

# Variabilité climatique et statistiques. Étude par simulation de la puissance et de la robustesse de quelques tests utilisés pour vérifier l'homogénéité de chroniques

Climatic variability and statistics.  
A simulation approach for estimating power  
and robustness of tests of stationarity

H. LUBÈS-NIEL<sup>1\*</sup>, J.M. MASSON<sup>2</sup>, J.E. PATUREL<sup>3</sup>, E. SERVAT<sup>3</sup>

Reçu le 20 mars 1997, accepté le 4 février 1998\*\*.

## SUMMARY

Statistical analysis of hydrometeorological time series is often used to identify climatic variations. Most often this analysis consists in applying and interpreting statistical tests of time series homogeneity. Hydrological time series (rainfall and runoff data) are often short and do not always comply with the hypotheses of the statistical methods. Through simulation we have investigated the power and the robustness of some tests which are widely used in the studies dealing with climatic variability. In each case studied, one hundred samples of fifty elements have been generated based on the main characteristics of natural rainfall series. A shift in the mean has been used to represent a possible climatic variation. The procedures used are the rank correlation test, Pettitt's test, Buishand's test, Lee and Heghinian's bayesian procedure, and Hubert and Carbonnel's segmentation procedure for hydrometeorological series.

Each simulation of one hundred samples is used to assess the performances of the methods considering a specific characteristic of the series: normality or non-normality, autocorrelation, trend, shift in the variance. First of all, stationary series have been simulated to evaluate the type I error of the tests. Then the series have been simulated with a break in the mean with different levels of amplitude, from 25% to 100% of the standard deviation value. The rank correlation test, Pettitt's test, Buishand's test and the segmentation procedure with a significance level of 1% (significance level of Scheffé's test) reject as heterogeneous less than ten series over one hundred homogeneous simulated series. This result is consistent with the type I error of a statistical test. On the other hand, Lee and Heghinian's bayesian method rejects about 40% of the series. This result means that this latter procedure must only be applied under

1. Laboratoire d'Hydrologie, ORSTOM, BP 5045, 34032 Montpellier cedex 1, France.

2. Laboratoire Géofluides-Bassins-Eau, UMR-CNRS 5569, Université Montpellier II, Place Eugène Bataillon, 34095 Montpellier cedex 5, France.

3. Antenne hydrologique ORSTOM, BP 1203 Cidex 1, Abidjan 06, Côte d'Ivoire.

\* Correspondance.

\*\* Les commentaires seront reçus jusqu'au 30 avril 1999.

the hypothesis of heterogeneity. The estimated power of the methods exceeds from 40% to 50% when the break in the mean is more than 75% of the standard deviation value.

Independent series have been simulated from normal, log-normal and Pearson distributions to compare the performances of the methods requiring the normality. The results show that normality has no significant impact on the performances of these methods. However, the simulations do show that the condition of independence of the successive elements of the series is essential to keep performances constant. Otherwise a trend in the series makes the tests inefficient, except for the rank correlation test for which the alternative is a trend. No method seems to be robust against both negative and positive autoregressive dependencies. The procedures requiring a constant variance are robust when the series keep a constant mean, but seem more or less slightly influenced by a break both in the mean and in the standard deviation.

**Key-words:** *time series, statistical tests, homogeneity, power, robustness.*

## RÉSUMÉ

L'analyse statistique de séries chronologiques de données hydrométéorologiques est un des outils d'identification de variations climatiques. Cette analyse consiste le plus souvent à la mise en œuvre et à l'interprétation de tests statistiques d'homogénéité des séries.

Les séries hydrologiques (données de pluie ou de débit) se caractérisent fréquemment par des effectifs faibles, et ne répondent que rarement aux conditions requises par l'application des tests statistiques dont certains sont paramétriques.

Nous avons cherché à évaluer, en terme de puissance et de robustesse, le comportement de quelques méthodes statistiques largement employées dans les études de variabilité climatique. Ce travail a été mené dans chaque cas étudié au moyen de procédures de simulations type Monte-Carlo de 100 échantillons de 50 valeurs conformes aux caractéristiques souvent rencontrées dans les séries naturelles.

La variabilité simulée est celle d'un changement brutal de la moyenne.

Les procédures concernées sont le test de corrélation sur le rang, le test de Pettitt, le test de Buishand, la procédure bayésienne de Lee et Heghinian, et la procédure de segmentation des séries hydrométéorologiques de Hubert et Carbonnel.

Des séries artificielles soit stationnaires, soit affectées par une rupture de la moyenne, normales, non-normales, autocorrélées, présentant une tendance linéaire ou un changement brutal de la variance ont été générées.

Les conclusions de ce travail doivent être nuancées selon la méthode considérée. D'une manière générale la puissance maximale estimée se situe autour de 50 % pour des taux de rupture de la moyenne de l'ordre de 75 % de la valeur de l'écart type. Par ailleurs il apparaît que l'autocorrélation et la présence d'une tendance dans les séries sont les deux caractéristiques qui pénalisent le plus les performances des procédures.

**Mots clés :** *séries temporelles, tests d'homogénéité, rupture en moyenne, puissance, robustesse.*

## 1 – INTRODUCTION

L'étude de la stationnarité de séries hydrométéorologiques est une des tâches incontournables des hydrologues (BERNIER, 1977), et depuis plusieurs années elle cherche à répondre aux questions qui relèvent de la caractérisation d'une variabilité climatique et de ses effets sur les ressources en eau. L'Organisation Mondiale de la Météorologie dans sa note technique numéro 79 « Climatic Change » (WMO, 1966), préconise ainsi la mise en œuvre de tests statistiques particuliers pour analyser la stationnarité des séries temporelles de nature hydro-météorologique.

Nous nous sommes intéressés de plus près à ces tests statistiques dans le cadre du programme ICCARE (Identification et Conséquences d'une variabilité du Climat en AfRIque de l'ouest non sahélienne) (SERVAT, 1994) dont la finalité est d'identifier des variations climatiques en Afrique de l'ouest et centrale non sahélienne, et d'appréhender leurs conséquences sur les ressources en eau.

L'Afrique de l'ouest sahélienne a connu autour des années 1970 une sécheresse tragique. L'hypothèse d'ICCARE se fonde sur cette observation pour arguer depuis la même période d'une baisse sensible des ressources en eau plus au sud en Afrique de l'ouest non sahélienne. Plusieurs études ont déjà validé cette hypothèse (SUTCLIFFE et KNOTT, 1987 ; NICHOLSON *et al.*, 1988 ; MAHÉ et OLIVRY, 1991), mais le but du programme ICCARE est de décrire plus finement la variation survenue dans les régimes pluviométriques et d'écoulement. Plusieurs hypothèses spécifiques ont ainsi été émises dont : décalage dans le temps des saisons des pluies, modification de la répartition des quantités précipitées dans l'année, diminution effective des événements importants, diminution du nombre de jours de pluie, sur l'année, la saison, décalage des crues dans le temps, modification des caractéristiques des hautes eaux, modification des phénomènes de décrue, renforcement des étiages, modification des réponses des bassins versants. Des séries chronologiques au pas de temps annuel portant sur des variables caractéristiques des différents aspects à traiter ont été constituées pour être soumises à une étude de stationnarité. Pour réaliser ce travail, une synthèse bibliographique (LUBÈS *et al.*, 1994) a d'abord permis de sélectionner des méthodes statistiques souvent utilisées dans les études de stationnarité de séries hydrométéorologiques, et de préférence déjà éprouvées dans le contexte africain.

Cinq tests ou procédures statistiques ont été retenus. Leur choix peut être controversé mais ce n'est pas l'objet du présent papier que d'en discuter dans la mesure où ils ont déjà été mis en œuvre dans ICCARE. Leur utilisation sur des séries chronologiques de données de pluie et de débit en de nombreux points de mesure de l'Afrique de l'ouest non sahélienne (du Sénégal à l'ouest à la République de Centrafrique à l'est) a permis de caractériser dans le temps et dans l'espace des changements significatifs de moyenne autour des années 1970, en phase donc avec la sécheresse sévissant en zone sahélienne (AKA *et al.*, 1996 ; SERVAT *et al.*, 1997 ; PATUREL *et al.*, 1997). Appliqués à chaque site de mesure, ces tests ont donné des résultats généralement concordants du moins au niveau de la reconnaissance d'une hétérogénéité dans la série, même si les estimations des ruptures sur la moyenne des séries données par plusieurs procédures ont différé parfois de quelques années. Toutefois les différentes méthodes ont quelquefois débouché sur des conclusions contradictoires sur une même série de

données. Par ailleurs l'interprétation spatio-temporelle des résultats s'est heurtée aux comportements particuliers de certaines stations situées dans des zones présentant en apparence un caractère uniforme non concordant – « stations à séries atypiquement stationnaires ou non stationnaires ». Les causes de nature « physique » (topographique, pluviométrique, appareillage) qui ont d'abord été recherchées n'ont généralement pas été jugées explicatives. Aussi il nous a semblé naturel de remettre en question les conclusions énoncées par les procédures statistiques mises en œuvre, ce qui nous a conduit à nous intéresser à leur puissance et à leur robustesse. Toutefois nous n'avons pas trouvé dans la littérature des développements à ce propos, et l'étude analytique de ces critères de qualité étant difficile, nous avons essayé de les apprécier en ayant recours à des simulations type Monte-Carlo dont le principe et les résultats font l'objet du présent article.

## 2 – LES MÉTHODES STATISTIQUES UTILISÉES (AKA *et al.*, 1996)

### 2.1 Présentation des méthodes

$X_i$  ( $i = 1, \dots, N$ ) désigne la série chronologique étudiée.

#### Le test de corrélation sur le rang

(KENDALL et STUART, 1943 ; WMO, 1966 ; OLANIRAN, 1991)

Le test de corrélation sur le rang est basé sur le calcul du nombre  $P$  de paires  $(X_i, X_j)$  pour lequel  $X_j > X_i$  ( $j > i$ ,  $i = 1, \dots, N - 1$ ). Sous l'hypothèse nulle ( $H_0$ ) de stationnarité de la série, la variable  $\omega$  définie par :

$$\omega = \frac{4P}{N(N-1)} - 1$$

suit une distribution normale de moyenne nulle et de variance égale à :

$$\sigma_t^2 = \frac{2(2N+5)}{9N(N-1)}.$$

Pour un risque  $\alpha$  de première espèce donné, l'acceptation de  $H_0$  est définie par l'appartenance de  $\omega$  à l'intervalle  $[-U_{1-\alpha/2} \sigma_t, U_{1-\alpha/2} \sigma_t]$ ,  $U_{1-\alpha/2}$  désignant la valeur de la variable normale réduite de probabilité de non dépassement  $1-\alpha/2$ .

