
CRUE : Un modèle d'estimation des probabilités des débits de crue

CRUE : A model for calculating the probability of river flooding

J. Miquel

Laboratoire national d'hydraulique
Direction des Etudes et Recherches d'Electricité de France

INTRODUCTION

Il existe de nombreuses méthodes d'estimation des risques de crue. Parmi les qualités qu'un utilisateur demande à une méthode, les suivantes sont importantes :

- exiger des hypothèses aussi peu nombreuses et aussi peu contraignantes que possibles ;
- exploiter au mieux l'information disponible, aussi diverse soit-elle en nature et en quantité ;
- être fiable, c'est-à-dire contrôler les hypothèses, évaluer les incertitudes et à l'usage donner de bons résultats ;
- être facile et rapide à mettre en oeuvre.

Il n'y a pas de méthode miracle, qui, pour résoudre n'importe quel problème, possède toutes les qualités énumérées ci-dessus. Il apparaît d'ailleurs que ce n'est pas seulement une méthode que l'utilisateur souhaite, mais un ensemble cohérent valable pour une classe de problèmes, qui de l'information, à l'édition des résultats, passe par l'analyse et la critique des données, la vérification des hypothèses, les procédures d'estimation, les analyses de sensibilité sur les paramètres incertains, et finalement l'évaluation des incertitudes.

Un tel ensemble a été élaboré au Laboratoire National d'Hydraulique au cours de ces dix dernières années, et appliqué à de nombreux bassins français, grands et petits, en vue d'estimer les événements dont la probabilité pouvait atteindre le millénaire. L'information minimale requise est une série de débits journaliers d'au moins 15 ans, qui est analysée statistiquement par une méthode de type Renouvellement, décrite plus loin.

Après la mise au point d'un outil informatique (Programme CRUE), un manuel pratique a été récemment rédigé [7] de façon à présenter les méthodes, et surtout à en faciliter l'application. Un tel manuel est forcément volumineux, et le résumer, même en quelques pages, serait fastidieux : on a donc préféré examiner ici quelques points importants et donner un exemple.

L'INFORMATION

Le premier type d'information nécessaire à une étude de crue est descriptif : météorologie, types de crues, propagation et concomitances, influences humaines, etc. Ces renseignements ont pour but de renseigner sur la physique des phénomènes, mais également de déceler d'éventuelles hétérogénéités, non stationnarités, ou bien vérifier la cohérence d'une estimation avec celles qui ont pu déjà être faites sur le même bassin.

Le second type d'information, le plus fréquemment employé, est une série de hauteurs observées chaque jour, en un ou plusieurs points du site, pendant quelques années. La conversion de ces hauteurs en débits est impérative : la sensibilité des hauteurs aux conditions locales d'écoulement est beaucoup plus forte que celle des débits. Une étude de crue basée sur des hauteurs est dangereuse (sauf contrôle soigneux de la stabilité des conditions d'écoulement des crues de l'échantillon) et fragile (car elle risque d'être périmée par une modification future).

Le troisième type d'information que nous appelons "historique" est l'ensemble des renseignements, plus ou moins précis, que l'on a pu collecter sur les plus fortes crues de ces 100 ou 200 dernières années. Cette information, qui éveille beaucoup de méfiance, du fait des incertitudes qui l'entourent, est paradoxalement l'une des plus importantes, surtout si la série des débits journaliers est courte. En effet le plus gros risque d'erreur réside moins dans l'estimation imparfaite d'un paramètre de loi (erreur d'échantillonnage) que dans le choix de la loi elle-même, (changements de phénomène entre les crues exceptionnelles et les crues faibles ou moyennes), ou bien dans une non stationnarité de très longue période (50 ans ou plus), indécélable avec les courtes séries de débits journaliers.

C'est une des caractéristiques de ce manuel, que d'avoir redonné son importance à ce type d'information, qui est directement intégré dans les estimations elles-mêmes.

D'autres informations peuvent être utilisées, la principale étant la pluviométrie. Bien que ce manuel ne la traite pas, plusieurs solutions existent pour l'intégrer dans les méthodes d'estimation : soit en travaillant sur les données (reconstitution de série), soit directement sur les paramètres des lois.

LES HYPOTHESES

En pratique deux sont essentielles : homogénéité et stationnarité.

L'*homogénéité*, qui garantit que tous les débits appartiennent bien à la même "famille" peut être mise en défaut :

- si les débits ne sont pas tous naturels ;
- si les régimes climatiques ou les concomitances aux confluent sont très marqués ;
- si les effets saisonniers sont prépondérants.

Dans de tels cas, on peut être amené :

- à reconstituer les débits naturels ;
- à séparer les échantillons en sous-familles ;
- à traiter séparément les crues d'Automne-Hiver et de Printemps-Eté, puis vérifier que :

$$F(Q)_{\text{annuel}} = F_{AH}(Q) \cdot F_{PE}(Q)$$

La *stationnarité* exprime que les propriétés statistiques de la variable étudiée sont indépendantes du temps.

Le contrôle de cette hypothèse est effectué automatiquement par le programme CRUE, de deux façons différentes :

- en découpant l'échantillon en deux parties, et en testant les moyennes ;
- en traçant la distribution des dates d'occurrence des événements inférieurs à un seuil, et en vérifiant qu'elle est uniforme sur la période d'observation.

RAPPEL DES PRINCIPALES METHODES

Nous écarterons les méthodes "empiriques" ou "déterministes" qui en maximisant les phénomènes n'évaluent pas clairement les risques. Quant aux méthodes probabilistes, elles peuvent se ranger en quatre catégories :

- les méthodes à échantillonnage fixe (Maxima Annuels),
- les méthodes du type Renouvellement,
- les méthodes "hydrométéorologiques",
- les méthodes stochastiques.

Les *méthodes stochastiques* modélisent statistiquement les processus à l'échelle journalière ou mensuelle, et sont très utiles pour étudier, par simulation, des variables complexes, mais elles contrôlent mal les incertitudes, ce qui limite la validité des extrapolations.

