

Les incertitudes hydrologiques dans les problèmes de dimensionnement d'ouvrages Valeur des informations locales et spatiales

Importance of an integrated analysis
of hydrological problems of estimation
and of decision in water resource fields

J. BERNIER¹

RÉSUMÉ

Le présent article a pour but de montrer l'importance d'une analyse intégrée des problèmes d'estimation hydrologique et des problèmes de décisions d'aménagements pour tenir compte rationnellement des diverses incertitudes intervenant dans ces problèmes. La méthode d'analyse est basée sur les techniques bayésiennes qui permettent des évaluations économiques comparées des divers types d'informations locales ou régionales mobilisées. Un cas d'application schématique est présenté.

Mots clés : hydrologie, décisions, incertitudes, informations, régionalisation, méthodes de Bayes.

SUMMARY

This paper aims to show the importance of an integrated analysis of hydrological problems of estimation and of decision in water resource fields. Such an analysis is necessary to take into account the various uncertainties involved. The analysis is based on bayesian techniques which allow comparisons between the economical values of various types of information. This information can be both local and regional. A schematic example is presented.

Key-words : hydrology, decisions, uncertainties, informations, regionalization, Bayes methods.

1. Laboratoire National d'Hydraulique (EDF-DER), 6, quai Watier, 78401 Chatou Cedex, France.

1 - INTRODUCTION

En matière de gestion des eaux ou de dimensionnement de structures hydrauliques il est d'usage de distinguer deux types de problèmes censés être de natures différentes :

- d'un côté les problèmes d'estimation des éléments hydrologiques pertinents comme la distribution de fréquences des crues par exemple,

- de l'autre côté le choix opérationnel (de nature économique en particulier) des décisions d'aménagement comme la fixation de valeurs de projet notamment, compte tenu des conséquences diverses résultant de ce choix.

Le traitement séparé de ces deux problèmes néglige l'aspect essentiel de la solidarité qui les lie, solidarité résultant de la propagation des diverses incertitudes hydrologiques de mesure des observations, de modélisation et d'échantillonnage.

Il est intéressant d'envisager l'ensemble de ces problèmes dans un contexte global pour mieux apprécier l'incidence de ces incertitudes. Par ailleurs ce contexte global est particulièrement indiqué pour comparer les diverses techniques de réduction de ces incertitudes par mobilisation des informations hydrologiques : informations locales disponibles sur chaque site étudié (souvent en quantité limitée), informations complémentaires spatiales.

Une approche globale peut être donnée par la théorie de la décision et les méthodes bayésiennes. S'il importe de distinguer les grandeurs aléatoires naturelles (débits, précipitations) d'un côté, les grandeurs fixées mais incertaines (invariants des bassins, paramètres des modèles) de l'autre côté, les méthodes bayésiennes caractérisent ces grandeurs entâchées d'incertitude, comme les aléas naturels, par des distributions de probabilité sur les champs de valeurs possibles. Le fondement de cette approche probabiliste des incertitudes est basée sur une conception décisionnelle où les probabilités sont définies par les conditions du pari que tout décideur est amené à prendre face à des conséquences incertaines de ces décisions (voir notamment BERGER, 1985). On conçoit donc que l'approche bayésienne fournisse un cadre adéquat à l'analyse globale dont il a été question plus haut.

Le propos de cet article est de montrer l'intérêt de cette approche pour apprécier la valeur des informations hydrologiques locales et régionales dans un contexte d'aménagement des eaux. Les problèmes réels (dimensionnement de protections contre les crues, prévision hydrologique, gestion des eaux en temps réel,...) peuvent être traités de cette façon. Pour faciliter l'exposition nous nous bornerons ici à présenter un cas, schématique pour certaines des hypothèses admises, appartenant à la première classe de problèmes cités ci-dessus.

2 - LE CAS SCHÉMATIQUE

Le dimensionnement des ouvrages de protection contre les crues est un problème opérationnel classique. Il n'en est pas moins important et actuel sous de nombreux aspects notamment en matière d'assainissement urbain pour déterminer les capacités d'évacuation des eaux pluviales. Schématiquement, pour un bassin donné, on considère des décisions de choix d'un débit de crue de projet a dont il résulte les conséquences suivantes :

- un coût d'investissement de protection égal à $C \cdot a$, supposé ainsi proportionnel au débit de projet. (On posera $C = 1$ en exprimant ainsi les coûts de dommages en proportion de ce coût d'investissement),

- des dommages d'inondation résiduels (après choix de a) égaux à $D \cdot (q - a)$, supposés ainsi proportionnels à l'écart par rapport au débit de projet du débit réel q d'une crue éventuelle si celui-ci est plus grand, et nul sinon. On considèrera un exemple numérique où $D = 4$, exprimant ainsi un dommage quatre fois plus grand que l'investissement par unité de débit.

Ainsi à chaque décision a est associé un coût total d'investissement et de dommages moyens exprimé par :

$$L(a; \lambda; \rho) = C \cdot a + k\lambda \cdot \int_a^{+\infty} D(q - a) \cdot f(q; \rho) \cdot dp \quad (1)$$

En effet les dommages étant aléatoires, le coût total doit faire intervenir la valeur moyenne, l'espérance mathématique du coût des dommages. Dans cette relation :

- k prend en compte le taux d'actualisation annuel nécessaire pour représenter l'effet de la répartition temporelle des crues dommageables (on utilisera un taux de 4 %),

- λ est le taux d'occurrence des crues dommageables (nombre moyen par année). Ce paramètre caractérise la distribution statistique du nombre annuel N d'occurrences représentée usuellement par la loi de Poisson :

$$\text{Prob.}[N = n] = e^{-\lambda} \cdot \frac{\lambda^n}{n!} \quad (2)$$

- $f(q, \rho)$ est la densité de probabilité des niveaux en débit de ces crues dommageables ; elle dépend d'un paramètre ρ et est représentée fréquemment par la loi exponentielle :

$$f(q, \rho) = \rho e^{-\rho(q - q_0)} \quad (3)$$

où q_0 est un seuil convenablement choisi inférieur au seuil de dommages.

