

REVUE DE STATISTIQUE APPLIQUÉE

ERIC PARENT

JACQUES BERNIER

**Une procédure bayésienne de sélection/validation
différentielle pour déterminer le domaine
d'attraction des valeurs extrêmes**

Revue de statistique appliquée, tome 52, n° 4 (2004), p. 5-31

http://www.numdam.org/item?id=RSA_2004__52_4_5_0

© Société française de statistique, 2004, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

*Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques*

<http://www.numdam.org/>

UNE PROCÉDURE BAYÉSIENNE DE SÉLECTION/VALIDATION DIFFÉRENTIELLE POUR DÉTERMINER LE DOMAINE D'ATTRACTION DES VALEURS EXTRÊMES

Eric PARENT et Jacques BERNIER

(Laboratoire de Gestion des Risques En Sciences de l'Environnement),
ENGREF, 19, Avenue du Maine, F-75732 Paris Cedex 15,
Parent@engref.fr; Jacques.Bernier2@wanadoo.fr

RÉSUMÉ

En statistique bayésienne le choix d'un modèle se pose comme un problème d'inférence ordinaire dont l'objet est de chercher la loi *a posteriori* d'un paramètre supplémentaire. Ce paramètre supplémentaire est une variable indicatrice qui détermine le numéro du modèle éventuellement retenu. Un des aspects les plus fructueux de cette approche est l'emploi d'une loi *a priori* qui, dans ce cas particulier de sélection de modèle, peut être interprété comme l'apport d'un échantillon d'apprentissage. On présente un exemple généralement considéré comme délicat, la sélection du domaine d'attraction des extrêmes d'une loi de probabilité, grâce à l'emploi des facteurs de Bayes en distinguant «échantillon d'apprentissage» et «échantillon de validation». Sur cet exemple de la neige à la Plagne où données d'apprentissage et données de validation ne sont pas indépendantes, on montre la souplesse de modélisation et les facilités de calcul apportées par les variables latentes et le raisonnement conditionnel bayésien.

Mots-clés : *Modélisation bayésienne en Environnement, Méthodes Monte Carlo par chaînes de Markov, Variables latentes, Sélection de modèle, Loi des extrêmes.*

ABSTRACT

In this paper we are concerned with the use of bayesian model selection. We emphasize the use of prior distribution, which in the case of model selection may be considered as the information stemming from a learning sample. We work on a rather tricky problem, the determination of the asymptotic domain of attraction for the univariate extremes. We make the distinction between a learning sample and a validation sample. The explicit introduction of latent variables and a conditional line of reasoning allows for an easy and realistic treatment of the example (snow falls at la Plagne) although the data for calibration and for validation are not independent.

Keywords : *Bayesian models for Environmental data, Markov Chain Monte Carlo methods, Latent variables, Extreme value analysis, Model selection.*

1. Introduction

Dans de nombreux domaines d'applications de la statistique, une attention particulière est dévolue au comportement des valeurs extrêmes : sous des conditions générales de régularité, la loi des maxima d'un phénomène aléatoire *iid* unidimensionnel sera attirée vers une des trois formes possibles de comportement (Coles, 2001). Comment identifier cette forme limite ? En pratique, quand on examine de nombreuses séries de données du domaine de l'environnement (pluies, neiges, vents...) une grande incertitude porte sur le coefficient qui détermine le domaine d'attraction : même si la littérature est abondante sur le sujet, il n'existe pas de test classique universellement accepté par la communauté des statisticiens pour trancher entre les comportements de type Gumbel, Weibull ou Fréchet (Chaouche et Bacro, 2003). Quant à l'approche bayésienne (Coles et Powell, 1996), elle introduit une loi *a posteriori* sur les valeurs probables du paramètre qui règle le domaine d'attraction. D'une certaine façon, cette attitude bayésienne traditionnelle conserve et ménage l'incertitude dans un «super» modèle qui englobe tous les types de comportement possibles des valeurs extrêmes. En rupture avec cette approche, l'article développe une analyse bayésienne originale afin de choisir un modèle d'extrêmes parmi les trois formes possibles. Cette analyse d'adéquation des modèles probabilistes exploite la séparation fonctionnelle des données entre un échantillon d'apprentissage et un échantillon test.

L'article est organisé de la façon suivante. La section 2 décrit les notations et présente l'exemple des hauteurs de neige de la Plagne. La troisième partie rappelle l'utilité du modèle «Peak Over Threshold» pour la modélisation et l'estimation de l'occurrence des fortes valeurs d'un phénomène naturel. La section 4 donne les principes, les avantages et les limites de la sélection bayésienne de modèles. La structure conditionnelle du modèle proposé est exploitée dans la section 4 pour réaliser l'estimation bayésienne des paramètres de ces modèles et des facteurs de Bayes par la technique dite de Monte Carlo par Chaînes de Markov. Dans la section 5, on propose une procédure originale de sélection du domaine d'attraction de la loi des valeurs extrêmes de chutes de neige, en distinguant un échantillon d'apprentissage formé des faibles valeurs et un échantillon de test formé des plus fortes valeurs. Les résultats sont exposés en section 6. En section 7, on conduit une analyse de sensibilité, on discute des résultats obtenus, des perspectives et des limites de la modélisation employée. La section 8 dresse une conclusion générale.

2. Chutes de neige à la Plagne

2.1. Les données

Dans un but de prévision des conséquences des avalanches en montagne, il importe d'étudier les événements déclenchants. Parmi ceux-ci figurent les hauteurs totales de neige extrêmes cumulées sur 3 jours consécutifs, considérées généralement comme très significatives (Ancey, 1996). C'est ainsi que pour caractériser le risque sur ces variables, les spécialistes en nivologie s'accordent le plus souvent sur l'utilisation du modèle de distribution de Gumbel, classique modèle des valeurs extrêmes que l'on applique sur le maximum à l'échelle annuelle des enregistrements de ces variables. Dans le cas de La Plagne, station des Alpes Françaises, on dispose des données sur les

chutes journalières de neige dont on a pu déduire celles de plus de 3 jours dont le total maximal sur 3 jours dépasse le seuil $u = 50 \text{ cm}$ au cours de la saison hivernale Octobre - Avril. Les tableaux 1 et 2 présentent respectivement les statistiques de répartition des nombres de dépassements de ce seuil ainsi que des intensités X observées au cours de cette période de 28 ans.

TABLEAU 1
Répartition des nombres de dépassements de 50 cm par les chutes de neige continues de plus de 3 jours à la Plagne

Nombre d'occurrences annuelles	0	1	2	3	4	5	6
Fréquence sur 28 ans	2	10	5	5	3	3	0

TABLEAU 2
Répartition en fréquences des intensités des chutes de neige continues $> 50 \text{ cm}$, de plus de 3 jours à la Plagne

Classe de dépassement	Fréquence sur 62 évènements
$50 < X \leq 60$	16
$60 < X \leq 70$	11
$70 < X \leq 80$	9
$80 < X \leq 90$	7
$90 < X \leq 100$	6
$100 < X \leq 110$	6
$110 < X \leq 120$	4
$120 < X \leq 130$	0
$130 < X \leq 140$	2
$140 < X \leq 150$	0
$150 < X$	1

2.2. Choix du comportement de la queue de distribution de la neige.

Pour traiter cet exemple, les pratiques courantes d'ingénierie s'appuieraient sur les seuls 28 maxima annuels, dont les ajustements au max de vraisemblance pour les modèles Gumbel et pour la loi généralisée des valeurs extrêmes souvent désignée par l'acronyme anglais GEV pour generalized extreme values (Coles, 2001), sont présentés sur la figure 1. Sur ce graphique, il est bien difficile de distinguer le meilleur modèle; ce ne serait guère gênant si leurs extrapolations ne fournissaient pas

des résultats si radicalement différents quant à l'intensité probable des évènements rares. Ces évènements rares, par exemple la chute de période de retour 1000 ans, sont communément utilisés comme valeurs de projet pour les calculs de dimensionnement d'ouvrages de protection ou d'aménagements : ainsi, en extrapolant les courbes de la figure 1, l'intensité associée à l'évènement millénal vaut ici 240 cm pour le premier modèle mais seulement 200 si c'est le second modèle qui décrit vraiment le phénomène.

On espère obtenir une meilleure possibilité de discrimination entre les modèles en concurrence en étudiant l'échantillon plus complet constitué de toutes les 62 valeurs extrêmes supérieures au seuil de 50 cm. C'est l'objet de cet article.

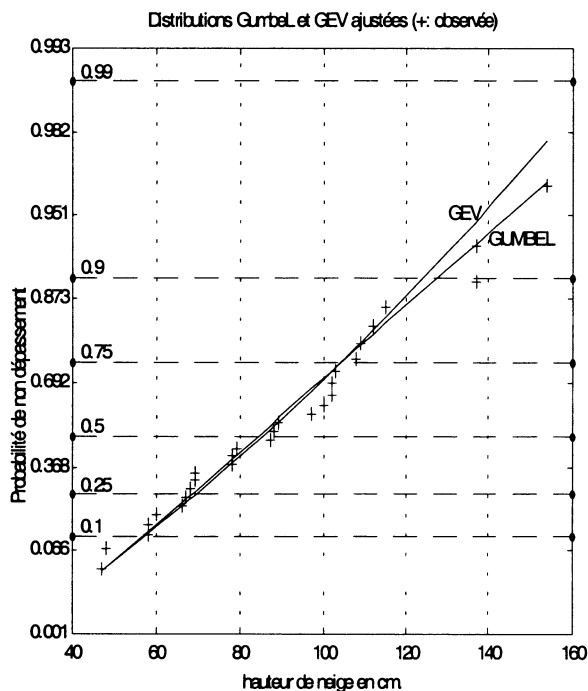


FIGURE 1

Distributions des hauteurs de neige maximales annuelles à la Plagne

3. Modèles d'extrêmes

3.1. Notations

On notera $[A]$ la probabilité de l'évènement aléatoire A et on utilisera indifféremment le signe d'intégration \int pour le calcul d'intégration avec des mesures discrètes ou diffuses. Ainsi la loi des probabilités conditionnelles s'écrira $[A, B] =$

$[A|B][B]$ et le calcul des probabilités totales s'exprime par $[A] = \int_z [A|z][z] dz$ que z soit une grandeur conditionnante à support discret partitionnant l'espace de définition ou une grandeur aléatoire à support réel.

