ANALYSE STATISTIQUE DES VALEURS EXTRÊMES DE PLUIE

T.P.T. NGUYEN[1] ET PH. BOIS[1]

RÉSUMÉ

Les méthodes paramétriques d'estimation des pluies maximales de fréquences rares nécessitent de connaître la forme de leurs distributions. Cette contrainte peut s'avérer être un obstacle majeur dans la mesure où les fortes pluies peuvent être issues de populations différentes de celle des pluies de moindre importance.

Dans cette communication, l'auteur, à partir de simulations, montre que la méthode non paramétrique de Villasenor est une des plus performante pour l'estimation des périodes de retour des pluies extrêmes.

[1] IMG - LTHE. INPG de Grenoble

INTRODUCTION

L'estimation des périodes de retour des valeurs extrêmes de pluie est un des problèmes majeurs posés aux concepteurs d'aménagements. Pour cela, on procède à l'analyse statistique des chroniques d'observations.

L'analyse peut porter sur la totalité des données disponibles. Mais les mécanismes météorologiques responsables des averses extrêmes sont très différents de ceux à l'origine des pluies de moindre importance. De même, les erreurs commises sur leurs mesures ne sont pas de même nature (elles sont beaucoup plus importantes pour les faibles averses que pour les fortes). C'est pourquoi il peut être préférable de ne retenir que les pluies supérieures à un seuil fixé ou les valeurs maximales observées chaque année.

L'estimation des distributions peut relever de deux types de méthodes. On peut se fixer *a priori* une loi mathématique. L'ajustement de cette loi aux observations permet d'en définir les paramètres. Il s'agit des méthodes paramétriques. Ces méthodes supposent donc que l'on connaisse la forme des distributions originelles, ce dont on n'est jamais sûr. On peut dans ces conditions refuser de faire une telle hypothèse et préférer estimer la distribution point par point. Il s'agit des méthodes non paramétriques.

On présente ici les principaux résultats d'une étude visant à évaluer l'efficacité de plusieurs de ces méthodes (paramétriques et non paramétriques). La démarche suivie est expérimentale. Par tirage aléatoire dans diverses lois de probabilité, on a généré un grand nombre d'échantillons de taille variable, simulant les différents types de chroniques citées plus haut. Pour chacune des méthodes, on procède sur chaque échantillon à une estimation de la distribution des pluies maximales. On obtient alors pour chaque quantile un ensemble de valeurs ajustées, ensemble que l'on peut caractériser par sa médiane et un intervalle de confiance. Une méthode sera d'autant plus efficace que les médianes des quantiles ajustés seront proches de ceux de la population originelle, et que les intervalles de confiance seront resserrés.

GÉNÉRATION DES ÉCHANTILLONS

Pour imiter la diversité des données pluviométriques, nous avons généré des échantillons à partir de plusieurs lois de probabilité, avec des paramètres différents et correspondant à des pas de temps différents.

Les lois de probabilité retenues sont les suivantes :

— pour les distributions complètes, la loi SEXP, somme de deux exponentielles ;

— pour les distributions des valeurs extrêmes ou seuillées, les lois de Gumbel et de Weibull.

Pour chaque loi, on a généré des séries de 100 échantillons de 15, 30 et 120 individus.

La loi Sexp

Cette loi peut s'écrire :

$$F(x) = 1 - b_1 \exp(-x/a) - b_2 \exp(-x/c) \qquad (1)$$

où :

— $F(x)$ est la fréquence au non-dépassement de la valeur x ;

— a et c sont les paramètres d'échelle des deux exponentielles ;

— b_1 et b_2 sont les fréquences des tirages dans l'une et l'autre des deux exponentielles.

On s'est imposé en outre les contraintes suivantes (1) :

— a voisin de $10.c$;

— b_2 voisin de $3.b_1$.

Dans ces conditions pour les fortes valeurs de x, la deuxième exponentielle devient négligeable devant la première et on peut admettre l'égalité suivante :

$$G(k,x) = (1 - b_1 \exp(-x/a))^k \qquad (2)$$

où, $G(k,x)$ est la probabilité que la valeur maximale d'un échantillon de k individus soit égale ou supérieure à x. Dans ce qui suit nous appellerons, MAXSEXP, la loi des valeurs maximales définie par (2).