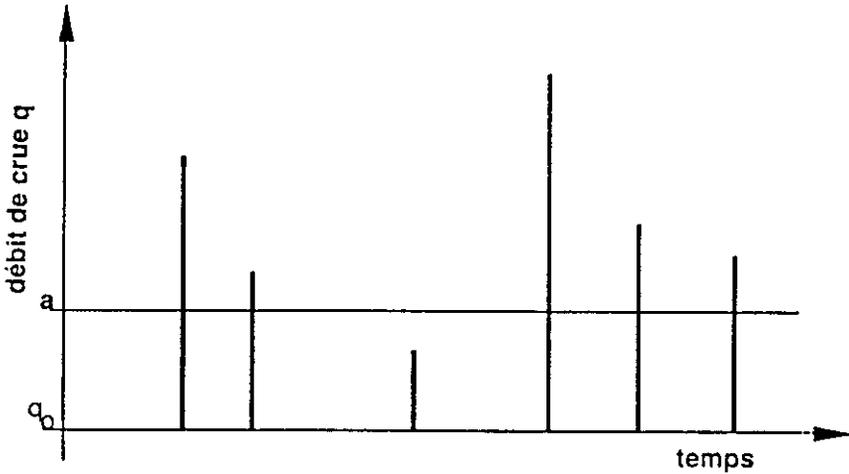


Figure 1 Occurrence des crues dommageables.

2.1 Limites des hypothèses précédentes

Le modèle probabiliste décrit par les équations précédentes représente les aléas naturels dans notre problème ; appelé modèle de dépassement (ASH-KAR et ROUSSELLE, 1981) ou modèle de renouvellement (MIQUEL, 1984), il est assez réaliste pour les crues ou même les précipitations génératrices de crues. Le caractère schématique de notre exemple ressort surtout des expressions données aux différents coûts économiques et à l'hypothèse sous-jacente que seuls les coûts monétaires sont pris en compte pour apprécier les conséquences des décisions. Nous reviendrons plus loin sur ces hypothèses.

2.2 Retour au problème de décision

Le calcul du coût total s'effectue simplement, on obtient :

$$L = C \cdot a + \frac{k\lambda D}{\rho} \cdot e^{-\rho(a - q_0)} \quad (4)$$

De ce résultat on peut déduire la décision optimale \hat{a} minimisant L et le coût minimal L_m :

$$\hat{a} = q_0 - \frac{1}{\rho} \text{Log} \frac{C}{\lambda k D} \quad (5)$$

$$L_m = L(\hat{a}, \lambda, \rho) = C \left(\hat{a} + \frac{1}{\rho} \right) \quad (6)$$

Il peut être intéressant de traiter le cas numérique d'un bassin où le seuil q_0 est égal à $100 \text{ m}^3/\text{s}$ et les paramètres λ et ρ respectivement estimés à 1,3 et $0,015 (\text{m}^3/\text{s})^{-1}$; on obtient alors, avec les hypothèses économiques précédentes :

$$\hat{a} = 424,5 \text{ m}^3/\text{s}$$

$$L_m = 491,17 \text{ C}$$

La crue de projet correspond ici exactement à la crue centennale.

3 - INCERTITUDES ET REGRET - L'APPROCHE BAYESIENNE

La crue de projet \hat{a} ci-dessus obtenue représente la solution du problème de décision dans le cas d'information parfaite (modèle de renouvellement parfaitement connu y compris les valeurs de ces paramètres λ et ρ). Sans nous poser ici de problèmes d'incertitudes sur la forme du modèle hydrologique ainsi que sur les hypothèses économiques d'ailleurs, il est important d'analyser l'effet des incertitudes hydrologiques sur les paramètres λ et ρ que nous nommerons dans la suite « état de la nature θ ». La méconnaissance de θ entraîne une certaine incertitude quant au choix de l'optimum de décision. Si donc une décision a différente de \hat{a} est effectivement prise, le décideur encourt un surcoût égal à :

$$OL(a, \theta) = L(a, \theta) - L(\hat{a}, \theta) \quad (7)$$

Ce surcoût, appelé **regret**, a une interprétation économique importante ; si la décision a est envisagée, il représente le prix maximum que le décideur doit être prêt à payer pour obtenir l'information parfaite (sans incertitude) sur θ . En effet le regret est alors égal à l'économie réalisée en changeant la décision a au profit de \hat{a} optimale si θ est connu parfaitement. Ainsi le regret s'interprète comme la valeur de l'information parfaite associée à a . Pour notre cas :

$$OL(a, \lambda, \rho) = C(a - \hat{a}) + \frac{k\lambda D}{\rho} \cdot e^{-\rho(a - q_0)} - \frac{C}{\rho} \quad (8)$$

L'approche bayésienne propose de traiter l'incertitude sur θ (ici le couple λ, ρ) par une distribution de probabilité de densité $\pi(\theta)$, tout comme on traite habituellement l'aléa naturel sur le débit. Notons cependant qu'aléa naturel et incertitude sont de natures différentes. Aucune procédure d'observation, aucune collecte de données ne peut permettre de réduire l'aléa naturel au long terme des crues notamment mais on peut toujours envisager une procédure d'observation, une collecte de données assez longue pour permettre une réduction de l'incertitude. C'est cette possibilité de modification de la « probabilité » $\pi(\theta)$ qui en permet une évaluation non arbitraire qui dépend de toute

information hydrologique x utile pour θ . Une telle information x est utile si sa densité de probabilité $f(x;\theta)$ dépend effectivement de l'état de la nature incertain. On obtient $\pi(\theta)$, notée en fait $\pi^x(\theta)$ pour faire apparaître cette information x , par la formule dite de Bayes :

$$\pi^x(\theta) = \frac{f(x;\theta) \cdot \pi^0(\theta)}{\int f(x;\theta) \cdot \pi^0(\theta) \cdot d\theta} \quad (9)$$

$\pi^x(\theta)$ est la distribution à postériori et $\pi^0(\theta)$ la distribution à priori élaborée par l'hydrologue à partir de ses connaissances et des informations complémentaires.