L'hypothèse alternative de ce test est celle d'une tendance.

Les autres procédures statistiques retenues dans le cadre du programme ICCARE pour analyser des séries chronologiques de données de pluie et de débit sont des méthodes de détection de rupture. Une rupture peut être définie de façon générale par un changement dans la loi de probabilité d'une série chronologique à un instant donné le plus souvent inconnu.

**Le test de Pettitt** (PETTITT, 1979 ; DEMAREE, 1990 ; SUTHERLAND *et al.*, 1991 ; VANNITSEM et DEMAREE, 1991)

Pour décrire le test de Pettitt, des extraits du papier original de Pettitt sont repris ici. Pettitt considère une séquence de variables aléatoires indépendantes

$X_1, X_2, \dots, X_N$ . La séquence est censée contenir un point de rupture à  $\tau$  si les  $X_t$  pour  $t = 1, \dots, \tau$  ont une distribution commune  $F_1(X)$ , et les  $X_t$  pour  $t = \tau + 1, \dots, N$  ont une distribution commune  $F_2(X)$ , différente de  $F_1(X)$ . L'hypothèse nulle de « non-rupture »,  $H_0 : \tau = N$  contre l'hypothèse alternative de « rupture »,  $H_1 : 1 \leq \tau < N$ , est testée au moyen d'un test statistique non paramétrique. Aucune condition particulière n'est requise pour les formes fonctionnelles de  $F_1$  et  $F_2$  exceptée la continuité.

Pettitt montre comment une formulation appropriée du test de Mann-Whitney (DAGNÉLIE, 1970) peut être utilisée pour tester  $H_0$  contre  $H_1$ .

Si  $D_{ij} = \text{sgn}(X_i - X_j)$  où  $\text{sgn}(X) = 1$  si  $X > 0$ ;  $0$  si  $X = 0$  et  $-1$  si  $X < 0$ , alors, la variable

$$U_{t,N} = \sum_{i=1}^t \sum_{j=t+1}^N D_{ij}$$

est équivalente à la statistique de Mann-Whitney pour tester l'appartenance des deux échantillons,  $X_1, \dots, X_t$  et  $X_{t+1}, \dots, X_N$  à la même population.

La statistique  $U_{t,N}$  est considérée pour les valeurs de  $t$  comprise entre 1 et  $N$ . Pour tester  $H_0$  contre  $H_1$ , Pettitt propose d'utiliser la variable

$$K_N = \max |U_{t,N}|$$

En utilisant la théorie des rangs, Pettitt donne la probabilité de dépassement approximative d'une valeur  $k$  par :

$$\text{Prob}(K_N > k) \sim 2\exp(-6k^2/(N^3+N^2))$$

Pour un risque  $\alpha$  de première espèce donné,  $H_0$  est rejetée si cette probabilité est inférieure à  $\alpha$ . Dans ce cas, la série présente une rupture au temps  $t = \tau$  définissant  $K_N$ .

Le test est plus particulièrement sensible à un changement de moyenne.

### La statistique de Buishand (BUISHAND, 1982, 1984)

La statistique de Buishand est dérivée d'une formulation originale donnée par GARDNER (1969). La statistique de Gardner utilisée pour un test bilatéral de rupture en moyenne à un instant inconnu s'écrit :

$$G = \sum_{k=1}^{N-1} P_k \{S_k / \sigma_x\}^2 \text{ avec } S_k = \sum_{i=1}^k (X_i - \bar{X})$$

$P_k$  désigne la probabilité *a priori* que la rupture survienne juste après la  $k^{\text{ième}}$  observation.

Cette formulation suppose que la variance  $\sigma_x^2$  est connue. Si elle est inconnue, elle peut être remplacée par la variance de l'échantillon  $D_x^2$  et si  $P_k$  est choisie uniforme, on obtient finalement la statistique  $U$  définie par

$$U = \frac{\sum_{k=1}^{N-1} (S_k / D_x)^2}{N(N+1)} \text{ avec } D_x^2 = \sum_{i=1}^N (X_i - \bar{X})^2 / N$$

Des valeurs critiques de la statistique U ont été d'abord données par BUIHAND (1982) à partir d'une procédure de Monte Carlo. De meilleures estimations sont parues ultérieurement (BUIHAND, 1984).

### La méthode bayésienne de Lee et Heghinian

(LEE et HEGHINIAN, 1977 ; BRUNEAU et RASSAM, 1983 ; CHAUCHE, 1988)

La méthode bayésienne de Lee et Heghinian ne s'exprime pas comme un test statistique classique. Toutefois son interprétation vise à confirmer ou à infirmer l'hypothèse d'un changement de moyenne dans la série. Il s'agit d'une approche paramétrique qui requiert une distribution normale des variables étudiées.

Le modèle de base de la procédure est le suivant :

$$X_i = \begin{cases} \mu + \varepsilon_i & i = 1, \dots, \tau \\ \mu + \delta + \varepsilon_i & i = \tau + 1, \dots, N \end{cases}$$

où les  $\varepsilon_i$  sont indépendants et normalement distribués, de moyenne nulle et de variance  $\sigma^2 \cdot \tau$  et  $\delta$  représentent respectivement la position de la rupture dans le temps et l'amplitude du changement sur la moyenne.

L'approche bayésienne est basée sur les distributions marginales *a posteriori* de  $\tau$  et  $\delta$ . L'étude est ici limitée à la distribution *a posteriori* de  $\tau$ .

Si  $p$ , la densité de probabilité *a priori* de  $\tau$  est uniforme, *a posteriori* elle est définie par :

$$p(\tau|x) \propto [N/(\tau(N-\tau))]^{1/2} [R(\tau)]^{-(N-2)/2}, \quad 0 \leq \tau \leq N-1 \text{ avec}$$

$$R(\tau) = \left[ \sum_{i=1}^{\tau} (x_i - \bar{x}_{\tau})^2 + \sum_{i=\tau+1}^N (x_i - \bar{x}_{N-\tau})^2 \right] / \sum_{i=1}^N (x_i - \bar{x}_N)^2,$$

$$\bar{x}_N = 1/N \sum_{i=1}^N x_i, \quad \bar{x}_{\tau} = 1/\tau \sum_{i=1}^{\tau} x_i, \quad \bar{x}_{N-\tau} = 1/(N-\tau) \sum_{i=\tau+1}^N x_i.$$

Quand la distribution est unimodale, l'instant de la rupture est estimé par le mode de la distribution *a posteriori* de  $\tau$ . L'estimation est d'autant plus précise que la dispersion de la distribution est faible. Pour les résultats qui seront présentés ici, nous avons retenu comme estimation du point de rupture, soit le mode de la distribution si celle-ci est unimodale, soit le mode maximum dans le cas d'une distribution multi-modale, les modes secondaires constituant des signaux non examinés.

### La procédure de segmentation de séries hydrométéorologiques

(HUBERT et CARBONNEL, 1987 ; HUBERT *et al.*, 1989 ;

HUBERT et CARBONNEL, 1993)

La procédure de segmentation de séries hydrométéorologiques n'est pas, en toute rigueur, un test statistique. Cette méthode est adaptée à la recherche de multiples changements de moyenne dans la série. Son principe est de « découper » la série en  $m$  segments ( $m > 1$ ) de telle sorte que la moyenne calculée sur tout segment soit significativement différente de la moyenne du (ou des) segment(s) voisin(s).

Toute série  $X_i$ ,  $i = i_1, i_2$  avec  $i_1 \geq 1$  et  $i_2 \leq N$  où  $i_1 < i_2$ , constitue un segment de la série initiale des  $(X_i)$ . Toute partition de la série initiale en  $m$  segments est une segmentation d'ordre  $m$  de cette série. À partir d'une segmentation particulière d'ordre  $m$  pratiquée sur la série initiale, on définit :

$i_k$ ,  $k = 1, 2, \dots, m$ , le rang dans la série initiale de l'extrémité terminale du  $k^{\text{ième}}$  segment ;

$\bar{x}_k$  la moyenne du  $k^{\text{ième}}$  segment ;

$D_m$ , l'écart quadratique entre la série et la segmentation considérée,

$$D_m = \sum_{k=1}^{k=m} d_k \text{ avec } d_k = \sum_{i=i_{k-1}+1}^{i=i_k} (x_i - \bar{x}_k)^2.$$

La segmentation retenue au terme de la mise en œuvre de la procédure doit être telle que pour un ordre  $m$  de segmentation donné, l'écart quadratique  $D_m$  soit minimum et que les moyennes de deux segments contigus soient significativement différentes. Cette dernière contrainte est satisfaite par application du test de Scheffé (KENDALL et STUART, 1943) qui repose sur le concept de contraste (DAGNÉLIE, 1970).

D'après les auteurs, cette procédure de segmentation peut être considérée comme un test de stationnarité. Si la procédure ne produit pas de segmentation acceptable d'ordre supérieur ou égal à 2, l'hypothèse de stationnarité de la série est acceptée. Toutefois aucun niveau de signification n'est attribué à ce pseudo-test.

## 2.2 Les conditions d'application des méthodes

Les procédures statistiques qui sont étudiées regroupent des tests statistiques classiques, une méthode bayésienne et une technique de segmentation de séries chronologiques.

Leurs conditions d'application sont ici précisées.

### Les tests classiques

Le test de corrélation sur le rang ne suppose aucune propriété particulière de la série chronologique étudiée. Le test de Pettitt s'applique à des séries non auto-corrélées et requiert implicitement que la variance de la série ne soit pas affectée par la rupture si une rupture en moyenne est prioritairement recherchée. La statistique de Buishand repose sur l'ensemble des hypothèses suivantes : normalité de la série, égalité des variances des distributions de part et d'autre du point de rupture, absence d'autocorrélation.

### Procédure bayésienne

La procédure bayésienne de Lee et Heghinian impose normalité, non-auto-corrélation et constance de la variance.

### Segmentation

La segmentation qui fait intervenir le test de Scheffé sous-entend implicitement la normalité de la série chronologique.

### 3 – MÉTHODOLOGIE

#### 3.1 Puissance et robustesse des méthodes

L'étude présentée ne cherche pas à comparer les performances des différentes méthodes présentées en terme d'estimation de puissance et de robustesse pour choisir « la meilleure » procédure parmi plusieurs. En effet, d'une part ces méthodes ne sont pas équivalentes puisqu'elles ne produisent pas les mêmes éléments de réponse à la question de la stationnarité (soit simple acceptation ou rejet de la stationnarité, soit estimation d'une ou plusieurs dates de rupture, soit interprétation d'une densité de probabilité), d'autre part elles ont déjà toutes été mises en œuvre dans le programme ICCARE. L'intérêt du présent travail est d'apprécier simplement, dans des conditions contrôlées, la valeur des conclusions auxquelles elles conduisent.