3.2. Modèle POT

La théorie des valeurs extrêmes est un des sujets d'étude les plus traditionnels de la Statistique depuis Frechet (1927) et Fisher-Tippett (1928). Sans vouloir retracer l'histoire de tous les travaux importants du domaine, nous nous appuyerons sur les résultats de Pickands (1975). Son théorème principal donne la distribution limite des excès $X - u$ d'une variable aléatoire X de loi donnée, dont les occurrences se répètent selon une série régulière $(X_1, X_2, \dots, X_t, \dots)$ quand le seuil u s'approche du point limite de la distribution de probabilité de X . Ce modèle asymptotique général vis-à-vis du comportement des dépassements (appelé Peaks Over Threshold ou POT) est aujourd'hui courant (Smith, 1984; Davidson et Smith, 1990; Coles, 2001). Il repose sur les concepts suivants :

Les dates de dépassements au-delà d'un seuil u fixé, suivent un processus homogène de Poisson, si bien que le nombre de dépassements sur une période T est une loi de Poisson de paramètre μT . Pour les données de la Plagne $T = 28$ ans. Le paramètre de ce modèle, μ , s'interprète comme l'espérance de la loi de Poisson pour $T = 1$, c'est donc le nombre annuel moyen de dépassements.

L'intensité du dépassement est une loi de Pareto généralisée :

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

β et ρ sont les coefficients d'échelle de la Pareto, $\rho > 0$. La densité de probabilité associée est :

$$f(x | \rho, \beta, u) = \frac{\partial F(x | \rho, \beta, u)}{\partial x} = \rho (1 - \beta(x - u))^{\frac{\rho}{\beta} - 1} \quad (2)$$

Ce modèle n'est pas régulier puisque la condition $1 - \beta(x - u) > 0$ lie le support de variation de la variable aléatoire aux paramètres, mais par continuité, on retrouve le modèle exponentiel simple lorsque β tend vers 0. Le comportement de la loi de Pareto généralisée $F(x | \rho, \beta, u)$ pour les grandes valeurs dépend notablement de β et la condition précédente implique que, si on dispose de X_1, X_2, \dots, X_l , répétitions *iid* de loi $F(x | \rho, \beta, u)$:

$$\max(X_1, X_2, \dots, X_l) < \frac{1}{\beta} + u \quad (\text{si } \beta > 0)$$

Le modèle Pareto généralisé se ramène au cas exponentiel par le changement de variable :

$$y = -\frac{\log(1 - \beta(x - u))}{\beta} \quad (3)$$

TABLEAU 3
Hypothèses du comportement des distributions de Pareto

<i>Fréchet</i>	<i>Gumbel</i>	<i>Weibull</i>
$\beta < 0$	$\beta = 0$	$\beta > 0$

Les domaines d'attraction du maximum (cf. Tableau 3) dépendent du signe de β : on obtient directement la loi des extrêmes (Generalized Extreme Value ou *GEV*) pour le maximum du dépassement sur T années (d'une certaine façon, le passage au comportement asymptotique a déjà été anticipé en supposant une structure POT pour le phénomène dont on veut étudier les extrêmes).

$$[\max(X_1, X_2, \dots) \leq x | T \text{ années}, x \geq u] = \sum_{k=0}^{+\infty} F(x | \rho, \beta, u)^k \frac{(\mu T)^k \exp(-\mu T)}{k!} \quad (4)$$

$$[\max(X_1, X_2, \dots) \leq x | T \text{ années}, x \geq u] = \exp((-\mu T)(1 - F(x | \rho, \beta, u))) \quad (5)$$

On prend $T = 1$ an pour obtenir :

$$[Max\ annual(X_i) \leq x | x \geq u] = \begin{cases} \exp\left(-\mu(1 - \beta(x - u))^{\frac{\rho}{\beta}}\right) & \text{pour } \beta \neq 0 \\ \exp(-\mu \exp(-\rho(x - u))) & \text{pour } \beta = 0 \end{cases} \quad (6)$$

$\xi = -\frac{\beta}{\rho}$ est sans dimension, c'est le paramètre traditionnellement utilisé auquel nous préférons β pour des raisons techniques que nous exposerons plus loin. ξ ou β déterminent le poids de la queue de distribution des extrêmes.

- $\beta = 0$, le max annuel suit une loi de Gumbel tronquée et l'accroissement de quantiles du centenal au millenal est égal à l'accroissement de quantiles du décennal au centenal. Le domaine d'attraction qui correspond à cette loi des maxima attire les lois courantes comme la loi Normale, lognormale, exponentielle, Gamma...
- $\beta > 0$ ou $\xi < 0$: loi de Weibull, à domaine de variation borné supérieurement.
- $\beta < 0$ ou $\xi > 0$ loi de Fréchet, où l'accroissement entre le quantile centenal et le quantile millenal est plus important que l'accroissement de quantiles du décennal au centenal.

La théorie nous enseigne que seul l'un des trois modèles du Tableau 3 est asymptotiquement valide. Nous rappelons d'abord les outils bayésiens d'inférence et de choix de modèle avant d'exposer notre angle d'attaque de cette question à la section 5.

3.3. *Vraisemblance*

Considérons l'ensemble des n observations supérieures au seuil u obtenues sur T années d'enregistrements, en associant la loi de Poisson de paramètre μT pour

le nombre d'occurrences et la densité de la Pareto généralisée (2), la vraisemblance s'écrit :

$$[\mathbf{x}, n | \mu, \rho, \beta, u, T] = \frac{(\mu T)^n \exp(-\mu T)}{n!} \prod_{i=1}^n f(x_i | \rho, \beta, u) \quad (7)$$

$$= \frac{(\rho \mu T)^n \exp(-\mu T)}{n!} \prod_{i=1}^n (1 - \beta(x_i - u))^{\frac{\rho}{\beta} - 1} \quad (8)$$

Un rôle pivot est joué par la quantité :

$$S_n(x, \beta) = \frac{1}{\beta} \sum_{i=1}^n \text{Log}(1 - \beta(x_i - u)) \quad (9)$$

(On pose, par continuité $S_n(x, 0) = -\sum_{i=1}^n (x_i - u)$). En effet :

$$[\mathbf{x}, n | \mu, \rho, \beta, u, T] = \left[\frac{(\mu T)^n \exp(-\mu T)}{n!} \right] [\rho^n \exp((\rho - \beta)S_n(\mathbf{x}, \beta))] \quad (10)$$

Si β était connu, le modèle appartiendrait à la famille exponentielle. Cette propriété, raison de notre choix de paramètres, est exploitée pour construire une loi *a priori* des paramètres (prior) semi-conjuguée, facilitant l'inférence par l'algorithme de Gibbs.

3.4. Prior partiellement conjugué

Selon l'approche bayésienne, la connaissance *a priori* sur $\theta = (\mu, \rho, \beta)$ est encodée grâce à une loi de probabilité. Nous choisissons de prendre ce prior dans la famille de lois suivantes :

$$[\theta | \pi_0(\cdot), \gamma(\cdot), \varphi(\cdot), \nu, \lambda] = \frac{\lambda^\nu}{\Gamma(\nu)} \mu^{\nu-1} \exp(-\lambda\mu) \frac{\varphi^\gamma}{\Gamma(\gamma)} \rho^{\gamma-1} \exp(-\varphi\rho) \pi_0(\beta) \quad (11)$$

Cette structure flexible couvre un grand ensemble de lois *a priori* et a prouvé son utilité pour introduire dans l'analyse hydrologique des priors informatifs (Parent et Bernier, 2001; 2003). Elle est construite de la façon suivante :

- la loi marginale *a priori* pour β , $\pi_0(\beta)$ est quelconque; conditionnellement à β , ρ est décrit par une loi Gamma d'hyperparamètres $(\gamma(\beta), \varphi(\beta))$.
- μ et (β, ρ) sont indépendants, ce qui traduit une pratique courante chez les experts qui portent généralement des jugements *a priori* sur le nombre de crues et sur leur intensité de façon indépendante. La loi *a priori* de μ est choisie dans la loi conjuguée de la distribution de Poisson, la loi gamma d'hyperparamètres (ν, λ) .

Les quantités $(\pi_0(\beta), \nu, \lambda, \gamma(\beta), \varphi(\beta))$ sont à tirer de la connaissance des experts en hydrologie.

3.5. La formule de Bayes donne la loi a posteriori de (μ, β, ρ)

Associant la vraisemblance (10) et le prior (11), la formule de Bayes procure la loi *a posteriori* (posterior) $[\mu, \beta, \rho | \mathbf{x}]$. Le posterior conserve la même structure que la loi *a priori*. En particulier :

- μ demeure *a posteriori* indépendant de (β, ρ) . Il suit une loi gamma d'hyperparamètres $(\nu + n, \lambda + T)$.
- Conditionnellement à β, ρ est aussi gamma d'hyperparamètres $(\gamma + n, \varphi - S_n(x, \beta))$.
- On connaît la marginale *a posteriori* de β à une constante de normalisation près :

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

La grandeur (aléatoire pour un bayésien) β étant scalaire, le calcul numérique de cette constante d'intégration ne pose pas de problème pratique.

