

Méthode des moments de probabilité pondérés

Application à la loi de Jenkinson

Hélène LUBES (1), Jean-Marie MASSON (2)

RÉSUMÉ

Cet article a pour objectif de présenter une méthode, relativement nouvelle, d'estimation des paramètres d'une distribution de probabilité : la méthode des moments de probabilité pondérés, introduite par GREENWOOD et al. (1979).

La facilité de mise en œuvre, la robustesse et l'efficacité de cette méthode ont retenu notre attention.

Nous faisons une synthèse des études menées sur l'application de cette méthode à l'estimation des paramètres de la loi de JENKINSON (Generalized Extreme Value distribution - GEV).

MOTS-CLÉS : Loi de probabilité — Estimation de paramètres — Méthode des moments de probabilité pondérés — Loi de JENKINSON — Loi généralisée des valeurs extrêmes.

ABSTRACT

PROBABILITY WEIGHTED MOMENTS METHOD. APPLICATION TO THE JENKINSON DISTRIBUTION

This paper presents a relatively new method of estimating parameters of some probability distributions. It is the probability weighted moments (PWM) method, introduced by GREENWOOD and al. (1979).

This method is interesting because of its robustness, its efficiency, and its easy application.

We give a summary of studies on parameter estimation of the JENKINSON distribution (Generalized Extreme Value distribution - GEV) by the PWM method.

KEY-WORDS : Probability distribution — Parameter estimation — Probability weighted moments — Jenkinson distribution — Generalized Extreme Value distribution.

INTRODUCTION

Les lois de valeurs extrêmes sont régulièrement utilisées dans toutes les études hydroclimatiques où il s'agit d'évaluer un risque. La loi de JENKINSON ou loi généralisée des valeurs extrêmes en fait partie.

Les quantiles de la loi de JENKINSON ont la particularité de s'exprimer de manière analytique en fonction des probabilités correspondantes. Cette propriété permet l'utilisation de la méthode des moments de probabilité pondérés pour estimer les paramètres de la loi à partir d'un échantillon de données.

Les documents consultés à la fois sur les lois de valeurs extrêmes et sur la méthode des moments de probabilité pondérés, nous ont semblé particulièrement intéressants. Nous avons donc jugé utile de faire le point sur cette recherche bibliographique dans un article de synthèse.

(1) Laboratoire d'hydrologie, centre Orstom, BP 5045, 34032 Montpellier cedex.

(2) Laboratoire d'hydrologie et modélisation, université Montpellier II, place Eugène-Bataillon, 34095 Montpellier cedex

DÉFINITION DES MOMENTS DE PROBABILITÉ PONDÉRÉS

GREENWOOD *et al.* (1979) définissent les moments de probabilité pondérés, notés PWM en anglais (Probability Weighted Moments), d'une distribution de probabilité par l'expression :

$$M(l, j, k) = E \left[X^l F^j (1 - F)^k \right] = \int_0^1 x(F)^l F^j (1 - F)^k dF$$

où : $x(F)$ est la valeur de la variable X correspondant à la valeur F de la fonction de répartition théorique ;
 l, j , et k sont des réels.

Pour $j = k = 0$ et l entier positif, $M(l, 0, 0)$ est le classique moment d'ordre l par rapport à l'origine.

Pour certaines distributions continues de probabilité où $x(F)$, fonction inverse de $F(x)$, peut être définie de manière analytique, il existe des relations simples entre les moments de probabilité pondérés $M(l, j, 0)$ ou $M(l, 0, k)$ (j et k entiers positifs) et les paramètres de la distribution. En d'autres termes, il est alors possible d'exprimer les paramètres de la loi de probabilité comme des fonctions des moments de probabilité pondérés.

On s'intéressera par la suite au moment de probabilité $M(1, j, 0)$, noté par commodité :

$$\beta_j = M(1, j, 0) = E \left[X F^j \right] = \int_0^1 x(F) F(x)^j dF$$

LOI DE JENKINSON OU LOI GÉNÉRALISÉE DES VALEURS EXTRÊMES

LES LOIS DE VALEURS EXTRÊMES

Les valeurs extrêmes des phénomènes hydrologiques, crues, sécheresses, présentent un intérêt tout particulier lorsqu'on parle en terme de risques, puisque ces phénomènes peuvent mettre en péril ouvrages et populations (DESBORDES, 1990). Très tôt les statisticiens ont essayé de trouver les distributions de ces seules valeurs extrêmes.

Le problème n'est pas simple. Les articles traitant du sujet font référence aux mêmes sources : FISHER et TIPPETT (1928), et JENKINSON (1955), mais en donnent, au-delà des pures divergences de notation, des interprétations sensiblement différentes ne faisant qu'accroître la complexité du problème.

Nous avons donc choisi d'avoir recours aux deux articles de référence précités pour essayer d'y voir plus clair.

Le problème posé est celui de la probabilité attachée à la plus grande valeur x_{\max} (ou à la plus petite x_{\min}) d'une variable hydrologique X dans un échantillon de taille N . Si l'on connaît la fonction de répartition $F(x)$ de la variable X , les distributions de x_{\max} et x_{\min} sont explicitement connues pour une taille d'échantillon donnée. Pour s'affranchir de la taille de l'échantillon et de la connaissance de $F(x)$ — aucune théorie n'est capable de déterminer la loi de probabilité suivie par la plupart des variables hydrométéorologiques —, il est apparu nécessaire de rechercher des approximations suffisamment satisfaisantes des distributions de x_{\max} et x_{\min} , indépendantes de $F(x)$ et de N . Ces approximations sont données par les formes limites de ces distributions quand N tend vers l'infini.

La définition du modèle probabiliste retenue par les spécialistes repose sur les considérations suivantes, extraites de FISHER et TIPPETT (1928).

La valeur extrême d'un échantillon de taille $N = m \cdot n$ est la valeur extrême d'un échantillon de n valeurs extrêmes d'échantillons de taille m .

Quand m croît indéfiniment, on conçoit que la probabilité attachée à la valeur extrême, déduite de l'étude d'un échantillon de taille $m \cdot n$, et celle attachée à la valeur extrême, déduite de l'étude de n valeurs extrêmes d'échantillons de taille m , tendent à devenir égales.

Les auteurs traduisent ceci de la façon suivante : si une « forme » limite existe, ces deux types de distribution tendent vers la « forme » limite quand m croît indéfiniment.

Il s'ensuit donc que l'on peut approcher, par l'étude d'un nombre donné (n) de valeurs extrêmes déduites d'échantillons de taille m (et ceci est d'autant plus vrai que m est grand), la connaissance en terme de probabilité de la valeur extrême d'un phénomène. Ainsi en hydrologie, pour estimer par exemple le débit d'une rivière relatif à un risque donné, on retient, pour chacune des n années observées, la valeur maximale du débit, et l'analyse statistique porte sur l'échantillon de ces n seules valeurs de débit maximum annuel.

FISHER et TIPPETT (1928) poursuivent de la manière suivante :

Soit un échantillon de n valeurs extrêmes supérieures (n maxima annuels par exemple), si $P(x)$ est la probabilité qu'une observation soit inférieure à x , la probabilité que la plus grande valeur de l'échantillon soit inférieure à x est $P^n(x)$, puisque cela revient à dire que toutes les valeurs de la série supposées indépendantes sont inférieures à x .

Dans les distributions limites, d'après les auteurs, $P^n(x)$ s'exprime sous la forme de l'équation fonctionnelle :

$$P^n(x) = P(a_n x + b_n) \quad (1)$$

a_n et b_n étant des coefficients fonction de n . L'équation fonctionnelle exprime que la distribution de la valeur extrême d'un phénomène se déduit de la distribution des n plus grandes valeurs de n échantillons par une transformation linéaire de la variable dépendant de n .

La résolution de cette équation repose sur la détermination des coefficients a_n et b_n , et de l'expression de P .

Le même type de raisonnement peut être appliqué aux valeurs extrêmes inférieures.

FISHER et TIPPETT (1928), puis GUMBEL (1966) ont montré que l'équation (1) admet trois types de solutions. FRÉCHET (1927) a l'antériorité de la découverte d'au moins l'une d'entre elles.

Il faut souligner que les justifications théoriques de l'équation fonctionnelle et des lois de distribution qui s'en déduisent sont loin d'être vérifiées en hydrologie, et que le choix d'une loi reste basé sur son bon ajustement à l'échantillon des observations.

LOI DE JENKINSON

JENKINSON (1955) a proposé une formulation unique des trois solutions de l'équation fonctionnelle, que nous développons ici uniquement dans le cas des valeurs extrêmes supérieures.

Il pose tout d'abord :

$$P(x) = \exp(-e^{-y})$$

ce qu'il justifie de la manière suivante.

Soit les N valeurs journalières x_i sur un an d'observations d'une variable météorologique.