L'étude s'attache donc à donner des estimations de la puissance et de la robustesse de chacune de ces procédures. Nous ferons référence à DAGNÉLIE (1970) pour définir ces deux notions.

La puissance désigne la probabilité de rejeter l'hypothèse nulle alors qu'elle est fausse, et cette probabilité est bien évidemment fonction du degré de fausseté de l'hypothèse nulle. Les conditions d'application des méthodes étant respectées, la relation entre la probabilité de rejet et le degré de fausseté de l'hypothèse nulle sera estimée.

La robustesse est la propriété d'un test dont le résultat s'avère être indépendant du fait que ses conditions théoriques d'application sont ou ne sont pas vérifiées.

Complémentaire de l'étude de puissance, l'analyse de robustesse peut être d'un apport particulièrement apprécié par l'utilisateur hydrologue qui le plus souvent rencontre des difficultés (temps, mise en œuvre...) pour vérifier la validité des conditions d'application d'une méthode, d'autant que des incertitudes diverses demeurent par ailleurs (qualité, suffisance des données...). Ce travail répond donc au souci de se placer dans des conditions quasi-réelles, les séries chronologiques naturelles, données de pluie (hauteurs de pluie sur différents pas de temps), de débits (maxima instantanés, valeurs moyennées sur différents pas de temps), mettant souvent en défaut les conditions d'indépendance, de normalité et de constance de la variance sans que cela soit toujours et rigoureusement vérifié.

Selon la méthode analysée, nous allons donc juger de l'influence de la non normalité des séries, de la présence d'une autocorrélation, d'une tendance ou d'un changement de la variance.

#### 3.2 Conditions générales des simulations

Puissance et robustesse sont étudiées à partir de séries chronologiques artificielles générées par des méthodes type Monte-Carlo et soumises aux différentes procédures statistiques examinées.

Tous les échantillons simulés ont un effectif de 50, ce qui correspond en général à la taille des séries hydrologiques analysées.

Comme une moyenne de 1 500 mm et un coefficient de variation de 0,2 correspondent aux valeurs observées sur des séries réelles de hauteurs annuelles de précipitations en Côte d'Ivoire entre 1950 et 1990, valeurs par ailleurs répan-



dues en Afrique de l'ouest non sahélienne, les séries stationnaires sont simulées à partir de variables ayant une espérance mathématique de 1 500 et un écart type de 300. Nous ne prétendons pas garantir la généralisation des résultats à toute combinaison de valeurs d'espérance mathématique et d'écart type, mais nous ferons état ultérieurement (§ 6) de la non-influence des caractéristiques statistiques choisies sur les performances obtenues dans un cas précis.

Les résultats de l'étude sont bien sûr étroitement dépendants de la taille des échantillons, et ils ne devront donc pas être généralisés à d'autres effectifs de séries.

Les tests statistiques classiques, test de corrélation sur le rang, test de Buis-hand, et test de Pettitt sont mis en œuvre avec un risque de première espèce  $\alpha$  de 10 %, identique à celui qui a été retenu pour l'analyse des séries du programme ICCARE. Pour la segmentation nous avons choisi le niveau de signification du test de Scheffé de 1 % conseillé par HUBERT *et al.* (1989) ; une justification de ce choix sera donnée au § 4.3.

Sur les séries perturbées avec changement de moyenne et/ou d'écart type, les ruptures sont artificiellement introduites entre les individus 10 et 40, et seules les ruptures détectées par les procédures entre les individus 6 et 45 compris sont retenues. Hors de ces limites les détections sont jugées douteuses étant donné la petite taille d'une des sous-séries pour estimer correctement les grandeurs statistiques que sont moyenne et variance, et les séries sont donc considérées comme homogènes. Concernant les ruptures sur la moyenne, quatre taux de rupture de 5, 10, 15 et 20 % de la moyenne théorique de 1 500, représentant respectivement 25, 50, 75 et 100 % de la valeur de l'écart type, sont systématiquement choisis. Pour les ruptures sur la variance, les effets d'une augmentation de 10, 20 et 50 % de l'écart type sont analysés.

Pour les procédures donnant une estimation de la date de rupture (test de Pettitt, méthode bayésienne et procédure de segmentation) une rupture est dite correctement estimée si elle est située au plus à 4 positions autour du vrai point de rupture. Cette convention est arbitraire mais elle correspond à une incertitude de l'ordre de 4 ans, tout à fait acceptable sur les séries naturelles.

### 3.3 Méthodes de simulation

L'étude nécessite de générer des séries artificielles normales, non normales, autocorrélées et présentant une tendance linéaire. Les méthodes de simulation utilisées sont ici précisées.

#### Simulation de variables normales indépendantes

La méthode de Box et Muller (CLARKE, 1973 ; KOTTEGODA, 1980) a été retenue car elle permet de générer à chaque lancement de la procédure de simulation deux variables normales réduites indépendantes.

Si  $U$  désigne une variable uniforme entre 0 et 1, et  $V$  une variable exponentielle de paramètre 1 générée à partir d'une variable uniforme  $U_v$  entre 0 et 1,  $V = -\ln U_v$ ,  $U$  et  $V$  étant indépendantes, les variables  $W_1$  et  $W_2$  définies par :

$$W_1 = (2V)^{1/2} \cos 2\pi U$$

$$W_2 = (2V)^{1/2} \sin 2\pi U$$

sont deux variables normales réduites indépendantes.

Les variables  $N(0, 1)$  ainsi simulées sont transformées ensuite en variables  $N(1\ 500, 90\ 000)$ .

### Simulation de variables lognormales indépendantes

La fonction densité de probabilité de la loi lognormale est la suivante :

$$f(x) = \frac{1}{\beta x \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{\text{Log}(x) - \alpha}{\beta} \right)^2}$$

Des variables aléatoires normales réduites  $Y_i$  générées par la méthode de Box et Muller sont transformées en variables lognormales  $X_i$  par :

$$X_i = \exp(\beta Y_i + \alpha)$$

$\alpha$  et  $\beta$  étant les deux paramètres de la loi qui se déduisent des valeurs d'espérance mathématique et d'écart type de la population-mère. Pour obtenir la moyenne et l'écart type souhaités (respectivement 1 500 et 300), les deux paramètres de la loi lognormale conduisent à une valeur théorique du coefficient de dissymétrie de 0,608.

### Simulation de variables gamma indépendantes

Nous avons décidé de simuler des échantillons tirés d'une loi gamma de coefficient de dissymétrie  $\gamma_1 = 0,7$ , légèrement supérieur à celui des séries simulées selon une loi lognormale.

La méthode qui est ici décrite, due à WHITTAKER (1973) génère une variable distribuée selon une loi gamma à 1 paramètre compris entre 0 et 1. Elle est du type acceptation-rejet (RIPLEY, 1987 ; ROSS, 1997), et passe par la génération de variables distribuées selon une loi beta.

La distribution gamma à 1 paramètre et la distribution beta ont pour fonctions de répartition respectives :

$$F(x) = \int_0^x \frac{e^{-u} u^{p-1}}{\Gamma(p)} du, \quad p > 0 ; x \geq 0$$

et

$$F(x) = \int_0^x \frac{u^{p-1} (1-u)^{q-1}}{B(p, q)} du, \quad p, q > 0 ; 0 \leq x \leq 1$$

$U_1, U_2, U_3$  sont trois variables uniformes indépendantes comprises entre 0 et 1. On définit :

$$S_1 = U_1^{1/p}, S_2 = U_2^{1/(1-p)}, \text{ avec } p \text{ compris entre } 0 \text{ et } 1.$$

Si la condition :  $Z = S_1 + S_2 \leq 1$  est vérifiée, on définit les variables suivantes :

$$Y = S_1 / (S_1 + S_2), X_1 = -Y \ln U_3, X_2 = -(1 - Y) \ln U_3$$

WHITTAKER (1973) montre que  $Y$  est une variable beta de paramètre  $p$  (avec  $q = 1-p$ ), et que  $X_1$  et  $X_2$  sont des variables gamma indépendantes de paramètre respectivement  $p$  et  $1-p$ .

Cette procédure est utilisée pour générer ensuite des variables distribuées selon une loi gamma à 3 paramètres, les valeurs de moyenne, écart type et coefficient de dissymétrie choisies fixant les trois paramètres de la distribution.

### Simulation de séries autocorrélées

Les échantillons sont générés à partir de variables normales dépendantes. La dépendance est celle d'un modèle autorégressif d'ordre 1. Nous avons pu observer en effet sur des séries naturelles de pluie et surtout de débit (modules annuels) une dépendance de cet ordre.

Le modèle de génération s'écrit :

$$U_i = \rho U_{i-1} + U \sqrt{1 - \rho^2}$$

$U_1$  et  $U$  sont des variables normales réduites générées par la méthode de Box et Muller.  $\rho$  est le coefficient d'autocorrélation avec retard de 1. Les variables réduites dépendantes ainsi générées sont ensuite transformées en variables  $N(1\ 500, 90\ 000)$ .

### Simulation de séries avec une tendance linéaire sur la moyenne

Des variables normales réduites  $U_i$  sont simulées par la méthode de Box et Muller avant d'être converties en variables normales par  $X_i = U_i \sigma + \mu_i$ .  $\sigma$  a pour valeur 300 tandis que  $\mu_i$  évolue linéairement autour de la valeur  $\mu = 1\ 500$  en fonction de  $i$  et d'un point « perturbateur »  $j$  tiré aléatoirement dans une loi uniforme entre les individus 10 et 40. Le modèle de génération imposé est le suivant :

$$\begin{aligned} \mu_{j/2} &= \mu \\ \mu_{j/2+25} &= \mu - k * \mu \end{aligned}$$

$k$  correspond à une variation sur l'espérance mathématique de 50, 75 ou 100 % de la valeur de l'écart type.

Ces deux équations conduisent à la relation linéaire :

$$\mu_i = \mu + k * \mu * j/50 - k * \mu * i/25$$

La moyenne des espérances mathématiques des variables simulées a pour valeur :

$$\mu + k * \mu * (j - 51)/50 < \mu \text{ puisque } 10 \leq j \leq 40.$$

## 4 – ÉTUDE DE PUISSANCE

Les séries définies ci-après comprennent 50 valeurs, et les conclusions énoncées résultent dans chaque cas de l'analyse des résultats obtenus sur 100 séries simulées.