La loi de Weibull

Cette loi peut s'écrire

$$F(x) = 1 - \exp(-r.(x-So)^p) \qquad (3)$$

Où, So est la valeur du seuil (paramètre de position) et r et p sont les paramètres d'échelle et de forme de la loi.

La loi de Gumbel

Cette loi peut s'écrire :

$$F(x) = \exp(-\exp(-(x-xo)/a)) \qquad (4)$$

Où, xo et a sont les paramètres de position et d'échelle de la loi.

ESTIMATION DES DISTRIBUTIONS DES VALEURS EXTRÊMES

ÉCHANTILLONS GÉNÉRÉS À PARTIR DE LA LOI SEXP

Chaque échantillon simule l'ensemble des observations d'une année. On peut fusionner plusieurs de ces échantillons pour simuler des chroniques entières (échantillons des chroniques). On peut isoler la valeur maximale de chaque échantillon et constituer ainsi des échantillons des maxima annuels (échantillons des maxima). On peut également ne retenir pour chacun d'entre eux que les valeurs supérieures à un seuil (échantillons seuillés).

Pour l'estimation des distributions des valeurs extrêmes, nous avons testé les méthodes suivantes :

— paramétriques :
  . ajustement d'une loi SEXP aux échantillons des chroniques,
  . ajustement d'une loi de Gumbel aux échantillons des maxima,
  . ajustement d'un modèle basé sur la théorie du renouvellement aux échantillons seuillés ;

— non paramétriques :
  . méthodes d'Adamowski appliquées aux échantillons seuillés,
  . méthode de Villasenor appliquée aux échantillons seuillés.

AJUSTEMENT D'UNE LOI SEXP (MÉTHODE A)

À chaque échantillon de chroniques, on ajuste par la méthode des moments (1) une loi SEXP et on estime les quantiles de la distribution des maxima par inversion de la formule (2).

AJUSTEMENT D'UNE LOI DE GUMBEL (MÉTHODE B)

À chaque échantillon des maxima, on ajuste par la méthode des moments une loi de Gumbel. Le calcul des quantiles se fait par inversion de la formule (4).

AJUSTEMENT D'UN MODÈLE BASÉ SUR LA THÉORIE DU RENOUVELLEMENT (MÉTHODE C)

Si k est le nombre de valeurs supérieures à un seuil *So* fixé, observées au cours d'une année, si $P(k)$ est sa loi de probabilité, la fréquence au non-dépassement, $F(x)$, des valeurs maximales annuelles peut s'écrire :

$$F(x) = \sum_{k=0}^{\inf} P(k).G^k(x)$$

où,

— $x > So$;

— $G(x)$ est la fréquence au non-dépassement des valeurs x, sachant qu'elles sont supérieures à So.

Si on admet que k suit une loi de Poisson et que x suit une loi de Weibull, on peut admettre que pour les fréquences rares (>.9) :

$$F(x) = 1 - E(k) \,.\, exp(- r.(x-So)p) \qquad (5)$$

où, $E(k)$ est l'espérance de k.

Sur un ensemble de n échantillons seuillés, représentant n années d'observation, on calcule $E(k)$. Sur la totalité des n échantillons, on ajuste par la méthode du maximum de vraisemblance la loi de Weibull (4). On peut alors par inversion de la formule (5) estimer les quantiles de la distribution des maxima annuels.

Méthodes d'Adamowski

La densité de probabilité d'une valeur x est estimée à partir des réalisations, x_i, obtenues en son voisinage, par l'intermédiaire d'une fonction, le noyau $K(x-x_i)$, décroissante de part et d'autre de x.

Adamowski propose pour le noyau, l'expression suivante :

— si $-1 < u_i < 1$

$$K(u_i) = \frac{3}{4}.(1-u_i^2) ;$$

— si $|u_i| > 1$

$$K(u_i) = 0$$

où, $u_i = (x-x_i)/h$, h étant un paramètre caractérisant l'étendue du voisinage.

L'estimation, $f(x)$, de la fonction de densité en x, s'écrit :

$$f(x) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{h}k(u)$$

où, N est la taille de l'échantillon.