Les N observations sont supposées indépendantes et la probabilité que l'une de ces valeurs soit inférieure à x est $1 - f(x) / N$, $f(x)$ désignant le nombre moyen annuel de valeurs journalières supérieures à x . Il s'ensuit que la probabilité que toutes les valeurs journalières d'une année soient inférieures à x , c'est-à-dire que le maximum annuel soit inférieur à x , est :

$$P(x) = [1 - f(x) / N]^N$$

soit si $f(x)$ est voisin de zéro, ce qui est le cas lorsqu'on s'intéresse aux valeurs extrêmes supérieures :

$$P(x) = e^{-f(x)}$$

En posant $y(x) = -\text{Log}[f(x)]$, il vient :

$$P(x) = \exp[-e^{-y(x)}]$$

Dans ces conditions l'équation (1) s'écrit :

$$[\exp(-e^{-y(x)})]^n = \exp[-e^{-y(a_n x + b_n)}]$$

ce qui, en prenant les logarithmes deux fois, donne :

$$\text{Log}(n) - y(x) = -y(a_n x + b_n)$$

La résolution de cette équation, établie par JENKINSON, conduit à :

$$y(x) = -1/k \text{Log}[1 - (x - x_0) / a]$$

avec : ak positif et $k \neq 0$

$$a_n = n^k$$

$$b_n = (x_0 + a) (1 - a_n)$$

Finalement :

$$P(x) = \exp[-e^{-y(x)}] = \exp\left[-\left(1 - \frac{x - x_0}{a}\right)^{1/k}\right]$$

soit si on pose $ak = s$:

$$P(x) = \exp[-e^{-y(x)}] = \exp\left[-\left(1 - \frac{k(x - x_0)}{s}\right)^{1/k}\right]$$

expression connue sous le nom de loi de JENKINSON ou loi généralisée des valeurs extrêmes (notée en anglais GEV distribution).

C'est de l'ajustement de la loi de JENKINSON sur les plus grandes valeurs de plusieurs échantillons (maxima annuels par exemple) que l'on déduit par extrapolation les probabilités afférentes aux valeurs extrêmes du phénomène étudié.

Selon le signe de k , la formulation proposée par JENKINSON permet de retrouver les trois types de lois déduits de l'équation (1) par FISHER et TIPPETT (1928).

En effet, $y(x)$ est toujours une fonction croissante de x , mais selon que dy/dx croît, décroît ou reste constant quand x croît, trois types de courbes peuvent être représentés.

La représentation de la fonction $P(x)$ est équivalente à celle de $y(x)$ puisqu'on passe, par la simple relation $\exp[-e^{-y(x)}]$, d'un axe gradué en y à un axe gradué en P .

L'étude de la fonction $y(x)$ est la suivante :

— $k > 0$: type de lois dont le domaine de variation des variables est borné dans le sens des extrêmes auxquels on s'intéresse :

$$\begin{array}{lll} y = -\infty & x = -\infty & dy/dx = 0 \\ y = 0 & x = x_0 & dy/dx = 1/s \\ y = +\infty & x = s/k + x_0 & dy/dx = +\infty \end{array}$$

— $k < 0$: type de lois dont le domaine de variation des variables est illimité dans le sens des extrêmes auxquels on s'intéresse, mais limité dans l'autre sens :

$$\begin{array}{lll} y = -\infty & x = s/k + x_0 & dy/dx = +\infty \\ y = 0 & x = x_0 & dy/dx = 1/s \\ y = +\infty & x = +\infty & dy/dx = 0 \end{array}$$

— quand $k = 0$:

$$\left[1 - \frac{k(x - x_0)}{s} \right]^{1/k} \text{ tend vers } \exp\left(-\frac{x - x_0}{s}\right)$$

d'où :

$$F(x) = \exp\left(-\exp\left(-\frac{x - x_0}{s}\right)\right), \text{ loi de GUMBEL}$$

et

$$y(x) = (x - x_0) / s$$

On a alors :

$$\begin{array}{lll} y = -\infty & x = -\infty & dy/dx = 1/s \text{ constante positive} \\ y = 0 & x = x_0 & dy/dx = 1/s \text{ constante positive} \\ y = +\infty & x = +\infty & dy/dx = 1/s \text{ constante positive} \end{array}$$

Dans le système de graduations fonctionnelles du papier de GUMBEL, les trois types de lois se distinguent par leur courbure.

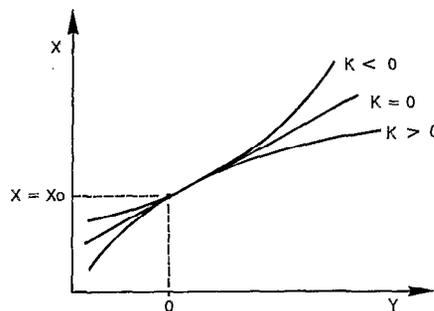


FIG. 1. — Représentation des trois types de lois de valeurs extrêmes sur un papier fonctionnel de Gumbel.

ESTIMATION DES PARAMÈTRES x_0 , s , k , DE LA LOI DE JENKINSON

INTRODUCTION

La méthode des moments est exposée dans Nerc (1975). On estime d'abord k qui est une fonction compliquée de γ_1 , estimateur du coefficient de dissymétrie, calculable à l'ordinateur ou tabulée. Connaissant k , on estime ensuite s et x_0 à partir de la moyenne et de l'écart-type de l'échantillon.

Nous n'avons pas testé la méthode des moments, mais elle ne semble pas performante puisque tous les ouvrages consultés en proposent d'autres (méthode graphique, méthode des sextiles, etc.).

La méthode du maximum de vraisemblance (JENKINSON, 1969) ne converge pas toujours. OTTEN et VAN MONTFORT (1980) ont proposé des modifications pour diminuer le nombre d'itérations et rendre la méthode plus fiable. HOSKING (1985) a publié un algorithme, comportant plusieurs pages de Fortran, qui semblait régler le problème. En fait MACLEOD (1988) y a apporté encore quelques améliorations.

Les problèmes de convergence semblent concerner surtout les petits échantillons ($n < 25$) avec des valeurs de k égales ou supérieures à 0,4.

Suite à l'article de HOSKING *et al.* (1985), nous avons jugé plus efficace et plus simple d'utiliser la méthode des moments de probabilité pondérés.

ESTIMATION DES PARAMÈTRES DE LA LOI DE JENKINSON PAR LA MÉTHODE DES MOMENTS DE PROBABILITÉ PONDÉRÉS

De l'expression de la fonction de répartition, on déduit pour $k \neq 0$:

$$x(F) = x_0 + \frac{s}{k} [1 - (-\text{Log } F)^k] \quad (2)$$

et donc :

$$\beta_j = \int_0^1 \left\{ x_0 + \frac{s}{k} [1 - (-\text{Log } F)^k] \right\} F^j dF$$

Dans le cas particulier où l , j et k sont des entiers positifs, $M(l, j, k)$ est proportionnel à $E[X_{j+1, k+j+1}^l]$, le $l^{\text{ième}}$ moment par rapport à l'origine de la statistique d'ordre $j+1$ d'un échantillon de taille $k+j+1$ (GREENWOOD *et al.*, 1979).

Plus précisément :

$$M(l, j, k) = E[X_{j+1, k+j+1}^l] B[j+1, k+1]$$

où $B[.,.]$ désigne la fonction Bêta (annexe 1).

Cette relation permet de calculer (LANDWEHR *et al.*, 1979), à partir d'un échantillon dont les observations sont classées par valeurs croissantes :

$$x_1 \leq x_2 \leq x_3 \leq \dots \leq x_i \leq \dots \leq x_n$$

un estimateur b_j de β_j (annexe 2) :

$$b_j = \frac{1}{n} \sum_{i=1}^n \frac{(i-1)(i-2)\dots(i-j)}{(n-1)(n-2)\dots(n-j)} x_i \quad (3)$$

$$\text{et } b_0 = \frac{1}{n} \sum_{i=1}^n x_i$$

LANDWEHR *et al.* (1979) montrent que cet estimateur est non biaisé.

HOSKING *et al.* (1985) préconisent dans la pratique l'utilisation d'un autre estimateur de β_j :

$$\hat{\beta}_j = \frac{1}{n} \sum_{i=1}^n [Fn(i)]^j x_i \quad (4)$$

$F_n(i)$ est la fréquence expérimentale de non dépassement de la valeur x_i de rang i (plotting position en anglais).