Le dénominateur de la formule (9) :

$$m(x) = \int f(x;\theta) \cdot \pi^0(\theta) \cdot d\theta \quad (10)$$

est la distribution de l'information x dite prépostérieure qui jouera un rôle important dans les transferts d'information.

Malgré la différence de nature soulignée précédemment entre incertitude et aléa naturel, il est justifié (cf. BERGER, 1985) d'utiliser la distribution *a posteriori* dans tout contexte décisionnel. La décision optimale \hat{a}^x compte tenu de l'incertitude sur θ modifiée par l'information x est obtenue en minimisant la moyenne du coût total pondérée par cette distribution *a posteriori* ou de façon équivalente l'espérance *a posteriori* du regret :

$$EOL^x(a) = \int OL(a,\theta) \cdot \pi^x(\theta) \cdot d\theta \quad (11)$$

La valeur minimale du regret représente alors la valeur intrinsèque de l'information parfaite *a posteriori* compte tenu de l'information (imparfaite) apportée par x , soit :

$$VIP(x) = EOL^x(\hat{a}^x) \quad (12)$$

Comme toute information efficace doit réduire l'incertitude et donc diminuer le prix maximal que le décideur est prêt à payer l'information parfaite, on peut mesurer la valeur spécifique de l'information x par l'écart des valeurs de l'information parfaite à priori et *a postériori* :

$$VI(x) = EOL^x(\hat{a}^x) - EOL^0(\hat{a}^0) \quad (13)$$

où $EOL^0(.)$ a la même expression que (11) mais calculée pour la distribution *a priori*.

Les calculs peuvent être aisément faits avec notre exemple. Pour références ultérieures nous donnerons ici des résultats seulement approchés, valables pour autant que l'information x ne soit pas trop réduite. Le débit optimal de la crue de projet s'écrit :

$$\hat{a}^x \approx \bar{a} + \frac{V(\rho/x)}{E(\rho/x)} \cdot \frac{(\bar{a} - q_0)^2}{2} \quad (14)$$

$$\text{avec } \bar{a} = q_0 - \frac{1}{E(\rho/x)} \cdot \text{Log} \frac{C}{kD \cdot E(\lambda/x)}$$

Quant à la valeur *a posteriori* approchée de l'information parfaite, elle s'écrit :

$$\text{VIP}(x) \approx C \left[\frac{(\bar{a} - q_0)^2}{2} \cdot \frac{V(\rho/x)}{E(\rho/x)} + \frac{1}{2} \frac{V(\lambda/x)}{E(\rho/x) \cdot E^2(\lambda/x)} \right] \quad (15)$$

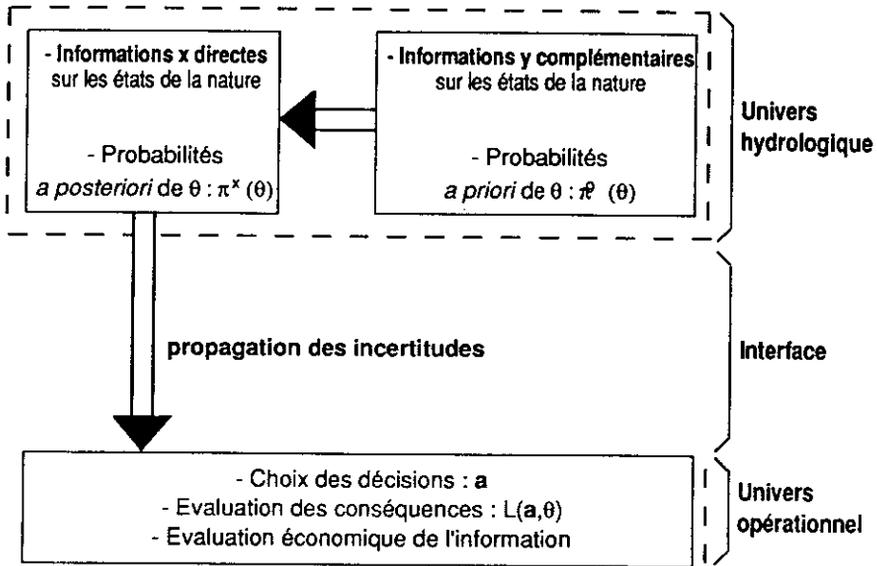
Dans ces expressions (14) et (15), $E(. / x)$ et $V(. / x)$ désignent les espérances et variances *a posteriori* des paramètres correspondants. Elles seront adaptées aux différents types d'informations analysés dans les paragraphes suivants.

4 - INFORMATIONS HYDROLOGIQUES ET PROBABILITÉS A POSTERIORI

Le diagramme de la page suivante présente la structure du problème de décision en illustrant la dualité des deux univers : l'univers opérationnel où se place le choix des décisions et l'univers hydrologique où se traite l'information pour « connaître » les états de la nature. Contrairement à l'usage fréquent qui tend à séparer ces deux univers, il importe d'insister sur leur solidarité étroite exprimée par la propagation des incertitudes sur les états de la nature. C'est le propre de la théorie de la décision de proposer une démarche intégrée. On constatera en particulier que la décision optimale \hat{a}^x , ainsi que la valeur économique de l'information (formules (14) et (15)) dépendent de ces incertitudes par l'intermédiaire des variances *a posteriori* de λ et ρ .

L'approche bayésienne intégrée prend donc correctement en compte la propagation des incertitudes par le calcul des probabilités *a posteriori* des états de la nature et leur introduction dans le processus de décision.

Cette approche intégrée diffère de la présentation usuelle de la méthode de Bayes en hydrologie (cf. KUCZERA, 1982, 1983) qui isole les probabilités à priori π^0 de toutes les informations possibles qui sont mobilisées par la formule de Bayes. Ceci a pour résultat fallacieux d'amplifier le caractère subjectif, qualifié trop souvent d'arbitraire, de l'évaluation de π^0 . Mais en hydrologie les probabilités *a priori* peuvent n'être pas arbitraires si on mobilise pour leur évaluation des informations pertinentes comme le montreront nos exemples.