- β et ρ sont *a posteriori* dépendants comme le montre la loi *a posteriori* de β conditionnellement à ρ où φ et γ sont des fonctions de β :

$$[\beta | \rho, \mathbf{x}] \propto \frac{\varphi^\gamma}{\Gamma(\gamma)} \rho^\gamma \exp(-\varphi \rho) \exp((\rho - \beta) S_n(x, \beta)) \pi_0(\beta) \quad (12)$$

Il est ainsi très facile de simuler directement des tirages Monte Carlo dans la loi *a posteriori* $[\mu, \beta, \rho | \mathbf{x}]$ en suivant la décomposition de Gibbs appliquée en séquence (Tanner, 1996) :

$$\text{simuler } [\mu | \mathbf{x}], \text{ simuler } [\beta | \rho, \mathbf{x}], \text{ simuler } [\rho | \beta, \mathbf{x}] \quad (13)$$

3.5.1. Quantiles

Pour le modèle POT, l'expression des valeurs de quantiles de probabilité p de non dépassement revêt une forme analytique particulièrement simple.

Ces valeurs de projets, utiles dans les analyses de risque, sont des fonctions explicites de μ, β, ρ et du niveau de défaillance acceptable $(1 - p)$. $1 - p$ est typiquement de l'ordre de $1/100$ or $1/1000$. L'inversion de la formule (6) donne :

$$q(p, \theta) = u + \frac{1}{\beta} \left[1 - \left(-\frac{\log(p)}{\mu} \right)^{\frac{\beta}{\rho}} \right] \quad (14)$$

Les distributions *a posteriori* de ces quantiles $[q(p, \theta) | \mathbf{x}]$ se calculent aisément en introduisant $q(p, \theta)$ comme variable supplémentaire dans l'algorithme de simulation donné par (13).

4. Sélection bayésienne de modèles

4.1. Principe

Pour choisir un modèle m_j parmi un ensemble fini $M = \{m_1, \dots, m_j, \dots, m_k\}$, l'approche bayésienne utilise la démarche suivante :

- On calcule la distribution *a posteriori* $[\theta_j | \mathbf{x}, m_j]$ des paramètres θ_j du modèle m_j et la vraisemblance moyenne *a priori* $[\mathbf{x} | m_j]$:

$$[\theta_j | \mathbf{x}, m_j] = \frac{[\mathbf{x} | \theta_j, m_j][\theta_j]}{[\mathbf{x} | m_j]}$$

avec $[\mathbf{x} | m_j] = \int [\mathbf{x} | \theta_j, m_j][\theta_j] d\theta_j$

La dénomination *vraisemblance moyenne a priori* est ici utilisée de préférence au terme technique usuel de *distribution prédictive a priori* : la vraisemblance moyenne est un nombre qui qualifie le modèle m_j pour notre perspective de sélection alors que nous réservons l'adjectif prédictif à la distribution de probabilité qui, elle, porte sur les observables \mathbf{x} .

- Considérant maintenant les modèles de l'ensemble M affectés chacun d'une probabilité *a priori* $[m_j]$, on calcule les probabilités *a posteriori* de chacun des modèles :

$$[m_{j_1} | \mathbf{x}] = \frac{[m_{j_1}][\mathbf{x} | m_{j_1}]}{\sum_{m_{j_1} \in M} [m_{j_1}][\mathbf{x} | m_{j_1}]} = \frac{1}{1 + \sum_{j_2 \neq j_1} ([m_{j_2}]/[m_{j_1}]) B_{j_2 j_1}(\mathbf{x})}$$

où les $B_{j_2 j_1}(\mathbf{x}) = \frac{[\mathbf{x} | m_{j_2}]}{[\mathbf{x} | m_{j_1}]}$ sont les rapports de Bayes des modèles m_{j_2} relativement au modèle m_{j_1} .

- Les $[m_j | \mathbf{x}]$ peuvent ainsi être la base d'une inférence sur le choix des modèles. Cette approche revient à introduire un paramètre d'occurrence de chaque modèle et à en faire l'inférence bayésienne ordinaire. Elle est entièrement générale. Elle s'applique que certains modèles soient emboîtés ou qu'ils soient de structures totalement différentes.

4.2. Interprétation décisionnelle des facteurs de Bayes

Une analyse bayésienne décisionnelle de choix entre deux modèles se baserait par exemple sur la table de coûts ci-après :

Les coûts d'une décision inadaptée à l'état de la nature C_{12} et C_{21} doivent être représentatifs des enjeux du problème : il est par exemple sûrement plus coûteux de ne pas appeler les pompiers quand la maison brûle que de les appeler à tort. L'analyse bayésienne préconise alors d'évaluer le coût moyen *a posteriori* $W_{\mathbf{x}}$ de

TABLEAU 4
Table de décision entre deux modèles

Etat de la nature \rightarrow	m_1	m_2
Décision \downarrow		
d_1 : choisir le modèle m_1	0	C_{12}
d_2 : choisir le modèle m_2	C_{21}	0

chaque décision :

$$W_{\mathbf{x}}(d_1) = [m_1 | \mathbf{x}] \times 0 + [m_2 | \mathbf{x}] \times C_{12}$$

$$W_{\mathbf{x}}(d_2) = [m_1 | \mathbf{x}] \times C_{21} + [m_2 | \mathbf{x}] \times 0$$

et de retenir la décision menant au coût minimal, soit d_1 si $[m_2 | \mathbf{x}] \times C_{12} \leq [m_1 | \mathbf{x}] \times C_{21}$ et d_2 sinon. En revenant au facteur de Bayes on trouve qu'il faut retenir le modèle 1 si $[m_2] [\mathbf{x} | m_2] \times C_{12} \leq [m_1] [\mathbf{x} | m_1] \times C_{21}$ soit :

$$B_{12}(\mathbf{x}) \geq \frac{C_{12}}{C_{21}} \frac{[m_1]}{[m_2]} \quad (15)$$

Kass et Raftery (1994) interprètent le rapport $B_{12}(\mathbf{x})$ comme l'évidence, apportée par l'information \mathbf{x} en faveur de m_1 vis-à-vis de m_2 . Comme l'évaluation (15) des priors et des coûts dépend du problème spécifique traité, ils proposent un barème permettant un choix indicatif entre deux modèles, utile en première approche. Par analogie avec le test classique de la déviance, ils expriment leur proposition en prenant deux fois le logarithme du rapport de Bayes :

TABLEAU 5
Barèmes des rapports de Bayes selon Kass et Raftery (1994)

Facteur de Bayes $B_{12}(\mathbf{x})$	$2 \log(B_{12})$	Evidence de M_1
1 à 3	0 à 2	Aucune
3 à 20	2 à 6	Positive
20 à 150	6 à 10	Forte
> 150	> 10	Très forte

Dans les problèmes de choix *a posteriori* global de modèles, ce barème est quelquefois utilisé pour écarter les modèles dans une première phase de sélection.

4.3. Avantages des facteurs de Bayes

Les facteurs de Bayes ne se manient pas sans précaution, notamment vis-à-vis de l'information *a priori* : le « paradoxe » de Lindley par exemple (Berger, 1985) est un avatar des raffinements mathématiques de passage à la limite vers des distributions *a priori* impropres dégénérées.

Mais, dans le contexte du choix de modèles, l'approche bayésienne présente de nombreux avantages. Par exemple, les techniques de tests de Neyman-Pearson se retrouvent immédiatement dans un cadre bayésien. D'autre part, choisir un modèle sur la base des facteurs de Bayes garantit que si le « vrai » modèle m^* se trouve dans la liste M des modèles en compétition, la procédure bayésienne retiendra à coup sûr m^* quand le nombre de données tend vers l'infini (Bernardo et Smith, 1994). Si le « vrai » modèle m^* ne figure pas dans la liste M des modèles en compétition, la procédure précédente sélectionne le modèle de M le plus proche de m^* quand le nombre de données augmente à l'infini (en mesurant la proximité par la distance construite sur la divergence de Kullback-Leibler).

De plus il est inutile de pénaliser le facteur de Bayes en fonction du nombre de paramètres : la procédure bayésienne tient compte directement de cet effet désirable. On montre d'ailleurs que le critère BICS qui pénalise la déviance (deux fois l'opposé de la logvraisemblance) par le produit du nombre de paramètres et du log de la taille d'échantillon est la limite asymptotique du facteur de Bayes (Gelfand et Dey, 1995).

Enfin, à l'inverse des problèmes de tests classiques bien souvent dichotomiques, la procédure bayésienne se généralise facilement au cas de plusieurs modèles. Dans l'exemple de la neige, on peut certes vouloir tester la proposition $\beta = 0$ (le domaine d'attraction est de type Gumbel) contre $\beta \neq 0$ (les distributions limites avec un paramètre additionnel Weibull ou Fréchet), mais il est plus rationnel de rechercher une affectation entre $\beta > 0$ ou bien $\beta = 0$ ou bien $\beta < 0$.

4.4. Mise en oeuvre algorithmique

Cependant, la mise en oeuvre du paradigme bayésien pose des problèmes particuliers de calcul de la vraisemblance moyenne : le passage de la vraisemblance $[y|\theta, m_j]$ à la prédictive $[y|m_j]$ nécessite l'intégration (bien souvent multidimensionnelle) du vecteur θ des paramètres et le calcul de la vraisemblance moyenne demande que cette quantité soit évaluée exactement en $y = \mathbf{x}$ (c'est-à-dire sur les valeurs observées). L'emploi aujourd'hui généralisé des méthodes MCMC pour l'estimation bayésienne des paramètres d'un modèle fournit des solutions efficaces pour le calcul des facteurs de Bayes par simulation (Chib et Jeliazkov, 1995, 2001).

