ARIMA

EPIsciences

Tests de Monte Carlo de détection de ruptures(s)-application à la détection d'un changement brusque du climat

Jean Cléophas ONDO*[1] and Taha BMJ Ouarda[2]

[1]Institut Sous-régional de Statistique et d'Economie Appliquée (ISSEA), B.P. 294, Cameroun
[2]Institut National de Recherche Scientifique (INRS-ETE), Université du Québec, Canada

*E-mail : j.cleophas.ondo@issea-cemac.org ou obotte@yahoo.com

DOI : 10.46298/arima.13719
Submitted on June 04, 2024 - Published on January 20, 2025
Volume : **43** - Year : **2025**

Editors : *Clémentin Tayou Djamegni, Mathieu Roche, Nabil Gmati*

Abstract

The statistical tests commonly used in break detection (s) are based on asymptotic approximations whose reliability on finite samples is not guaranteed (see Diebold and Chen (1996)[15]). The resulting problems in empirical applications include: lack of good critical values, level control problems and the possibility of detecting erroneous break dates. These drawbacks have contributed to the emergence of a generation of tests based on Edgeworth and Bootstrap simulations, for example (Bai and Perron, 1998[3], Shao and Zhang, 2010[43]). However, the methods proposed by the latter do not perform level control.

In this paper, we propose two exact tests (the risk of error of the first type is equal to the significance level) for detecting break dates. The first is a Monte Carlo (MC) test for detecting a break date. The second is a sequential test for detecting one or more break dates. The proposed procedures provide accurate marginal significance levels (p-values). We evaluate the performance of the proposed MC test by simulation. Our results show that the tests widely used in climate change studies in West and Central Africa present serious level distortions, whereas the MC tests control the level and have good power.

Keywords

Time series, detection of break(s) on average, Monte Carlo test, abrupt change in climate.

I INTRODUCTION

La littérature des tests de changement de régime, tests de changement structurel, tests de stabilité temporelle, ou encore tests de détection de rupture(s) dans une série chronologique est particulièrement abondante (voir Truong et al., 2020[45]). Pour les modèles de régressions linéaires, les travaux pionniers de Chow (1960)[13], Quandt (1958)[41] et Barten et Bronsard (1970)[5] ont considéré la date de rupture comme étant connu. Les recherches se sont dirigées vers une modélisation où cette date de rupture est inconnue. Page (1954, 1955 et 1957)[33–35] et Hinkley (1969, 1971)[24, 25] ont considéré l'estimation d'une date de changement dans

une séquence de variables indépendantes. Le travail de Brown, Durbin et Evans (1975)[8] a généralisé les tests de type cusum aux modèles de régressions linéaires. Broemeling et Choy (1981) et Bhatti et Wang (1998)[6] présentent un survol des développements en statistique. Des études récentes ont aussi abordé le problème des tests de changement structurel dans le cadre de modèles économétriques avec comme régresseurs des variables avec tendance, racine unitaire, et/ou corrélation temporelle. A cet effet, nous pouvons citer les travaux de Hansen (1992)[23], Andrews (1993) [1], Andrews et Ploberger (1994)[2], Kuan et Hornik (1995)[28] et Lumsdaine et Papell (1997)[32]. Dufour et Ghysels (1996)[18] présentent une revue de littérature. Bien que la littérature soit vaste, les études qui se sont directement intéressées au problème relatif aux changements structurels multiples sont peu nombreuses mais croissantes. Notons en particulier les travaux de Christiano (1992)[14], Bai et Perron (1998)[3], Banerjee et al. (1998)[4], Qu et Perron (2007)[40] et Shao and Zhang (2010)[43], Karavias et al. (2023)[27] et Lumsdaine et al. (2023)[31]. La majorité de ces tests est fondée sur des approximations asymptotiques dont la fiabilité sur des échantillons finis n'est pas garantie (voir Diebold et Chen (1996)[15], Vogelsang (1997)[47] et Perron (2006)[38]). Les problèmes qui en découlent dans les applications empiriques incluent : absence de bonnes valeurs critiques, problème de contrôle de niveaux et possibilité de détecter des dates de ruptures erronées. Ces inconvénients ont contribué à l'émergence d'une génération de tests basés sur des simulations de type bootstrap c'est le cas dans les travaux de Christiano (1992)[14], Diebold et Chen (1996)[15], Banerjee et al. (1998)[4] et Shao and Zhang (2010)[43]. Pourtant, les tests proposés par ces derniers ne réalisent pas le contrôle de niveau.

Le problème de la détection de rupture(s) a également reçu beaucoup d'attention dans la littérature hydrométéorologique. L'organisation météorologique mondiale a présenté une synthèse d'un certain nombre de tests et de procédures utilisées pour la détection des modifications climatiques (WMO, 1988[49]). Cavadias (1992, 1993)[11, 12] et Ryberg et al. (2020)[42] présentent une revue plus détaillée des approches statistiques disponibles et une liste d'applications dans le domaine. La détection de ruptures est aussi une source de préoccupation pour les gestionnaires des ressources en eau et les analystes de données, car ils peuvent être la preuve de changements naturels ou anthropiques dans les processus climatiques, hydrologiques ou paysagers, tels que des changements brusques dans les précipitations, le ruissellement, le stockage des réservoirs, les dérivations, l'urbanisation ou les pratiques agricoles (voir Ryberg et al., 2020[42]). Bien que la littérature soit vaste, les études qui ont traités le problème relatif à plusieurs ruptures sont peu nombreuses. Notons, en particulier, la procédure de segmentation des chroniques hydrométéorologiques proposée par Hubert et al. (1989)[26] qui n'est pas en toute rigueur un test statistique mais plutôt une méthode d'estimation du nombre de ruptures. De plus la majorité des tests sont unidimensionnels et utilisent des points critiques simulés. Ces tests reposent le plus souvent sur l'hypothèse que les chroniques sont indépendantes alors qu'en réalité, elles sont autocorrélées. Les problèmes qui en découlent dans les applications empiriques incluent : absence de bonnes valeurs critiques, problème de contrôle de niveaux et possibilité de détecter des dates de ruptures erronées. A ce propos, Lubes-Niel et al. (1998)[30] ont montré que, pour les tests largement employés dans les études de modification climatique le test de Pettitt (1979)[39], le test de Buishand (1984)[10], la procédure bayésienne de Lee et Heghinian (1977)[29], et la procédure de segmentation de Hubert et al. (1989)[26], la persistance (l'autocorrélation) et la présence d'une tendance dans les chroniques sont les deux caractéristiques qui pénalisent le plus leur performance.