On peut admettre que h est une constante (estimation FKE) ou une variable (estimation VKE)

Estimation FKE (Fixe Kernel Estimator) (méthode D)

Adamowski propose pour le calcul de h, si N est la taille de l'échantillon des x_i rangés :

$$h = \frac{\sum_{i=2}^{N}\sum_{j=1}^{i-1}(x_i - x_j)}{\sqrt{5}N\left(N-\frac{10}{3}\right)}$$

L'estimation de la fréquence au non-dépassement $F(x)$ s'exprime par :

$$F(x)=\frac{j-1}{N}+\frac{1}{hN}\sum_{k=j}^{j+m}\int_{x\min}^{x}K\left(\frac{x-x_k}{h}\right)dx$$

où,

— j est la première valeur satisfaisant la condition $((x_j - x) < h)$;

— $xmin = (x_k - h)$;

— m = le nombre de noyaux non entiers de part et d'autre de x.

Soit

$$F(x)=\frac{j-1}{N}+\frac{3}{4N}\sum_{k=j}^{j+m}\left(\frac{x-x_k}{h}-\frac{1}{3}\left(\frac{x-x_k}{h}\right)^3+\frac{2}{3}\right) \qquad (6)$$

par inversion de la formule (6), on calcule l'estimation des quantiles des pluies maximales annuelles.

Estimation VKE (Variable Kernel Estimator) (méthodes E)

Prendre une valeur de h constante, présente l'inconvénient de « sur-lisser » la distribution expérimentale dans les parties où la densité est grande et au contraire de la « sous-lisser » dans celle où la densité est faible. C'est pourquoi il est préférable de faire varier h en fonction de la densité. On peut, pour cela, poser :

$$h=\frac{1}{c_k d_{jk}} \qquad (7)$$

où, c_k est une constante et $d_{jk} = |x_k - x_j|$, x_k étant, dans l'échantillon rangé, le k[ème] point le plus proche de x_j.

Pour déterminer c_k et k, Breiman *et al.* (1977) minimisent la variance de $\left(\sum_{i=1}^{k} c_i d_{ji}\right)$ et posent :

$$O=c_k-\frac{1}{N}\sum_{i=1}^{N}\frac{\displaystyle\sum_{j\neq k, j=1}^{N}\frac{(x_i-x_j)K\left(\frac{x_i-x_j}{c_k d_{jk}}\right)}{d_{jk}^2}}{\displaystyle\sum_{j\neq k, j=1}^{N}K\left(\frac{\frac{x_i-x_j}{c_k d_{jk}}}{d_{jk}}\right)}$$

La fréquence au non-dépassement s'exprime par la formule (6) en y remplaçant h par l'expression (7).

MÉTHODE DE VILLASENOR (MÉTHODE F)

Comme la méthode (D), cette méthode se fonde sur la théorie du renouvellement. En effet selon cette théorie, pour les fréquences rares, on peut admettre (3) que la fréquence au non-dépassement, $F(x)$, du maximum annuel peut s'écrire :

$$F(x) = 1 - E(k) \cdot [1 - GSo(x)] \qquad (8)$$

où:

— $E(k)$ est l'espérance du nombre annuel, k, des valeurs supérieures à un seuil So fixé ;

— $GSo(x)$ est la fréquence au non-dépassement des valeurs (x) supérieures à So.

$E(k)$ est estimé par sa valeur expérimentale :

$$\hat{E}(k) = \frac{\sum_{i=1}^{N} k_i}{N}$$

où :

— N est le nombre d'années (= le nombre d'échantillons seuillés) ;

— k_i est le nombre de valeurs $>So$ observées au cours de l'année i (= l'échantillon i).

À l'inverse de la méthode (D), on ne fait pas d'hypothèse quant à la forme de $GSo(x)$, mais on l'estime de façon non paramétrique à partir des réalisations de la distribution à deux dimensions du doublet (x, k). En effet, pour un seuil fixé, si on note $H(x, k)$, la probabilité que la valeur maximale soit inférieure à x et que le nombre de valeurs au-dessus du seuil soit inférieur ou égal à k, on a :

$$H(x,k) = \sum_{j=1}^{k} GSo^{j}(x) P(j) \qquad (9)$$

où $P(j)$ est la probabilité que le nombre annuel d'observations au-dessus du seuil soit égal à j.