L'approche naïve : tirer aléatoirement des paramètres $\theta^{(g)}$, (l'indice de simulation g allant de 1 à L) selon la loi *a priori* et sommer sur la vraisemblance, fondée sur la définition :

$$[y|H] = \int_{\theta} [y|\theta, H][\theta|H]d\theta \approx \frac{1}{L} \sum_{g=1}^L [y|\theta^{(g)}, H],$$

est inefficace car les deux fonctions de θ , prior et vraisemblance, peuvent répartir l'essentiel de leur masse dans des régions très séparées de l'espace des paramètres. H désigne ici l'ensemble des hypothèses afférentes à un état de connaissance donné, notamment l'adoption d'un modèle donné (par exemple $H = m_j$), mais aussi tout autre conditionnement possible. On oublie d'ailleurs souvent d'indiquer ce conditionnement par commodité (par exemple les variables sont *iid*, etc.).

L'approche de Raftery emploie la loi *a posteriori* et la vraisemblance au lieu du prior et de la vraisemblance et, intuitivement, on comprend qu'il obtienne ainsi un meilleur recouvrement des deux fonctions qui interviennent dans un calcul d'intégration par simulation. Avantage supplémentaire, il récupère les résultats des simulations MCMC obtenues dans la phase d'estimation des paramètres. Le résultat suivant, directement issu de la règle de Bayes fonde sa démarche : la vraisemblance moyenne est l'espérance harmonique *a posteriori* de la vraisemblance.

$$[y | H]^{-1} = \int_{\theta} [y | \theta, H]^{-1} [\theta | y, H] d\theta \quad (16)$$

L'estimateur $\left[\frac{1}{L} \sum_{g=1}^L [y | \theta^{(g)}, H]^{-1} \right]^{-1}$, où les paramètres $\theta^{(g)}$, $g = 1..L$, sont cette fois générés selon la loi *a posteriori*, est meilleur que le précédent et nous l'utiliserons dans le cas d'application. Notons toutefois qu'il peut être déstabilisé par de fortes contributions issues de valeurs faibles de la vraisemblance quand apparaissent dans de longues séries de simulation des valeurs peu crédibles de la loi *a posteriori*. Hoeting *et al.* (1999) ont proposé des modifications pour améliorer son comportement, mais sur le cas d'application, nous nous contenterons en pratique de vérifier sa stabilité sur plusieurs chaînes MCMC lancées en parallèle.

Les autres approches de calcul évitent le calcul des facteurs de Bayes. Elles évaluent directement les probabilités *a posteriori* $[m_j | y]$ en introduisant explicitement l'indice de chaque modèle qu'elles considèrent comme un paramètre catégoriel supplémentaire. Quand on aborde le problème sous cet angle, une difficulté technique de construction des algorithmes d'estimation MCMC est de réaliser des transitions entre espaces de paramètres de dimensions éventuellement différentes. Carlin et Chib (1995) montrent comment contourner cette difficulté en travaillant sur l'espace produit des espaces de paramètres de chacun des modèles en compétition. On trouvera dans Han et Carlin (2000) une revue très complète des techniques de sélection bayésienne de modèles.

5. Sélection ou validation ?

Les problèmes d'adéquation de modèles peuvent être envisagés selon deux perspectives :

- la perspective *a posteriori*, où, après inférence sur échantillon, une décision « terminale » de choix de modèle est prise (sans autre examen d'un jeu de données de validation). La théorie de la décision esquissée dans le paragraphe précédent donne le cadre général méthodologique qui permet la prise en compte des conséquences de ce choix.

- la perspective dite *prédictive* où, après calage préalable, l'adéquation des modèles est analysée en se référant à une source d'information de validation différente de la source utilisée pour le calage. Cette dernière est en général indépendante de l'information de calage, mais les sections suivantes montrent que cette indépendance n'est ni nécessaire ni forcément pertinente vis-à-vis d'un objectif de sélection de modèle.

Pour l'approche bayésienne, les deux perspectives s'inscrivent de fait dans le même cadre probabiliste qui en garantit la cohérence. Grâce à la propriété de mise à jour séquentielle des connaissances sur les paramètres que permet la formule de Bayes, quand on obtient des informations supplémentaires, on peut repousser la phase de décision terminale et tirer profit de ces éventuelles informations additionnelles dans une phase de validation.

5.1. La sélection bayésienne de modèle demande un prior informatif

Le modélisateur bayésien souhaite bien souvent éviter les difficultés d'encoder l'expertise recueillie sur le problème dans un prior (Kadane *et al.*, 1998), et l'analyse *a priori* non informative (Berger, 1985) semble un bon moyen de se rapprocher du point de vue classique, où le statisticien refuse de prendre en compte toute information extérieure aux données mesurées. Il faut alors utiliser un prior de densité impropre pour lequel on a pris la précaution de vérifier que la loi *a posteriori* soit, elle, intégrable sur l'espace des paramètres. On peut alors réaliser l'inférence bayésienne ordinaire à partir d'un prior «non informatif» sur les paramètres. Cette approche est malheureusement inacceptable pour le choix de modèle, car le résultat du calcul de $[x | m_j]$ n'est pas une intégrale définie quand le prior est impropre. C'est bien là le plus gros défaut du choix de modèle par facteurs de Bayes.

L'approche généralement utilisée pour pallier à ce problème consiste à utiliser un échantillon d'apprentissage pour mettre à jour un prior impropre en une loi *a posteriori* propre. Cette loi sera à son tour utilisée comme prior – cette fois intégrable – dans une phase d'analyse inférentielle et de sélection de modèle. C'est, au fond, raisonner dans le cadre prédictif où sont distingués : échantillon de calage (ou d'apprentissage) et échantillon de validation constitué par le reste des données. Par exemple, Berger et Perrichi (1996) proposent d'utiliser la médiane d'un facteur de Bayes empirique *intrinsèque* de la façon suivante :

- Prendre un prior non informatif impropre,
- choisir dans les données, un sous-échantillon de taille aussi petit que possible de telle sorte que le posterior soit propre,
- calculer le facteur de Bayes sur le reste des données
- recommencer cette opération sur tous les sous-échantillons possibles et prendre la médiane des facteurs de Bayes ainsi calculés. Cette procédure permet d'ailleurs, lorsque le nombre de sous-échantillons possible est grand, d'être effectuée grâce à un échantillonnage aléatoire.

O'Hagan (1997) développe une idée similaire avec les *Fractional Bayes Factors* : on n'utilise plus cette fois-ci une fraction de l'échantillon pour rendre le prior propre, mais une fraction de la vraisemblance.

5.2. L'échantillon d'apprentissage est formé par les plus petites valeurs

Dans l'approche que développe ce papier, on propose également de s'appuyer sur un échantillon d'apprentissage. Mais ce n'est pas uniquement la commodité mathématique (rendre le prior « propre » est indispensable au choix de modèle par facteurs de Bayes) qui nous guide, mais dans une certaine mesure, c'est également la physique du problème étudié qui permet de faire ici d'une pierre deux coups : l'échantillon d'apprentissage ne contiendra que les valeurs *ordinaires* de l'échantillon. L'échantillon qui servira à calculer les facteurs de Bayes, contiendra par contre toutes les plus hautes valeurs, et permettra ainsi de faire une distinction entre les modèles vis-à-vis de l'emploi qu'on souhaite en faire : décrire au mieux les valeurs extrêmes. Compte tenu de sa connotation d'indépendance, l'emploi du terme *échantillon*, adapté quand il s'agit de deux périodes temporelles différentes de mesures d'un même phénomène *iid*, peut sembler ici abusif. Par les dépendances introduites lors du passage aux statistiques d'ordre, échantillon de calage et échantillon de validation sont nécessairement dépendants. Cette difficulté, réglée par conditionnement dans le cadre bayésien, ne soulève pas de problème technique insurmontable ainsi que le montrent les deux paragraphes suivants.

5.3. Les informations de calage et validation sont liées via les statistiques d'ordre

Quand on travaille avec le modèle POT, le nombre n de dépassements sur une période donnée suit une loi de Poisson. On réalise dans cette section une inférence conditionnelle où le nombre n est fixé. Nous ne considérons ici que les niveaux de dépassement du seuil u : on est en présence d'un n -échantillon indépendant de la distribution de Pareto donnée par l'équation (1).

Soient $x_1^* \leq x_2^* \leq \dots \leq x_{n-N}^* \leq \dots \leq x_n^*$ les observations rangées par ordre croissant.

Dans la phase de validation de nos hypothèses, nous séparons l'information complète en deux parties :

- un *échantillon de calage* \mathbf{x}_{n-N}^* composé des $n - N$ plus petites valeurs enregistrées : $x_1^* \leq x_2^* \leq \dots \leq x_{n-N}^*$,
- un *échantillon de validation* \mathbf{x}_N^* formé des N plus fortes valeurs de chute de neige enregistrées sur la période étudiée : $x_{n-N+1}^* \leq \dots \leq x_n^*$.