Dans ce papier, nous proposons des tests exacts de détection de rupture dans les modèles de régression linéaire. Premièrement, nous proposons un test de monte carlo (mc) de détection d'une

rupture. Ce test est apparenté à ceux proposés par Christiano (1992)[14] et Dufour et Kiviet (1996)[19]. Deuxièmement, nous proposons deux tests séquentiels de détection de plusieurs ruptures dans des chroniques hydrométéorologiques (données de pluie ou de débit). Ces tests sont généralisés sous différents angles : nous fournissons des p-valeurs exactes et nous proposons pour les tests séquentiels de contrôler le niveau global. Nous illustrons l'utilité de ces méthodes par une étude de simulation et une application empirique portant sur des chroniques hydrologiques (les débits moyens annuels d'un ensemble de 3 stations préalablement choisies dans trois pays de l'Afrique Centrale : Garoua au Cameroun, Oubangui en République Centrafricaine (RCA) et Lambaréné au Gabon). Ces chroniques ont été publiées dans le cadre du projet UNESCO (1993)[46].

II MODELE ET STATISTIQUES DE TEST

2.1 Modèle de base

Les méthodes que nous présentons sont valides dans le cadre du modèle de régression linéaire général de Bai et Perron (1998)[3] :

$$x_t = y_t'b + z_t'd_j + \varepsilon_t, \quad t = T_{j-1}+1, \ldots, T_j, \quad j = 1, \ldots, m+1 \tag{2.1}$$

c'est-à-dire :

$$\begin{aligned}
x_t &= y_t'b + z_t'd_1 + \varepsilon_t, \quad t = 1, \ldots, T_1 \\
x_t &= y_t'b + z_t'd_2 + \varepsilon_t, \quad t = T_1+1, \ldots, T_2 \\
&\vdots \qquad\quad \vdots \qquad\quad \vdots \\
x_t &= y_t'b + z_t'd_{m+1} + \varepsilon_t, \quad t = T_m+1, \ldots, T
\end{aligned}$$

où x_t est le terme générique de la série chronologique $\{x_t\}_{t=1,\ldots,T}$, z est le vecteur des régresseurs $(q,1)$, (exemple: $z_t = \{1\}$ pour un changement sur la moyenne, $z_t = \{t\}$ pour un changement de tendance), y_t est un vecteur $(p,1)$, b et $d_j (j = 1, ..., m+1)$ sont les vecteurs de coefficients correspondants, et ε_t est le terme d'erreur au temps t. Les dates de ruptures $T_1, ..., T_m$ sont inconnues et sont estimées avec les coefficients $(b, d_1, ..., d_{m+1})$ lorsque les T données de la chronique $\{x_t\}_{t=1,\ldots,T}$ sont disponibles. La distribution du terme d'erreur est supposée connue à un paramètre d'échelle près. L'étude de Monte Carlo se concentre sur les cas où les erreurs sont indépendantes et identiquement distribuées selon une loi normale standard : $\varepsilon_t \overset{\text{iid}}{\sim} N(0,1)$. Nous maintenons (à cette phase du travail) l'hypothèse d'indépendance et d'homoscédasticité.

Dans la littérature hydrométéorologique Cavadias (1992, 1993)[11, 12] identifie deux grands types de modifications climatiques à suivre: (i) changement de la moyenne (incluant la tendance en saut si l'on accepte l'existence de tel changement); et (ii) changement de la variance (augmentation ou diminution). Nous nous intéressons au premier type de modification. C'est pourquoi, sans perte de généralité, et pour simplifier la présentation des tests et l'étude de Monte Carlo subséquente, nous considérons le cas où dans (2.1), $b = 0$, et z_t inclut une constante et/ou une tendance. Ce modèle présente, peut être, un aspect dynamique simplifié. Il n'en demeure pas moins que les tests de détection de rupture en hydrométéorologie sont utilisés dans ce contexte (Lee et Heghinian, 1977[29]; Pettitt, 1979[39] et Buishand, 1984[10]).

Pour présenter le résultat de base qui justifie le caractère exact des procédures de tests que nous proposons, il est utile de commencer par le cas d'une rupture où le modèle de génération de la chronique $\{x_t, t = 1, ..., n\}$ peut se réécrire comme suit :

$$x_t = \beta_0 + \beta_1 ID(t) + \varepsilon_t, \quad t = 1, ..., n \tag{2.2}$$

où,

- la rupture se fait au moyen de la variable indicatrice $ID(t_b)$ qui répond à l'une ou l'autre des définitions suivantes:

$$ID(t_b) = \begin{cases} 0 & \text{si } t \leq t_b \\ 1 & \text{autrement} \end{cases} \quad \text{(rupture brusque à la date } t = t_b\text{)}$$

$$ID(t_b) = \begin{cases} 0 & \text{si } t \leq t_b \\ t - t_b & \text{autrement} \end{cases} \quad \text{(rupture avec continuité à la date } t = t_b\text{)}$$

- β_0 est l'ordonnée à l'origine dans le modèle de régression et β_1 est l'amplitude (ou la pente) de la rupture à la date $t = t_b$.