Les expressions les plus courantes pour $F_n(i)$, qui sont telles que $\hat{\beta}_j$ soit asymptotiquement équivalent à b_j , sont :

$$F_n(i) = (i - a) / n \text{ avec } 0 < a < 1$$

ou :

$$F_n(i) = (i - a) / (n + 1 - 2a) \text{ avec } -1/2 < a < 1/2$$

Si on pose $u = -\text{Log}(F)$, il en résulte :

$$\beta_j = \int_0^{\infty} \left\{ x_0 + \frac{s}{k} [1 - u^k] \right\} e^{-u(j+1)} du$$

soit :

$$\beta_j = \left(x_0 + \frac{s}{k} \right) \int_0^{\infty} e^{-u(j+1)} du - \frac{s}{k} \int_0^{\infty} u^k e^{-u(j+1)} du$$

donc :

$$\beta_j = \frac{1}{j+1} \left\{ x_0 + \frac{s}{k} \left[1 - (j+1)^{-k} \Gamma(k+1) \right] \right\}, k > -1 \quad (5)$$

De cette expression on déduit :

$$\beta_0 = x_0 + \frac{s}{k} (1 - \Gamma(1+k)) \quad (6)$$

$$2\beta_1 - \beta_0 = \frac{s}{k} \Gamma(1+k) (1 - 2^{-k}) \quad (7)$$

$$\frac{3\beta_2 - \beta_0}{2\beta_1 - \beta_0} = \frac{1 - 3^{-k}}{1 - 2^{-k}} \quad (8)$$

Pour estimer les paramètres k , x_0 et s de la loi de JENKINSON, à partir d'un échantillon aléatoire d'effectif n tiré de la population, on utilise les expressions précédentes en remplaçant β_0 , β_1 et β_2 par leurs estimateurs b_0 , b_1 , b_2 ou $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$, calculés sur l'échantillon.

HOSKING *et al.* (1985) recommandent les estimateurs $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$, avec pour expression de fréquence empirique de non dépassement :

$$F_n(i) = (i - 0,35) / n$$

Il faut noter toutefois que SINCLAIR et AHMAD (1988) ont remarqué qu'un changement d'origine se répercute sur la valeur du paramètre de forme k , quand on utilise cette expression. Pour corriger ce défaut, ils proposent l'expression :

$$\begin{aligned} F_n(i) &= (i + a) / (n + d) \\ a &= [-(n + 1) + (n^2 - 1)^{1/2}] / 2 \\ d &= 1 + 2a \end{aligned}$$

Pour NOPHADOL et VAN THANH (1989), la meilleure expression de la fréquence empirique pour la loi de JENKINSON serait :

$$F_n(i) = \frac{n - i + 0,05 \gamma_1 + 0,65}{n - 0,08 \gamma_1 + 0,38}$$

avec γ_1 = coefficient de dissymétrie.

Mais cette formule est la meilleure par référence à la valeur moyenne de l'observation de rang i d'un échantillon aléatoire d'effectif n tiré de la loi de JENKINSON, sans considérer l'estimation des paramètres par les moments de probabilité pondérés.

$\beta_0, \beta_1, \beta_2$ étant remplacés par leurs estimateurs, par itération on peut obtenir un estimateur de k à partir de l'expression (8). Cependant, comme les valeurs de k se situent en pratique entre $-1/2$ et $+1/2$, HOSKING *et al.* (1985) proposent une approximation polynomiale :

$$\hat{k} = 7,8590 c + 2,9554 c^2$$

avec :

$$c = \frac{2\beta_1 - \beta_0}{3\beta_2 - \beta_0} - \frac{\text{Log } 2}{\text{Log } 3}$$

Dans cette gamme de valeurs de k , l'erreur induite par cette approximation ne dépasse pas 9.10^{-4} . k étant estimé, on peut calculer ensuite :

$$\hat{s} = \frac{(2\beta_1 - \beta_0)\hat{k}}{\Gamma(1 + \hat{k})(1 - 2^{-\hat{k}})} \quad (9)$$

en remplaçant β_0 et β_1 par leurs estimateurs, et :

$$\hat{x}_0 = \beta_0 + \hat{s} \frac{\Gamma(1 + \hat{k}) - 1}{\hat{k}} \quad (10)$$

en remplaçant de même β_0 par son estimateur.

TEST DE LA VALEUR DE k DANS UNE LOI DE JENKINSON

Pour décider si une loi de GUMBEL suffit pour décrire la distribution de valeurs extrêmes supérieures, on teste l'hypothèse nulle $H_0 : k = 0$ contre l'hypothèse alternative $k \neq 0$.

Le test le plus simple est proposé par HOSKING *et al.* (1985). Selon ces auteurs, sous l'hypothèse H_0 , l'estimateur \hat{k} du paramètre k suit asymptotiquement une loi normale de moyenne 0 et de variance $0,5633/n$ (n effectif de l'échantillon). Le test consiste donc à comparer :

$$U_{\text{obs}} = \hat{k} (n/0,5633)^{1/2}$$

aux valeurs critiques de la loi normale réduite pour accepter ou rejeter H_0 .

Les tests proposés par OTTEN et VAN MONTFORT (1978) sont beaucoup plus compliqués et font appel à des valeurs critiques, calculées par simulation en fonction de n et de α (α = risque de première espèce, c'est-à-dire probabilité de rejeter l'hypothèse H_0 alors qu'elle est vraie).

INTERVALLES DE CONFIANCE DES QUANTILES D'UNE LOI DE JENKINSON DONT LES PARAMÈTRES SONT ESTIMÉS PAR LA MÉTHODE DES MOMENTS DE PROBABILITÉ PONDÉRÉS

On peut montrer que le vecteur b des estimateurs des moments de probabilité pondérés $(b_0, b_1, b_2)^T$ suit asymptotiquement une distribution d'échantillonnage normale à trois dimensions, de moyenne $\beta = (\beta_0, \beta_1, \beta_2)^T$ et de matrice de variance-covariance $n^{-1}V$. Les éléments de V ont été calculés par HOSKING *et al.* (1985).

La distribution asymptotique des estimateurs des paramètres de la loi de JENKINSON par la méthode des moments de probabilité pondérés découle de la distribution des estimateurs des moments de probabilité pondérés.

Si on pose $\hat{\theta} = (\hat{x}_0, \hat{s}, \hat{k})^T$, le vecteur de ces paramètres, les équations (8), (9) et (10) définissent la relation $\hat{\theta} = f(b)$. Asymptotiquement $\hat{\theta}$ suit une loi normale à trois dimensions, de moyenne $\theta = f(\beta)$ et de matrice de variance-covariance :

$$n^{-1} G V G^T$$

G est une matrice 3×3 , dont les éléments g_{ij} correspondent aux dérivées des relations entre les paramètres et les moments :

$$g_{ij} = \frac{\delta f_i}{\delta b_j}$$

La matrice de variance-covariance est de la forme :

$$n^{-1}GVG^T = n^{-1} \begin{bmatrix} \hat{s}^2 w_{11} & \hat{s}^2 w_{12} & \hat{s} w_{13} \\ \hat{s}^2 w_{12} & \hat{s}^2 w_{22} & \hat{s} w_{23} \\ \hat{s} w_{13} & \hat{s} w_{23} & w_{33} \end{bmatrix}$$

Les éléments w_{ij} sont des fonctions compliquées de k . Elles ont été tabulées par HOSKING (1985) pour des valeurs de k allant de $-0,4$ à $0,4$ au pas de $0,1$.

Dans cette matrice, les indices 1, 2 et 3 se rapportent aux paramètres x_0 , s et k respectivement. Les éléments diagonaux représentent les variances, les autres les covariances.

A partir de la matrice précédente, on calcule facilement la variance $\text{var}(x_p)$ d'un quantile x_p , puis son intervalle de confiance, en utilisant les méthodes qui s'appliquent habituellement aux paramètres estimés par la méthode du maximum de vraisemblance (MASSON, 1983).

On a :

$$\begin{aligned} \text{var}(x_p) \approx & \left[\frac{\delta x_p}{\delta x_0} \right]^2 \text{var}(x_0) + \left[\frac{\delta x_p}{\delta s} \right]^2 \text{var}(s) + \left[\frac{\delta x_p}{\delta k} \right]^2 \text{var}(k) \\ & + 2 \frac{\delta x_p}{\delta x_0} \frac{\delta x_p}{\delta s} \text{cov}(x_0, s) + 2 \frac{\delta x_p}{\delta x_0} \frac{\delta x_p}{\delta k} \text{cov}(x_0, k) + 2 \frac{\delta x_p}{\delta s} \frac{\delta x_p}{\delta k} \text{cov}(s, k) \end{aligned}$$

Les dérivées se déduisent facilement de l'expression du quantile :

$$x_p = x_0 + \frac{s}{k} - \frac{s}{k} e^{-ky_1}$$

$$\text{avec } y_1 = -\text{Log}[-\text{Log}(p)] \text{ et } p = \text{Prob}(X \leq x_p) = F(x_p)$$

On obtient donc :

$$\frac{\delta x_p}{\delta x_0} = 1$$

$$\frac{\delta x_p}{\delta s} = \frac{1}{k} - \frac{1}{k} e^{-ky_1}$$

$$\frac{\delta x_p}{\delta k} = -\frac{s}{k^2} + \frac{y_1 s}{k} e^{-ky_1} + \frac{s}{k^2} e^{-ky_1}$$

Par la suite l'intervalle de confiance est obtenu par :

$$x_p \pm U_{1-\alpha/2} \sqrt{\text{var}(x_p)}$$

$U_{1-\alpha/2}$ est la variable normale réduite de probabilité de non dépassement $1 - \alpha/2$, et α est la probabilité que l'intervalle ne contienne pas la vraie valeur du quantile.