Plusieurs vraisemblances nous seront utiles dans la suite. La *vraisemblance partielle marginale* de l'échantillon de calage \mathbf{x}_{n-N}^* , est la loi marginale conjointe des $(n - N)$ premières statistiques d'ordre $[\mathbf{x}_{n-N}^* | \rho, \beta, n]$. On l'obtient à partir de la vraisemblance (10) par intégration et changement de variable. Comme le Jacobien de la transformation des variables initiales \mathbf{x} aux statistiques d'ordre \mathbf{x}^* vaut $n! 1_{x_1^* \leq x_2^* \leq \dots \leq x_n^*}$:

$$[\mathbf{x}^* | \rho, \beta, n] = n! [\mathbf{x} | \rho, \beta, n] 1_{x_1^* \leq x_2^* \leq \dots \leq x_n^*} \quad (17)$$

$$[\mathbf{x}^* | \rho, \beta, n] = n! \rho^n \prod_{i=1}^n (1 - \beta(x_i^* - u))^{\beta-1} 1_{x_1^* \leq x_2^* \leq \dots \leq x_n^*}$$

Par commodité et abus de notation, on n'indiquera plus la nullité des densités hors de leur domaine de définition par la contrainte $1_{x_1^* \leq x_2^* \leq \dots \leq x_n^*}$ quand, pour les paragraphes suivants, on travaillera avec des statistiques d'ordre. On intègre ensuite (17) sur chaque composante de $\mathbf{x}_{\mathbf{N}}^*$:

$$[\mathbf{x}_{\mathbf{n}-\mathbf{N}}^* | \rho, \beta, n] = n! \rho^{n-N} \prod_{i_1=1}^{n-N} (1 - \beta(x_{i_1}^* - u))^{\frac{\rho}{\beta}-1} \times \left(\iiint \prod_{i_2=n-N+1}^n (\rho(1 - \beta(x_{i_2}^* - u))^{\frac{\rho}{\beta}-1}) d\mathbf{x}_{\mathbf{N}} \right)$$

Pour effectuer cette intégration, on retourne à une transformation des dernières statistiques d'ordre en N variables initiales non ordonnées (mais toutes supérieures à la $(n - N)^{\text{ème}}$ statistique d'ordre). Le Jacobien de cette transformation est cette fois $\frac{1}{N!}$ et l'intégration s'effectue alors sans problème, pour donner finalement la vraisemblance partielle marginale de l'échantillon de calage :

$$[\mathbf{x}_{\mathbf{n}-\mathbf{N}}^* | \rho, \beta, n] = \frac{n!}{N!} \cdot \rho^{n-N} \prod_{i=1}^{n-N} (1 - \beta(x_i^* - u))^{\frac{\rho}{\beta}-1} [1 - F(x_{n-N}^* | \rho, \beta, u)]^N \quad (18)$$

D'autre part la vraisemblance conditionnelle $[\mathbf{x}_{\mathbf{N}}^* | \mathbf{x}_{\mathbf{n}-\mathbf{N}}^*, \rho, \beta, n]$ de l'échantillon de validation $\mathbf{x}_{\mathbf{N}}^*$ connaissant l'échantillon de calage $\mathbf{x}_{\mathbf{n}-\mathbf{N}}^*$ et les paramètres, s'obtient selon la définition des probabilités conditionnelles à partir du rapport de (17) par (18) :

$$[\mathbf{x}_{\mathbf{N}}^* | \mathbf{x}_{\mathbf{n}-\mathbf{N}}^*, \rho, \beta, n] = N! \rho^N \prod_{i=n-N+1}^n (1 - \beta(x_i^* - u))^{\frac{\rho}{\beta}-1} [1 - F(x_{n-N}^* | \rho, \beta, u)]^{-N} \quad (19)$$

Cette vraisemblance ne dépend de $\mathbf{x}_{\mathbf{n}-\mathbf{N}}^*$ que par l'intermédiaire de son plus grand élément x_{n-N}^* . F (de densité f) est donné par l'équation (1). $[\mathbf{x}_{\mathbf{N}}^* | \mathbf{x}_{\mathbf{n}-\mathbf{N}}^*, \rho, \beta, n]$ est la vraisemblance d'un N -échantillon ordonné de la distribution tronquée de F sur $[x_{n-N}^*, +\infty]$ et de densité $\frac{f(x)}{1 - F(x_{n-N}^* | \rho, \beta, u)}$.

5.4. La structure semi-conjuguée, commode pour l'inférence, se conserve

5.4.1. Vraisemblances

Compte tenu de l'expression (1) de la fonction de répartition F , l'équation (18) se réécrit sous la forme :

$$[\mathbf{x}_{\mathbf{n}-\mathbf{N}}^* | \rho, \beta, n] = \frac{n!}{N!} \rho^{n-N} \exp((\rho - \beta) \sum_{i=1}^{n-N} S_1(x_i^*, \beta) + N \rho S_1(x_{n-N}^*, \beta)) \quad (20)$$

avec $S_1(x, \beta) = \frac{1}{\beta} \text{Log}(1 - \beta(x - u))$.

La vraisemblance marginale de l'échantillon de calage $[\mathbf{x}_{n-N} | \rho, \beta]$ est de structure tout à fait semblable à celle de la vraisemblance complète. On va profiter de cette similarité de structure pour l'utiliser ultérieurement avec le même modèle de prior.

La vraisemblance conditionnelle de l'échantillon de validation présente également cette propriété, puisqu'elle s'écrit :

$$[\mathbf{x}_N | \mathbf{x}_{n-N}, \rho, \beta] = N! \rho^N \exp((\rho - \beta) \sum_{i=n-N+1}^n S_1(x_i^*, \beta) - N\rho S_1(x_{n-N}^*, \beta))$$

5.4.2. Loi a posteriori avec information de calage

Compte tenu de la seule information incomplète de calage \mathbf{x}_{n-N}^* les distributions *a posteriori* de (ρ, β) sont également de formes connues analogues à (11) :

$$[\rho, \beta | \mathbf{x}_{n-N}^*] \propto \frac{\varphi^\gamma}{\Gamma(\gamma)} \rho^{\gamma-1} \exp(-\varphi\rho) \pi_0(\beta) \rho^{n-N} \exp((\rho - \beta) \sum_{i=1}^{n-N} S(x_i^*, \beta) + N\rho S(x_{n-N}^*, \beta))$$

- On dispose de la marginale *a posteriori* de β à une constante de normalisation près :

$$[\beta | \mathbf{x}_{n-N}^*] \propto \frac{\pi_0(\beta) \exp(-\beta(\sum_{i=1}^{n-N} S(x_i^*, \beta)))}{(\varphi - \sum_{i=1}^{n-N} S(x_i^*, \beta) - N S(x_{n-N}^*, \beta))^{\gamma+n-N}}$$

- Conditionnellement à β , ρ est gamma d'hyper-paramètres $(\gamma + n - N, \varphi - \sum_{i=1}^{n-N} S(x_i^*, \beta) - N S(x_{n-N}^*, \beta))$.
- la loi *a posteriori* de β conditionnellement à ρ fixé est :

$$[\beta | \rho, \mathbf{x}_{n-N}^*] \propto \frac{\varphi^\gamma}{\Gamma(\gamma)} \rho^{\gamma-1} \exp(-\varphi\rho) \exp[(\rho - \beta) \sum_{i=1}^{n-N} S(x_i^*, \beta) + N\rho S(x_{n-N}^*, \beta)] \pi_0(\beta) \quad (21)$$

5.5. Le calcul des vraisemblances moyennes prédictives s'effectue plus facilement grâce à une formule de Raftery réduite

5.5.1. Mise en oeuvre de la formule de Raftery

Appliquée à notre cas, l'évaluation des rapports de Bayes demande le calcul préalable des vraisemblances moyennes prédictives $[\mathbf{x}_N^* | \mathbf{x}_{n-N}^*, m_j]$ de chacun des

modèles m_j . L'indice j prend les valeurs 1, 2, 3 du Tableau 3. L'échantillon complet est $\mathbf{x}_N^* \cup \mathbf{x}_{n-N}^*$. La loi $[\rho_j, \beta_j | \mathbf{x}_{n-N}^*, m_j]$, est propre grâce à l'emploi de l'échantillon de calage. Elle sert de prior quand on passe à l'analyse de l'échantillon de validation \mathbf{x}_N^* .

$$[\mathbf{x}_N^* | \mathbf{x}_{n-N}^*, m_j] = \int [\mathbf{x}_N^* | \mathbf{x}_{n-N}^*, \rho_j, \beta_j, m_j] [\rho_j, \beta_j | \mathbf{x}_{n-N}^*, m_j] d\rho_j d\beta_j \quad (22)$$

La formule de Raftery (16), appliquée avec $\theta = (\beta, \rho)$ et $H = (m_j, \mathbf{x}_{n-N}^*)$ – c'est-à-dire pour la distribution de probabilités $[y | \mathbf{x}_{n-N}^*, m_j]$ – montre que cette intégrale double pourrait se calculer en parallèle aux simulations MCMC utilisées pour évaluer la distribution conjointe *a posteriori* de $\theta_j = \{\rho_j, \beta_j\}$

$$[\rho_j, \beta_j | \mathbf{x}_N^*, \mathbf{x}_{n-N}^*, m_j] = \frac{[\mathbf{x}_N^* | \mathbf{x}_{n-N}^*, \rho_j, \beta_j, m_j] [\rho_j, \beta_j | \mathbf{x}_{n-N}^*, m_j]}{[\mathbf{x}_N^* | \mathbf{x}_{n-N}^*, m_j]} \quad (23)$$

En fait on peut profiter de la structure spécifique du modèle pour construire un algorithme d'estimation beaucoup plus simple.