2.2 Statistique de test : cas d'une rupture

Désignons par $F(t_b)$ la statistique de Fisher pour tester une rupture à la date t_b. Le *test du supremum* est construit de la façon suivante: pour chaque instant d'observation susceptible de correspondre avec une date de rupture, par exemple $t_b = 3, \ldots, (n-1)$, on calcule la statistique de test $F(t_b)$. La date de rupture est estimée à la date t_b pour laquelle la statistique $F(t_b)$ atteint sa valeur maximale que nous dénotons

$$F_{max} = \max_{t_b} \{F(t_b)\} \tag{2.3}$$

La statistique de test à la date t_b s'écrit alors de la façon suivante :

$$F(t_b) = \frac{\varepsilon' M_c \varepsilon - \varepsilon' M_{(t_b)} \varepsilon}{\varepsilon' M_{(t_b)} \varepsilon} \times \frac{n-2}{1} \tag{2.4}$$

C'est-à-dire,

$$F(t_b) = \frac{\left(\frac{\varepsilon}{\sigma}\right)' M_c \left(\frac{\varepsilon}{\sigma}\right) - \left(\frac{\varepsilon}{\sigma}\right)' M_{(t_b)} \left(\frac{\varepsilon}{\sigma}\right)}{\left(\frac{\varepsilon}{\sigma}\right)' M_{(t_b)} \left(\frac{\varepsilon}{\sigma}\right)} \times \frac{n-2}{1} \tag{2.5}$$

où M_c et $M_{(t_b)}$ sont les matrices de projections usuelles sur l'espace contraint et non contraint engendré par les colonnes des régresseurs pertinents. Il est clair que sous l'hypothèse principale d'absence de rupture ($H_0 : \beta_1 = 0$), la statistique de test, encore appelée *statistique du supremum* (F_{max}) est pivotale, c'est-à-dire que sa distribution sous H_0 ne dépend d'aucun paramètre inconnu. Rappelons pour mémoire que la distribution des termes d'erreur est supposée connue à un paramètre d'échelle près que nous avons réussi à « évacuer » dans l'équation (2.5).

2.3 p-valeur de Monte Carlo : cas d'une rupture

La technique des tests de MC (voir Dwass (1957)[20] et Dufour (1995)[16]) sera appliquée sur le test d'une rupture et les tests séquentiels que nous présenterons par la suite. Comme ces derniers sont eux-mêmes fondés sur des p-valeurs MC associées à des tests d'une rupture, nous préférons présenter cette technique à ce niveau du travail. Cependant, pour accommoder les cas des diverses statistiques de test que nous considèreront, nous ne nous contraindrons pas au cas de F_{max} . Nous présenterons la technique générale associée à un test unilatéral à droite fondé sur une statistique pivotale dénotée T.

Afin d'avoir une notation claire, nous désignerons par T_{Obs} la valeur observée de la statistique de test, et par $T_1, \ldots, T_N$ ses contreparties simulées sous l'hypothèse principale H_0 d'absence de rupture. La p-valeur du test de MC notée $\hat{p}_N(T_{Obs})$ est basée sur le rang de la statistique observée dans la série $T_1, \ldots, T_N$ telle que

$$\hat{p}_N(T_{Obs}) = \frac{N\hat{G}_N(T_{Obs}; T_1, \ldots, T_N) + 1}{N + 1} \tag{2.6}$$

où $N\hat{G}_N(T_{Obs}; T_1, \ldots, T_N)$ est le nombre de statistiques simulées qui sont plus grandes ou égales à T_{Obs}.

La région critique du test MC de détection d'une rupture correspond alors à

$$\hat{p}_N(T_{Obs}) \leq \alpha \tag{2.7}$$

La question qui se pose alors est la suivante : sous quelles conditions cette procédure simulée contrôle-t-elle le niveau exact du test de MC? La réponse est apportée par Dufour (1995)[16] qui a démontré qu'une région critique de la forme de l'équation (2.7) contrôle le niveau dans le sens où

$$P[\hat{p}_N(T_{Obs}) \leq \alpha / H_0] = \alpha \tag{2.8}$$

pourvu que $\alpha(N+1)$ soit un entier. Cette démonstration repose sur des résultats distributionnels fondamentaux concernant la distribution des rangs associés à un ensemble fini de statistiques échangeables. Dans le cas du test de MC de détection d'une rupture, une condition (suffisante) qui garantit le contrôle du niveau est une pivotalité de la statistique de test. En d'autres termes un test de MC tel que décrit ci-dessus satisfait l'équation (2.8) lorsque la distribution de la statistique de test sous H_0 ne dépend pas de paramètre de nuisance (Dufour, 1995[16]).

En somme, la procédure de test de MC est très semblable à un Bootstrap paramétrique, mais avec une différence fondamentale. Dans le test de MC, le nombre de tirages est explicitement pris en compte de sorte que des tests randomisés exacts peuvent être facilement obtenus. Le fait que la procédure soit randomisée joue un rôle prépondérant dans la détermination de la taille du test. Par contre, dans un test de type Bootstrap le nombre N d'échantillons simulés détermine la précision du test. En d'autres mots, on fait comme si le nombre de tirages était infini $(N \to \infty)$.

2.4 Statistiques de test et p-valeurs : cas de plusieurs ruptures

Dans le cas de la détection de plusieurs ruptures dans un modèle de régression linéaire, nous proposons deux tests séquentiels pour estimer le nombre de ruptures à un niveau de signification donné. Ces deux tests ont un point en commun dans la détermination de la statistique de test et du calcul de la p-valeur : nous considérons m statistiques significatives F_i^{max} non indépendantes, chacune retenue pour tester l'hypothèse d'absence de rupture H_{0i}, et leurs *p-valeurs* MC respectives $pv(F_i^{max}), i = 1, ..., m$. Afin de simplifier l'exposé (et sans perte de généralité) nous supposons que H_{0i} est rejetée au niveau de signification α lorsque F_i^{max} est « grand », c'est-à-dire $F_i^{max} \geq k_{\alpha i}$, où F_i^{Obs} est sa valeur observée dans l'échantillon et $k_{\alpha i}$ est un point critique tel que $P(F_i^{max} \geq k_\alpha) \leq \alpha$ lorsque H_{0i} est vraie. De manière équivalente, H_{0i} est rejetée au niveau de signification α lorsque $pv(F_i^{max}) \leq \alpha$.