### 4.1 Les tests classiques

#### Le test de corrélation sur le rang

Aucune condition d'application n'est spécifiée pour ce test. Les simulations ont porté dans un premier temps sur des séries respectant l'hypothèse nulle de stationnarité puis dans un deuxième temps sur des séries rejetant cette hypothèse.

Les séries stationnaires sont des séries artificielles non autocorrélées suivant une loi normale. Les séries non-stationnaires correspondent à trois types de situation envisagés :

- Soit présence d'une rupture sur la moyenne : une diminution brutale de la moyenne a été simulée sur des séries générées à partir de variables normales indépendantes de même variance.

- Soit présence d'une tendance.

- Soit présence d'un changement brutal de variance : celui-ci a été simulé sur des séries générées à partir de variables normales indépendantes.

Les résultats obtenus montrent que la proportion de rejets de l'hypothèse nulle sur les séries stationnaires simulées égale à 14 % est de l'ordre du risque de première espèce de 10 % adopté.

En présence d'une rupture sur la moyenne des séries il apparaît que les performances du test augmentent avec l'amplitude de la rupture simulée. De l'ordre de 30 % pour une baisse brutale sur la moyenne de 50 % de l'écart type, le pourcentage de séries déclarées non stationnaires se situe autour de 80 % lorsque la moyenne subit un décrochement de 100 % de la valeur de l'écart type comme le montre la figure 1.

Sur les séries simulées avec une tendance linéaire, le test de corrélation sur le rang rejette l'hypothèse de stationnarité avec des taux de 61 %, 90 % et 99 % en fonction du degré de la tendance « mesuré » par  $k$  égal respectivement à 50 %, 75 % et 100 % de la valeur de l'écart type. Ces résultats tendent à montrer qu'il est justifié d'attribuer à ce test l'hypothèse alternative d'une tendance.

Le test n'est pas sensible par contre à un changement brutal de l'écart type puisque pour une rupture de ce type, quel que soit le taux de rupture variant de 10 % à 50 % de l'écart type, la proportion de séries déclarées non stationnaires se situe autour de 10 %.

Le test apparaît par conséquent puissant lorsque la non-stationnarité relève d'une tendance ou d'une rupture significative sur la moyenne.

### **Le test de Pettitt**

Pour estimer la proportion de rejets erronés de l'hypothèse nulle de « non-rupture », des séries sans rupture en moyenne ont été générées à partir de variables normales indépendantes de même variance.

La fonction puissance a été estimée en simulant par ailleurs deux types de séries :

- Avec rupture sur la moyenne : une diminution brutale de la moyenne a été simulée sur des séries générées à partir de variables normales indépendantes de même variance.

- Avec une rupture sur la variance simulée sur des séries issues de variables normales indépendantes de même moyenne.

Sur les séries stationnaires simulées, la proportion de rejets de l'hypothèse nulle égale à 2 % est par conséquent inférieure à la valeur du risque de première espèce choisie. Ce fait est peut-être imputable à la taille des échantillons simulés et au nombre de séries analysées. Mais il est raisonnable d'admettre aux fluctuations d'échantillonnage près que le respect du risque de première espèce est assuré.

En présence d'une rupture sur la moyenne, la proportion de rejets de l'hypothèse nulle associés à une estimation correcte de la date de rupture est faible tant que le taux de rupture ne dépasse pas 75 % de la valeur de l'écart type : cette proportion est de 16 % et 37 % pour une rupture respectivement de 50 % et 75 % de la valeur de l'écart type. Par contre dans 60 % des cas, la rupture est bien identifiée lorsque la moyenne diminue de 100 % de la valeur de l'écart type. Ces résultats sont portés sur la figure 1.

Les séries de variables normales indépendantes simulées avec rupture sur l'écart type mettent en évidence la faiblesse du test pour détecter ce type d'hétérogénéité. Pour une rupture représentant de 10 % à 50 % de la valeur de l'écart type, le pourcentage de rejets de l'hypothèse nulle de stationnarité avec une détection correcte de la date de rupture varie dans le même temps entre 0 et 5 %.

Il semble par conséquent que le test n'ait pas la même puissance selon la caractéristique statistique affectée par une rupture. Au vu des simulations, seule une rupture sur la moyenne peut être raisonnablement estimée.

### **Le test de Buishand**

L'évaluation du risque de première espèce repose sur la simulation de séries stationnaires générées selon des variables normales indépendantes de même variance.

L'hypothèse alternative du test étant un changement brutal de la moyenne, la fonction puissance est estimée en générant des séries à partir de variables normales indépendantes de même variance mais présentant une rupture de la moyenne à partir d'un individu choisi aléatoirement.

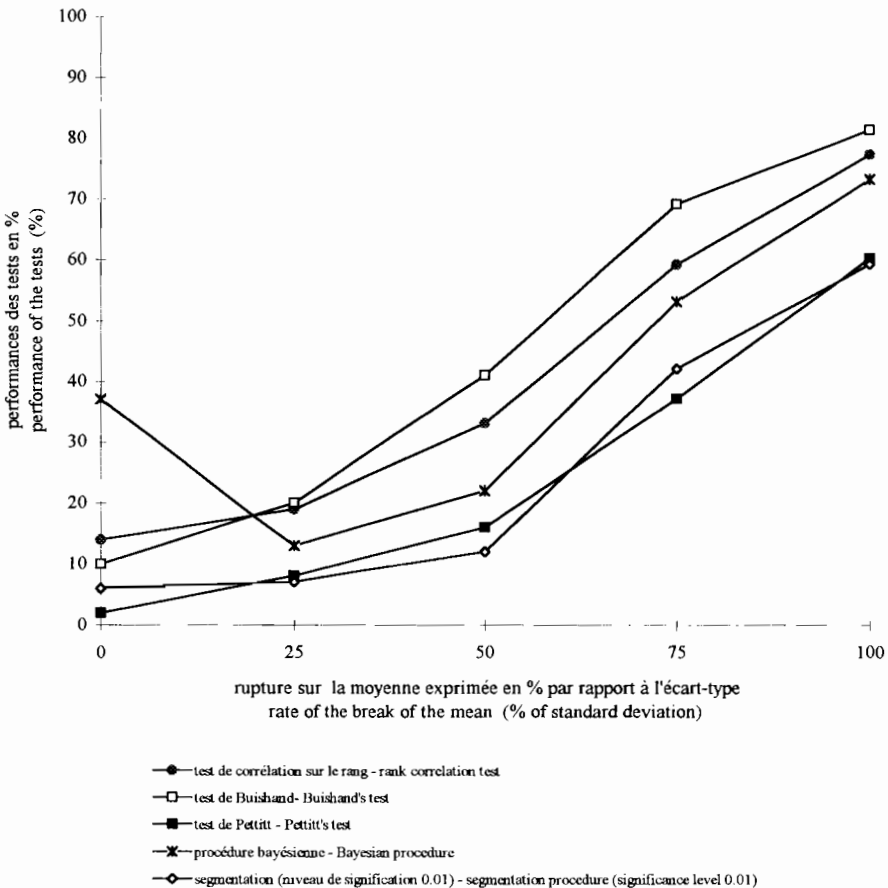
Le risque de première espèce est correctement estimé par la proportion de rejets de l'hypothèse nulle sur les séries stationnaires simulées. En présence d'une rupture sur la moyenne la proportion de séries déclarées avec rupture croît linéairement avec le taux de rupture (*figure 1*) : elle est de l'ordre de 20 %, 40 % et 80 % pour des baisses de moyenne de 25 %, 50 % et 100 % de la valeur de l'écart type respectivement. Le test apparaît donc relativement puissant pour des ruptures sur la moyenne d'amplitude même modérée.

## **4.2 La procédure bayésienne**

La procédure bayésienne de Lee et Heghinian n'est pas mise en œuvre comme un test classique visant à confirmer ou à infirmer une hypothèse nulle. Toutefois elle est assimilée à une démarche de prise de décision qui en l'occurrence consiste à déclarer la série stationnaire ou non par analyse de la densité de probabilité *a posteriori* de la position du point de rupture. Il est donc possible de mesurer les performances de la méthode sous les conditions d'application requises en considérant des séries stationnaires puis des séries présentant une rupture sur la moyenne avec différents niveaux de rupture. Toutes les séries sont générées à partir de variables normales indépendantes de même variance.

Nous conviendrons que la méthode bayésienne de Lee et Heghinian reconnaît effectivement la stationnarité d'une série stationnaire simulée lorsque la fonction densité de probabilité *a posteriori* du point de rupture a une forme en U.

Dans ces conditions il apparaît une proportion élevée de rejets de la stationnarité de 37 % sur les séries stationnaires simulées. Par contre en présence



**Figure 1** Fonctions puissance ( $\mu = 1\,500$ ,  $\sigma = 300$ ).  
*Power functions ( $\mu = 1\,500$ ,  $\sigma = 300$ ).*

d'une rupture simulée sur la moyenne des séries, la proportion de ruptures bien identifiées par la méthode est de l'ordre de 20 % et 70 % pour des baisses de la moyenne respectivement de 50 % et 100 % de la valeur de l'écart type comme l'illustre la figure 1. Toutefois il faut retenir le caractère subjectif de l'interprétation des fonctions de distribution *a posteriori* sur laquelle se fondent les résultats énoncés.

### 4.3 La segmentation

D'après les auteurs, la segmentation peut être considérée comme un test de stationnarité. Aussi sa puissance a été estimée en considérant des séries stationnaires puis des séries présentant une rupture sur la moyenne. Comme la condition de normalité est recommandée par le test de Scheffé, les séries ont été générées à partir de variables normales indépendantes de même variance. Sur les séries perturbées la méthode ne devrait identifier que la seule rupture

simulée ; néanmoins comme la probabilité que la méthode détecte plusieurs ruptures n'est pas nulle, nous ferons état du nombre de détections correctes données par la méthode issues de segmentations d'ordre 2 ou supérieur à 2.

Sur les séries stationnaires la procédure de segmentation partitionne de manière significative, au regard du test de Scheffé, 6 % des séries pour le niveau de signification 1 % du test. Nous avons constaté que plus de 40 % des séries sont partitionnées lorsque ce niveau de signification est de 5 %. Ces résultats sont conformes à ceux obtenus par HUBERT (HUBERT *et al.*, 1989) au cours d'une simulation de 100 séries stationnaires de 50 valeurs, avec une proportion de rejets de l'hypothèse nulle de 47 % et 11 % pour un niveau de signification respectif de 5 % et 1 % du test de Scheffé. Nous avons tenu à vérifier que ces résultats ne dépendent pas du nombre de séries simulées. L'analyse de 1 000 échantillons stationnaires a confirmé l'ordre de grandeur des pourcentages précédents puisque 53 % et 16 % des séries ont été déclarées non stationnaires pour les niveaux de signification respectifs de 5 % et 1 % du test de Scheffé.