Les *méthodes "hydrométéorologiques"* sont fiables et opérationnelles pour des bassins versants limités et homogènes (par exemple le "gradex"). Elles ont également l'avantage d'analyser les événements générateurs, et d'être ainsi plus proches de la physique des phénomènes. Par contre, pour de grands bassins versants

(> 20 000 km²), ou en cas de mécanismes hydrométéorologiques complexes (situations météorologiques très variées, neige, crues en série, concomitances nombreuses, ...), les difficultés de combinaisons des événements élémentaires conduisent actuellement à des incertitudes encore beaucoup trop grandes. Il n'en reste pas moins que ces méthodes nous paraissent prometteuses, malgré leur complexité.

Ces deux types de méthodes ont été rappelés pour mémoire car ils ne sont pas utilisés dans le guide.

Les *méthodes du type Maxima Annuel*, sont à la fois les plus simples (ajustement direct d'une loi statistique au débit le plus fort de chaque année), et les plus répandues dans le monde, aux variantes d'estimations près.

Les lois les plus fréquemment employées sont les lois log. Normale, Gumbel, Frechet, Pearson ou log. Pearson dont les procédures d'estimation, au demeurant très classiques, sont rappelées dans le manuel.

A signaler quelques défauts que les méthodes du type Renouvellement ont l'avantage de résoudre :

1) L'aspect "processus" de la série des débits d'une rivière est complètement ignoré : on néglige toutes les fortes crues qui n'ont pas eu la "chance" d'être le maximum de l'année où elles se sont produites. Cette information "oubliée" fait évidemment défaut lorsque l'échantillon est court.

2) Certaines années la crue maximale annuelle n'a pas été forte : ce type de crue faible n'a pas grand chose à voir avec les crues importantes dont on veut connaître la probabilité. L'hypothèse d'homogénéité de l'échantillon est donc mal vérifiée.

3) La prise en compte d'information historique est délicate : on peut certes placer les points représentatifs de ce type d'information sur les graphiques d'ajustement, mais il est difficile de disposer d'un test objectif de rejet ou d'acceptation de la loi ajustée en fonction des écarts, et ensuite de modifier les paramètres de la loi en tenant compte de cette information particulière.

Les *méthodes du type Renouvellement* considèrent l'ensemble du processus des débits journaliers et retiennent tous les événements qui dépassent une certaine amplitude. De la sorte les événements importants figurent bien dans l'échantillon, et ceux qui sont mineurs ne s'y trouvent pas, ce qui améliore l'homogénéité des séries. La suite de ces événements forme alors un processus de Renouvellement, comme, en fiabilité, les pannes de machines.

On combine ainsi deux aléas : l'occurrence dans les temps des événements dépassant un seuil S_0 , et l'amplitude des événements qui dépassent ce seuil (Fig. 1).

Le calcul de la probabilité $F(Q)$ pour qu'une crue Q^* , la plus forte de l'année, ne dépasse pas une valeur Q , s'effectue en appliquant les règles de composition des probabilités d'événements indépendants :

$$\text{Prob}[Q^* < Q] = \sum_{k=0}^{\infty} \text{Prob}$$

[Au cours de l'année, $\exists k$ crues $> S_0$ et toutes $< Q$]

Si $P(k) = \text{prob}[\exists k \text{ crues } > S_0 \text{ au cours de l'année}]$

et $G(Q) = \text{prob}[q < Q \mid q > S_0]$

alors :

$$F(Q) = \sum_{k=0}^{\infty} P(k) \cdot G(Q)^k$$

$P(k)$ est habituellement bien représenté par une loi de Poisson (parfois Binomiale Négative).
et $G(Q)$ est une loi de type exponentiel, choisie parmi les suivantes :

$$G(Q) = 1 - e^{-\rho(Q-S_0)}$$

$$G(Q) = 1 - e^{-\rho(\log Q - \log S_0)}$$

$$G(Q) = 1 - e^{-\rho(Q-S_0)^p} \text{ (Weibull)}$$

$$G(Q) = 1 - e^{-\rho(Q^2-S_0^2)}$$

$$G(Q) = 1 - \theta e^{-\rho_1(Q-S_0)} - (1 - \theta) e^{-\rho_2(Q-S_0)}$$

Si $G(Q)$ proche de 1, alors :

$$F(Q) = 1 - N(1 - G(Q))$$

avec :

$$N = \sum_{k=0}^{\infty} k \cdot P(k) = \text{nombre moyen annuel de crue.}$$

Habituellement le nombre de paramètre à estimer est de 2 ou 3 :

- un pour la loi de Poisson (nombre moyen annuel de crue),
- un pour la loi exponentielle simple (ρ) ou deux pour la loi de Weibull (ρ et p).

A ces paramètres il faut ajouter le seuil choisi S_0 : bien que n'étant pas un paramètre, il faut le choisir bas pour avoir un nombre suffisant de crues, mais pas trop bas pour avoir une bonne indépendance entre les crues. Une analyse de sensibilité est donc nécessaire, ne serait-ce que pour s'assurer que les résultats ne sont pas sensibles à une légère variation de ce seuil.

LES ESTIMATIONS ET LEURS INCERTITUDES

Après le recueil et la critique des données, les procédures d'ajustements passent par les trois étapes suivantes :

- choisir la loi statistique à ajuster aux crues observées,
- estimer ses paramètres,
- évaluer les incertitudes sur le résultat.

Le choix de la loi ne s'effectue pas a priori, mais par comparaison des qualités d'ajustement d'un éventail de lois qui correspondent à des comportements asymptotiques différents (4 pour les "Maxima annuels", 7 pour le "Renouvellement") : examen graphique d'une part, test d'adéquation (X^2) d'autre part.

