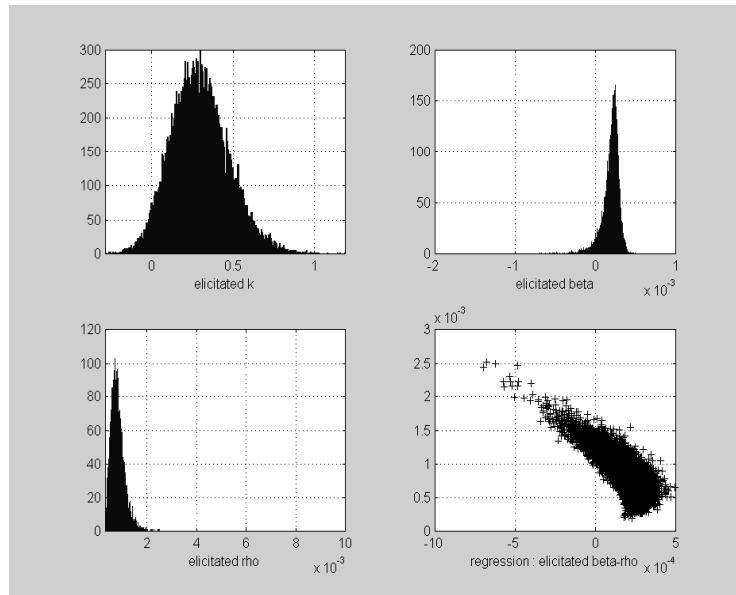


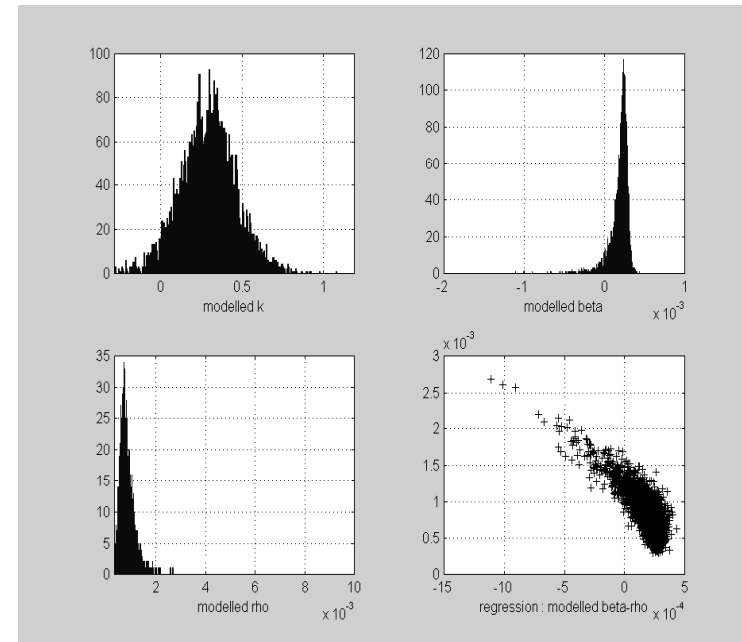
Figure 7: The two parameters of the gamma pdf for $p(\rho|\beta)$ as a function of β