5.5.2. Adaptation de la formule de Raftery

Laissons de côté pour l'instant l'indice j du modèle (en supposant qu'il conditionne toutes les probabilités de ce paragraphe) et considérons le jeu de paramètres $\theta = \{\rho, \beta\}$. A partir de la formule (23), intégrons par rapport à ρ . On trouve que :

$$\frac{[\beta | \mathbf{x}_{n-N}^*]}{[\mathbf{x}_N^* | \mathbf{x}_{n-N}^*]} = \frac{[\beta | \mathbf{x}_N^*, \mathbf{x}_{n-N}^*]}{\int [\mathbf{x}_N^* | \mathbf{x}_{n-N}^*, \rho, \beta] [\rho | \mathbf{x}_{n-N}^*, \beta] d\rho} = \frac{[\beta | \mathbf{x}_N^*, \mathbf{x}_{n-N}^*]}{[\mathbf{x}_N^* | \mathbf{x}_{n-N}^*, \beta]} \quad (24)$$

En intégrant le premier et le deuxième terme sur β , il vient :

$$\frac{1}{[\mathbf{x}_N^* | \mathbf{x}_{n-N}^*]} = \int \frac{[\beta | \mathbf{x}_N^*, \mathbf{x}_{n-N}^*]}{[\mathbf{x}_N^* | \mathbf{x}_{n-N}^*, \beta]} d\beta \quad (25)$$

C'est donc une formule de Raftery de type (16) de moindre dimension puisque $\theta = \beta$ et $H = (m_j, \mathbf{x}_{n-N}^*)$

De façon générale, cette opération s'applique à tout modèle où on effectue une bipartition des paramètres et une bipartition des données en deux sous-ensembles indépendants ou non. Son intérêt très pratique sur le plan algorithmique est de réduire le calcul numérique de $[\mathbf{x}_N^* | \mathbf{x}_{n-N}^*]$ à une seule intégration numérique lors du calcul employant les simulations MCMC!

En effet, conditionnellement à β le modèle (7) est exponentiel et l'on a justement choisi un prior de structure conjuguée gamma; par conséquent, l'intégration analytique par rapport à ρ (à β fixé) est facile (on connaît l'expression des coefficients

normalisateurs d'une loi gamma); elle donne :

$$[\mathbf{x}_N^* | \mathbf{x}_{n-N}^*, \beta] = N! \frac{\Gamma(\nu + n)}{\Gamma(\nu + n - N)} \frac{[\varphi - \sum_{i=1}^{n-N} S_1(x_i^*, \beta) - N S_1(x_{n-N}^*, \beta)]^{\nu+n-N}}{[\varphi - \sum_{i=1}^n S_1(x_i^*, \beta)]^{\nu+n}} \\ \times \exp(-\beta \sum_{i=n-N+1}^n S_1(x_i^*, \beta))$$

on notera que $\sum_{i=1}^n S_1(x_i^*, \beta) = S_n(\mathbf{x}, \beta)$

6. Résultats sur le cas de la Plagne

6.1. Mise en place de l'algorithme

La procédure que nous avons appelée « validation différentielle », choisit un modèle sur sa capacité à extrapoler pour les grandes valeurs observées les estimations données par les valeurs inférieures. On suppose pour simplifier que l'on considère le nombre d'évènements $n = 62$ comme fixé, auquel cas on ne s'intéresse plus à l'inférence sur μ . On scinde l'échantillon en deux parties : un sous-échantillon d'apprentissage \mathbf{x}_{n-N}^* constitué des $n - N$ plus petits x_i , l'autre sous-échantillon \mathbf{x}_N^* est formé des N plus grandes valeurs de l'échantillon et servira d'échantillon test. La procédure a été répétée pour chacune des 5 valeurs de N

$$N = 10, 15, 20, 25, 30$$

Nous avons adopté par ailleurs les valeurs non informatives $\gamma = \varphi = 0$ et β uniforme pour les distributions *a priori* (hors information $\mathbf{x}_N^* \cup \mathbf{x}_{n-N}^*$). Ce prior initial est mis à jour par la prise en compte du premier sous-échantillon \mathbf{x}_{n-N}^* excluant les N extrêmes : il fournit la loi $[\mu, \beta, \rho | \mathbf{x}_{n-N}^*, m_j]$. Le calcul des rapports de Bayes est effectué grâce à la formule de Raftery sur le second sous-échantillon \mathbf{x}_N^* , ce qui permet un calcul de $[\mathbf{x}_N^* | \mathbf{x}_{n-N}^*, m_j]$ d'après la formule (25). Les simulations ont été menées en parallèle pour chacun des 3 modèles repérés par l'indice 1 pour Fréchet, 2 pour Weibull, 3 pour Gumbel. Les $[\mathbf{x}_N^* | \mathbf{x}_{n-N}^*, m_j]$ sont les distributions prédictives des éléments de \mathbf{x}_N^* *a posteriori* compte tenu des \mathbf{x}_{n-N}^* sous l'hypothèse d'adoption du modèle m_j . Pour le calcul des rapports de Bayes prédictifs sur ces bases, l'algorithme de simulation a été appliqué sur 12000 tirages de $\{\beta, \rho\}$ dont les 8000 premiers ont été éliminés pour les estimations. D'ailleurs la distribution conditionnelle complète de β n'est pas de forme standard aisément manipulable. Nous l'avons discrétisée sur une grille de valeurs équidistantes de 1500 points sur le domaine de β dans chaque cas. Rappelons qu'*a posteriori* ce domaine est limité supérieurement par $\frac{1}{\max(\mathbf{x}_n) - u}$. C'est une procédure de traitement habituelle des séquences markoviennes simulées dont la première phase peut dépendre fortement des conditions initiales arbitraires. Nous ne donnons pas ici la vérification pour ces résultats de contrôle de convergence, effectuée par le test visuel de Gelman et Rubin (1992). De multiples essais avec les mêmes conditions confirment également la stabilité des résultats, notamment quant

au calcul numérique des vraisemblances moyennes, dont les difficultés d'estimation sont connues (Kass et Raftery, 1994).

6.2. La loi a posteriori des deux paramètres de la loi de Pareto

A titre de première information, la figure 2 donne les distributions *a posteriori* des paramètres des 3 modèles compte tenu de l'information complète $\mathbf{x}_N^* \cup \mathbf{x}_{n-N}^*$. On notera la troncature et la forme non standard des distributions de β que le calcul bayésien prend en compte aisément.

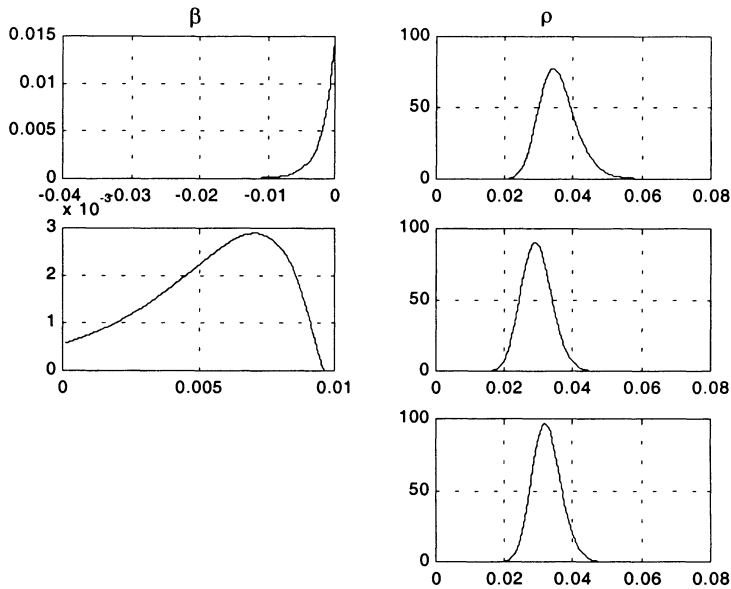


FIGURE 2
 Distributions a posteriori de β à gauche et ρ à droite
 De haut en bas : $\beta < 0, \beta > 0, \beta = 0$

6.3. Les rapports de Bayes prédictifs retiennent le domaine d'attraction de Weibull

Le tableau 6 et la figure 3 montrent les valeurs des rapports de Bayes prédictifs définis selon les principes précédents et leurs évolutions en fonction de N (les indices j pour les modèles m_j réalisent la correspondance avec les domaines d'attractions selon l'ordre FRECHET : 1 - WEIBULL : 2 - GUMBEL : 3).

TABLEAU 6
Rapports de Bayes prédictifs pour les 3 modèles d'extrêmes

$B \setminus N$	10	15	20	25	30
B_{21}	7.59	21.34	26.00	38.51	32.08
B_{23}	3.42	5.84	5.92	6.70	5.93
B_{31}	2.22	3.66	4.39	5.75	5.41

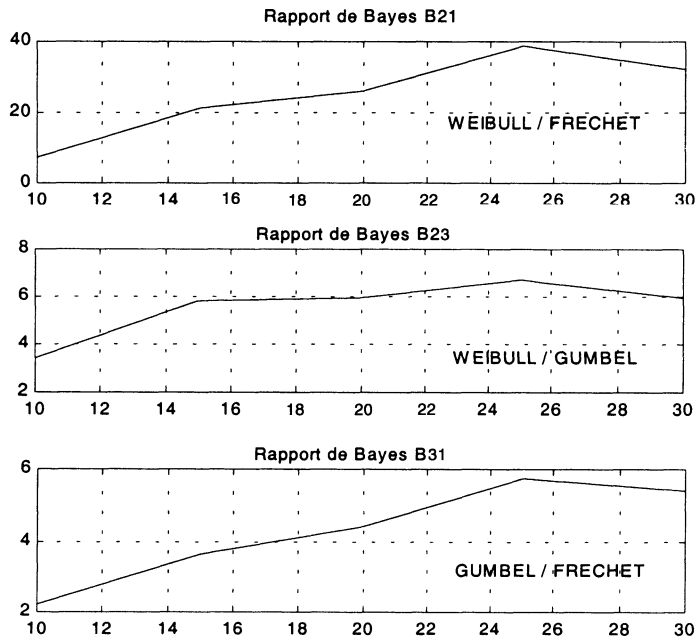


FIGURE 3
Evolution des valeurs des rapports de Bayes prédictifs
selon la taille N de l'échantillon de validation

Si *a priori* les 3 modèles ont des crédibilités (probabilités subjectives) égales, ces rapports estiment alors les «ratios» de crédibilité *a posteriori* correspondants («odd ratios» selon la terminologie anglaise). Sans même se référer au barème de Raftery on note que, quel que soit N , l'augmentation de crédibilité apportée par l'information relative (x_N^* vis à vis de x_{n-N}^*) pour le modèle de Weibull, est très nette en comparaison avec les autres modèles. Vient ensuite Gumbel puis Fréchet. On notera sur la figure 3 que ces «odds ratio» augmentent régulièrement avec N , ce qui traduit «une augmentation d'efficacité de la sélection» jusqu'à une limite cependant qui serait estimée ici à $N = 25$. Il est en effet clair que l'augmentation de taille de l'échantillon de validation est faite aux dépens de la taille de l'échantillon de calage avec notre méthode et donc de la précision du calage.