Pour illustrer cela nous avons décomposé l'univers hydrologique en deux blocs : le premier bloc représente le « sous-univers » où la seule information mobilisée est constituée d'informations directes du processus aléatoire en jeu dans l'univers opérationnel : c'est-à-dire les dépassements observés du seuil q_0 pour les crues de la station hydrométrique de référence pour le site étudié ; le second bloc représente le « sous-univers » qu'il faut considérer pour mobiliser les informations complémentaires (observations sur un site voisin, données régionales, etc.)

4.1 Informations directes

Il s'agit ici de l'ensemble des observations du processus des dépassements de crue à la station de référence étudiée et que l'on résume habituellement pour ce modèle avec les deux grandeurs statistiques suivantes :

$$\bullet \bar{N}_1 = \frac{N_1}{p} \quad \text{où } N_1 \text{ est le nombre total de dépassements du seuil } q_0 \text{ observés sur } p \text{ années,}$$

$$\bullet \bar{p}_1 = \frac{N_1}{\sum_{i=1}^{N_1} (q_i - q_0)} \quad \text{où } \sum_{i=1}^{N_1} (q_i - q_0) \text{ est la somme des écarts au seuil pour les } N_1 \text{ dépassements observés}$$

Conjointement à ces données, les connaissances *a priori* de l'hydrologue peuvent être assez floues.

Cette situation très classique est prise en compte dans l'approche bayésienne par l'utilisation d'une distribution *a priori* dite peu informative pour l'état de la nature θ comme l'illustre la figure 2.

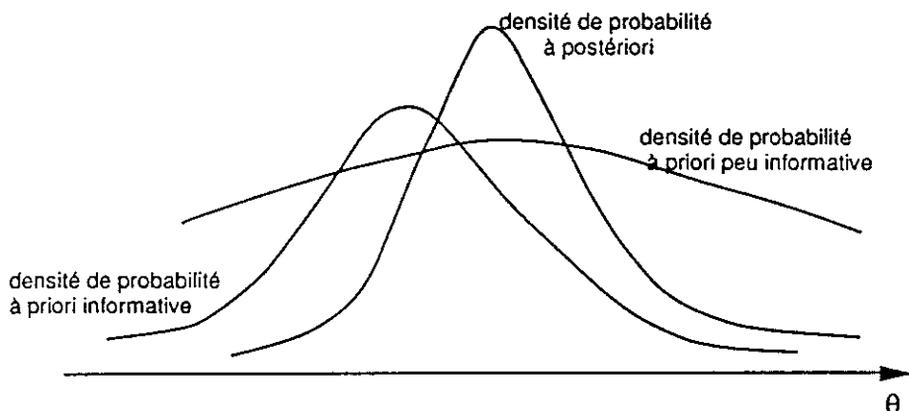


Figure 2 Distributions *a priori* et *a posteriori*.

Avec ces données la règle de Bayes (9) permet le calcul des paramètres *a posteriori* :

$$E(\lambda/x) = \bar{N}_1 \quad \text{et} \quad \text{Var}(\lambda/x) = \frac{\bar{N}_1}{p^2} \quad (16)$$

$$E(\rho/x) = \bar{\rho}_1 \quad \text{et} \quad \text{Var}(\rho/x) = \frac{\bar{\rho}_1}{N_1}$$

Une application est illustrée par la figure 3. Les valeurs des paramètres économiques sont celles du paragraphe 3 où l'optimum $\hat{\lambda}^x$ (égal ici à \bar{a} - cf la formule (14)) correspond à la crue centennale dans le cas d'information parfaite. Cette figure montre la variation du surdimensionnement $\hat{\lambda}^x - \bar{a}$ rendu nécessaire par l'incertitude sur le modèle dont les paramètres sont en fait estimés sur les données ; la variation est exprimée en fonction du nombre d'années p d'observations. Ce surdimensionnement est comparé à celui obtenu en utilisant la solution classique de la limite supérieure de l'intervalle de confiance à 95 % et qui apparaît comme très nettement supérieur.

Ce résultat doit être commenté ; le surdimensionnement « optimal » semble assez faible, c'est qu'il ne tient compte que d'un seul type d'incertitude liée aux erreurs d'échantillonnage pour un modèle donné alors que les incertitudes de choix du modèle pourtant souvent plus sévères ne sont pas incluses dans cet exemple schématique. Par ailleurs le surdimensionnement « optimal » serait certainement plus grand pour un coût de dommage de crue croissant plus rapidement que linéairement avec le débit. Utiliser la limite supérieure d'un intervalle de confiance pour un seuil de confiance assez large c'est, d'une certaine façon vouloir intégrer forfaitairement l'ensemble des incertitudes dans le calcul. Mais cette intégration est faite à l'aveugle sans

distinguer la part des diverses incertitudes comme le ferait une analyse bayésienne plus complète et qui est parfaitement possible ici. Il faut surtout retenir de l'exemple numérique l'existence d'un surdimensionnement lié à l'incertitude plutôt que les valeurs numériques absolues.