Le problème qui nous intéresse ici est le suivant: comment peut-on combiner ces différents tests de MC de détection d'une rupture pour évaluer le niveau global, c'est-à-dire de façon à contrôler la probabilité de rejeter à tort l'hypothèse conjointe d'absence de rupture H_0 (« toutes les hypothèses H_{0i} sont vraies ») ? La réponse que nous donnons à cette question est d'appliquer la technique de tests de MC. La démarche est la suivante. En premier lieu, on obtient une statistique de test combinée que nous dénotons SQ. En second lieu, nous appliquons la procédure de test MC décrite en (2.3) pour obtenir la *p-valeur* MC.

Pour simplifier la présentation et sans perte de généralité, nous supposons que le test basé sur la statistique SQ rejette H_0 lorsque sa valeur observée dans l'échantillon est « grande ». Ce faisant, plusieurs choix sont possibles pour SQ incluant :

1. Les tests de type Tippett (1931)[44] fondés sur le minimum des p-valeurs,

$$SQ = 1 - \min_{i=1,\cdots,m} \{pv(F_i^{max})\} \tag{2.9}$$

   H_0 est rejetée lorsque SQ est grand.
2. La procédure de Fisher (1932)[21] et Pearson (1933)[37] construite sur le produit des p-valeurs,

$$SQ^* = \prod_{i=1}^{m} pv(F_i^{max})$$

   ou encore une des transformations suivantes de ce produit :

$$SQ^{**} = 1 - \prod_{i=1}^{m} \{pv(F_i^{max})\} \tag{2.10}$$

$$SQ^{***} = -2 \sum_{i=1}^{m} ln(pv(F_i^{max})) \tag{2.11}$$

   on rejette H_0 lorsque SQ^* est petit (de manière équivalente, lorsque SQ^{**} ou SQ^{***} est grand). Nous retenons la statistique de l'équation (2.10),

$$SQ = 1 - \prod_{i=1}^{m} \{pv(F_i^{max})\} \tag{2.12}$$

à partir de laquelle nous appliquerons un test de MC pour avoir un contrôle global. Bien évidemment, dans chaque cas où nous calculons $pv(F_i^{max})$, le modèle qui sert à générer les échantillons simulés est adapté en conséquence, mais ceci n'affecte pas le caractère pivotal des statistiques impliquées.

2.4.1 Test sequentiel 1: statistique de test

Le test appelé « *séquentiel 1* » s'inspire Bai et Perron (1998)[3]. Leur méthode, bien connue sous le nom de procédure de segmentation binaire, a été pour la première fois proposée par Vostrikova (1981)[48] qui a aussi prouvé sa convergence. La procédure de test séquentiel consiste à appliquer le test de MC de détection d'une rupture sur la série temporelle complète, ensuite diviser la chronique en deux à la date de rupture détectée, et appliquer itérativement le test de MC de détection de rupture sur les deux segments jusqu'à ce que plus aucune date de rupture ne soit détectée. Cette procédure consiste donc en l'application de tests pour k versus $k+1$ changements pour une séquence avec $k = 0, 1, 2, \ldots$. L'estimation du nombre de changements est la première valeur de k qui correspond à un non-rejet de l'hypothèse principale.

Nous appliquerons la procédure de segmentation binaire basée sur le critère du test de MC de détection d'une rupture pertinent à chaque étape de recherche. Soient $F_i^{max}, \ldots, F_m^{max}$, les statistiques significatives retenues, et $pv(F_i^{max}), i = 1, \ldots, m$, leurs p-valeurs MC respectives. À partir de ces dernières, nous construisons la statistique de test combinée suivante :

$$SQ1 = 1 - \prod_{i=1}^{m} pv(F_i^{max}) \tag{2.13}$$

2.4.2 Test sequentiel 2: statistique de test et p-valeur

Le deuxième test que nous proposons s'appelle « *séquentiel 2* » avec la modification suivante. Au lieu de répéter la détection séquentielle des dates de rupture jusqu'à un maximum (k) prédéfini, nous nous arrêterons au premier test non-significatif (cette décision sera basée sur la p-valeur MC pertinente à l'étape en question). Soient $F_i^{max}, \ldots, F_q^{max}$ les statistiques significatives $q \leq m$ retenues et $pv(F_i^{max}), i = 1, \ldots, q$, leurs p-valeurs MC respectives. À partir de ces dernières, nous construisons, comme ci-haut, la statistique de test combinée :

$$SQ2 = 1 - \prod_{i=1}^{q} pv(F_i^{max}) \tag{2.14}$$

III ÉTUDE DE PERFORMANCE PAR SIMULATION

Nous étudions la performance du test de MC de détection d'une date de rupture et du test « séquentiel 1 » qui ont été proposés dans ce papier. Quatre cas nous intéressent. Premièrement, celui où nous cherchons à comparer le niveau du premier test à celui des tests habituellement utilisés dans la littérature hydrométéorologique : Buishand (1984)[10], le test de Worsley (1986)[50] et le test de Pettitt (1979)[39]. Deuxièmement, celui où nous cherchons à comparer sa puissance à celle de ces mêmes tests. Troisièmement celui où nous cherchons à connaître sa

robustesse, c'est-à-dire s'il est performant pour des chroniques contenant une persistance (situation habituellement rencontrée en hydrométéorologie). Quatrièmement celui où nous cherchons à connaître le niveau et la puissance du test « séquentiel 1 ». Pour des raisons de commodité, Nous adoptons la notation suivante : MC (test de MC de détection d'une rupture), Bui (test de Buishand), Wor (test de Worsley), Pet (test de Pettitt) et MC_krup (test séquentiel 1).