7. Discussion

7.1. Les avantages à utiliser la formule de Raftery réduite

Dans ce problème il y a plusieurs avantages à utiliser la formule de Raftery réduite (25) qui aide à surmonter les difficultés souvent rencontrées dans le calcul numérique des vraisemblances moyennes :

- La vraisemblance moyenne est calculée par intégration sur la distribution des paramètres *a posteriori* conditionnée par l'information complète $\mathbf{x}_N^* \cup \mathbf{x}_{n-N}^*$. On doit s'attendre à ce que le calcul numérique d'une moyenne par rapport à une distribution moins dispersée que le posterior pour \mathbf{x}_{n-N}^* fixé seulement soit plus précis à effort de calcul donné.
- On sait que si $\beta > 0$, le domaine de la variable de Pareto est limité par le paramètre; de fait $\max(x_n - u) \leq \frac{1}{\beta}$. Le calcul de la vraisemblance moyenne est assez fortement dépendant de la borne de β qui est plus aisément estimée et plus stable avec l'information complète.
- Dans cette étude nous faisons varier la taille N de l'échantillon de validation. La mise en oeuvre de la formule de Raftery (16) sur notre proposition de selection/validation bayésienne de modèle ne nécessite que le calcul d'une seule séquence de l'algorithme de Gibbs appliqué à $[\rho, \beta | \mathbf{x}_N^*, \mathbf{x}_{n-N}^*]$ alors que la simulation par la méthode naïve directe aurait dû être recommencée pour chaque N , puisque le prior en dépend.

7.2. Vers une validation/sélection ouverte

Que ce soit dans l'optique de la théorie bayésienne générale du choix de modèle ou même dans le cadre de la vision différentielle proposée ici (calage effectué sur les données inférieures et validation mise en place sur les valeurs extrêmes de l'échantillon complet), le critère de sélection est très global et ne permet pas d'analyser le résultat en détail. Cette critique tient essentiellement au fait que ces procédures s'inscrivent dans une perspective de choix dite fermée dans le sens où ce choix ne s'effectue que sur un nombre fini de modèles spécifiés. D'autres modèles ne seraient-ils pas préférables? Bernardo et Smith (1994) définissent dans cet esprit *la perspective ouverte* : on travaille encore avec une famille M de modèles, mais on ne suppose plus que les données soient des réalisations de l'un d'entre eux dans l'absolu. Au contraire, on imagine que ces modèles ne peuvent être que des approximations d'un autre modèle que l'analyste ne peut spécifier, mais il doit cette fois considérer la possibilité de rejeter tout modèle de M au profit d'un modèle extérieur non déterminé. Cette formulation est beaucoup plus nuancée et ne débouche pas sur une règle fondée sur les rapports de Bayes où l'analyste limite son choix aux éléments de M . La perspective ouverte est éclairante pour critiquer la méthode développée dans cet article. Nous nous contenterons d'illustrer sur quelques graphiques heuristiques quelques réflexions autour de cette « validation ouverte ».

- Les courbes de la figure 4 donnent pour chaque x_i^* du N -échantillon de validation où $N=25$, l'espérance prédictive $EF_i^{(m)}$ de la probabilité de non dépassement

de x_i^* pour chaque modèle m_j .

$$EF_i^{(m_j)} = \int F(x_i^* | \beta_j, \rho_j, m_j) [\beta_j, \rho_j | \mathbf{x}_{n-N}, m_j] d\beta_j d\rho_j \quad (26)$$

- Supposons que i soit le rang de x_i^* parmi les N valeurs de validation croissantes. Dans une optique non paramétrique d'inférence sur sa probabilité $F(x_i^*)$ de non dépassement, les seules informations utilisables sont n, N, i . L'estimation de $F(x_i^*)$ est possible dans le cadre du modèle Dirichlet-Polya (Ferguson, 1973) classique pour les analyses bayésiennes non paramétriques. Pour ce modèle on peut utiliser pour chaque $F(x_i^*)$ un prior non informatif propre qui est uniforme sur $[0, 1]$. Dans ce cas, un calcul simple donne :

$$EF_i = \frac{n - N + i + 1}{n + 2}$$

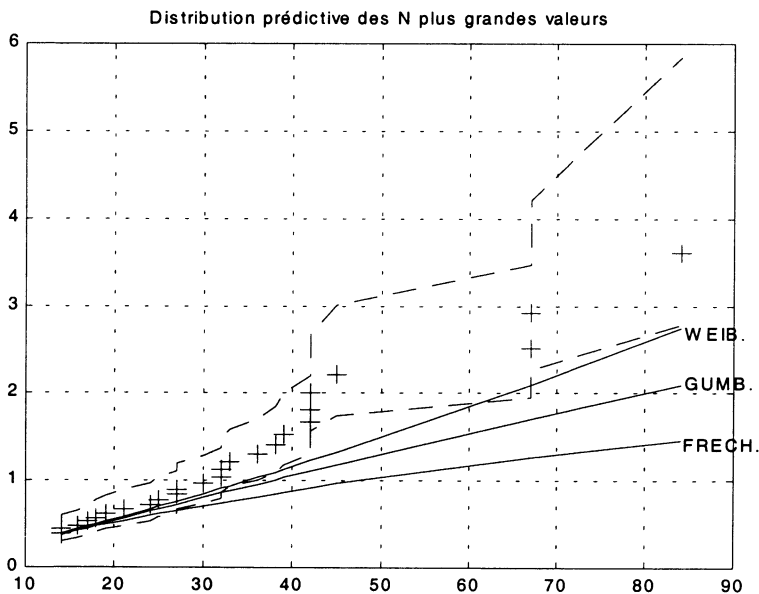


FIGURE 4

Espérance prédictive de la probabilité de non-dépassement de chacune des 25 valeurs de l'échantillon de validation

C'est l'espérance prédictive de la probabilité de non-dépassement pour chaque x_i^* en ne supposant donc aucune forme de distribution pour l'extrapolation proposée en son temps par Laplace (1847). Ces espérances non paramétriques (Fortin *et al.*, 1997) sont représentées par des «+» sur le graphique ensemble avec les limites d'une « bande de crédibilité à 90% » (tiretés – –) elle-même non paramétrique calculée à

partir de la distribution *a posteriori* de $F(x_i^*)$ qui est une $B\hat{e}ta(n - N + i, N - i)$. Cette bande, assez large, est naturellement sensible à la succession erratique des observations croissantes.

Ce graphique montre un certain décalage, dont on ne peut toutefois donner la signification statistique ici, par rapport au modèle Weibull, privilégié par la sélection fermée. On constate cependant que celui-ci est le plus proche de la fourchette de crédibilité inférieure.

7.3. Analyse de sensibilité sur le seuil u

7.3.1. Perspective ouverte

La méthode graphique d'analyse non paramétrique peut être étendue aux deux échantillons $x_N \cup x_{n-N}$ réunis :

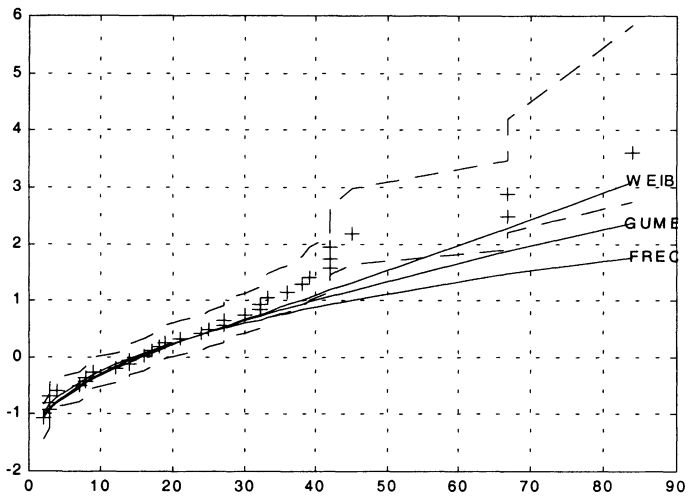


FIGURE 5
Distributions des extrêmes calées sur un échantillon complet de $n = 32$ (valeurs $> 70cm$)

Le graphique de la figure 5 montre l'ajustement, toujours représenté par les espérances prédictives des probabilités. Il s'agit de l'échantillon complet des chutes de neige supérieures au seuil u plus élevé égal à 70 cm. La comparaison de ce graphique avec le précédent où $N = 25$ semble nous indiquer que l'écart précédent avec la loi de Weibull pourrait être imputable au choix du seuil à $u = 50cm$.

7.3.2. Perspective fermée

Nous ajoutons ici le tableau 7 des rapports de Bayes différentiels avec $u = 70$ ($n = 32$) pour 3 valeurs de $N = (10, 15, 20)$,

TABLEAU 7
Rapports de Bayes prédictifs pour les 3 modèles d'extrêmes avec $u > 70\text{cm}$

$B \setminus N$	10	15	20
B_{21}	5.74	14.30	17.15
B_{23}	2.03	2.81	2.86
B_{31}	2.83	5.09	5.99

La comparaison des courbes et des points permet ainsi un jugement visuel heuristique sur la valeur d'un modèle en validation. Nous employons le qualificatif heuristique car la comparaison ici des espérances prédictives associées à des intervalles de crédibilité indépendants ne permet qu'un jugement partiel, cependant assez éclairant dans notre exemple. Cette présentation graphique se rapproche d'une présentation de la Statistique traditionnelle comme celles de la figure 1 mais elle a l'avantage sur celle-ci d'être interprétée en termes prédictifs par un jugement probabiliste direct. Pour les applications de la méthode à cet exemple, nous ne pouvons aller guère plus loin dans les essais avec différents seuils et différentes valeurs de N , compte tenu de l'information limitée disponible.