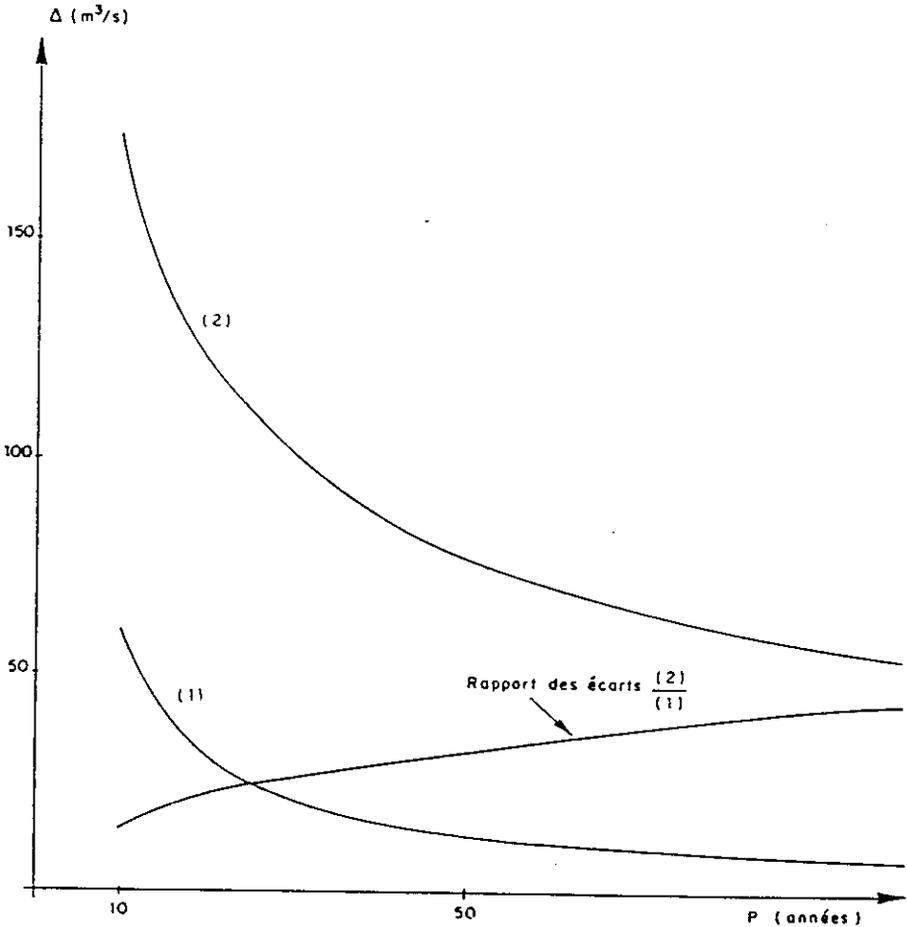


Figure 3 Surdimensionnement $\Delta = \bar{a}^x - \bar{a}$ dû à l'incertitude.

(1) Solution bayésienne approchée

(2) Limite supérieure de l'intervalle de confiance à 95 %

$$q_0 = 0$$

$$= 0,01; \bar{N}_1 = 1,3; \bar{\rho}_1 = 0,015$$

$$\frac{C}{kD}$$

$$\bar{a} = 424,5 \text{ m}^3/\text{s}$$

} valeurs de référence
de la crue de projet

4.2 Informations complémentaires

Des informations complémentaires sont nécessaires toutes les fois où les observations directes sont insuffisantes pour modifier très notablement la distribution de probabilité *a priori* de l'état de la nature. Nous allons donner deux exemples où l'utilisation de la règle de Bayes permet la mobilisation d'une information objective complémentaire Y.

4.2.1 Données incomplètes

Ici l'information complémentaire Y est celle d'une station hydrologique voisine où on a observé des dépassements de crues (au-dessus d'un seuil adéquat) sur une plus longue série de P années dont p sont communes avec les observations de la première station.

Pour mobiliser statistiquement cette information complémentaire il est nécessaire de représenter la dépendance stochastique entre les crues des deux stations sous la forme d'un modèle de dépassement à deux variables où interviennent :

- le processus des crues concomitantes aux deux stations et non concomitantes pour chacune d'entre elles (avec des taux d'occurrence annuels respectifs $\lambda_0, \lambda_1, \lambda_2$ qui vérifient la relation $\lambda = \lambda_0 + \lambda_1$),

- la liaison statistique entre les amplitudes (débits ou niveaux) des crues concomitantes estimée par le coefficient de corrélation r entre les amplitudes des crues concomitantes sur la série commune.

Ce type de modèle a déjà été utilisé dans le domaine océanographique (BERNIER et GRAFF, 1983). Il peut s'appliquer tout aussi bien aux problèmes hydrologiques pour autant que la liaison entre stations soit significative (taux

de concomitance $\frac{\lambda_0}{\lambda_0 + \lambda_1}$ et coefficient de corrélation assez forts).

Dans le cadre de ce modèle l'information Y permet une évaluation préalable de la distribution à priori π_0 par l'utilisation préliminaire de la formule de BAYES avec Y et une première distribution *a priori* peu informative. La méthode de calcul utilise généralement une estimation particulière intermédiaire des données manquantes Y' de la première station pour la période complémentaire de P - p années (cf. DEMPSTER *et al.*, 1977). On obtient les résultats suivants :

$$\left. \begin{aligned} E(\lambda/x,y) &= \bar{N}_1 + \frac{n_0}{n_0 + n_2} [\bar{N}_2 - (\bar{n}_0 + \bar{n}_2)] = \hat{\lambda} \\ \text{Var}(\lambda/x,y) &= \frac{\hat{\lambda}}{p} \left[1 - \frac{P-p}{p} \frac{n_0^2}{(n_0 + n_1)(n_0 + n_2)} \right] \end{aligned} \right\} \quad (17)$$

$$\left. \begin{aligned} E(\rho/x,y) &= \bar{\rho}_1 \left[1 - r + r \frac{\bar{\rho}_2}{\bar{\rho}_2^c} \right] = \hat{\rho} \\ \text{Var}(\rho/x,y) &= \frac{\hat{\rho}^2}{N_1} \left[1 - \frac{(N_2 - N_1)}{N_2} r_2 \right] \end{aligned} \right\} \quad (18)$$

Pour les notations on se reportera à la figure 4. Les n ou N barrés représentent les moyennes annuelles (division par p des nombres correspondants) et les ρ barrés les inverses des moyennes des écarts au seuil q_0 des amplitudes.