En présence d'une rupture sur la moyenne des séries, les proportions de rejets de la stationnarité sont plus élevées lorsque le niveau de signification du test de Scheffé est de 5 % comparativement à celui de 1 %. Ainsi pour une rupture en moyenne de 100 % de la valeur de l'écart type, 60 % des séries sont déclarées non stationnaires lorsque le niveau de signification du test de Scheffé est de 1 % contre 70 % quand ce niveau de signification est de 5 %. Ces résultats, fonction du niveau de signification du test de Scheffé, peuvent surprendre. D'après HUBERT (1989), il faut distinguer nettement le risque de première espèce de la procédure de segmentation, du risque de première espèce du test de Scheffé, car quoique le premier soit dépendant du second, cette relation de dépendance n'est pas simple à exprimer. De la même façon il n'y a pas de liaison évidente entre le risque de deuxième espèce de la procédure et la valeur du risque de première espèce du test de Scheffé.

Comme annoncé précédemment, nous retiendrons le niveau de signification de 1 % du test de Scheffé conseillé par HUBERT *et al.* (1989) dans la suite de l'étude.

## 5 – ÉTUDE DE ROBUSTESSE

Comme dans l'étude précédente, les séries définies ci-après comprennent 50 valeurs, et les conclusions énoncées résultent dans chaque cas de l'analyse des résultats obtenus sur 100 séries simulées.

### 5.1 Les tests classiques

#### Le test de corrélation sur le rang

Dans la mesure où ce test est de nature non paramétrique, il n'est soumis à aucune condition particulière d'application qui pourrait être mise en défaut. Toutefois comme nous avons observé parfois une autocorrélation d'ordre 1 significative sur les séries réelles étudiées dans le cadre du programme ICCARE, nous avons

souhaité analyser le comportement du test vis-à-vis de séries normales autocorrélées.

Cette étude a été réalisée à partir de la simulation de 1 000 échantillons, et nous avons vérifié que les résultats énoncés étaient respectés par l'interprétation de 100 séries simulées, ce qui nous a permis de valider ce choix de référence.

Dans un premier temps nous avons étudié le cas de séries stationnaires. Pour des coefficients d'autocorrélation des séries simulées de  $-0,9$ ,  $-0,5$ ,  $-0,2$ ,  $0,2$ ,  $0,5$  et  $0,9$ , la proportion de rejets de l'hypothèse nulle de stationnarité est estimée respectivement à  $0\%$ ,  $0,7\%$ ,  $5\%$ ,  $17\%$ ,  $32\%$  et  $68\%$ . Ces résultats révèlent une différence de comportement du test selon le signe du coefficient d'autocorrélation. Les proportions de rejets de l'hypothèse nulle de stationnarité sont inférieures au risque de première espèce du test  $\alpha$  égal à  $10\%$  pour des valeurs négatives du coefficient d'autocorrélation, et supérieures à  $\alpha$  pour des valeurs positives de ce coefficient. Lorsque la mesure d'autocorrélation est supérieure à  $0,5$  le risque de première espèce est largement surestimé.

Une étude similaire a été menée sur des séries normales autocorrélées présentant une rupture de la moyenne. Des échantillons comportant un taux de rupture sur la moyenne de  $50\%$ ,  $75\%$  ou  $100\%$  de la valeur de l'écart type ont été simulés pour des coefficients d'autocorrélation de  $-0,9$ ,  $-0,5$ ,  $-0,2$ ,  $0$  (hypothèse d'indépendance),  $0,2$ ,  $0,5$ , et  $0,9$ . Le test est influencé différemment selon le signe du coefficient d'autocorrélation et le taux de rupture comme le montre la figure 2. Pour une rupture sur la moyenne de  $50\%$  de la valeur de l'écart type, le pourcentage de rejets de la stationnarité est inférieur à celui estimé sur les séries indépendantes lorsque le coefficient d'autocorrélation est négatif, et supérieur lorsque celui-ci est positif, de telle sorte que cette proportion de rejets croît de manière continue avec le coefficient d'autocorrélation. Au contraire pour un taux de rupture sur la moyenne de  $100\%$  de la valeur de l'écart type, le pourcentage de rejets est supérieur à celui estimé sur les séries indépendantes lorsque le coefficient d'autocorrélation est négatif, et inférieur quand il est positif. Dans ce cas-là la relation entre pourcentage de rejets de la stationnarité et coefficient d'autocorrélation est décroissante. Il faut souligner toutefois qu'une augmentation de l'autocorrélation pour des coefficients supérieurs à  $0,2$  n'a pas d'incidence significativement sensible sur la proportion de rejets de l'hypothèse nulle. Pour un taux de rupture sur la moyenne de  $75\%$  de la valeur de l'écart type, le pourcentage de rejets est quasiment stable pour toutes les valeurs du coefficient d'autocorrélation étudiées à l'exception de celle de  $0,9$  pour laquelle il connaît une augmentation de l'ordre de  $10\%$ .

Par ailleurs sur les séries présentant un coefficient d'autocorrélation de  $0,9$ , le pourcentage de rejets de l'hypothèse nulle est pratiquement le même quel que soit le taux de rupture de la moyenne considéré : il est de l'ordre de  $75\%$ .

Ce type de comportement a été observé également par STAUDTE et SHEATHER (1990) au cours d'une étude de robustesse du test  $t$  réalisée en présence d'une autocorrélation des observations. Ils montrent effectivement que les performances du test  $t$  varient en fonction du signe du coefficient d'autocorrélation, et ils notent que ce comportement a été observé pour d'autres tests « vérifiant » un changement de moyenne.

Le test de corrélation sur le rang est donc influencé par des observations dépendantes ; cette influence est différente selon le signe du coefficient d'autocorrélation et le taux de rupture.



## Le test de Pettitt

Le comportement du test dans des conditions d'application ne respectant plus indépendance et constance de la variance si une rupture sur la moyenne est recherchée est ici étudié. L'analyse est faite à partir de séries simulées soit avec une autocorrélation d'ordre 1, soit avec une tendance, soit avec un changement de variance.

En présence de séries stationnaires autocorrélées, le risque de première espèce estimé est d'autant plus élevé que l'autocorrélation est positive (coefficient d'autocorrélation positif) et forte. Il est estimé à 13 %, 44 %, et 92 % pour des coefficients d'autocorrélation respectifs de 0,2, 0,5 et 0,9. Il apparaît nettement une dégradation de la puissance du test. Lorsque le coefficient d'autocorrélation est négatif, le pourcentage de rejets de l'hypothèse nulle de stationnarité est inférieur ou égal à 1 %, ce qui pourrait laisser croire à une amélioration des performances du test, mais en fait traduit un mauvais fonctionnement.

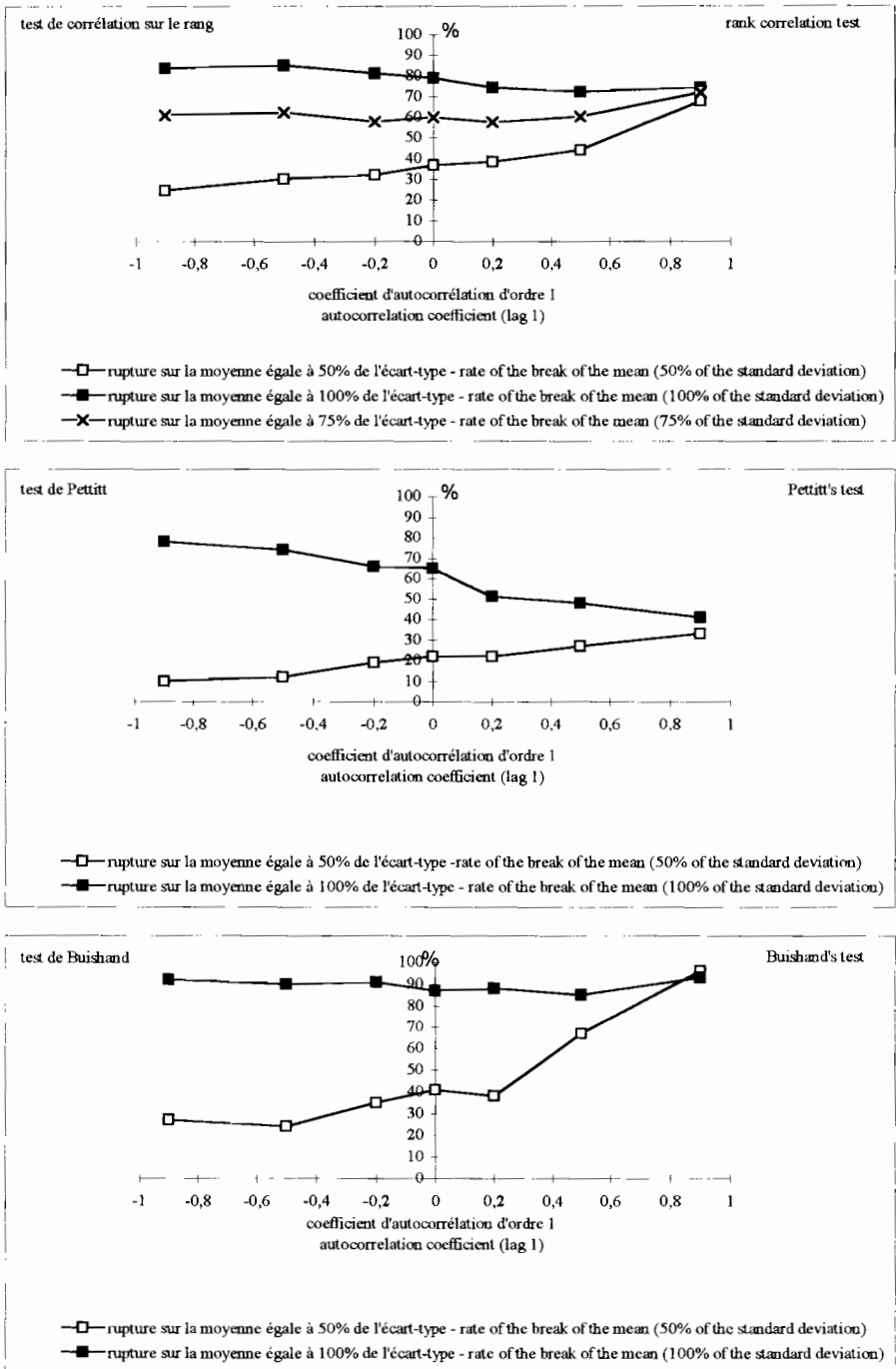
Lorsque l'autocorrélation est conjuguée à une rupture sur la moyenne, les observations concernant le test de corrélation sur le rang restent valides. Plus précisément le test de Pettitt affiche une proportion de ruptures correctement détectées qui passe de 10 % à 33 % pour un pourcentage de rupture sur la moyenne égal à 50 % de la valeur de l'écart type quand le coefficient d'autocorrélation varie de  $-0,9$  à  $0,9$  (figure 2). Lorsque le taux de rupture sur la moyenne est égal à 100 % de l'écart type, le pourcentage de ruptures bien détectées passe de 78 % à 41 % quand le coefficient d'autocorrélation varie de  $-0,9$  à  $0,9$ . Ce comportement est conforme à celui décrit par STAUDTE et SHEATHER (1990) à propos des tests de changement de moyenne.