L'estimation des paramètres, classique pour les méthodes de "Maxima Annuels" est par contre améliorée pour le "Renouvellement", par l'utilisation des techniques, bayésiennes, qui offrent l'avantage de pouvoir intégrer des informations très diverses (opinions d'expert), de se prêter à des réévaluations successives, et permet de calculer de façon plus simple et plus com-

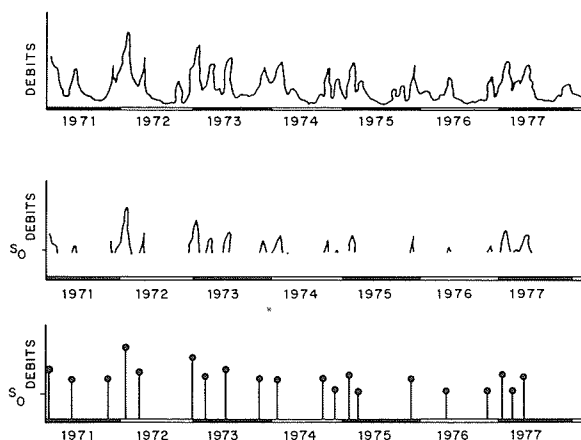


Figure 1 – Processus de renouvellement constitué par les crues dépassant un seuil S_0 fixé.

plète les incertitudes d'échantillonnage. D'autre part l'information "historique", c'est-à-dire celle constituée par l'évaluation du débit des crues exceptionnelles, a été intégrée dans les calculs, au même titre que l'information régulière habituelle constituée par les débits journaliers. On trouvera en annexe les principales étapes de ces calculs.

Les incertitudes qui entourent le résultat final sont de deux types :

- les incertitudes "non calculables"
- les incertitudes "calculables".

Les incertitudes "non calculables" sont par exemple :

- des erreurs systématiques, sur les données (détarages, extrapolations) ;
- des hétérogénéités de données, ou bien de phénomènes (saison, concomitances) ;
- des défauts d'exhaustivité (oubli de crues exceptionnelles) ;
- des défauts de stationnarité de très longues périodes (> 50 ans) ;
- inadéquation de la loi choisie, (visible sur graphique) dans la zone des crues observées.
- inadéquation de la loi choisie dans la zone extrapolée (changement de phénomène).

Nous avons déjà indiqué comment nous nous protégeons de ces incertitudes : contrôles systématiques des données et des hypothèses pour les premières, utilisation de l'information "historique" pour les secondes.

Les incertitudes "calculables" sont :

- les incertitudes d'échantillonnage, lors de l'estimation des paramètres de la loi retenue,
- les écarts aléatoires sur les données dus à la dispersion des jaugeages,
- l'imprécision entourant les crues "historiques",
- les défauts de stationnarité de courtes périodes,
- le choix du seuil dans la méthode type Renouvellement.

Seules les incertitudes d'échantillonnage sont à proprement parler calculables : c'est le classique intervalle de confiance, fourni également en annexe.

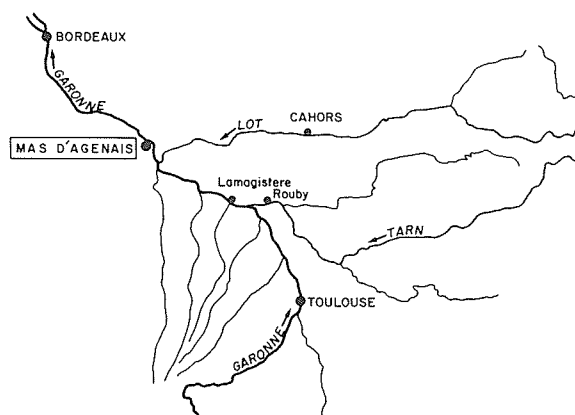


Figure 2 – Bassin de la Garonne à Mas d'Agenais.

Les autres incertitudes peuvent être cependant assez bien appréciées par des analyses de sensibilité :

Les écarts dus aux jaugeages, après critique des courbes hauteur-débit, sont en général faibles dans la zone des débits moyens, mais leur impact peut devenir important pour les débits les plus forts, qui ont dû être extrapolés : on rejoint le problème de l'imprécision sur les crues historiques et il est rare qu'on ne puisse donner des limites probables et exploitables aux débits litigieux, et tester leur impact sur l'estimation.

Le choix du seuil, lui, pose en général peu de problème : en le faisant varier, on trouve une plage assez large pour laquelle les variations des résultats sont faibles. Lorsque ce n'est pas le cas, c'est un signal d'alarme qui peut signifier que le comportement de la loi choisie n'est pas conforme aux données, ou bien que parmi celles-ci, un problème d'homogénéité se pose.

Enfin, lorsque le contrôle des hypothèses a mis en évidence de légers défauts de stationnarité, des découpages de l'échantillon peuvent permettre de situer leur importance au niveau des estimations finales.

Du point de vue pratique, le programme CRUE effectue le calcul des intervalles de confiance, ainsi que la plupart des contrôles d'hypothèse. Il offre enfin à l'utilisateur la possibilité de faire toutes les analyses de sensibilité qu'il juge utile.

EXEMPLE : LA GARONNE. A MAS D'AGENAIS

La station de Mas d'Agenais est située sur la Garonne, juste à l'aval du confluent du Lot. Le bassin versant contrôlé est d'environ 53 000 km² (Fig. 2).

L'information est abondante : de nombreuses études, dont principalement celles de Parde, permettent de décrire les conditions météorologiques, le déroulement des fortes crues, et de chiffrer approximativement les débits des plus forts événements depuis 1770 environ. L'information régulière en débit journalier débute en 1913. Il existe en outre, d'autres séries de débits journaliers à l'amont : Lamagistère sur la Garonne (1915), Cahors sur le Lot (1937), Rouby sur le Tarn (1923).

La description qualitative des phénomènes peut être schématisée ainsi :

– il existe trois grands types de crues : pyrénéen (arro-

sage des Pyrénées, vents de Nord Ouest), océanique (arrosage de la Garonne aval, vents d'Ouest-Sud Ouest), méditerranéen (arrosage du Tarn et Lot amont, vents de Sud – Sud Est). Certaines crues sont de type mixte, surtout océanique – pyrénéen.

– des crues en série, surviennent assez fréquemment et sont dangereuses.

– les concomitances temporelles aux confluits sont assez fréquentes. Par contre la simultanéité de débits très élevés sur la Garonne amont, le Tarn et le Lot, ne s'est pas produite. Ce dernier point, ainsi que l'alternance des types de crue, explique, qu'à l'aval des confluits, il n'y ait pas d'hétérogénéité statistique de la série des débits, comme on verra plus loin.

La critique des données a conduit à effectuer quelques corrections au-delà du débordement, où les extrapolations, basées sur les observations d'écoulement en lit mineur, ne coïncidaient pas avec les évaluations des débits de débordement trouvés dans la littérature. En dessous du débordement ($Q < 3\,800\text{ m}^3/\text{s}$) la courbe de tarage est d'assez bonne qualité. Quant à l'influence des barrages amont (2 990 Hm³) elle est inconnue, mais estimée presque négligeable au niveau du Mas d'Agenais, situé très à l'aval.