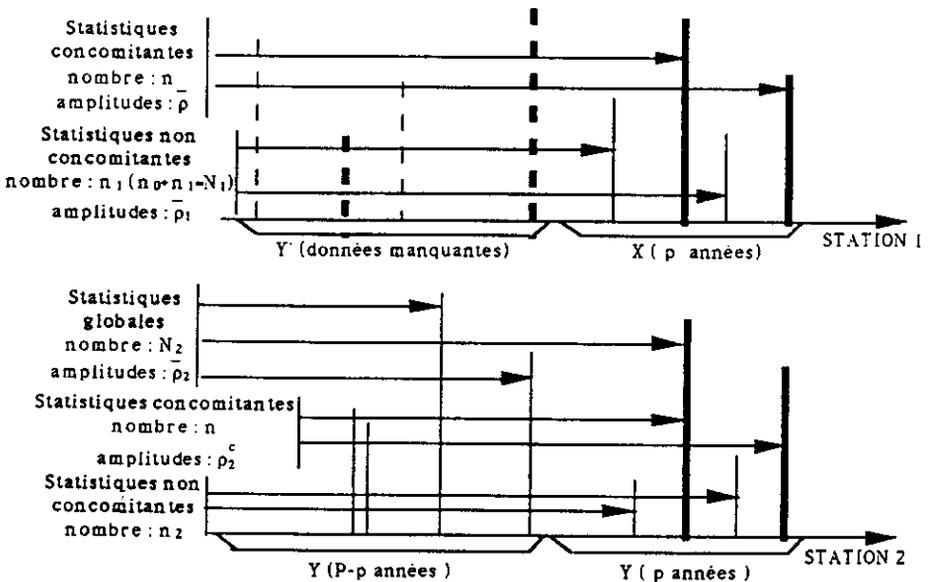


Figure 4 Configuration d'informations concomitantes et non concomitantes à deux stations hydrologiques.

Il suffit ensuite d'inclure ces résultats dans les formules (14) et (15) pour obtenir d'une part le surdimensionnement dû à l'incertitude résiduelle et d'autre part la valeur de l'information parfaite VIP. Le gain apporté par une information spécifique est mesuré par la différence des VIP avant et après mobilisation de cette information. La VIP est exprimée en pourcentage du coût d'investissement pour notre exemple numérique dans la figure 5. On y montre également la VIP afférente à l'utilisation de la seule information x de la station 1. Pour mobiliser les informations complémentaires y la méthode hydrologique traditionnelle consiste à estimer les données manquantes Y' par une

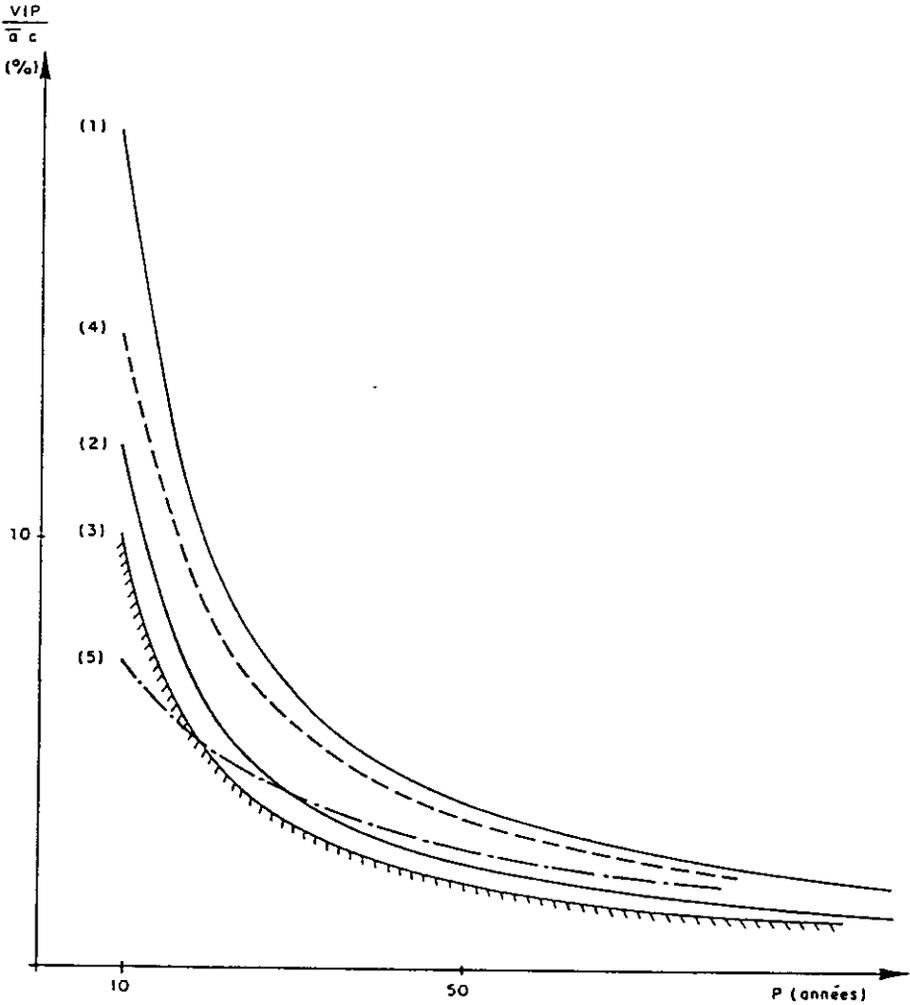


Figure 5 Valeur de l'information parfaite exprimée en pourcentage du coût d'investissement (paramètres numériques identiques à ceux de la figure 2).

(1) Information directe (corrélation et concomitance nulles)

(2) Information complémentaire sur une station voisine ($r = 0,9$; $\frac{\lambda_0}{\lambda_0 + \lambda_1} = 0,40$)

avec $P = 2p$ et $N_2 = 2,5p$

(3) Information complémentaire ($P = 2p$; $N_2 = 2,5p$) utilisée de façon déterministe $\left(r = \frac{\lambda_0}{\lambda_0 + \lambda_1} = 1 \right)$

(4) Transfert d'information régionale (grande dispersion spatiale $\frac{s}{p} = 0,5$; $\frac{\bar{N}}{S^2} = 0,5$)

(5) Transfert d'information régionale (homogénéité spatiale plus importante $\frac{s}{p} = 0,20$; $\frac{\bar{N}}{S^2} = 0,10$)

corrélation classique en considérant ces observations **reconstituées** comme des données réelles et négligeant l'incertitude de cette reconstitution ; en effet ceci revient à admettre que la liaison statistique est une liaison déterministe caractérisée par une corrélation et une concomitance absolues.