Ce calcul de l'intervalle de confiance, basé sur un comportement asymptotique, peut être utilisé dès que l'effectif de l'échantillon est supérieur ou égal à 50.

COMPARAISON DES MÉTHODES D'ESTIMATION DES PARAMÈTRES D'UNE LOI DE JENKINSON

Les comparaisons concernent la méthode des moments de probabilité pondérés et celle du maximum de vraisemblance.

Les conclusions de ces comparaisons varient évidemment en fonction du critère considéré.

On mesure traditionnellement la qualité d'un estimateur G d'un paramètre γ par son efficacité asymptotique E :

$$E = \lim_{n \rightarrow \infty} (\text{variance minimale de l'estimateur de } \gamma / \text{variance de } G)$$

Un estimateur G est asymptotiquement efficace quand E tend vers 1. C'est le cas des estimateurs obtenus par la méthode du maximum de vraisemblance.

On peut donc juger de la qualité des estimateurs obtenus par la méthode des moments de probabilité pondérés, en comparant leur variance à celle des estimateurs obtenus par le maximum de vraisemblance.

D'après HOSKING *et al.* (1985), pour des valeurs de k comprises entre $-0,2$ et $+0,2$, leur efficacité est supérieure à $0,7$. Cependant quand k tend vers $-1/2$, elle tend vers zéro !

Mais d'une manière générale, si la méthode du maximum de vraisemblance conduit asymptotiquement à une variance plus faible, elle conduit aussi à un biais plus important.

HOSKING *et al.* (1985) montrent ainsi que quand $k < -0,4$, la méthode des moments de probabilité pondérés perd de ses qualités : le biais qui, pour les valeurs de k supérieures à $-0,4$, avait l'ordre de grandeur de $1/n$ (n effectif de l'échantillon) devient alors très important. Toutefois il en est de même pour la méthode du maximum de vraisemblance pour $k > 1/3$.

Les considérations sur le comportement asymptotique des estimateurs ne suffisent pas pour juger une méthode. Il faut aussi étudier ce qui se passe quand on ajuste des lois sur de petits échantillons, et cette comparaison s'effectue par des simulations. HOSKING *et al.* (1985) montrent ainsi que la méthode des moments de probabilité pondérés donne la plus petite variance au paramètre de forme k , quand les échantillons sont petits ($n < 25$), et qu'il en résulte une meilleure estimation des quantiles supérieurs.

Ils soulignent donc que la relative inefficacité asymptotique de la méthode des moments de probabilité pondérés par rapport à la méthode du maximum de vraisemblance n'apparaît pas sur les résultats de la simulation, concernant des échantillons avec $n < 100$.

QUELQUES TESTS RELATIFS A LA MÉTHODE DES MOMENTS DE PROBABILITÉ PONDÉRÉS APPLIQUÉE A LA LOI DE JENKINSON

Avant de mettre en œuvre la méthode des moments de probabilité pondérés pour estimer les paramètres d'une loi de JENKINSON dans le cadre de variables hydrologiques, nous avons jugé nécessaire de procéder à quelques tests dont les résultats sont présentés ci-après.

CHOIX DES ESTIMATEURS DES MOMENTS DE PROBABILITÉ PONDÉRÉS

Nous avons généré par simulation 100 échantillons de $n = 30$ valeurs (ordre de grandeur de l'effectif des échantillons le plus souvent disponibles pour l'étude de valeurs extrêmes en hydrologie), tirés d'une loi de JENKINSON de paramètres connus :

$$x_0 = 105,8, \quad s = 42,5, \quad k = -0,13$$

Nous avons procédé par tirage aléatoire dans une loi uniforme $(0, 1)$ de valeurs de probabilité, dont nous avons estimé les quantiles correspondants en utilisant l'expression (2).

Les moments de probabilité pondérés $\beta_0, \beta_1, \beta_2$ sont donnés par l'équation (5) fonction de x_0, s et k :

$$\begin{aligned} \beta_0 &= 105,8 \\ \beta_1 &= 85,14 \\ \beta_2 &= 63,82 \end{aligned}$$

Nous avons testé différentes formules d'estimation de ces moments sur les 100 échantillons générés :

— la formule (3) des estimateurs non biaisés b_0, b_1, b_2 ;

— la formule (4) des estimateurs $\hat{\beta}_j$ avec $F_n(i) = (i - 0,35) / n$, nous l'appellerons formule (4.1.) ;

— la formule (4) des estimateurs $\hat{\beta}_j$ avec $F_n(i) = (i + a) / (n + d)$, nous l'appellerons formule (4.2.)

$$\begin{aligned} a &= [-(n + 1) + (n^2 - 1)^{1/2}] / 2 \\ d &= 1 + 2a \end{aligned}$$

L'estimateur de β_0 est la moyenne des valeurs de l'échantillon, quelle que soit la formule utilisée.

Le biais et la variance des estimateurs de β_1 , β_2 obtenus à partir des trois formules étudiées sur l'ensemble des 100 échantillons sont les suivants.

TABLEAU I
Moments de probabilité pondérés - Biais des estimateurs de β_1 et β_2

formule	biais estimateur de β_1	biais estimateur de β_2
3	-0,5275	-0,2256
4.1.	-0,4174	-0,1096
4.2.	-1,0845	-0,9303

TABLEAU II
Moments de probabilité pondérés - Variance des estimateurs de β_1 et β_2

formule	variance estimateur de β_1	variance estimateur de β_2
3	69,6381	49,7463
4.1.	68,9520	48,7446
4.2.	68,0717	47,6485

Ces tableaux montrent que les trois formules testées fournissent des estimateurs de β_1 et β_2 dont la moyenne (dont se déduit le biais par rapport à la vraie valeur) et la variance calculées sur les 100 échantillons sont sensiblement du même ordre de grandeur.

Ils ne mettent pas en évidence le fait que la formule (3) donne des estimateurs non biaisés. Les valeurs de biais déduites de l'utilisation de cette formule ne sont pas les plus faibles.

La formule (4.2.) donne les valeurs maximales de biais et minimales de variance.

A priori il conviendrait donc de choisir entre les deux premières formules, la deuxième apparaissant ici au vu des résultats obtenus comme le meilleur compromis au regard des critères de biais et de variance étudiés.

Ce choix est de plus conforme aux recommandations faites par HOSKING *et al.* (1985) et reprises par CUNNANE (1989) sur l'utilisation d'estimateurs biaisés des moments de probabilité pondérés (avec une préférence pour ceux obtenus à partir de la formule (4.1)) pour l'estimation des quantiles dans l'étude de variables hydrologiques.

Toutefois, il conviendrait de confirmer ces conclusions sur la base de simulations, réalisées à partir d'autres jeux de paramètres de la loi de JENKINSON.

MÉTHODE DES MOMENTS DE PROBABILITÉ PONDÉRÉS OU MÉTHODE DU MAXIMUM DE VRAISEMBLANCE ?

Nous avons tenu à étudier le comportement de la méthode des moments de probabilité pondérés et de la méthode du maximum de vraisemblance pour des valeurs de k inférieures à $-0,4$ et supérieures à $+0,4$, sur des échantillons d'effectif inférieur à 50.

Comme précédemment, nous avons généré par simulation 100 échantillons de 30 valeurs tirés d'une loi de JENKINSON de paramètres :

— Cas 1 :

$$\begin{aligned}x_0 &= 51,68 \\s &= 20,50 \\k &= 0,52\end{aligned}$$

Les échantillons générés ont un coefficient de dissymétrie négatif, supérieur à $-1,81$.

— Cas 2 :

$$\begin{aligned}x_0 &= 51,68 \\s &= 20,50 \\k &= -0,52\end{aligned}$$

Les échantillons générés ont un coefficient de dissymétrie positif inférieur à 3 dans 59 % des cas, et atteignant la valeur 5,22 sur un échantillon. Ces échantillons sont comparables à ceux de débits maxima annuels. En effet d'après CUNNANE (1989), le coefficient de dissymétrie estimé sur des séries de débits maxima annuels, est en général positif et compris entre 0 et 5 ou plus, avec *majoritairement* des valeurs allant de 0,5 à 3,0.

Dans les deux cas, nous avons mis en œuvre sur chacun des 100 échantillons la méthode des moments de probabilité pondérés et la méthode du maximum de vraisemblance pour estimer les paramètres de la loi.