Sur les séries simulées avec une tendance linéaire, le test de Pettitt déclare 57 %, 20 % et 4 % des séries stationnaires pour un degré de tendance respectivement de 50 %, 75 % et 100 % de la valeur de l'écart type. La proportion de rejets de l'hypothèse nulle de stationnarité est donc d'autant plus importante que le degré de tendance est élevé. Toutefois, l'hypothèse alternative du test étant celle d'une rupture d'une caractéristique de la population, le rejet de la stationnarité est interprété comme une rupture et non comme une tendance. Il en résulte une estimation incorrecte du modèle statistique de la variable chronologique étudiée.

Il conviendrait par conséquent d'éliminer toute forme de tendance des séries avant de les soumettre au test de rupture.

Des séries ont également été simulées à partir de variables normales indépendantes présentant une rupture sur la moyenne et une rupture sur la variance imposées à partir d'un même individu afin de contrôler l'effet d'un changement brutal de variance sur la détection d'une rupture sur la moyenne.

L'analyse d'une rupture sur la moyenne (de 50 % ou 100 % de la valeur de l'écart type) conjuguée à une rupture sur l'écart type montre que les performances du test, illustrées par la figure 5, restent pratiquement constantes quelle que soit la variation imposée à l'écart type, de 10 %, 20 % ou 50 %, pour un taux de rupture sur la moyenne donné. La proportion de rejets est du même ordre de grandeur que celle observée en présence d'une rupture seule sur la moyenne : autour de 15 % et 55 % pour des taux de rupture respectifs sur la moyenne de 50 % et 100 % de la valeur de l'écart type.



**Figure 2** Influence d'une autocorrélation d'ordre 1 sur les performances des tests classiques.  
*Influence of an autocorrelation of lag 1 on the performances of the standard tests.*

## Le test de Buishand

L'étude concerne les performances du test appliqué à des séries mettant en défaut les conditions d'application recommandées. Elle a pour but d'apprécier l'influence de la non-normalité, d'une autocorrélation d'ordre 1, d'une tendance ou d'un changement de variance sur les performances du test.

Des échantillons tirés d'une loi lognormale puis d'une loi gamma sans et avec rupture sur la moyenne ont été soumis au test. Les résultats obtenus ne sont pas significativement différents selon le type de loi dissymétrique imposé à la population. Ils ne sont pas non plus très différents de ceux obtenus avec les échantillons générés à partir d'une loi normale. La figure 4 illustre la robustesse du test vis-à-vis de la non-normalité des séries.

Sur les séries stationnaires normales autocorrélées le risque de première espèce estimé est inférieur à la valeur fixée de 10 % pour des coefficients d'autocorrélation négatifs. Par contre lorsque l'autocorrélation est positive la proportion de rejets de l'hypothèse nulle augmente avec le coefficient d'autocorrélation : elle est estimée à 22 %, 49 % et 95 % pour des valeurs respectives du coefficient d'autocorrélation de 0,2, 0,5 et 0,9. L'influence de l'autocorrélation sur les performances du test est donc manifeste.

Sur les séries simulées avec rupture sur la moyenne, comme pour les tests précédents le comportement du test de Buishand varie en fonction du taux de rupture appliqué, et du signe du coefficient d'autocorrélation. Pour un taux de rupture sur la moyenne de 50 % de la valeur de l'écart type, le pourcentage de ruptures correctement détectées passe de 27 % à 96 % lorsque le coefficient d'autocorrélation varie de  $-0,9$  à  $0,9$  (figure 2). Quand la rupture sur la moyenne est égale à 100 % de l'écart type, le taux de détections correctes reste relativement stable quelle que soit la valeur du coefficient d'autocorrélation entre  $-0,9$  et  $0,9$ .

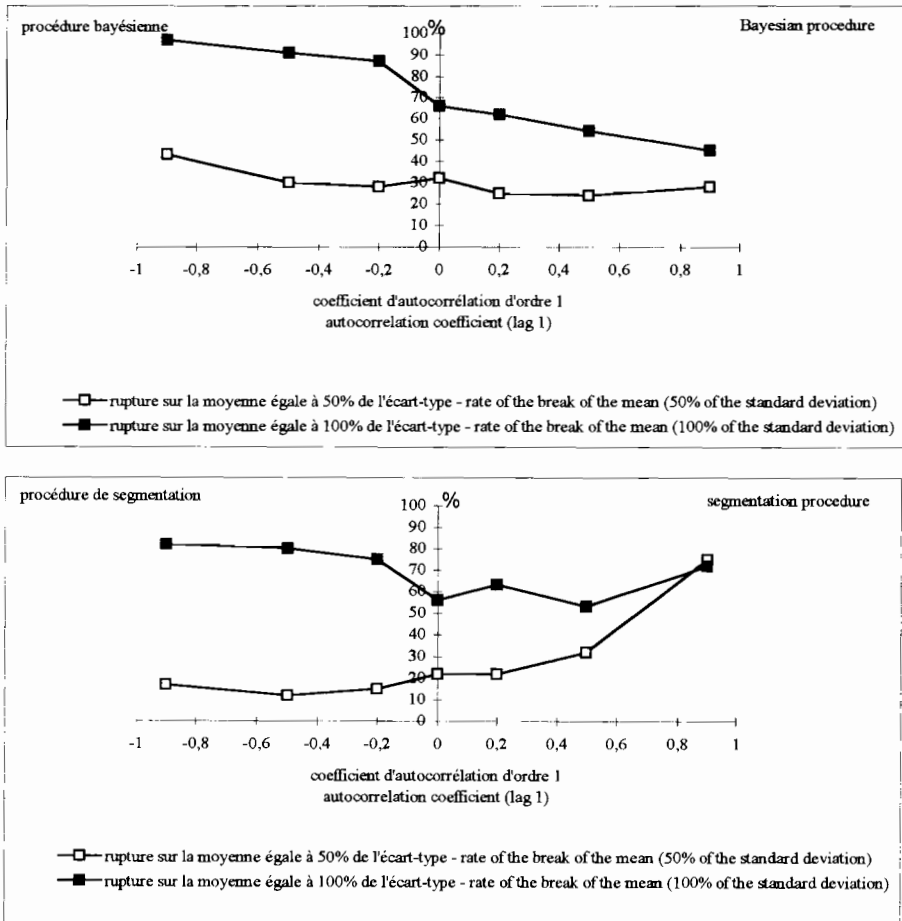
Le test de Buishand rejette l'hypothèse nulle de stationnarité sur les séries simulées présentant une tendance linéaire avec des pourcentages de 63 %, 92 % et 99 % en fonction du degré de tendance, respectivement de 50 %, 75 % et 100 % de la valeur de l'écart type. Le rejet de l'hypothèse nulle est correct, par contre l'hypothèse alternative retenue qui est celle d'une rupture brutale de la moyenne est ici erronée. Il apparaît donc nécessaire d'éliminer une tendance éventuelle sur les séries avant de recourir au test.

En présence d'un changement brutal de l'écart type, les performances du test évaluées sur des séries normales, non autocorrélées et stationnaires en moyenne, montrent que la proportion de rejets de l'hypothèse de stationnarité de la moyenne est de l'ordre du risque de première espèce fixé, pour les trois niveaux de rupture de l'écart type considérés de 10 %, 20 % et 50 %, ce qui traduit la robustesse du test dans ce cas.

Lorsqu'une rupture sur la moyenne est simulée simultanément à celle sur l'écart type, les pourcentages de détection de ruptures correctes sur la moyenne varient quand le taux de rupture sur l'écart type augmente (figure 5), et le sens de cette variation semble dépendre du taux de rupture de la moyenne.

## 5.2 La procédure bayésienne

L'étude de robustesse de la procédure bayésienne de Lee et Heghinian est plus délicate que celle réalisée sur les tests classiques du fait d'une difficulté d'interprétation de la densité de probabilité *a posteriori* produite par la méthode.



**Figure 3** Influence d'une autocorrélation d'ordre 1 sur les performances des procédures bayésienne et de segmentation.

*Influence of an autocorrelation of lag 1 on the performances of the Bayesian and segmentation procedures.*

L'influence de la non-normalité des séries, d'une autocorrélation des observations, d'une tendance linéaire et d'un changement de la variance des séries sur le comportement de la méthode a été analysée.

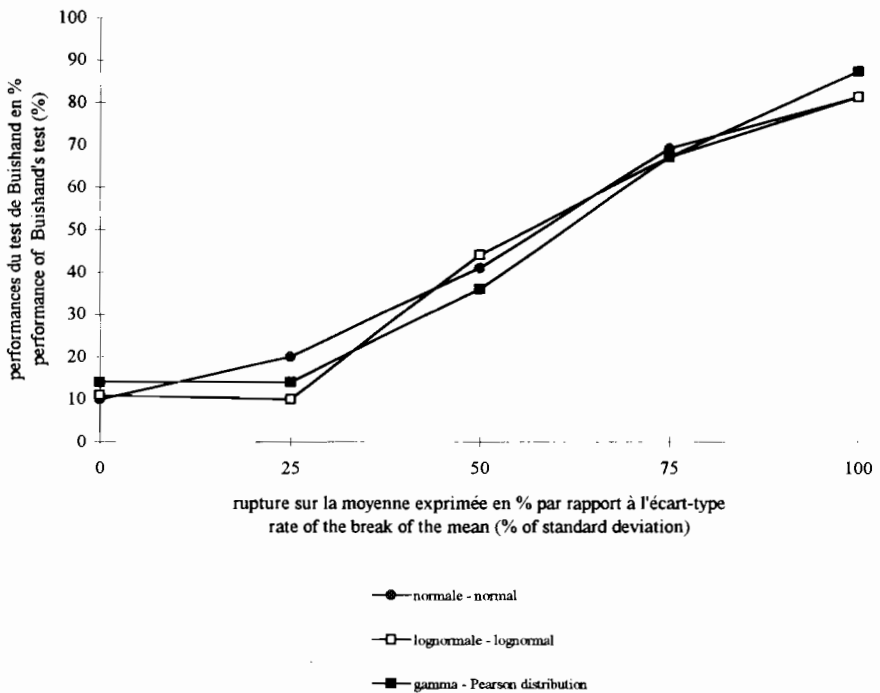
La présence d'une dissymétrie significative dans les séries simulées avec ou sans perturbation sur la moyenne n'est pas sensible sur les résultats de la méthode.