Le contrôle des hypothèses a donné les résultats suivants :

– homogénéité : l'effet des concomitances Lot-Garonne (Fig. 3) ainsi que l'effet saisonnier (Fig. 4) modifie peu les estimations.

– Stationnarité : le découpage en deux sous échantillons et le test des moyennes est favorable (3 550 m³/s contre 3 640 m³/s pour un écart type d'environ 1 000 m³/s), alors que la distribution des occurrences de crues n'est pas tout à fait uniforme (Fig. 5), en partie à cause des années 1946 à 1949, particulièrement "sèches".

L'estimation par une méthode de Maxima Annual est bonne, (Fig. 6), grâce au nombre important d'années d'observations (65 ans). On retiendra les valeurs suivantes :

$$Q_{10\text{ ans}} = 5\,650\text{ m}^3/\text{s}, \quad Q_{100\text{ ans}} = 8\,100\text{ m}^3/\text{s},$$

$$Q_{1\,000\text{ ans}} = 10\,500\text{ m}^3/\text{s}.$$

L'estimation par une méthode type Renouvellement a été effectuée en retenant toutes les crues d'un débit supérieur à 2 500 m³/s. L'indépendance entre les crues séparées de plus de 10 jours a été contrôlée ($\rho = -0,14$).

L'ajustement du nombre de crues supérieures à 2 500 m³/s de chaque année à une loi de Poisson est satisfaisante (Fig. 7) et l'ajustement des débits supérieurs à 2 500 m³/s à une loi de Weibull est également bon (Fig. 8). On obtient ainsi une première estimation :

$$Q_{10\text{ ans}} = 5\,670\text{ m}^3/\text{s}, \quad Q_{100\text{ ans}} = 7\,670\text{ m}^3/\text{s},$$

$$Q_{1\,000\text{ ans}} = 9\,590\text{ m}^3/\text{s}$$

L'introduction de l'information "historique", c'est à dire 12 crues, supérieures à 6 200 m³/s en 143 ans, a peu modifié les résultats précédents (Fig. 9). L'estimation finale sera :

$$Q_{10\text{ ans}} = 5\,770\text{ m}^3/\text{s}, \quad Q_{100\text{ ans}} = 7\,730\text{ m}^3/\text{s},$$

$$Q_{1\,000\text{ ans}} = 9\,570\text{ m}^3/\text{s}.$$

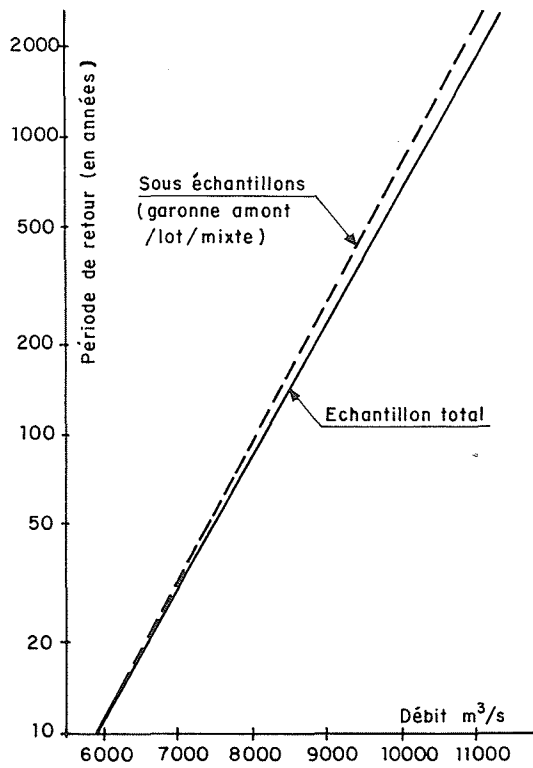


Figure 3 – Mas d'Agenais – Etude de l'effet de la confluence Garonne amont-Lot.

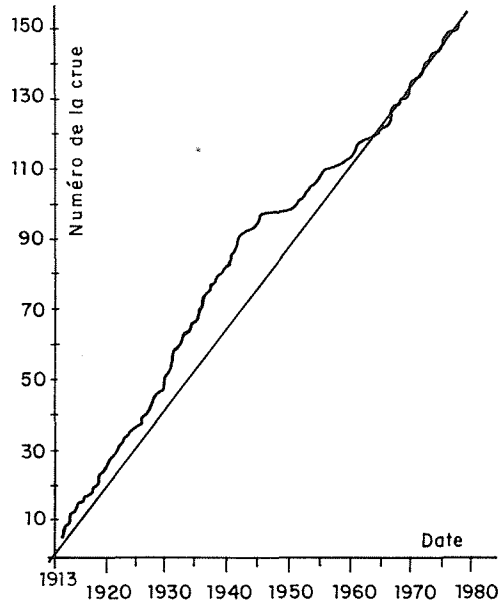


Figure 5 – Mas d'Agenais – Test de répartition uniforme des dates d'occurrence des crues dépassant 2 500 m³/s.

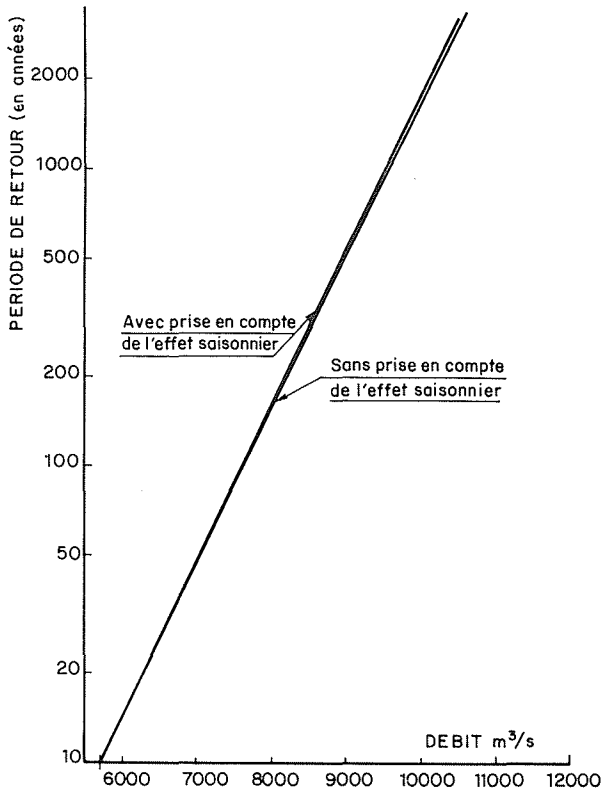


Figure 4 – Mas d'Agenais – Etude de l'effet saisonnier.