L'algorithme de calcul de la méthode du maximum de vraisemblance programmé est celui proposé par HOSKING (1985) et revu par MACLEOD (1988).

Les estimateurs des moments de probabilité pondérés β_0 , β_1 et β_2 ont été calculés en utilisant la formule (4.1).

Pour chacune des deux méthodes, l'estimation de chacun des paramètres en fonction de l'échantillon a été représentée.

Cas 1

Il apparaît sur les graphiques (fig. 2 à 5) que pour x_0 et s , les deux méthodes d'ajustement conduisent sur chaque échantillon à des estimations sensiblement égales (tabl. III).

Il s'ensuit que les estimations de ces paramètres dépendent davantage des caractéristiques propres de chaque échantillon que de la méthode utilisée. En ce qui concerne le paramètre k , les résultats sont les suivants (fig. 6 et 7) (tabl. IV).

Il apparaît ici que la méthode du maximum de vraisemblance a tendance à surestimer la valeur du paramètre k , tandis que la méthode des moments de probabilité pondérés la sous-estime. La méthode des moments de probabilité

TABLEAU III
Comparaison des méthodes des moments de probabilité pondérés et du maximum de vraisemblance
Influence sur les paramètres x_0 et s de la loi de JENKINSON - Cas 1

	moments de probabilité pondérés		maximum de vraisemblance	
	x_0	s	x_0	s
moyenne	50,090	20,487	51,641	20,415
variance	15,863	8,316	15,351	9,550
minimum	39,214	12,746	40,489	11,326
maximum	59,972	29,261	61,247	30,292

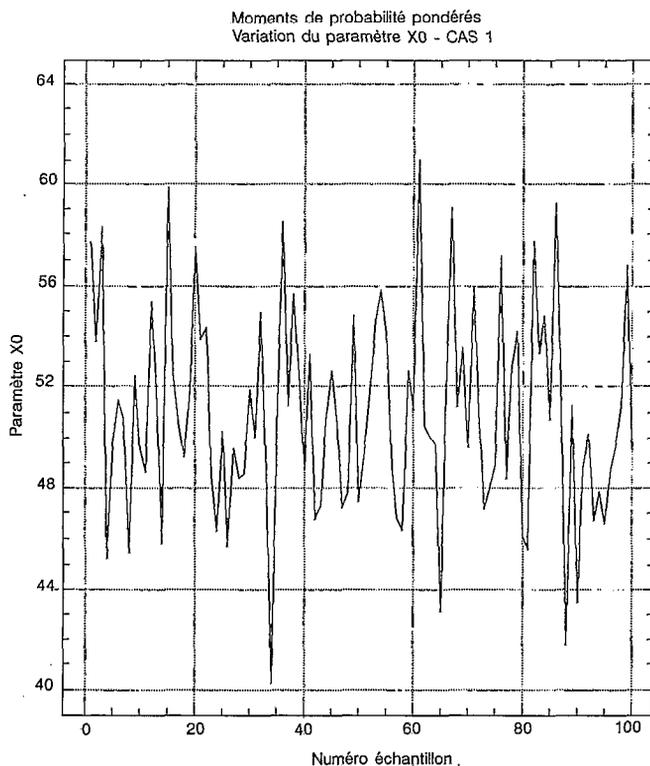


FIG. 2. — Méthode PWM. Paramètre x_0 .

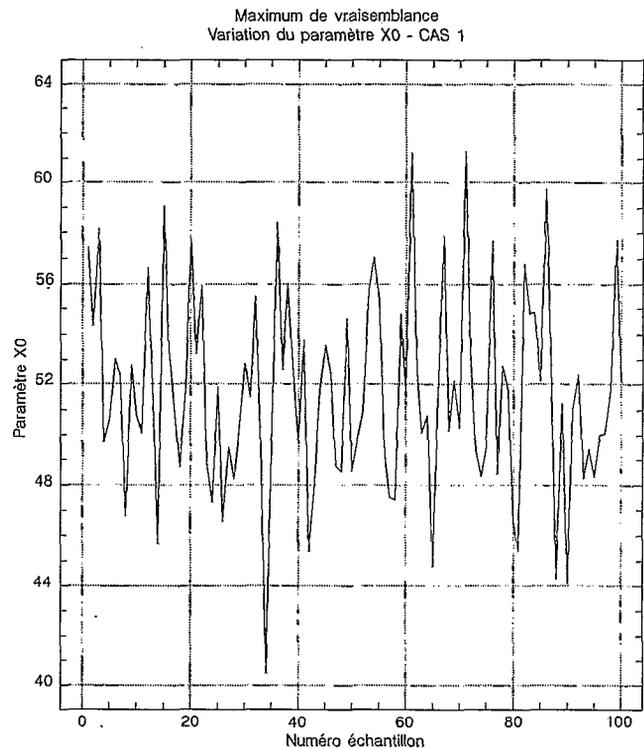


FIG. 3. — Maximum de vraisemblance. Paramètre x_0 .

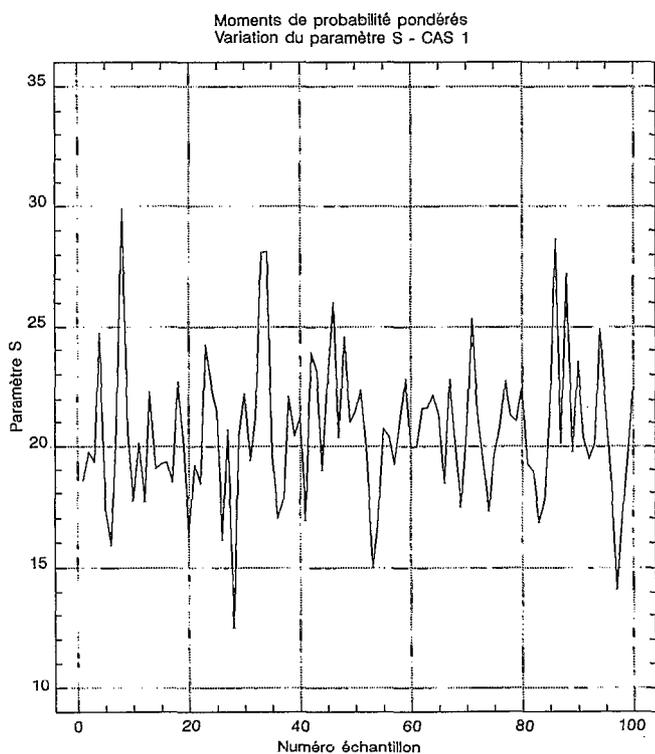


FIG. 4. — Méthode PWM. Paramètre s.

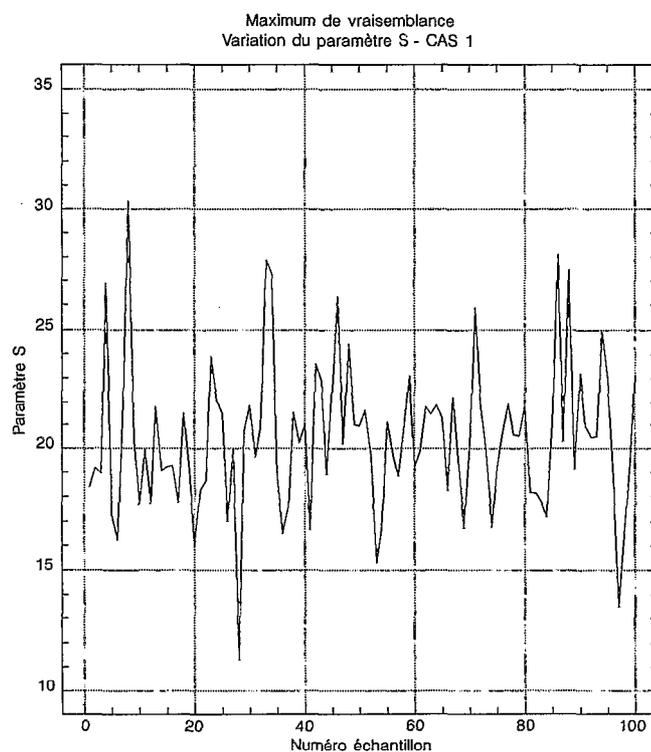


FIG. 5. — Maximum de vraisemblance. Paramètre s.