3.1 Présentation des tests de Buishand, de Worsley et de Pettitt

Le test de Buishand est de nature bayésienne. Il détecte un saut brusque de la moyenne survenu à la date t_0. En supposant une distribution *a priori* uniforme pour la position de t_b, sa statistique de test U est définie par :

$$U = [n(n+1)]^{-1} \sum_{k=1}^{n-1} (S_k^*/D_x)^2 \tag{3.1}$$

où $S_0^* = 0$, $S_k^* = \sum_{t=1}^{k}(x_t - \bar{x})$ pour $k = 1, ..., n$ et $D_x = n^{-1}\sum_{t=1}^{n}(x_t - \bar{x})^2$. Les points critiques sont dérivés à partir de la méthode de simulation de type MC (Buishand, 1982[9]). De meilleures estimations de ces points sont parues ultérieurement (Buishand, 1984[10]).

Le test de Worsley est un test paramétrique pour détecter une rupture brusque dans une chronique. En utilisant les sommes cumulées et le rapport de vraisemblance de Worsley, sa statistique de test est :

$$W = \max_{1 \leq k \leq n-1} \left\{ |S_k^*|/D_x \sqrt{k(n-k)} \right\} \tag{3.2}$$

Pour des échantillons excédant 10, les points critiques sont obtenus à partir des méthodes de simulation de MC (Worsley, 1986[50]).

Finalement, le test de Pettitt est une version modifiée du test non-paramétrique de Mann-Whitney qui vérifie que les deux échantillons formés avant et après la date de rupture t_0, c'est-à-dire $\{x_1, ..., x_{t_b}\}$ et $\{x_{t_b+1}, ..., x_n\}$ proviennent d'une même population. La statistique de test est :

$$K_n = \max_{1 \leq t \leq n-1} \left\{ \sum_{i=1}^{t} \sum_{j=t+1}^{n} D_{ij} \right\} \tag{3.3}$$

où :

$$D_{ij} = \text{sgn}(x_i - x_j)$$

tel que :

$$\text{sgn}(x_i - x_j) = \begin{cases} 1 & \text{si } x_i > x_j \\ 0 & \text{si } x_i = x_j \\ -1 & \text{si } x_i < x_j \end{cases}$$

Le calcul de la probabilité de dépassement p de la valeur observée de cette statistique, k_n^{obs}, est approximativement (Pettitt, 1979[39]):

$$p = P\left(K_n > k_n^{obs}\right) \approx 2\exp\left\{-6(k_n^{obs})^2/(n^3+n^2)\right\} \tag{3.4}$$

p peut être calculée à l'aide de la méthode bootstrap.

3.2 Etudes comparatives : cas d'une rupture

3.2.1 Etude de niveau

La présente étude nécessite de générer des chroniques artificielles sous l'hypothèse principale d'absence de rupture H_0 . Le modèle de génération s'écrit :

$$x_t = \mu_0 + \varepsilon_t \tag{3.5}$$

où, μ_0 est la moyenne régionale des chroniques réelles de la base de données à notre disposition et les erreurs $\varepsilon_t \overset{\text{iid}}{\sim} N(0,1)$. Les méthodes de simulation MC sont utilisées pour générer M=1000 chroniques de taille n=30, 50 ou 100 valeurs. Chaque chronique est soumise aux tests de MC , Buishand, Worsley et Pettitt pour tester H_0 au niveau de signification théorique α=1%, 5% ou 10%.

Les résultats obtenus sont consignés dans la figure 1 ci-dessous. En général, pour les trois tailles des chroniques et les trois niveaux de signification retenus, ces résultats déclarent performants les tests de Pettitt, de Buishand et de MC. En effet, le niveau empirique de chacun de ces tests est, aux fluctuations d'échantillonnage près, conforme aux niveaux de signification théoriques α=1%, 5% et 10% . Ces tests ne présentent donc aucune distorsion de niveau. Par contre le test de Worsley présente un léger problème de distorsion de niveau pour les trois niveaux théoriques α=1%, 5% ou 10% retenus. Nous observons également que les niveaux de signification estimés diminuent à mesure que la taille de la chronique augmente. Nous sommes donc en droit de penser que pour les niveaux de significations retenus, la performance des tests étudiés est meilleure pour des chroniques de grande taille.

3.2.2 Etude de puissance

Pour estimer la puissance d'un test, l'étude de simulation MC nécessite de générer des chroniques artificielles sous la contre-hypothèse H_1 de la présence de rupture à la date t_b . Comme l'un des problèmes cruciaux de détection de rupture est posé par les effets des bords (les points de changement intervenant aux bords de l'intervalle d'observations), il est opportun de considérer trois dates de rupture : $t_b = [n/3]$ (pour une rupture survenue au début de la chronique), $t_b = [n/2]$ (pour une rupture survenue au milieu de la chronique) et $t_b = [2n/3]$ (pour une rupture survenue à la fin de la chronique). Dans le cas d'une rupture brusque, les études de simulation sont menées pour différentes amplitudes de rupture Δ . Les valeurs de Δ représentent respectivement un taux de rupture de 0.5, 1, 1.5, 2 et 2.5% de la moyenne μ_0:

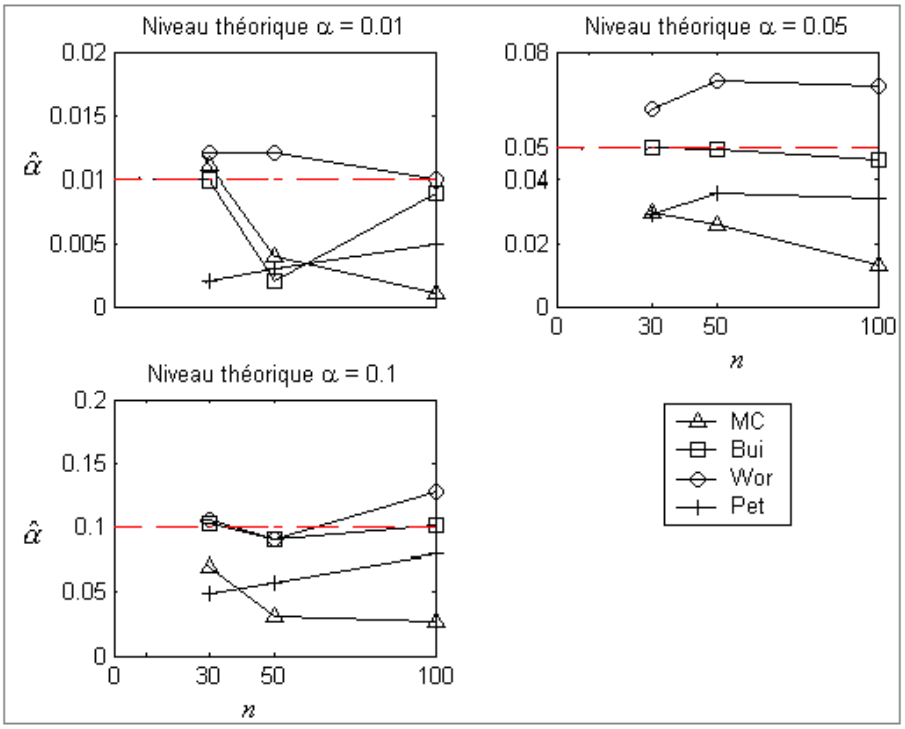

Figure 1: Étude comparative de niveau dans des chroniques sans persistance.

$$x_t = \begin{cases} \mu_0 + \varepsilon_t & t = 1, \ldots, t_b \\ (\mu_0 + \Delta) + \varepsilon_t & t = t_b + 1, \ldots, n \end{cases} \tag{3.6}$$

La figure 2 présente les résultats obtenus. Les quatre tests sont performants pour détecter une rupture brusque lorsque celle-ci survient au milieu de la chronique. Le test de MC s'avère être plus puissant que les tests de Pettitt, de Buishand et de Worsley pour détecter une rupture au début ou vers la fin de la chronique.

3.2.3 Étude de robustesse

L'étude de robustesse consiste à évaluer le comportement des tests vis à vis de chroniques normales autocorrélées. Dans cette étude, le modèle de génération de données est simplement un modèle autorégressif d'ordre 1, *AR(1)* . Il s'écrit comme suit :

$$x_t = \rho x_{t-1} + \varepsilon_t \tag{3.7}$$

Le paramètre autorégressif ρ prend les valeurs ρ=0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8 et 0.9 . Les méthodes de simulation MC sont utilisées pour générer des chroniques de différentes tailles (n=30, 50 ou 100). Comme dans les expériences précédentes, chaque chronique simulée est soumise aux tests de MC, Buishand, Worsley et Pettitt pour tester H_0 au niveau de signification théorique α=1%, 5% ou 10% . La figure 3 résume les résultats obtenus. Il ressort que les tests habituellement employés dans les études de rupture en hydrométéorologie sont fortement influencés par la présence de persistance dans l'échantillon de données. Il est donc clair que, lorsqu'ils sont appliqués à des chroniques hydrométéorologiques, ces tests peuvent conduire à des rejets fallacieux de l'hypothèse principale d'absence de rupture.

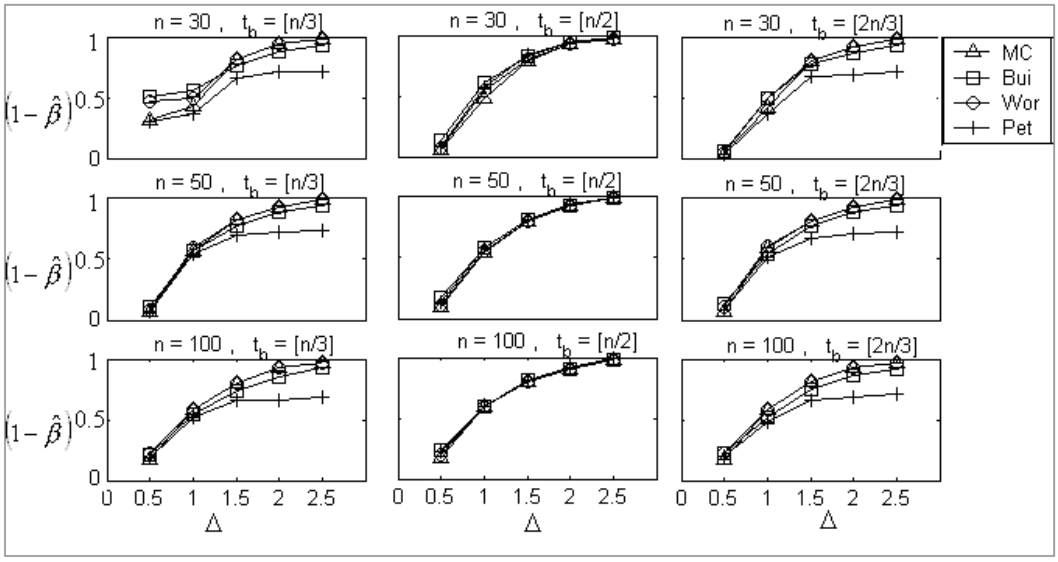

Figure 2: Étude comparative de puissance dans des chroniques sans persistance et contenant une rupture(α=0.01).