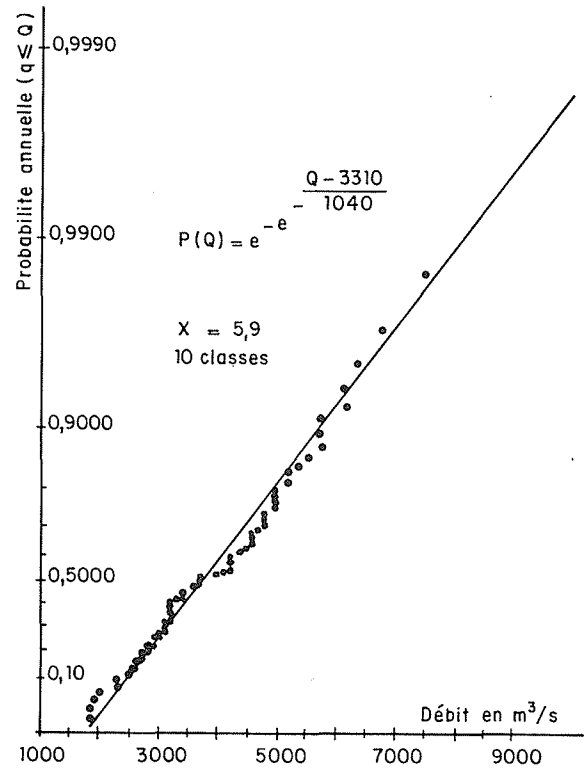


Figure 6 – Mas d'Agenais – Méthode des maxima annuels – Lois des valeurs extrêmes.

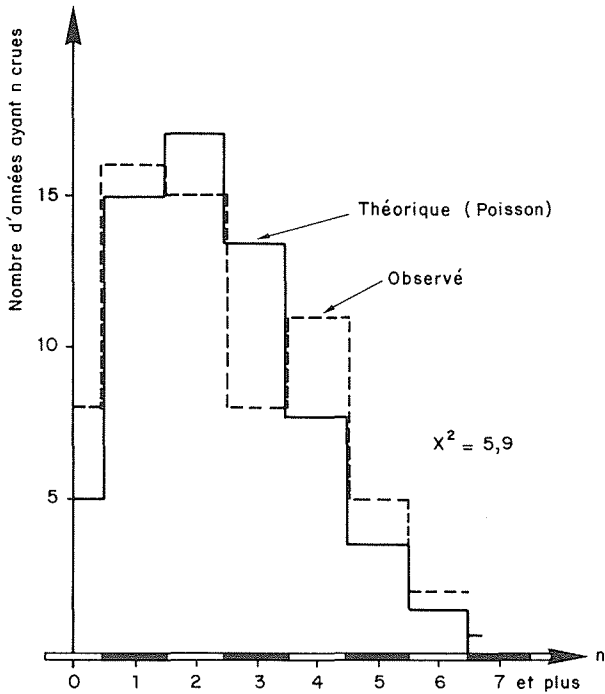


Figure 7 – Mas d’Agenais – Ajustement à une loi de Poisson du nombre de crues dont le débit dépasse 2 500 m³/s.

la probabilité de la plus forte crue annuelle s’écrit :

$$\text{Prob}(Q^* < Q) = 1 - 2,36 e^{-0,00025(Q-2500)^{1,17}}$$

Le contrôle des incertitudes “non calculables”, effectué par les divers tests mentionnés précédemment (stationnarité, homogénéité saisonnière et de concom-

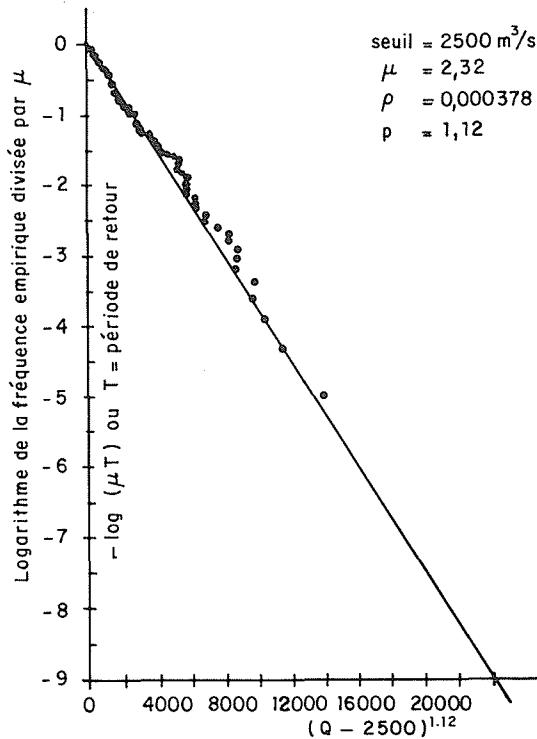


Figure 8 – Mas d’Agenais – Ajustement à une loi de Weibull des débits supérieurs à 2 500 m³/s.