La comparaison, illustrée dans la figure 5, des résultats corrects avec ceux de cette dernière procédure fallacieuse montre le rôle important joué par la nature non déterministe de la relation entre les deux stations hydrologiques.

4.2.2 Transfert d'information régionale

Dans le cas de données hydrologiques localement limitées l'information régionale peut être très précieuse. L'hydrologie traditionnelle a élaboré des modèles régionaux, dans le cas des crues notamment. On connaît le succès rencontré par la méthode de transposition spatiale des averses en particulier. Ces méthodes souffrent de ne pas posséder de cadre rationnel où la mesure de la précision et donc de leur intérêt serait possible. L'approche bayésienne fournit un tel cadre. Elle a été notamment proposée pour ce type de problème hydrologique par KUCZERA (1982-1983). Nous en donnerons ici un aperçu partiel pour le cas de dépassements de crue.

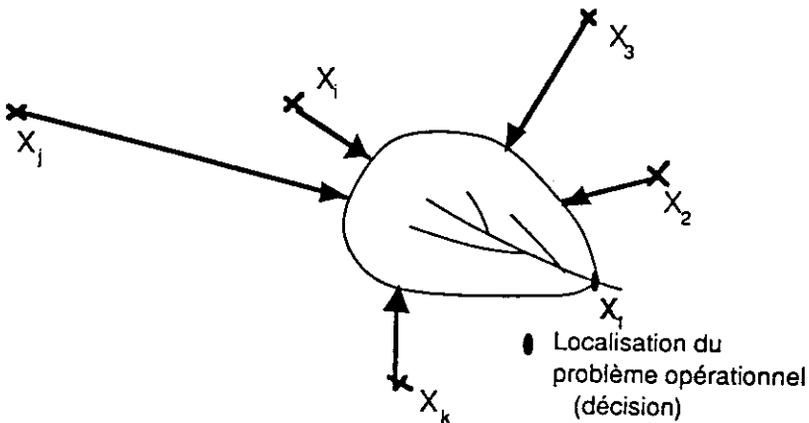


Figure 6 Transposition d'information régionale.

Nous noterons $X_2, X_3, \dots, X_i, X_k$ les observations de $k-1$ processus de dépassements de crues dans $k-1$ stations régionales environnant le premier site et pouvant éventuellement enrichir l'information disponible pour celui-ci. Sans tenir compte de liaisons stochastiques spatiales du genre évoqué dans le paragraphe précédent, il est possible d'utiliser ces données régionales dans le cas où les occurrences de crue présentent une certaine homogénéité spatiale relativement aux états de la nature respectifs θ_i .

Dans le contexte hydrologique traditionnel admettre cette homogénéité c'est admettre l'hypothèse d'égalité de ces θ_i ; dans le contexte bayésien la notion d'homogénéité est beaucoup plus souple, elle implique seulement que les informations *a priori* sont échangeables pour ces paramètres c'est-à-dire

que leur distribution de probabilité conjointe *a priori* est invariante par permutation quelconque des θ_i . Ainsi toutes les informations X peuvent-elles être mobilisées dans ce cas. Une condition suffisante d'échangeabilité est l'hypothèse *a priori* qu'un θ_i quelconque dans l'ensemble régional est une variable aléatoirement distribuée avec une densité $\pi^0(\theta; \alpha, \gamma)$ dépendant de paramètres régionaux α et γ . Il importe alors d'utiliser cette densité *a priori* avec une estimation spatiale de ces paramètres régionaux.

La méthode bayésienne complète utilise la notion de modèles hiérarchiques (BERGER, 1985) où les paramètres α et γ jouent le rôle d'états de la nature de deuxième niveau. Une approximation pratique de ces méthodes hiérarchiques est la méthode de Bayes empirique basée sur la distribution prépostérieure $m(x)$ de l'ensemble des X_i (voir la formule (10)) qui sont ainsi utilisés pour l'estimation des paramètres.

Il faut choisir une distribution *a priori* commode et assez réaliste pour représenter la variabilité spatiale des paramètres. Dans le cas des modèles de dépassement une telle distribution peut être la loi gamma :

$$\pi^0(\theta; \alpha; \gamma) = \frac{\alpha^\gamma}{\Gamma(\gamma)} e^{-\alpha\theta} \theta^{\gamma-1}$$

pour les paramètres respectifs $\theta = \lambda$ et $\theta = \rho$.

On obtient finalement les espérances *a posteriori* suivantes :

$$E(\lambda/X_1, X_2 \dots X_k) = (1 - C_1) \bar{N}_1 + C_1 \bar{\bar{N}} \tag{19}$$

$$\text{avec } \bar{\bar{N}} = \sum_{i=1}^k \frac{\bar{N}_i}{k}$$

$$\text{et } C_1 = \frac{\bar{\bar{N}}}{S_2} \text{ où } S_2 = \sum_{i=1}^k \frac{(\bar{N}_i - \bar{\bar{N}})^2}{k}$$

$$E(\rho/X_1, X_2 \dots X_k) = (1 - C_2) \bar{\rho}_1 + C_2 \bar{\bar{\rho}} \tag{20}$$

$$\text{avec } \bar{\bar{\rho}} = \sum_{i=1}^k \frac{\bar{\rho}_i}{k}$$

$$\text{et } C_2 = \frac{\bar{\bar{\rho}} \bar{\rho}_1}{\bar{\bar{\rho}} \bar{\rho}_1 + N_1 s^2} \text{ où } s^2 = \sum_{i=1}^k \frac{(\bar{\rho}_i - \bar{\bar{\rho}})^2}{k}$$

Les variances *a posteriori* respectives sont :

$$\text{Var}(\lambda/X_1, X_2 \dots X_k) = \frac{E(\lambda/X_1, X_2 \dots X_k)}{\rho} (1 - C_1) \tag{21}$$

$$\text{Var}(\rho/X_1, X_2, \dots, X_k) \approx E^2(\rho/X_1, X_2, \dots, X_k) \frac{s^2}{N_1 s^2 + \rho^2} \quad (22)$$

On constate que les moyennes *a posteriori* s'expriment comme des moyennes pondérées entre l'estimation locale (\bar{N}_1 ou $\bar{\rho}_1$) et une estimation spatiale moyenne sur toutes les stations (\bar{N} ou $\bar{\rho}$) avec des pondérations fonction des

dispersions régionales relatives $\frac{s^2}{\bar{N}}$ ou $\frac{s^2}{\bar{\rho}}$ des paramètres estimés.