3.3 Etudes comparatives : cas de plusieurs ruptures

Les résultats dont nous disposons à cette phase du travail portent sur le test « *séquentiel 1* ». Dans un premier temps, nous nous intéressons à la performance de ce test du point de vue du niveau. Dans un deuxième temps, notre intérêt se porte sur l'exploration de sa puissance

3.3.1 Etude de niveau

Les chroniques artificielles sont générées sous l'hypothèse principale H_0 d'absence de rupture, à partir du modèle décrit à l'équation (2.13). Les simulations de Monte Carlo sont alors utilisées pour produire M=1000 chroniques de taille n=30, 50 ou 100 valeurs. Chaque chronique générée est ensuite soumise au test MC_krup pour vérifier H_0 au niveau de signification α=5% . La longueur minimale de segmentation admise dans la procédure séquentielle est h=10 et $N = 19$ chroniques simulées.

Le tableau 1 donne les résultats obtenus. Il appert que pour les différentes tailles des chroniques considérées, le test de MC de détection de ruptures contrôle parfaitement le niveau. Il n'a donc pas de problème de distorsion de niveau.

Table 1: Étude de niveau dans des chroniques sans persistance (MC_krup).

n	**Niveau de signification α = 5%**
30	0.043
50	0.028
100	0.016

*C*hroniques générées par $x_t = \mu_0 + \varepsilon_t$, $\varepsilon \overset{\text{iid}}{\sim} N(0,1)$, $N = 19, M = 1000$.

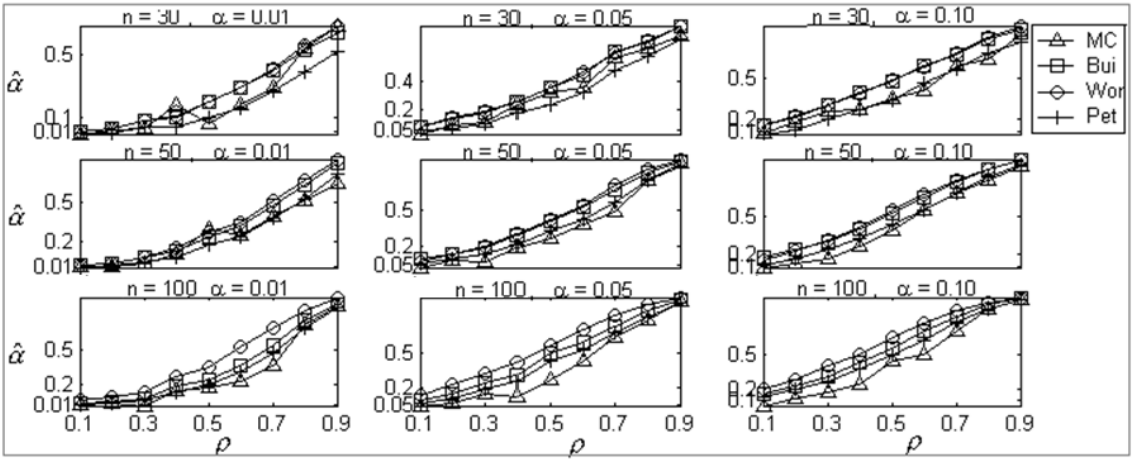

Figure 3: Étude comparative de niveau dans des chroniques autocorrélées.

3.3.2 Etude de puissance

Les chroniques artificielles indépendantes présentant une rupture brusque au début ($t_{b_1} = [n/3]$) et à la fin de la chronique ($t_{b_2} = [2n/3]$) sont générées par le processus

$$x_t = \begin{cases} \mu_0 + \varepsilon_t & t = 1, \ldots, t_{b_1} \\ (\mu_0 + \Delta) + \varepsilon_t & t = t_{b_1} + 1, \ldots, t_{b_2} \\ \{(\mu_0 + \Delta) + \Delta) + \varepsilon_t\} & t = t_{b_2} + 1, \ldots, n \end{cases} \tag{3.8}$$

les paramètres μ_0, Δ et les erreurs ε_t sont analogues à ce qui a été fait précédemment. À partir des méthodes de simulation MC, nous générons M=1000 chroniques de n=30, 50 ou 100 valeurs. Chaque chronique est ensuite soumise au test MC _krup pour vérifier H_0 au niveau de signification α=5% . Nous explorons ici deux longueurs minimales de segmentation : $h = 5$ et 10 .

Les résultats obtenus sont présentés dans les tableaux 2 et 3. Nous observons que la puissance du test croît avec le taux de rupture. Le test arrive à détecter les deux dates de rupture simulées dans des proportions très élevées. Celles-ci atteignent la valeur de 0.80 pour des taux de rupture de 1.5%, 2% et 2.5%, et ce sans égard à la taille de la chronique et à la longueur minimale de segmentation h. D'autre part, le test estime correctement les deux dates de rupture simulées avec un pourcentage plus qu'acceptable : près de 50% des cas pour un taux de rupture de 1.5% ! En revanche, la puissance du test et sa proportion de rejets de H_0 associés à une estimation correcte de la date de rupture sont moins grandes pour un taux de rupture de 0.5%.

Table 2: Étude de puissance dans des chroniques sans persistance et contenant une rupture (MC_krup)

Niveau de signification $\alpha = 5\%$ $(h = 10)$						
n	**k**	$\Delta = 0.5\%$	$\Delta = 1\%$	$\Delta = 1.5\%$	$\Delta = 2\%$	$\Delta = 2.5\%$
30	0	0.687	0.077	0.000	0.000	0.000
	1	0.313	0.923	1.000	1.000	1.000
50	0	0.290	0.000	0.000	0.000	0.000
	1	0.710	1.000	1.000	1.000	1.000
100	0	0.042	0.000	0.000	0.000	0.000
	1	0.958	1.000	1.000	1.000	1.000

Chroniques générées par $x_t = \begin{cases} \mu_0 + \varepsilon_t, & t = 1, \ldots, t_b, \\ (\mu_0 + \Delta) + \varepsilon_t, & t = t_b + 1, \ldots, n', \end{cases}$

où $t_b = [n/2]$, $\varepsilon \overset{\text{iid}}{\sim} N(0,1)$.