8. Conclusions

Cet article développe une méthode de choix du domaine d'attraction des valeurs extrêmes selon l'idée qu'il faut retenir un modèle en fonction de sa capacité à prendre en compte les grandes valeurs alors que le calage n'est effectué que sur les plus basses valeurs. Cette validation « différentielle » est essentiellement différente de celle qui utiliserait tout l'échantillon car ici les extrêmes ne participent pas au calage initial du modèle. C'est un aspect particulier d'un problème de validation que permet l'analyse bayésienne.

En appliquant cette méthode au cas de la station de la Plagne, nous concluons que le modèle théorique des extrêmes basé sur l'hypothèse Weibull est réaliste pour la prévision des risques de fortes chutes de neige sur 3 jours. Cependant il apparaît que le seuil $u = 50\text{cm}$ initialement choisi pourrait être trop petit. Malheureusement l'information disponible est trop limitée pour faire une étude complète de la sensibilité de la validation au choix du seuil.

Sur le plan méthodologique il faut signaler que notre proposition ne se heurte à aucune difficulté technique de prise en compte des incertitudes d'échantillonnage et de choix de modèles. Ceci est d'autant plus important que le modèle Weibull, privilégié par les résultats précédents, est particulièrement difficile à estimer en statistique classique compte tenu de l'existence d'une borne supérieure paramétrique pour les observations. Le choix d'un prior semi-conjugué facilite le calcul par simulation de la formule de Raftery qui fonctionne alors sur un seul paramètre au lieu de deux, mais c'est loin d'être une contrainte technique de commodité indispensable à la bonne application de notre méthode.

Ainsi que le suggérait déjà Lehman (1986), une procédure bayésienne donne aussi une méthode générale pour construire un estimateur ponctuel souvent raisonnable et admissible sous des conditions générales. C'est pourquoi un statisticien

«de l'Ecole fréquentiste» pourra sans difficulté tirer parti de la technique bayésienne présentée dans cet article pour mettre en place une procédure de test statistique du paramètre de forme de la distribution de Pareto généralisée, concurrente de la méthode présentée par Chaouche et Bacro (2003), par exemple. A taille n fixée, il cherchera à déterminer quel est le couple approprié de valeurs (B_{seuil}, N) permettant de tester l'hypothèse H_0 (le modèle est m_3 , celui de Gumbel), contre H_1 (le modèle est m_1 , celui de Fréchet) selon la règle de décision : «ne pas rejeter H_0 si $B_{31}(\mathbf{x}) > B_{seuil}$ », avec $B_{31}(\mathbf{x})$ calculé selon l'équation (25). Par répétitions de simulation de n données sous H_0 , on peut en effet évaluer le risque de première espèce et, par simulations sous H_1 , le risque de seconde espèce au point de la contre-hypothèse $\beta = 1, \rho = 1$.

Enfin, si la sélection de modèles par l'approche bayésienne peut sembler plus complexe que la sélection par l'approche classique, elle est aussi plus riche et plus nuancée. Accepter un modèle par l'application trop mécanique d'un test ou d'une « p -value» comme le propose la plupart des méthodes ou logiciels standards peut être extrêmement fallacieux. L'approche bayésienne a l'avantage de pouvoir considérer l'usage qui doit être fait d'un modèle (Bernier *et al.* 2000) quand on cherche à le sélectionner. Par exemple, l'espérance prédictive utilisée dans l'équation (26), peut, selon les principes bayésiens, être remplacée par tout autre caractéristique choisie en fonction de conséquences définies dans une optique décisionnelle. C'est sans doute pour ces raisons que la critique et la sélection bayésiennes de modèles dans le cadre prédictif font aujourd'hui l'objet de nombreux développements originaux et prometteurs (Gelman *et al.* 1996).

Remerciements

Notre reconnaissance va à Pierre Cazes pour ses corrections d'épreuves méticuleuses et pour ses conseils, ainsi qu'à Jean Noël Bacro qui a contribué à l'amélioration de la première version de cet article.

9. Bibliographie

- ANCEY C. (1996), *Guide Neige et Avalanche : Connaissances, pratiques, sécurité*. Éditions Édisud.
- BERGER J.O. (1985), *Statistical decision theory and Bayesian analysis*. Springer Verlag, New York.
- BERGER J.O. and PERICCHI L. (1996), The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association*, 91, 109-122.
- BERNARDO J.M. and SMITH A.F.M. (1994), *Bayesian theory*. John Wiley and Sons, Londres.
- BERNIER J., PARENT E., BOREUX J.J. (2000), *Statistique pour l'environnement. Traitement bayésien de l'incertitude*. Lavoisier, Tec § Doc.
- CARLIN B. P. and CHIB S. (1995), Bayesian Model Choice via Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society*, B 57,3 473-484.

- CHAOUICHE A. and BACRO J.N. (2003), A statistical test procedure for the shape parameter of the generalized Pareto distribution, *Comp. Stat. Data An.*
- CHIB S. and JELIAZKOV I. (1995), Marginal Likelihood from the Gibbs Output, *J. Amer. Stat. Assoc.*, 90, 1313-1321.
- CHIB S. and JELIAZKOV I. (2001), Marginal Likelihood from the Metropolis-Hastings Output, *J. Amer. Stat. Assoc.*, 96, 270-281.
- COLES S. G. (2001), *An introduction to statistical modeling of extreme values.* Springer Verlag, New York.
- COLES S. G. and POWELL E. A. (1996), Bayesian methods in extreme value modelling : a review and new developments. *Internat. Stat. Rev.* 64, 119-136.
- DAVISON A. C. and SMITH R. L. (1990), Models for exceedances over high thresholds (with discussion). *Journal of the Royal Statistical Society*, B 52, 393-442.
- FERGUSON T. (1973), A Bayesian Analysis of Some Non Parametric Problems. *Annals Statist.* 1, 209-230.
- FISHER R. A., TIPPETT L. H. C. (1928), Limiting forms of the frequency distribution of the largest or the smallest member of a sample. *Proc. Cambridge Philo Soc.* 24, 180-190.
- FORTIN V., BERNIER J. et BOBÉE B. (1997), Simulation, Bayes, and bootstrap in statistical hydrology. *Water Resour. Res.*, 33 (3) : 439-448.
- FRÉCHET M. (1927), Sur la loi de probabilité de l'écart maximum. *Annals of Polish Math. Soc.* 6, 93.
- GELFAND A. E. and DEY D. K. (1995), Bayesian Model Choice : Asymptotics and exact calculations. *Journal of the Royal Statistical Society*, B 56,3 501-514.
- GELFAND A. and SMITH A. F. M. (1990), Sampling-based approaches to calculating marginal densities. *Journ. Amer. Stat. Ass.*, 85 : 398-409.
- GELMAN A., CARLIN J.B., STERN H.S. and RUBIN D.B. (1995), *Bayesian data analysis.* Chapman and Hall, Londres.
- GELMAN A. and RUBIN D.B. (1992), Inference from iterative simulation using multiple sequences, *Statistical Science*, 7, 457-511.
- GELMAN A., MENG X. and STERN H. (1996), Posterior Predictive Assessment of Model Fitness via Realized Discrepancies. *Statistica Sinica* 6, 733-807
- HAN C. and CARLIN B.P. (2000), MCMC methods for computing Bayes factors : a comparative review. Division of Biostatistics, University of Minnesota. *MCMC preprint*, available at <http://www.stat.duke.edu>.
- HOETING J.A., MADIGAND., RAFTERY A.E., VOLINSKY C.T. (1999), Bayesian model averaging : A Tutorial, *Statistical Science*, 14, 4, p. 382-417.
- KADANE J. B., WOLSON L.J., O'HAGAN A., CRAIG R. (1998), Papers on 'Elicitation' and Discussions. *The Statistician* 47 (1), 3-53.
- KASS R.E. and RAFTERY A.E. (1994), Bayes factors. *J. Amer. Stat. Assoc.*, 90, 773-795.

- LAPLACE P. S. (1847), Œuvres, Tome VII, *Théorie analytique des probabilités*. Reprint 1995. Editions Jacques Gabay, Paris.
- LEHMAN E.L. (1986), *Theory of point estimation*. 2nd edition. John Wiley and Sons, New York.
- O'HAGAN A. (1997), Properties of intrinsic and fractional Bayes factors. *Test* 6, 101-118.
- O'HAGAN A. (1998), Elicitation expert beliefs in substantial practical applications - *The Statistician* 47 (1), 21-35.
- PARENT E., BERNIER J. (2001), Méthodes bayésiennes et modélisation des risques géophysiques extrêmes. *La Revue de Modulad*. 28 : 1-26. INRIA.
- PARENT E., BERNIER J. (2002), Bayesian P.O.T. modeling for historical data. *Journal of Hydrology*, 274 : 95-108.
- PICKANDS J. (1975), Statistical inference using extreme order statistics. *Ann. Stat.*, 3, 11-130.
- SMITH R. L. (1984), Threshold model for sample extreme in T. de Oliveira (ed.), *Statistical Extremes and Applications*, 621-638. Reidel Dordrecht.
- TANNER M.A. (1996), *Tools for statistical inference : methods for the exploration of posterior distribution and likelihood functions*. Springer Verlag, New York.