BOYLES et SAMANIEGO (1986) montre que l'on peut maximiser la vraisemblance d'un échantillon de N réalisations du doublet (x, k) tirées dans une distribution définie par (9), en prenant pour estimateur de $GSo(x)$:

— *si* $x < x1$

$$GSo(x) = 0 ;$$

— si $x_i < x < x_{i+1}$ (avec i compris entre 1 et $N-1$)

$$GSo(x) \prod_{j=1}^{N-1} \frac{\sum_{j=1}^{i} K_j}{\sum_{j=1}^{i+1} K_j} ;$$

— $x_N < x$

$GSo(x) = 1$;

où (x_i, k_i) représentent les N valeurs du doublet, classées par ordre croissant des x.

L'estimateur présente l'inconvénient d'être discontinu et notamment d'être égal à 1 pour les valeurs supérieures à x_N. Pour y remédier nous avons effectué un lissage de la courbe $(1 - GSo(x))$ sur sa partie inférieure. Nous avons testé les quatre modèles donnés ci-après et retenu celui qui minimisait l'erreur quadratique moyenne.

Les quatre modèles testés sont :

a) $1 - GSo(x) = a / [1 + exp\ (c.x + b)]$;

b) $1 - GSo(x) = a / [1 + exp\ (b.x^c)]$;

c) $1 - GSo(x) = a / [1 + exp(exp\ (b.x^c))]$;

d) $1 - GSo(x) = a\ .\ [1 - exp\ (-exp(-(c.x + b)))]$;

à partir des estimateurs de $E(k)$ et de $(1- GSo(x))$, on peut alors calculer par la formule (8) les estimations des différents quantiles.

Échantillons générés à partir des lois de Weibull et de Gumbel

Les échantillons générés ici simulent soit des chroniques annuelles des pluies supérieures à un seuil, soit des chroniques des maxima annuels.

Dans le premier cas, toutes les méthodes vues plus haut ont été appliquées, sauf la méthode (A).

Dans le second cas, seule la méthode (B) a pu être appliquée.

Résultats et conclusion

Les figures 1.1 à 1.5 résument les résultats obtenus par ces différentes méthodes sur des séries d'échantillons générés par la loi SEXP. Pour chacune d'entre elles on a représenté la queue de la distribution des maxima annuels, réelle et estimée. On y a fait figurer également les bornes des intervalles de confiance à 80%, des estimations.

On notera :

— que les meilleurs résultats sont obtenus par ajustement de la loi SEXP (figure 1.1), ce qui n'est pas étonnant ;

— que la méthode des noyaux (figure 1.2) s'avère peu précise pour les fréquences rares ;

— que, des trois autres méthodes, c'est celle de Villasenor (figure 1.3) qui donne les meilleures estimations qui sont par ailleurs très proches de celles obtenues par la première méthode.

D'une façon plus générale, quel que soit le type de données générées, si on excepte le cas trivial de l'ajustement aux échantillons de la loi dont ils sont issus, les meilleurs résultats sont ceux de la méthode de Villasenor.

Cette méthode paraît donc pouvoir être recommandée pour l'analyse statistique des pluies maximales, puisqu'il n'est pas toujours possible dans la réalité de connaître la forme de leurs distributions.

BIBLIOGRAPHIE

ADAMOWSKI K., 1989. A Monte-Carlo comparison of parametric and non-parametric estimation of flood frequencies. *J. Hydrology*, 108, 295 - 308.

BOIRET P., 1987. Analyse des précipitations de 6 minutes à 24 heures par une méthode du type renouvellement.

DUANE C. BOES., 1989. Regional flood quantile estimation for a Weibull model. *Water resources research*, vol. 25, N°5, 979 - 990.

MIQUEL J., 1984. Guide pratique d'estimation des probabilités de crues, Eyrolle.

SLIMANI M., 1984. Étude des pluies de fréquence rare à faibles pas de temps sur la région Cévennes-Vivarais : estimation, relation avec le relief et cartographie synthétique, Thèse Docteur-Ingénieur, IMG - INPG, Grenoble.

SLIMANI M. et LEBEL Th., 1983. Cartographie du gradex de la pluie décennale et de la pluie centennale sur les Cévennes, Conférences « Climat méditerranéen et ressources en eau » - Marseille.