Les variances *a posteriori* ainsi que la valeur de l'information parfaite résultante (voir la figure 5) sont fonction de ces dispersions relatives. Plus celles-ci sont faibles plus la valeur de l'information régionale augmente. Par ailleurs si on compare les deux types d'informations complémentaires (site voisin fortement corrélé ou ensemble régional de sites échangeables) on constate que le second type peut être plus intéressant, pour autant que l'hypothèse d'échangeabilité soit réaliste bien sûr. Toutefois dans la réalité hydrologique cette hypothèse a des limites évidentes ; des facteurs de variabilités régionales systématiques existent. Il importe soit de restreindre l'échantillon des sites échangeables soit de représenter les variabilités systématiques par un modèle exprimant les θ_1 en fonction de variables physiographiques, climatologiques... expliquant les spécificités régionales. Ce sont alors les écarts constatés à ce modèle régional qui sont supposés échangeables.

5 - CONCLUSIONS

Les résultats numériques de nos exemples ne doivent pas être jugés dans l'absolu ; le surdimensionnement dû à l'incertitude et la valeur de l'information sont certainement sous-estimés par les hypothèses du calcul effectué (BERNIER, 1985). Ces hypothèses concernent les particularités des cas hydrologiques étudiés, la seule prise en compte des incertitudes d'échantillonnage (choix du modèle hydrologique non remis en question et l'on sait bien qu'un mauvais choix peut être très sensible), les approximations statistiques utilisées et les hypothèses économiques adoptées pour ces exemples illustratifs. Pour être plus réalistes nos exemples devraient notamment tenir compte de coûts marginaux de dommages croissant avec le débit. Il reste cependant que la réduction des effets des incertitudes due à l'utilisation d'informations complémentaires peut être économiquement importante dans le cas de données locales limitées. Les méthodes de Bayes sont particulièrement adaptées et assez souples pour mobiliser les types d'information les plus divers. Notons à ce propos qu'elles ont été utilisées pour prendre en compte des données historiques épisodiques observées dans le passé sur le site étudié (MIQUEL, 1984, BERNIER *et al.*, 1986).

Le cas du dimensionnement des protections contre les crues n'est pas le seul où la prise en compte d'informations hydrologiques diverses est importante pour le décideur. C'est aussi et peut-être surtout le cas des gestions de ressources en eau en temps réel et des prévisions intervenant comme éléments d'information dans les décisions de gestion. La méthodologie exposée succinctement ici permet l'introduction et la comparaison opérationnelle de diverses procédures de prévision mobilisant des informations temporelles et spatiales d'origines hydrologiques variées.

RÉFÉRENCES BIBLIOGRAPHIQUES

- ASHKAR F. et ROUSSELLE J., 1981. Design Discharge as a Random Variable. A Risk Study. *Water Resour. Res.*, 17(3) : 577-591.
- BERGER J.O., 1985. *Statistical Decision Theory and Bayesian Analysis*. Springer Verlag, New York.
- BERNIER J., 1985. Elements of Bayesian Analysis of Uncertainty in Hydrological Reliability and Risk Models. In : *Engineering Reliability and Risk in Water Resources*, L. DUCKSTEIN and E. PLATE (editors), M Nijhoff, Dordrecht, The Netherlands.
- BERNIER J., 1985. Bayesian Analysis : Further Advances and Applications. In : *Engineering Reliability and Risk in Water Resources*, L. DUCKSTEIN and E. PLATE (editors), M Nijhoff, Dordrecht, The Netherlands.
- BERNIER J. et GRAFF M., 1983. Méthode d'estimation de la houle de projet avec données incomplètes, Rapport EDF-LNH : HE42/83.34.
- BERNIER J. et MIQUEL J., 1979. Incertitudes de l'information et décisions en situation de risque - Cas des problèmes de protection contre les crues. *La Houille Blanche (Paris)* 4-5.
- BERNIER J., LEBOSSE A., MIQUEL J. et GRIFFET M., 1986. Use of Additional Historical Information for Estimation and Goodness of Fit of Flood Frequency Models, *Proceedings of the International Symposium on Flood Frequency and Risk Analysis*, Baton Rouge, USA.
- DAVIS D.R., KISIEL C.C., DUCKSTEIN L., 1972. Bayesian Decision Theory Applied to Design in Hydrology. *Water Resour. Res.*, 8(1) : 33-42.
- DEMPSTER A.R., LAIRD N.N., RUBIN D.B., 1977. Maximum Likelihood from Incomplete Data via the E.M. Algorithm, *J. Roy. Stat. Soc. série B*, 39(1).
- DUCKSTEIN L., PLATE E.J., 1985. A System Framework for Hydrologic Reliability with Applications to the Design of Structures. *Int. Ass. Hydro-sciences J.* 14.
- DUCKSTEIN L. et BERNIER J., 1985. Critères associés au risque en ressources hydriques, un modèle mathématique de système avec applications, Atelier sur l'analyse de risque dans la gestion des ressources hydriques, Ecole Polytechnique de Montréal, Québec, nov.
- KUCZERA G., 1982. Combining Site-Specific and Regional Information : an Empirical Bayes Approach, *Water Resour. Res.*, 18(18).
- KUCZERA G., 1983. A Bayesian Surrogate for Regional Skew in Flood Frequency Analysis. *Water Resour. Res.* 18(3).
- MIQUEL J., 1984. *Guide pratique d'estimation des probabilités de crues*. Eyrolles, Paris.
- TODOROVIC P., ZELENHASIC E., 1970. A Stochastic Model for Flood Analysis. *Water Resour. Res.* 6(6).