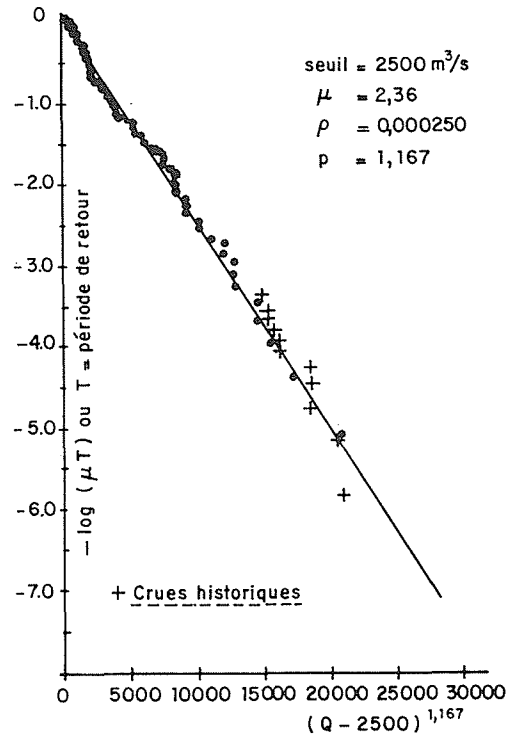


Figure 9 – Mas d’Agenais – Ajustement à une loi de Weibull des débits supérieurs à 2 500 m³/s avec prise en compte de l’information historique.

mitante, répartition des dates d’occurrence, indépendance des crues successives), peut être considéré comme satisfaisant. On a cependant vérifié que l’impact des années 1946-1950 n’est pas trop important sur les estimations (quelques pour cents).

En ce qui concerne les incertitudes “calculables” les intervalles de confiance ont été calculés conformé-

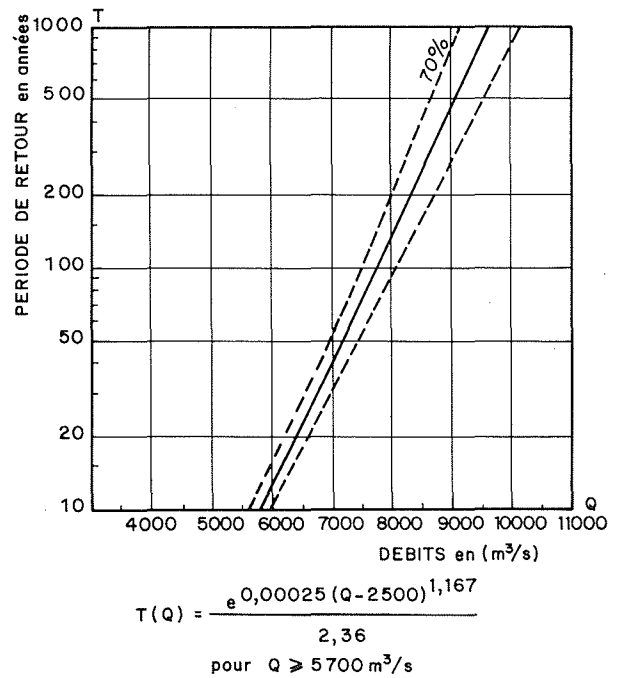


Figure 10 – Périodes de retour des débits de crue de la Garonne à Mas d’Agenais.

ment à l'annexe : pour la crue millénaire, ils passent de $\pm 900 \text{ m}^3/\text{s}$ à $\pm 500 \text{ m}^3/\text{s}$ lorsqu'on utilise l'information historique.

Enfin diverses analyses de sensibilité ont été effectuées, parmi lesquelles :

– effet du découpage en sous échantillon :

	1913-1977	1913-1945	1946-1977
Crue décennale	5 630	5 900	5 210
Crue centennale	7 570	7 770	7 210
Crue millénaire	9 410	9 530	9 150

– effet du seuil :

Seuil	2 000	2 250	2 500	2 750	3 000	3 250	3 500
Crue millénaire	9 430	9 670	9 540	9 760	9 690	9 120	8 890

– effet de la correction des débits journaliers :

	Correction minorée de 10 %	Correction standard	Correction minorée de 10 %
Crue millénaire	9 434	9 544	9 655

Le résultat final prend la forme de la figure 10 :

Il est intéressant de comparer les principales incertitudes qui entourent la crue millénaire ($9 570 \text{ m}^3/\text{s}$) :

Echantillonnage sans information historique	1 800 m^3/s
Echantillonnage avec information historique	1 000 m^3/s
Non stationarité	380 m^3/s
Correction des données	220 m^3/s
Effet de confluence Lot-Garonne	250 m^3/s
Choix du seuil entre 2 000 et 3 000 m^3/s	330 m^3/s

ASPECTS PRATIQUES

Le programme CRUE effectue tous les calculs ci-dessus en quelques secondes d'ordinateur IBM 3081, à partir du fichier de la série des débits journaliers et de la liste des débits des crues historiques. Une version conversationnelle facilite les analyses de sensibilité. Des moyens informatiques légers (mini ou micro) sont suffisants pour recevoir un programme analogue.

*

Nous tenons à souligner l'aide précieuse que nous ont apporté M. Bernier, ainsi que Mme Roux-Guiheneuf, et à remercier, à travers eux, tous ceux qui ont participé à la réalisation de cet outil.

BIBLIOGRAPHIE

- [1] BERNIER. – Sur la théorie du renouvellement et son application en hydrologie, *Rapport E.D.F. HYD 67 n° 10*, 1967.
- [2] BERNIER et MIQUEL. – Exemple d'application de la théorie de la décision statistique au dimensionnement d'ouvrages hydrauliques : prise en compte de l'information hétérogène. *XVII^e Congrès de l'A.I.R.H.*
- [3] BERNIER. – Eléments de Statistique Bayésienne. *Rapport E.D.F. HE 40/81.09*.
- [4] BOBEE. – Contribution à l'étude statistique des débits maximaux annuels de crue par les distributions de Pearson III et Log. Pearson III. Thèse de Docteur Ingénieur, Toulouse, 1976.
- [5] COX. – *Renewall Theory*. Wiley, New-York, 1962.
- [6] GUMBEL. – *Statistics of extremes*. Columbia University Press, New-York, 1958.
- [7] MIQUEL. – *Guide d'estimation des probabilités des de crue*. A paraître aux Editions Eyrolles en 1983.
- [8] NATURAL ENVIRONMENT RESEARCH COUNCIL. – *Flood studies report*. Londres, 1975.
- [9] MORLAT. – *Introduction à l'emploi des méthodes bayésiennes pour l'étude des événements rares*. Rapport du comité sur la sûreté des installations nucléaires sur le thème : Task force on problems of rare events in the reality analysis on nuclear power plants. Paris, 1978.
- [10] US WATER RESOURCES COUNCIL. – Guidelines for determining flood flow frequency. *Bulletin n° 17A of the Hydrology Committee*, 1977.
- [11] YEVJEVICH. – *Probability and statistics in hydrology*. Water Ressources Publications. Fort Collins, 1972.