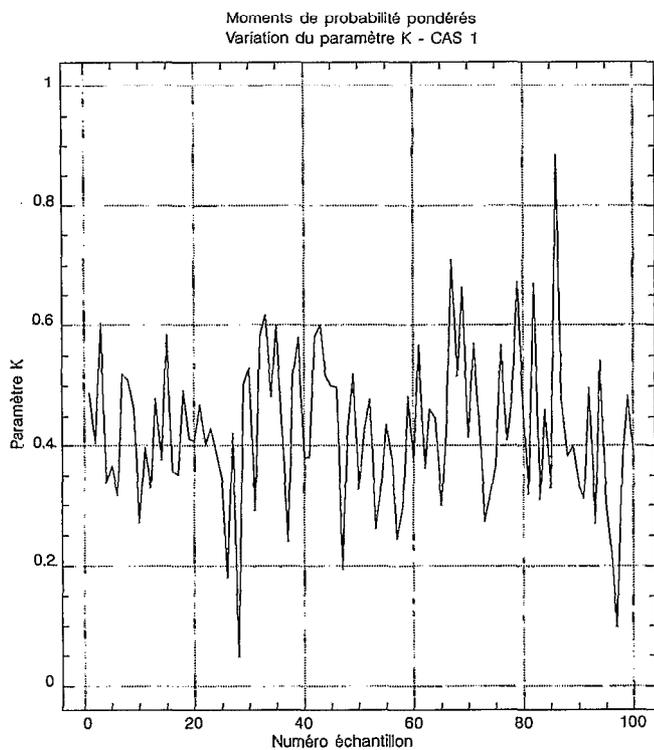


FIG. 6. — Méthode PWM. Paramètre k.

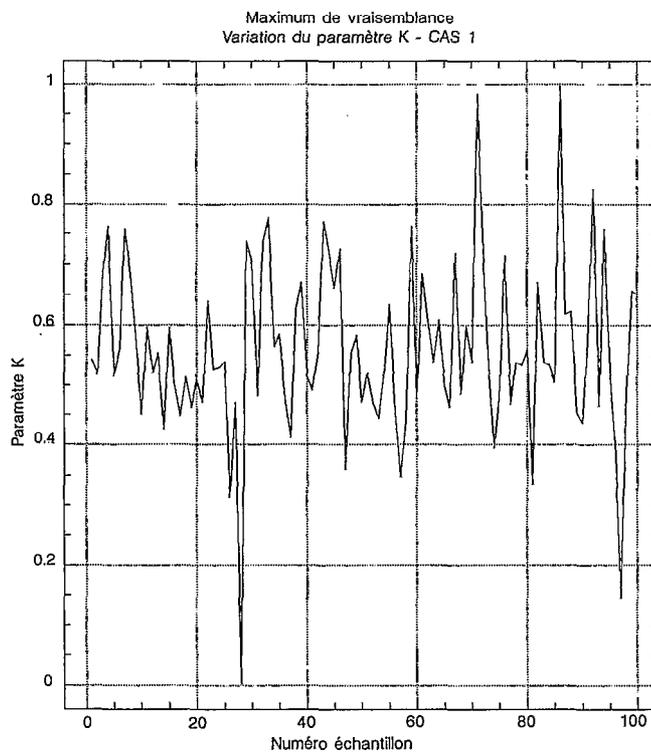


FIG. 7. — Maximum de vraisemblance. Paramètre k.

TABLEAU IV
 Comparaison des méthodes des moments de probabilité pondérés et du maximum de vraisemblance
 Influence sur le paramètre k de la loi de JENKINSON - Cas 1

	moments de probabilité pondérés	maximum de vraisemblance
	k	k
moyenne	0,426	0,558
variance	0,017	0,021
minimum	0,0477	0,0027
maximum	0,8864	0,998

pondérés met en évidence un biais plus important, mais une variance d'échantillonnage plus faible que la méthode du maximum de vraisemblance. Il est par conséquent difficile, sur ce cas, de départager ces deux méthodes.

Cas 2

Les graphiques (fig. 8 à 13) et les résultats (tabl. V) mettent en évidence que la méthode du maximum de vraisemblance donne dans ce cas des estimateurs dont la variance d'échantillonnage est très grande. La méthode des moments de probabilité pondérés conduit à des estimations tout à fait satisfaisantes.

On ne retrouve donc pas sur des échantillons de faible effectif (ici $n = 30$) les conclusions relatives au comportement asymptotique des estimateurs, en particulier en ce qui concerne la mise en défaut de la méthode des moments de probabilité pondérés au profit de la méthode du maximum de vraisemblance pour des valeurs de k inférieures à $-0,4$.

Le test réalisé conduit à la conclusion inverse. Cependant, ce test n'a porté que sur un cas particulier. Compte tenu, par conséquent, des qualités asymptotiques des estimateurs du maximum de vraisemblance, cette méthode mérite tout de même de retenir notre attention.

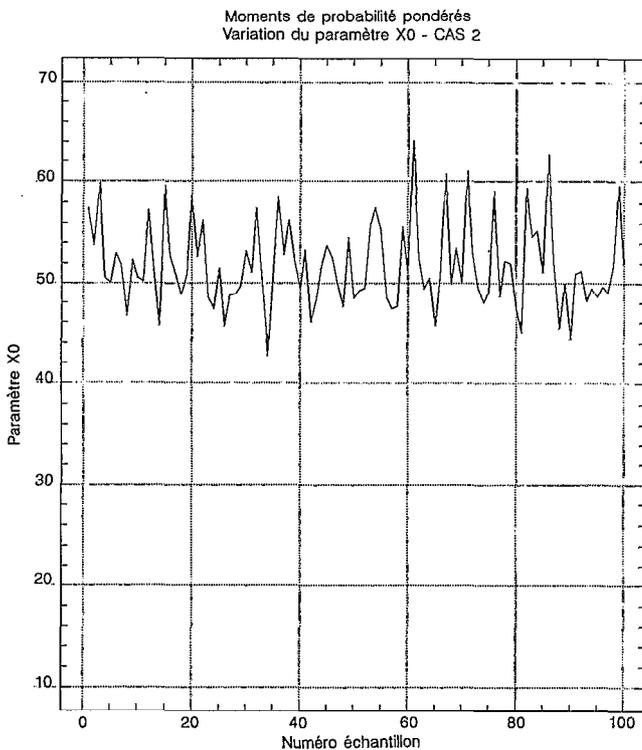


FIG. 8. — Méthode PWM. Paramètre x_0 .

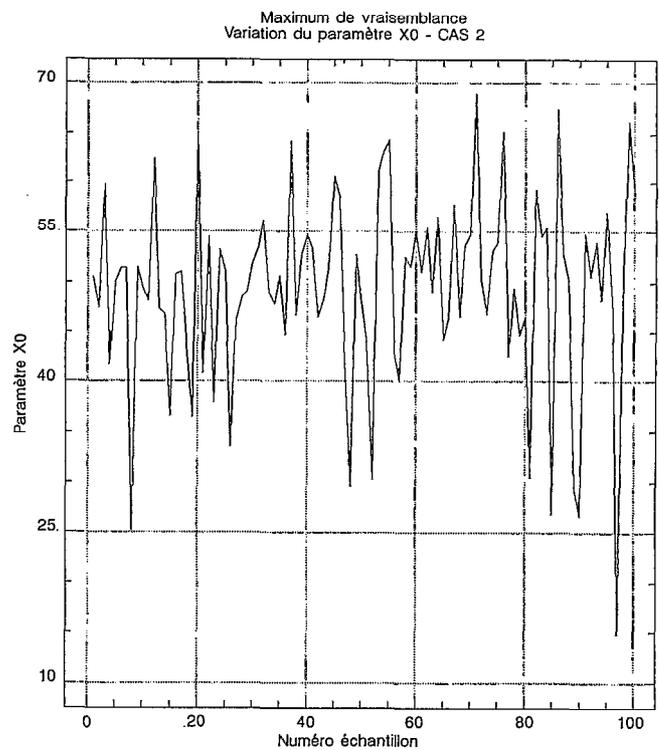


FIG. 9. — Maximum de vraisemblance. Paramètre x_0 .

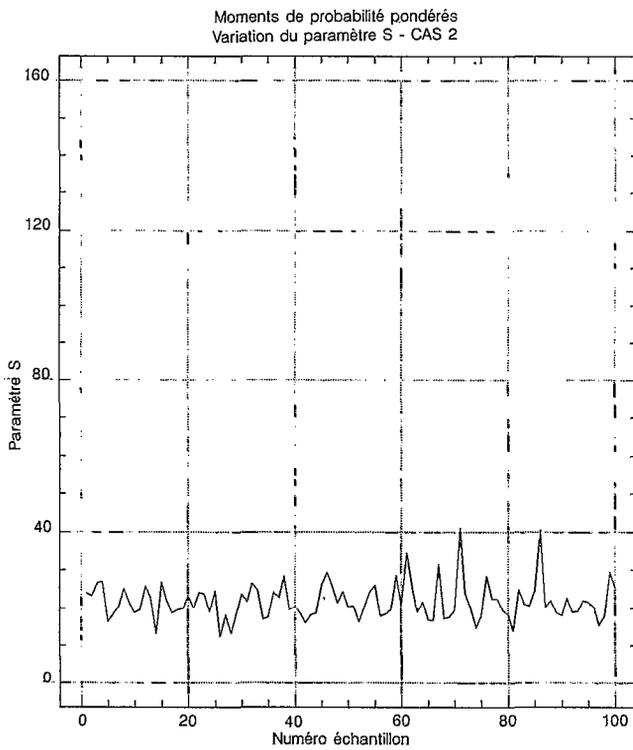


FIG. 10. — Méthode PWM. Paramètre s.