Des séries normales sans perturbation et présentant une autocorrélation du premier ordre ont également été générées. Le pourcentage de séries déclarées non stationnaires après interprétation de la fonction densité de probabilité *a posteriori* d'un point de rupture augmente avec le coefficient d'autocorrélation : il est de 0 % et de 95 % pour un coefficient d'autocorrélation respectivement de -0,9 et 0,9. Sur les séries simulées avec une rupture sur la moyenne égale à 50 % de

la valeur de l'écart type, l'effet de l'autocorrélation mesuré en pourcentage de détections correctes est sensiblement le même quelle que soit la valeur du coefficient d'autocorrélation supérieure à  $-0,9$  comme l'illustre la figure 3. Par contre lorsque le taux de rupture sur la moyenne est égal à 100 % de la valeur de l'écart type, la proportion de ruptures bien détectées passe de 97 % à 45 % quand le coefficient d'autocorrélation varie de  $-0,9$  à  $0,9$ . Il apparaît donc qu'une autocorrélation a une influence marquée sur le comportement du test quand les séries sont stationnaires ou présentent une rupture très significative sur la moyenne.

Sur les séries simulées avec une tendance linéaire, le pourcentage de détections erronées d'une rupture brutale sur la moyenne augmente avec le degré de la tendance. Il est de 71 %, 88 % et 97 % pour les degrés de tendance respectifs de 50 %, 75 % et 100 % de la valeur de l'écart type. Il est par conséquent fortement recommandé d'éliminer la tendance linéaire des séries avant de les soumettre à la procédure.

L'étude d'un changement brutal de l'écart type a permis de montrer sur des séries simulées sans perturbation de la moyenne que l'influence d'une variation brutale de l'écart type de 10 %, 20 % ou 50 % de sa valeur n'était pas sensible sur les performances du test. Le pourcentage de ruptures détectées à tort est du même ordre de grandeur que celui obtenu sur des séries stationnaires. En présence d'une rupture simultanée sur la moyenne et sur l'écart type, les performances diminuent sensiblement lorsque le taux de rupture sur l'écart type augmente (figure 5).



**Figure 4** Influence de la non normalité sur les performances du test de Buishand.  
*Influence of non-normality on the performances of Buishand's test.*

### 5.3 La segmentation

L'étude de robustesse de la procédure de segmentation ne concerne que l'influence de la non-normalité des séries sur la qualité des résultats, mais nous avons souhaité apprécier aussi le comportement de la méthode en présence de séries autocorrélées.

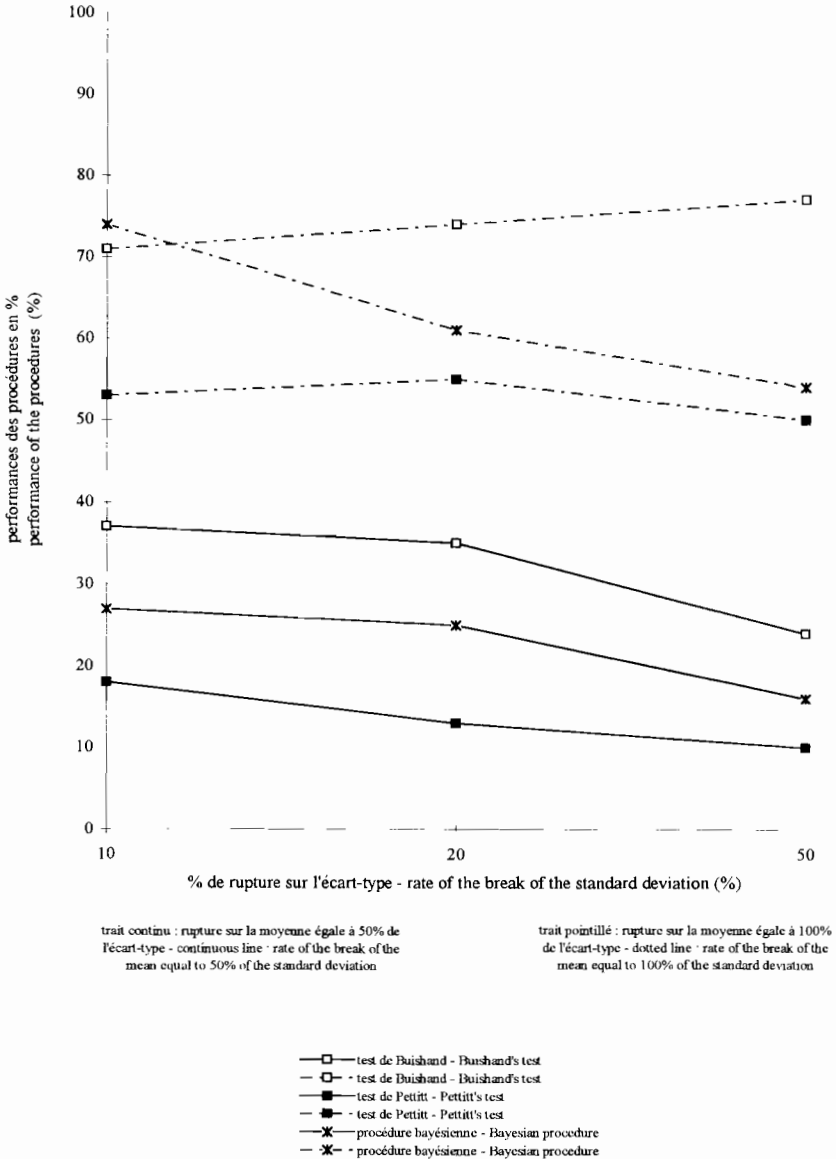


Figure 5

Influence d'une rupture de l'écart type sur les performances des tests.

*Influence of a break on the standard deviation in the performances of the tests.*

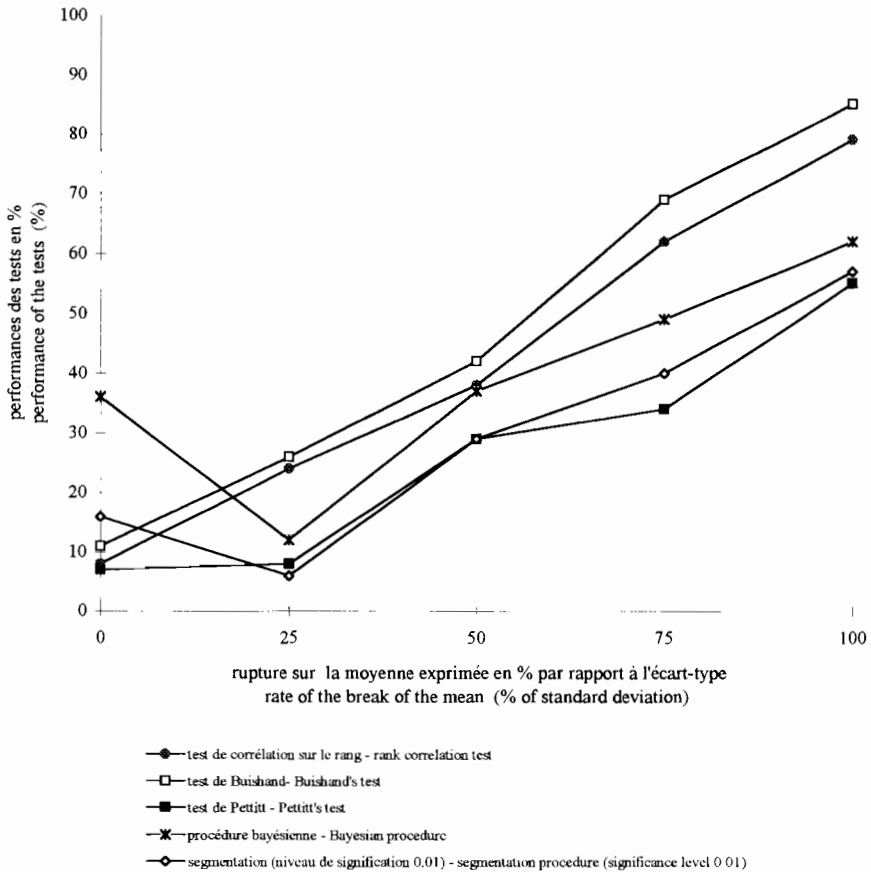
Avec un niveau de signification de 1 % du test de Scheffé recommandé par les auteurs, les séries normales ou non-normales conduisent aux mêmes performances.

Sur les séries stationnaires simulées autocorrélées la procédure de segmentation déclare les séries stationnaires dans 100 % des cas environ lorsque le coefficient d'autocorrélation est inférieur à 0,2, et seulement dans moins de 1 % des cas quand ce même coefficient est égal à 0,9. Sur les séries simulées avec rupture sur la moyenne, il apparaît avec un taux de rupture de 50 % de la valeur de l'écart type, que le pourcentage de ruptures correctement détectées (à partir de segmentations d'ordre 2 ou supérieur) par la méthode augmente avec la valeur du coefficient d'autocorrélation (*figure 3*). Avec un taux de rupture sur la moyenne égal à 100 % de la valeur de l'écart type, il semblerait que le comportement de la méthode soit moins net : la proportion de ruptures bien identifiées (toujours à partir de segmentations d'ordre 2 ou supérieur) diminue quand le coefficient d'autocorrélation varie de  $-0,9$  à  $0$  puis reste sensiblement voisine de 60 % pour les valeurs positives de ce même coefficient.

La présence d'une autocorrélation influence par conséquent les performances de la méthode.