ANNEXE

ESTIMATION DES PARAMETRES DE LA METHODE DU RENOUVELLEMENT

1) Rappel concernant la formule de Bayes

Elle permet de réestimer la densité d'une variable θ , compte tenu de l'arrivée d'une nouvelle information x :

si $g(\theta)$ est la densité de θ , estimée avant l'information x (a priori)

$g(\theta/x)$ est la densité de θ , estimée après l'information x (a posteriori)

$f(x/\theta)$ est la densité de x , pour θ fixé,

alors :

$$g(\theta/x) = \frac{g(\theta) \cdot f(x/\theta)}{\int g(t) \cdot f(x/t) dt}$$

Cette formule est la généralisation du cas discret bien connu :

$$\text{Prob (A/B)} = \frac{\text{Prob (A)} \cdot \text{Prob (B/A)}}{\text{Prob (B)}}$$

2) Rappel concernant la méthode du Renouvellement

A partir de la suite des débits observés sur NA années, on construit deux sous échantillons :

$$\{n_j\} \quad j = 1, \dots, NC$$

L'échantillon du nombre de crues supérieures à S_0 chaque année.

$$\{Q_j\} \quad j = 1, \dots, NC$$

l'échantillon des débits maxima de crues supérieures à S_0 (il y en a NC au total), au cours des NA années.

La méthode du Renouvellement permet de calculer la probabilité de la crue maximale annuelle Q^* , de la façon suivante :

$$F(Q) = \text{Prob}(Q^* < Q) = \sum_{k=0}^{\infty} P(k) \cdot G(Q)^k$$

avec :

$$\begin{cases} P(k) = e^{-\mu} \frac{\mu^k}{k!} \\ G(Q) = 1 - e^{-\rho(Q-S_0)^p} \end{cases}$$

Le problème est d'estimer μ , ρ et p à partir de l'échantillon ci-dessus.

3) Vraisemblance d'un échantillon de crues observées

La vraisemblance d'un échantillon n'est rien d'autre que la probabilité d'observer cet échantillon de NC crues avec exactement la répartition temporelle qui a été la sienne.

Autrement dit :

$$\begin{aligned} V = & \text{Prob}(\text{d'avoir observé } n_1 \text{ crues la 1}^{\text{ère}} \text{ année}) && P(n_1) \\ & \times \text{Prob}(\text{d'avoir observé } n_2 \text{ crues la 2}^{\text{ème}} \text{ année}) && P(n_2) \\ & \cdot && \cdot \\ & \cdot && \cdot \\ & \times \text{Prob}(\text{d'avoir observé } n_{NA} \text{ crues la NA}^{\text{ième}} \text{ année}) && P(n_{NA}) \\ & \times \text{Prob}(\text{la 1}^{\text{ère}} \text{ crue} = Q_1) && \frac{\partial G}{\partial Q}(Q_1) \\ & \cdot && \cdot \\ & \cdot && \cdot \\ & \times \text{Prob}(\text{la NC}^{\text{ième}} \text{ crue} = Q_{NC}) && \frac{\partial G}{\partial Q}(Q_{NC}) \end{aligned}$$

ceci suppose évidemment l'indépendance des événements.

On peut encore écrire, en utilisant les expressions de $P(k)$ et $G(Q)$:

$$V = \lambda e^{-\mu NA} \cdot \mu^{NC} \cdot p^{NC} \cdot \rho^{NC} \cdot \left[\prod_{i=1}^{NC} (Q_i - S_0) \right]^{p-1} e^{-\rho \sum_{i=1}^{NC} (Q_i - S_0)^p}$$

avec $\lambda = \frac{1}{n_1! n_2! \dots n_{NA}!}$ indépendant de ρ, μ et p .

4) Vraisemblance de ρ, μ et p par l'analyse bayésienne

Nous considérons maintenant ρ, μ et p comme des variables aléatoires dont on désire connaître les lois de probabilité.

Pour cela nous allons commencer par faire une hypothèse sur la forme de leur loi : des lois Gamma. Puis nous allons calculer la vraisemblance des paramètres ρ, μ et p , en tenant compte de ces lois a priori et en utilisant la formule de Bayes.

$$f(\mu) = a e^{-\mu a} \frac{(\mu a)^{a'-1}}{(a'-1)!}$$

$$g(\rho) = b e^{-\rho b} \frac{(\rho b)^{b'-1}}{(b'-1)!}$$

$$h(p) = c e^{-pc} \frac{(pc)^{c'-1}}{(c'-1)!}$$

Avant d'avoir observé l'échantillon, les probabilités a priori des paramètres ρ, μ et p sont $f(\mu), g(\rho)$ et $h(p)$. Lorsqu'arrive l'échantillon, dont la vraisemblance V a été calculée ci-dessus, on peut recalculer les probabilités de ces paramètres : on obtient leurs probabilités a posteriori, c'est-à-dire la densité de ces paramètres compte tenu de l'échantillon. Il importe donc de bien faire la différence entre :

- la densité V d'un échantillon, pour un jeu de paramètres μ, ρ et p fixés (c'est le V calculé précédemment),
- la vraisemblance V' des paramètres compte tenu de l'échantillon qui combine à la fois la vraisemblance de l'échantillon et les probabilités a priori des paramètres.

Ce calcul s'effectue au moyen de la formule de Bayes :

$$V' = \frac{V \cdot f(\mu) \cdot g(\rho) \cdot h(p)}{\int \int \int V \cdot f(\mu) \cdot g(\rho) \cdot h(p) d\mu d\rho dp}$$

$$V' = \lambda' e^{-\mu(NA+a)} \cdot \mu^{NC+a'-1} \cdot \rho^{NC+b'-1} \cdot p^{NC+c'-1} \cdot \left[\prod_{i=1}^{NC} (Q_i - S_0) \right]^{p-1} \cdot e^{-pc - \rho b - \rho \sum_{i=1}^{NC} (Q_i - S_0)^p}$$

Si on ne dispose d'aucune information a priori sur ρ, μ et p , on prend $a = b = c = 0$ et $a' = b' = c' = 1$.