VILLASENOR. J.A et Bois.Ph., 1990. Predicciôn de avenidas con periodo de retorne conocido. Memoires du IV Foro de Estadistica. Monterrey, México, 259 - 268, Edité par Universidad Na1. Autonoma de México.

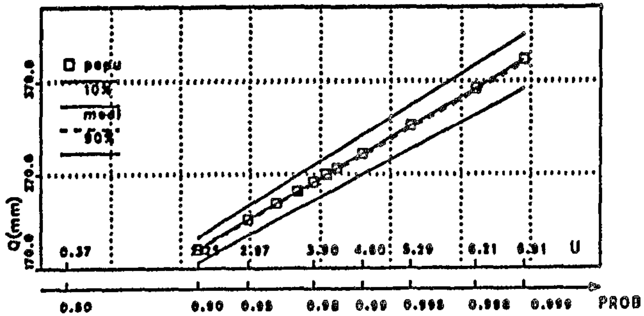

fig. 1.1 AJUSTEMENT DE MAX SEXP

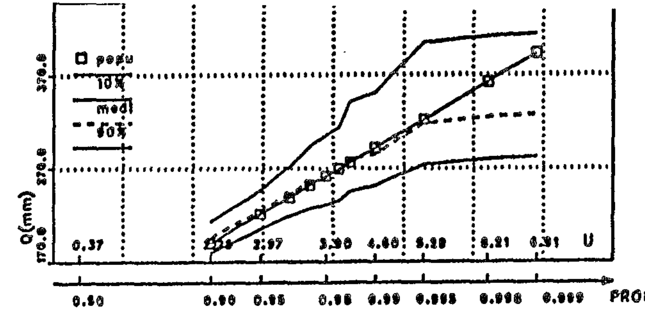

fig. 1.2 AJUSTEMENT EN FONC. NOYAUX

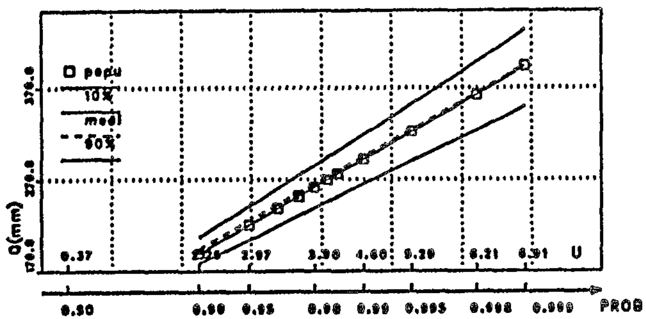

fig. 1.3 AJUSTEMENT VILLASENOR

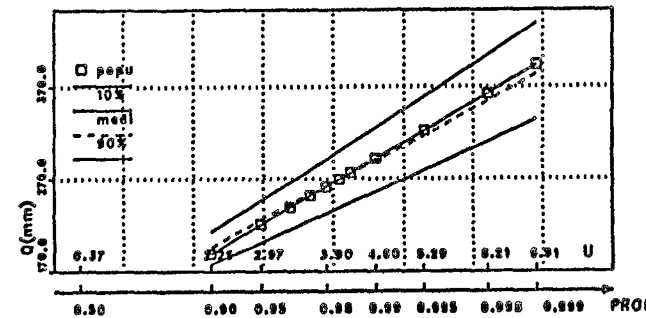

fig. 1.4 AJUSTEMENT GUMBEL

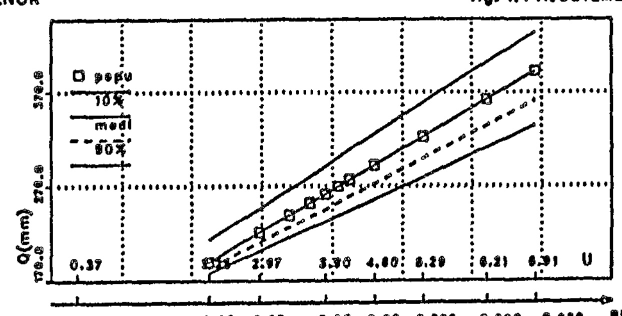

fig. 1.5 METHODE RENOUVELLEMENT

AJUSTEMENT DES ECHANTIL. GENERES PAR LA LOI SEXP

t [illegible] :12 [illegible] =45