## 6 – COMMENTAIRE GÉNÉRAL

Les résultats qui viennent d'être présentés pour rendre compte de la puissance et de la robustesse des méthodes retenues dans le programme ICCARE, ont été obtenus par l'analyse de séries simulées avec une espérance mathématique de 1 500 et un écart type de 300. Faut-il en conséquence faire dépendre les conclusions auxquelles nous ont conduits les diverses simulations à ce couple de caractéristiques statistiques, ou peut-on prétendre à une certaine généralisation ? Pour répondre à cette question, des séries de variables normales indépendantes stationnaires ou présentant une rupture sur la moyenne, ont été générées avec un coefficient de variation de 0,13, espérance mathématique de 2 600 et écart type de 340, correspondant aux ordres de grandeur de séries de totaux précipités annuels de stations camerounaises. Les fonctions puissance déduites de ces simulations sont représentées sur la figure 6. Leur allure est très voisine de celles estimées à partir des séries simulées avec un coefficient de variation de 0,2, portées sur la figure 1. Les différences de comportement qui sont sensibles pour un taux de rupture de la moyenne égal à 50 % de la valeur de l'écart type, sur le test de Pettitt, la procédure bayésienne et la méthode de segmentation, sont vraisemblablement dues aux fluctuations d'échantillonnage. Ces brefs essais ne prétendent pas valider la généralisation de tous les résultats qui ont été présentés, une étude complète devrait être spécialement réalisée dans ce but, mais ils justifient au moins de faire l'hypothèse d'une généralisation.



**Figure 6** Fonctions puissance ( $\mu = 2\,600$ ,  $\sigma = 340$ ).  
 Power functions ( $\mu = 2\,600$ ,  $\sigma = 340$ ).

## 7 – CONCLUSION

La simulation type Monte-Carlo a permis d'apprécier la puissance et la robustesse de méthodes statistiques recensées dans des études de variabilité climatique.

Pour les trois méthodes qui sont véritablement des tests statistiques, test de corrélation sur le rang, test de Buishand, test de Pettitt, nous avons d'abord vérifié sur 100 simulations de séries stationnaires, simples et aléatoires, que le nombre d'échantillons rejetant l'hypothèse nulle de stationnarité est conforme au risque de première espèce choisi. La méthode bayésienne déclare autour de 40 % des séries non stationnaires. Nous en déduisons que cette méthode ne doit être utilisée que sur des séries réelles pour lesquelles l'hypothèse *a priori* d'une rupture est forte. Pour la procédure de segmentation, lorsque le niveau de signification du test de Scheffé est égal 1 %, le pourcentage de séries rejetées inférieur à 10 %



est alors conforme à un pseudo risque de première espèce acceptable pour la procédure.

Pour l'ensemble des méthodes, les séries qui ont été simulées en retenant l'hypothèse alternative d'une rupture sur la moyenne ont permis de montrer que la puissance estimée est supérieure à 50 % pour des degrés de fausseté de l'hypothèse de stationnarité significatifs, tels qu'un taux de rupture sur la moyenne dépassant 75 % de la valeur de l'écart type.

L'étude de robustesse a montré que les méthodes exigeant l'indépendance des observations successives ne sont pas robustes vis-à-vis d'une tendance dans les séries. En présence d'une autocorrélation d'ordre 1, les résultats montrent que les procédures se comportent différemment selon le signe du coefficient d'autocorrélation, et qu'aucune d'entre elles n'est robuste vis-à-vis à la fois d'une autocorrélation positive et négative.

Les procédures qui requièrent la normalité sont robustes dans les limites ici testées, lorsque cette condition n'est pas respectée. De même les méthodes qui exigent l'invariance de l'écart type sont relativement robustes quand ce paramètre de dispersion est affecté par un changement, alors que la stationnarité en moyenne est assurée. En présence d'une rupture sur la moyenne, il semble que ces méthodes, et notamment la procédure bayésienne, soient relativement sensibles à une modification de l'écart type.

## REMERCIEMENTS

Nous tenons à remercier Monsieur P. Hubert pour nous avoir cédé les modules informatiques de la procédure de segmentation.

## RÉFÉRENCES BIBLIOGRAPHIQUES

- AKA A.A., LUBÈS H., MASSON J.M., SER-VAT E., PATUREL J.E., KOUAMÉ B., 1996. Analysis of the temporal variability of runoff in Ivory Coast: statistical approach and phenomena characterization. *Hydrol. Sci. J.*, 41(6), 959-970.
- BERNIER J., 1977. Étude de la stationnarité des séries hydrométéorologiques. *La Houille Blanche*. N° 4. 313-319.
- BRUNEAU P., RASSAM J.C., 1983. Application d'un modèle bayésien de détection de changements de moyennes dans une série. *Hydrol. Sci. J.*, 28(3), 341-354.
- BUISHAND T.A., 1982. Some methods for testing the homogeneity of rainfall records. *J. Hydrol.*, 58, 11-27.
- BUISHAND T.A., 1984. Tests for detecting a shift in the mean of hydrological time series. *J. Hydrol.*, 73, 51-69.
- CHAOUICHE A., 1988. Structure de la saison des pluies en Afrique Soudano-Sahélienne. Thèse de doctorat de l'École des Mines de Paris. Paris. 263 p.
- CLARKE R.T., 1973. Mathematical models in hydrology. Irrigation and drainage paper n° 19. FAO. 282 p.
- DAGNÉLIE P., 1970. *Théorie et Méthodes Statistiques*. Les Presses Agronomiques de Gembloux. Vol. 2 : 463 p.
- DEMARÉE G.R., 1990. Evidence of abrupt climate change from the rainfall data of a

- Mauritanian station. *IRM. Série A. N° 124.* 68-74.
- GARDNER L.A., 1969. On detecting changes in the mean of normal variates. *Ann. Math. Statist.*, 40, 116-126.
- HUBERT P., CARBONNEL J.P., 1987. Approche statistique de l'aridification de l'Afrique de l'Ouest. *J. Hydrol.*, 95, 165-183.
- HUBERT P., CARBONNEL J.P., CHAOUICHE A., 1989. Segmentation des séries hydro-météorologiques – Application à des séries de précipitations et de débits de l'Afrique de l'Ouest. *J. Hydrol.*, 110, 349-367.
- HUBERT P., CARBONNEL J.P., 1993. Segmentation des séries annuelles de débits de grands fleuves africains. *Bulletin de liaison du CIEH.*, 92, 3-10.
- KENDALL S.M., STUART A., 1943 (1977 edn). *The Advanced Theory of Statistics.* Charles Griffin, London, UK. Vol. 1, 2 & 3, 472 p, 723 p, 585 p.
- KOTTEGODA N.T., 1980. *Stochastic Water Resources Technology.* The Macmillan Press Ltd. 384 p.
- LEE A.F.S., HEGHINIAN S.M., 1977. A shift of the mean level in a sequence of independent normal random variables: A Bayesian approach. *Technometrics*, 19(4), 503-506.
- LUBÈS H., MASSON J.M., SERVAT E., PATUREL J.E., KOUAMÉ B., BOYER J.F., 1994. Caractérisation de fluctuations dans une série chronologique par applications de tests statistiques. Étude bibliographique. Programme ICCARE, rapport 3. ORSTOM, Montpellier, France. 21 p.
- MAHÉ G., OLIVRY J.C., 1991. Changements climatiques et variations des écoulements en Afrique occidentale et centrale du mensuel à l'interannuel. In: Hydrology for the Water Management of Large River Basins (Proc. Vienne Symp. August 1991), 163-172. IASH Publ. n° 201.
- NICHOLSON S.E., KIM J., HOOPINGARNER J., 1988. Atlas of African rainfall and its interannual variability. Department of Meteorology, Florida State University Tallahassee, Florida, USA.
- OLANIRAN O.J., 1991. Evidence of climatic change in Nigeria based on annual series of rainfall of different daily amounts, 1919-1985. *Climatic change.*, 19, 319-341.
- PATUREL J.E., SERVAT E., KOUAMÉ B., LUBÈS H., OUEDRAOGO M., MASSON J.M., 1997. Climatic variability in humid Africa along the Gulf of Guinea. Part II: an integrated regional approach. *J. Hydrol.* 191, 16-36.
- PETTITT A.N., 1979. A non-parametric approach to the change-point problem. *Appl. Statist.*, 28(2), 126-135.
- RIPLEY B.D., 1987. *Stochastic simulation.* Wiley Series in Probability and Mathematical Statistics. 237 p.
- ROSS S.M., 1997. *Simulation.* Second Edition. Academic Press, Inc. 282 p.
- SERVAT E., 1994. Programme ICCARE (Identification et Conséquences d'une variabilité du Climat en Afrique de l'ouest non sahéenne) – Présentation du Programme. Orstom, Abidjan, Côte d'Ivoire.
- SERVAT E., PATUREL J.E., LUBÈS H., KOUAMÉ B., OUEDRAOGO M., MASSON J.M., 1997. Climatic variability in humid Africa along the Gulf of Guinea. Part I: detailed analysis of the phenomenon in Côte d'Ivoire. *J. Hydrol.*, 191, 1-15.
- STAUDTE R.G., SHEATHER S.J., 1990. *Robust Estimation and Testing.* John Wiley & Sons, Inc. 351 p.
- SUTCLIFFE J.V., KNOTT D.G., 1987. Historical variations in African water resources. In: *The Influence of Climate Change and Climatic Variability on the Hydrologic Regime and Water Resources* (Proc. Vancouver Symp. August 1987), 463-475. IASH Publ. n° 168.
- SUTHERLAND R.A., BRYAN R.B., OOSTWOUW WIJENDES D., 1991. Analysis of the monthly and annual rainfall climate in a semi-arid environment, Kenya. *Journal of Arid Environments.*, 20, 257-275.
- VANNITSEM S., DEMARÉE G., 1991. Détection et modélisation des sécheresses au Sahel. *Hydrologie Continentale.* Vol. 6, n° 2, 155-171.
- WHITTAKER J., 1973. *Floods and Droughts,* Proceedings of the Second International Symposium in Hydrology, September 11-13, 1972, Fort Collins, Colorado, USA Water Resources Publications.
- WMO (World Meteorological Organization) (1966). *Climatic change by a working group of the commission for climatology.* W.M.O. 195, TP 100, Tech. note n° 79. WMO Geneva, Switzerland.

Lubès Niel Hélène, Masson J.M., Paturel Jean-Emmanuel,  
Servat Eric (1998)

Variabilité climatique et statistiques : étude par simulation de  
la puissance et de la robustesse de quelques tests utilisés  
pour vérifier l'homogénéité de chroniques

Revue des Sciences de l'Eau, 3, 383-408

ISSN 0992-7158