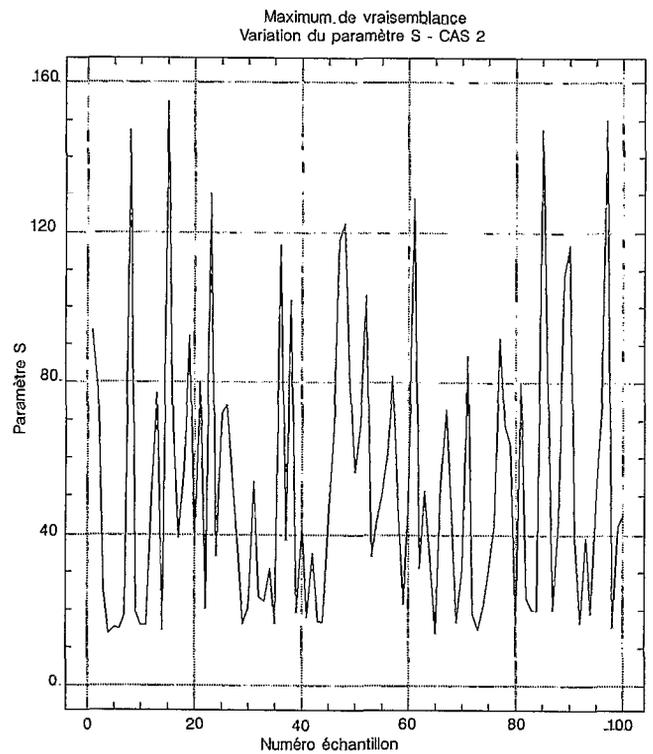


FIG. 11. — Maximum de vraisemblance. Paramètre s.

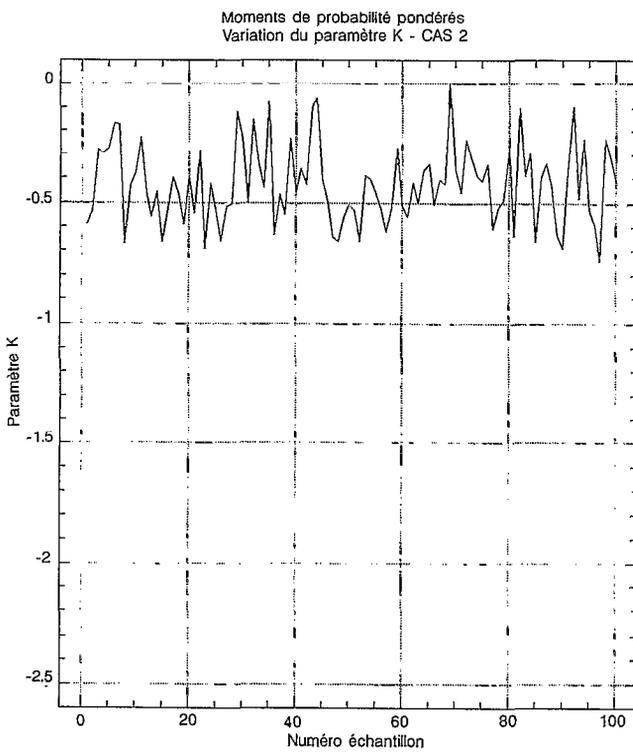


FIG. 12. — Méthode PWM. Paramètre k.

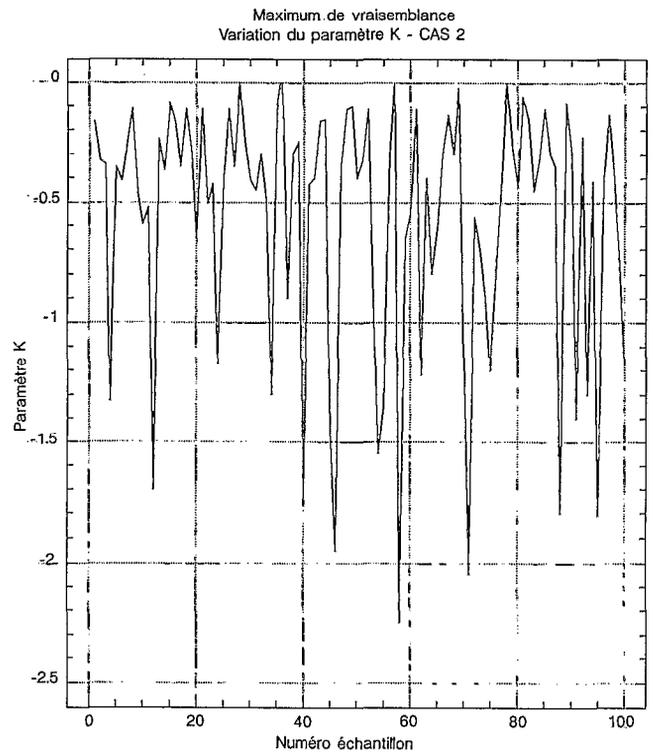


FIG. 13. — Maximum de vraisemblance. Paramètre k.

TABLEAU V
 Comparaison des méthodes des moments de probabilité pondérés et du maximum de vraisemblance
 Influence sur les paramètres x_0 , s , et k de la loi de JENKINSON - Cas 2

	moments de probabilité pondérés			maximum de vraisemblance		
	x_0	s	k	x_0	s	k
moyenne	51,745	21,686	-0,426	49,415	52,624	-0,561
variance	17,966	23,721	0,027	93,029	1350,94	0,277
minimum	42,708	12,488	-0,743	14,771	13,516	-2,249
maximum	64,162	40,930	0,005	68,835	154,837	0,051

EXEMPLE DE CALCUL DE L'INTERVALLE DE CONFIANCE A 90 % DES QUANTILES D'UNE LOI DE JENKINSON, DONT LES PARAMÈTRES SONT ESTIMÉS PAR LA MÉTHODE DES MOMENTS DE PROBABILITÉ PONDÉRÉS

Nous avons établi le programme informatique relatif à la procédure de calcul exposée précédemment sur les données de Nerc (1975) de 35 débits maxima annuels de la rivière Nidd à Hunsigore (Yorkshire), données auxquelles HOSKING *et al.* (1985) font référence dans leur article, et que nous avons pu nous procurer. L'intervalle de confiance à 90 % est présenté ci-dessous (fig. 14).

Malgré le faible effectif de l'échantillon, qui est inférieur à la limite théorique de validité de la méthode de calcul ($n \geq 50$), cet intervalle de confiance semble tout à fait acceptable, ce qui valide, sur cet exemple, la procédure de calcul.

CONCLUSION

La loi de JENKINSON présente l'avantage de traiter d'un seul coup les trois types de lois concernant les valeurs extrêmes supérieures. Grâce à la méthode des moments de probabilité pondérés, l'estimation des paramètres est facile et fiable.

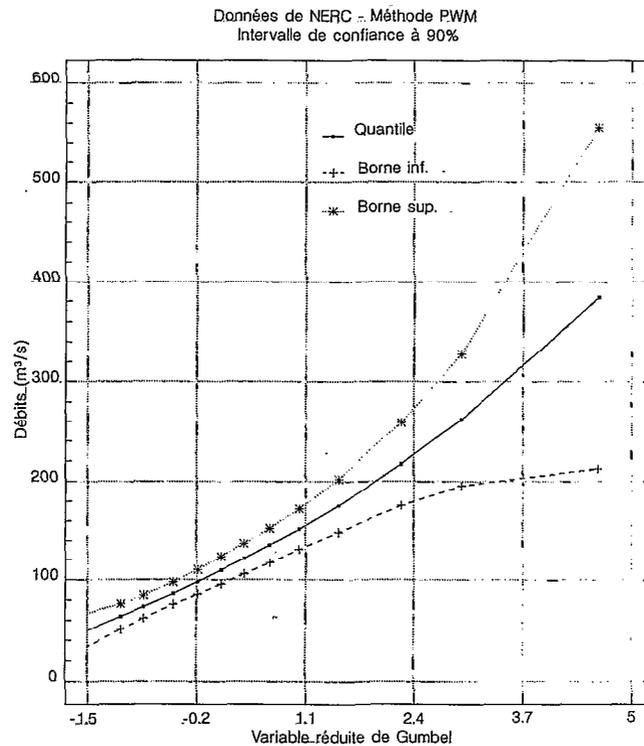


FIG. 14. — Intervalle de confiance à 90 %.