ρ sachant β est gamma

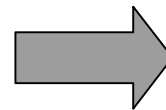
CALAGE ET VALIDATIONS DU PRIOR MODELISE



**Simulation (20000 réplicats)
de k, β, ρ élicités**



**Simulation (5000 réplicats)
de k, β, ρ modélisés**



Comparaison de paramètres hydrologiques

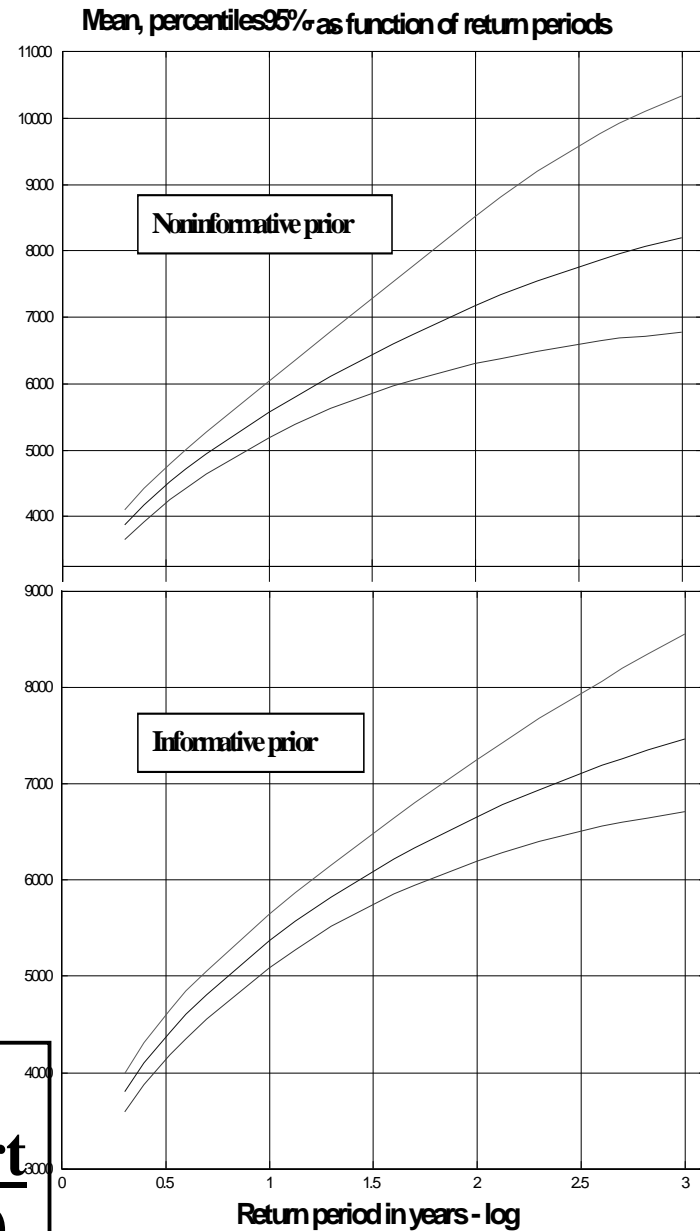
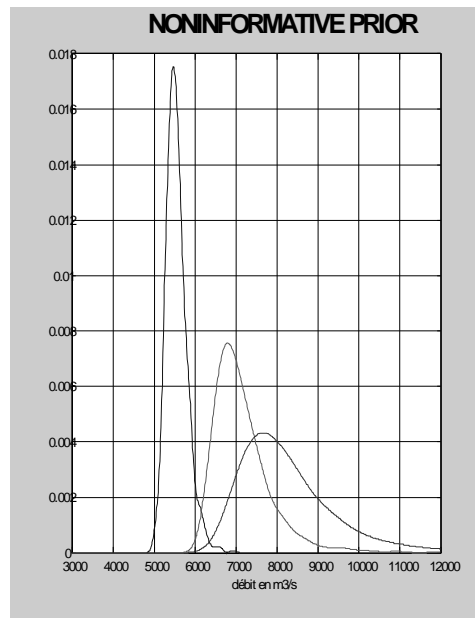
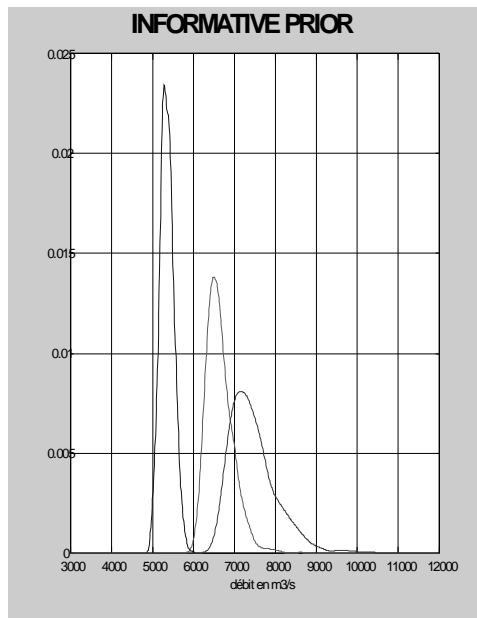
	q _{2₅}	q _{2₉}	r ₅	r ₉	Q _{10₅}	Q _{10₉}	Q _{100₅}	Q _{100₉}	Q _{1000₅}	Q _{1000₉}
élicité	1001	1609	2.00	3.51	4987	5550	6062	6708	6596	7842
modélisé	987	1654	2.02	3.30	5001	5473	6028	6755	6515	7962

ANALYSE A POSTERIORI

priors informatif et non informatif

Distributions a posteriori de

Q10, Q100, Q1000



Exemple

Q100	Med.	05lim.	95lim.	<u>Ecart</u>
non info.	7000	6290	8500	<u>2210</u>
info.	6590	6190	7240	<u>1050</u>

CONCLUSIONS

Avantages d 'un prior modélisé semi- conjugué naturel

***mise en œuvre aisée des techniques MCMC**

(échantillonnage de Gibbs avec une étape M. H)

ce qui permet :

- introduction de données complémentaires (historiques...)**
- analyse prédictive des décisions de protection contre les extrêmes**
- extension des modèles pour prendre en compte la variabilité spatiale (modèles hiérarchiques)**

*** avantage conceptuel d 'un cadre rationnel unique et complet.**

** Mais le modèle de prior semi conjugué est il assez souple pour représenter toute croyance quantifiée d 'un expert ?*

Validation partielle de la chaîne complète de traitement :

expertise subjective/elicitacion/modelisation/analyse a posteriori