5) Estimation des paramètres ρ, μ, p par le Maximum de Vraisemblance

On calcule les estimations modales par dérivation de V' :

$$\frac{\partial V'}{\partial \mu} = \frac{\partial V'}{\partial \rho} = \frac{\partial V'}{\partial p} = 0 \quad (\text{Maximum de Vraisemblance})$$

on obtient une estimation $\hat{\mu}$, $\hat{\rho}$ et \hat{p} de μ , ρ et p en résolvant ce système par approximations successives.

6) Incertitudes d'échantillonnage

Nous allons d'abord montrer que les variables aléatoires μ , ρ et p , suivent une loi normale, dont la matrice des variances et covariances contient les dérivées secondes de V' ($\hat{\mu}$, $\hat{\rho}$, \hat{p}) :

$$V' = e^{\log V'}$$

Posons : $W = \log V'$, donc $V' = e^W$

Développons W autour de $\hat{\mu}$, $\hat{\rho}$ et \hat{p} :

$$W = W(\hat{\mu}, \hat{\rho}, \hat{p}) + \frac{\partial W}{\partial \mu}(\hat{\mu}, \hat{\rho}, \hat{p})(\mu - \hat{\mu}) + \text{idem pour } \rho \text{ et } p + \frac{\partial^2 W}{\partial \mu^2}(\mu, \rho, p)(\mu - \hat{\mu})^2 + \text{idem pour } \rho \text{ et } p + \frac{\partial^2 W}{\partial \mu \partial \rho}(\hat{\mu}, \hat{\rho}, \hat{p})(\mu - \hat{\mu})(\rho - \hat{\rho}) + \text{autres termes croisés}$$

mais $\frac{\partial V'}{\partial \mu}(\hat{\mu}, \hat{\rho}, \hat{p}) = 0 \Rightarrow \frac{\partial W}{\partial \mu}(\mu, \rho, p) = 0$
 ↑
 Maximum de vraisemblance
 idem pour ρ et p

donc :

$$V' = V'(\hat{\mu}, \hat{\rho}, \hat{p}) \cdot e^{\left\{ \frac{\partial^2 W}{\partial \mu^2}(\hat{\mu}, \hat{\rho}, \hat{p}) \cdot (\mu - \hat{\mu})^2 + \text{etc.} \right\}}$$

On reconnaît une loi normale(*), dont les variances, covariances s'expriment en fonction de $\frac{\partial^2 W}{\partial \mu^2}$, etc. qui n'est rien d'autre que $\frac{\partial^2}{\partial \mu^2}(\log V')$, ce qui permet de calculer $\text{Var } \rho$, $\text{Var } \mu$, $\text{Var } p$ et leurs covariances. Le détail des calculs est fourni en référence [7].

7) Estimation finale

Nous connaissons maintenant une estimation de la moyenne et des variances et covariances des paramètres μ , ρ et p .

Nous pouvons donc calculer la crue de période de retour T fixée : sa valeur ponctuelle, mais aussi sa variance.

(*) La loi Normale à 3 dimensions est de la forme :

$$f(x, y, z) = Ke^{-\frac{1}{2}P(x, y, z)}$$

où $P(x, y, z)$ est un polynôme du 2^e degré et $K = \text{constante}$.

Sa valeur ponctuelle sera dans le cas où T est assez grand :

$$Q_T = S_0 + \left[\frac{\log(\hat{\mu}T)}{\hat{\rho}} \right]^{1/\hat{p}}$$

Sa variance se calcule à partir des variances et covariances de μ , ρ et p . Pour cela on commence par faire un développement limité de la vraie valeur inconnue Q_T au voisinage de \hat{Q}_T :

$$Q_T - \hat{Q}_T = \frac{\partial Q}{\partial \rho} d\rho + \text{etc.}$$

$$\text{Var } Q_T = \left[\frac{\partial Q}{\partial \rho}(\hat{\rho}, \hat{\mu}, \hat{p}) \right]^2 \cdot \text{Var } \rho + \text{etc.}$$

8) Prise en compte de l'information historique

Soient Q'_1, \dots, Q'_{NP} les NP crues les plus fortes en NAS années supplémentaires.

La vraisemblance de cet échantillon s'écrit :

$$V'' = \sum_{n=NP}^{\infty} \text{Prob} \{ \text{d'avoir observé } (n - NP) \text{ crues inférieures à } Q'_{NP} \text{ et supérieures à } S_0 \text{ en NAS années} \} \times \frac{\partial G}{\partial Q}(Q'_1) \times \dots \times \frac{\partial G}{\partial Q}(Q'_{NP})$$

Si Q'_{NP} est la plus faible de ces crues historiques :

$$V'' = \lambda' \cdot \hat{\mu}^{NP} \cdot \rho^{NP} \cdot p^{NP} \cdot \left[\prod_{i=1}^{NP} (Q'_i - S_0) \right]^{p-1} \cdot e^{-\rho \sum_{j=1}^{NP} (Q'_j - S_0)^p - \mu_{NAS}} \cdot e^{-\rho(Q'_{NP} - S_0)^p}$$

Il ne reste plus qu'à reprendre les mêmes calculs que précédemment en prenant pour vraisemblance de l'échantillon, la vraisemblance conjointe des échantillons régulier et historique : $V''' \equiv V' \cdot V''$.

Ce qui donne :

$$V''' = \lambda' \mu^{NC+NP+a'-1} \cdot \rho^{NC+NP+b'-1} \cdot p^{NC+NP+c'-1} \cdot e^{-\rho \left[b + \sum_{i=1}^{NC} (Q_i - S_0)^p + \sum_{j=1}^{NP} (Q'_j - S_0)^p \right]} \cdot \prod_{i=1}^{NC} (Q_i - S_0)^{p-1} \cdot \prod_{j=1}^{NP} (Q'_j - S_0)^{p-1} \cdot e^{-\mu_{NAS}} \cdot e^{-\rho(Q'_{NP} - S_0)^p} \cdot e^{-\mu(NA+a)-pc}$$

En maximisant ces vraisemblances comme précédemment on obtient de nouvelles estimations des paramètres ρ , μ , p et de leurs variances.