Nous disposons par ailleurs d'un test simple pour décider si une loi de GUMBEL ($k = 0$) suffit à décrire la distribution de fréquences des valeurs extrêmes supérieures.

La loi de JENKINSON est utilisée systématiquement dans certains pays (Grande-Bretagne) pour l'étude des crues. En France, la loi de FRÉCHET (loi de JENKINSON avec $k < 0$), après avoir été utilisée pour l'étude des crues dans les années cinquante, est tombée en désuétude en même temps que se généralisait l'utilisation de la méthode du gradex. Cependant, les hypothèses mêmes de la méthode du gradex devraient réhabiliter la loi de FRÉCHET.

En effet, dans la mesure où le gradex des débits, faible pour les petites périodes de retour, tend à rejoindre celui beaucoup plus élevé des pluies pour les fortes périodes de retour, les échantillons d'au moins cinquante années d'observations devraient donner sur un papier fonctionnel de GUMBEL, des distributions de fréquence à concavité tournée vers le haut (fig. 1).

Compte tenu de la flexibilité de la loi de JENKINSON, il nous semble nécessaire de la faire figurer dans le menu d'un logiciel de lois statistiques destinées à l'hydrologie, même si quelques vérifications restent encore à faire à son sujet.

BIBLIOGRAPHIE

- ABRAMOVITZ (M.), STEGUN (I.A.), eds, 1972. — Handbook of mathematical functions. Dover Publications. New York, 1046 p.
- BERGER (A.), MELICE (J.L.), DEMUTH (C.), 1982. — Statistical distributions of daily and high atmospheric SO₂ concentrations. *Atmospheric Environment*, vol. 16, n° 12 : 2863-2877.
- CUNNANE (C.), 1989. — Statistical distributions for flood frequency analysis. *World Meteorological Organization*, n° 718. Operational Hydrology Report n° 33, 118 p.
- DESBORDES (M.), 1990. — Risques de défaillance des ouvrages d'assainissement urbain : un concept révisable ? *La Houille Blanche*, n° 7/8 : 491-494.
- FISHER (R.A.), TIPPETT (L.H.C.), 1928. — Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*, vol. 24 : 180-190.
- FRÉCHET (M.), 1927. — Sur la loi de probabilité de l'écart maximum. *Ann. Soc. Polonaise Math.*, vol. 6 : 93-116.
- GREENWOOD (J.A.), LANDWEHR (J.M.), MATALAS (N.C.), WALLIS (J.R.), 1979. — Definition and relation to parameters of several distributions expressible in inverse form. *Water Resources Research*, vol. 15, n° 5 : 1049-1054.
- GUMBEL (E.J.), 1966. — Extreme value analysis of hydrologic data. *Statistical Methods in hydrology. Proceedings of Hydrology symposium*, n° 5 : 147-181.
- HOSKING (J.R.M.), 1985. — Maximum-likelihood estimation of the parameters of the generalized extreme-value distribution. *Algorithm AS 215. Applied Statistics*, vol. 34, n° 3.
- HOSKING (J.R.M.), WALLIS (J.R.), WOOD (E.F.), 1985. — Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, vol. 27, n° 3 : 251-261.
- JENKINSON (A.F.), 1955. — The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, vol. 81 : 158-171.
- JENKINSON (A.F.), 1969. — Statistics of extremes. Technical note 98, World Meteorological Office.
- KENDALL (M.G.), STUART (A.), 1977. — The advanced theory of statistics, vol. 1. Charles Griffin, Londres, 472 p.
- LANDWEHR (J.M.), MATALAS (N.C.), WALLIS (J.R.), 1979. — Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles. *Water Resources Research*, vol. 15, n° 5 : 1055-1064.
- MACLEOD (A.J.), 1988. — A remark on Algorithm AS 215. Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. Remark AS R76. *Applied statistics*.
- MASSON (J.M.), 1983. — Méthode générale approchée pour calculer l'intervalle de confiance d'un quantile. Application à quelques lois de probabilité utilisées en hydrologie. Note interne 6/1983. Laboratoire d'hydrologie et modélisation. Université Montpellier II, 41 p.
- NATURAL ENVIRONMENT RESEARCH COUNCIL (Nerc), 1975. — Flood Studies Report, Hydrological Studies, vol. 1, 550 p.
- NOPHADOL (I.N.), VAN THANH (V.T.), 1989. — An unbiased plotting position formula for the general extreme value distribution. *Journal of Hydrology*, vol. 106 : 193-209.
- OTTEN (A.), VAN MONTFORT (M.A.J.), 1978. — The power of two tests on the type of distributions of extremes. *Journal of Hydrology*, vol. 37 : 195-199.
- OTTEN (A.), VAN MONTFORT (M.A.J.), 1980. — Maximum-likelihood estimation of the general extreme-value distribution parameters. *Journal of Hydrology*, vol. 47 : 187-192.
- ROCHE (M.), 1963. — Hydrologie de surface. Gauthier-Villars, Paris, 430 p.
- SINCLAIR (C.D.), AHMAD (M.I.), 1988. — Location-invariant plotting position for PWM estimation of the parameters of the GEV distribution. *Journal of Hydrology*, vol. 99 : 271-279.
- YEVEVICH (V.), 1972. — Probability and statistics in hydrology. Water resources publications, Fort Collins, Colorado, 302 p.

ANNEXE 1

$$M(l, j, k) = E [X_{j+1, k+j+1}^l] B [j + 1, k + 1]$$

$$E [X_{j+1, k+j+1}^l] = \int_{Dx} x.^l g(x.) dx. \text{ avec } X. = X_{j+1, k+j+1}$$

Sachant que (KENDALL et STUART, 1977) :

$$g(x.) dx. = \frac{(k + j + 1)!}{j! k!} [F(x.)]^j [1 - F(x.)]^k dF., dF. = f(x.) dx.,$$

$$E [X_{j+1, k+j+1}^l] = \frac{(k + j + 1)!}{j! k!} \int_0^1 x.^l [F(x.)]^j [1 - F(x.)]^k dF$$

$$E [X_{j+1, k+j+1}^l] = \frac{(k + j + 1)!}{j! k!} M(l, j, k)$$

si l, j, k entiers positifs :

$$E [X_{j+1, k+j+1}^l] = \frac{\Gamma(j + k + 2)}{\Gamma(j + 1) \Gamma(k + 1)} M(l, j, k)$$

$$E [X_{j+1, k+j+1}^l] = \frac{M(l, j, k)}{B [j + 1, k + 1]}$$

B désigne la fonction Bêta : $B(z, w) = \frac{\Gamma(z) \Gamma(w)}{\Gamma(z + w)} = B(w, z)$

ANNEXE 2

Soit un échantillon de taille n dont les observations sont classées par valeurs croissantes :

$$x_1 \leq x_2 \leq x_3 \leq \dots \leq x_i \leq \dots \leq x_n$$

Le nombre de façons de former un échantillon de taille $j + 1$ à partir de n observations est C_n^{j+1}

On cherche un estimateur de $E [X_{j+1, j+1}^l]$. On s'intéresse donc à la distribution de la variable qui occupe la $(j + 1)$ ième position dans un échantillon de taille $j + 1$, dont les valeurs sont classées par ordre croissant.

Le nombre de sous-échantillons où la $(j + 1)$ ième valeur est :

$$x_n \text{ est } C_{n-1}^j$$

$$x_{n-1} \text{ est } C_{n-2}^j$$

$$x_{n-i} \text{ est } C_{n-i-1}^j$$

Un estimateur de $M(l, j, 0)$ est donc donné par :

$$b_j = \widehat{M(l, j, 0)} = B [j + 1, 1] E \widehat{[X_{j+1, j+1}^l]}$$

$$b_j = B [j + 1, 1] \sum_{i=0}^{n-j-1} x_{n-i} \frac{C_{n-i-1}^j}{C_n^{j+1}}$$

$$\begin{aligned}
 b_j &= \frac{1}{j+1} \sum_{i=0}^{n-j-1} x_{n-i} \frac{C_{n-i-1}^j}{C_n^{j+1}} \\
 b_j &= \frac{1}{n} \sum_{i=0}^{n-j-1} x_{n-i} \frac{(n-i-1)!}{(n-i-1-j)!} \frac{(n-j-1)!}{(n-1)!} \\
 b_j &= \frac{1}{n} \sum_{i=0}^{n-j-1} x_{n-i} \frac{(n-i-j)(n-i-j+1)\dots(n-i-1)}{(n-j)(n-j+1)\dots(n-1)} \\
 b_j &= \frac{1}{n} \sum_{i=1}^n x_i \frac{(i-1)(i-2)\dots(i-j)}{(n-1)(n-2)\dots(n-j)}
 \end{aligned}$$

Cette démonstration s'appuie sur celle proposée par LANDWEHR *et al.* (1979), relative à la détermination d'un estimateur non biaisé du moment de probabilité pondéré $M(1, 0, k)$.