$N = 19$, $M = 1000$, ***k** : nombre de ruptures détectées.*

Table 3: Etude de puissance dans des chroniques sans persistance et contenant deux ruptures (MC_krup)

Niveau de signification $\alpha = 5\%$ h=5						
n	**k**	$\Delta = 0.5\%$	$\Delta = 1\%$	$\Delta = 1.5\%$	$\Delta = 2\%$	$\Delta = 2.5\%$
30	0	0.908	0.000	0.000	0.000	0.000
	1	0.037	0.444	0.063	0.006	0.000
	2	0.052	0.502	0.808	0.824	0.868
	>2	0.003	0.054	0.129	0.170	0.132
50	0	0.005	0.000	0.002	0.000	0.000
	1	0.904	0.144	0.000	0.000	0.000
	2	0.088	0.763	0.796	0.809	0.831
	>2	0.003	0.093	0.202	0.191	0.169
100	0	0.000	0.000	0.000	0.000	0.000
	1	0.676	0.000	0.000	0.000	0.000
	2	0.317	0.787	0.793	0.801	0.876
	>2	0.007	0.213	0.207	0.199	0.124

Chroniques générées par $x_t = \begin{cases} \mu_0 + \varepsilon_t, & t = 1, \ldots, t_{b_1}, \\ (\mu_0 + \Delta) + \varepsilon_t, & t = t_{b_1} + 1, \ldots, t_{b_2}, \\ \{(\mu_0 + \Delta) + \Delta\} + \varepsilon_t, & t = t_{b_2} + 1, \ldots, n \end{cases}$

$\varepsilon \overset{\text{iid}}{\sim} N(0,1)$, $N = 19$, $M = 1000$, **k** : nombre de ruptures détectées.

IV ILLUSTRATION EMPIRIQUE

Nous voulons illustrer dans cette partie, l'applicabilité des tests que nous proposons dans ce papier.

4.1 Données

Nous disposons d'un ensemble de données sur les débits moyens annuels enregistrés dans trois stations d'Afrique Centrale : *Garoua* au Cameroun, *Oubangui* en République centrafricain (RCA) et *Lambaréné* au Gabon. La figure 4 présente une répartition de ces stations sur la carte de l'Afrique.

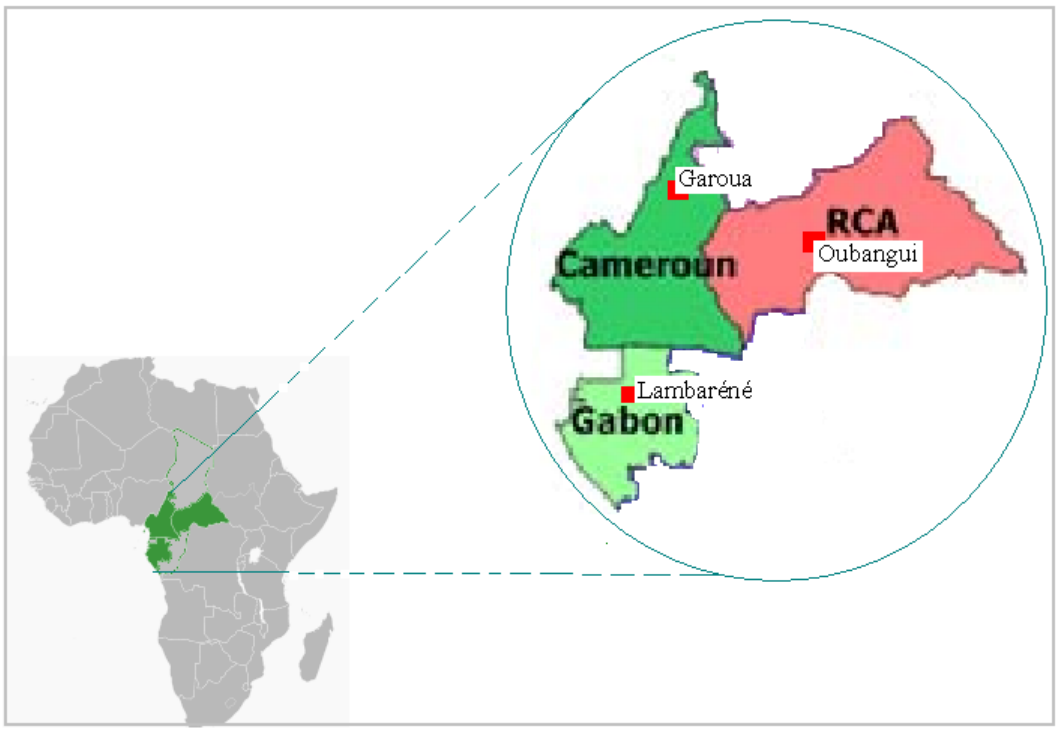

Figure 4: Représentation des stations de l'étude.

Les caractéristiques des trois stations sont présentées dans le tableau 4

Table 4: Caractéristiques des stations et des chroniques

	Rivière	**Station**	**Superficie**	**Période**
Cameroun	*Bénoué*	*Garoua*	*64000* km^2	*1943 - 1979*
RCA	*Oubangui*	*Bangui*	*500000* km^2	*1935 - 1984*
Gabon	*Ogooué*	*Lambaréné*	*205000* km^2	*1954 - 1975*

Source : UNESCO (1993)[46]

Nous avons reproduit sur la figure 5 ci-dessous l'évolution relative de ces trois chroniques. Une analyse de l'ensemble de ces chroniques permet d'envisager comme modèle de génération de

données, un modèle de régression intégrant une ou plusieurs ruptures brusques de la moyenne des chroniques.

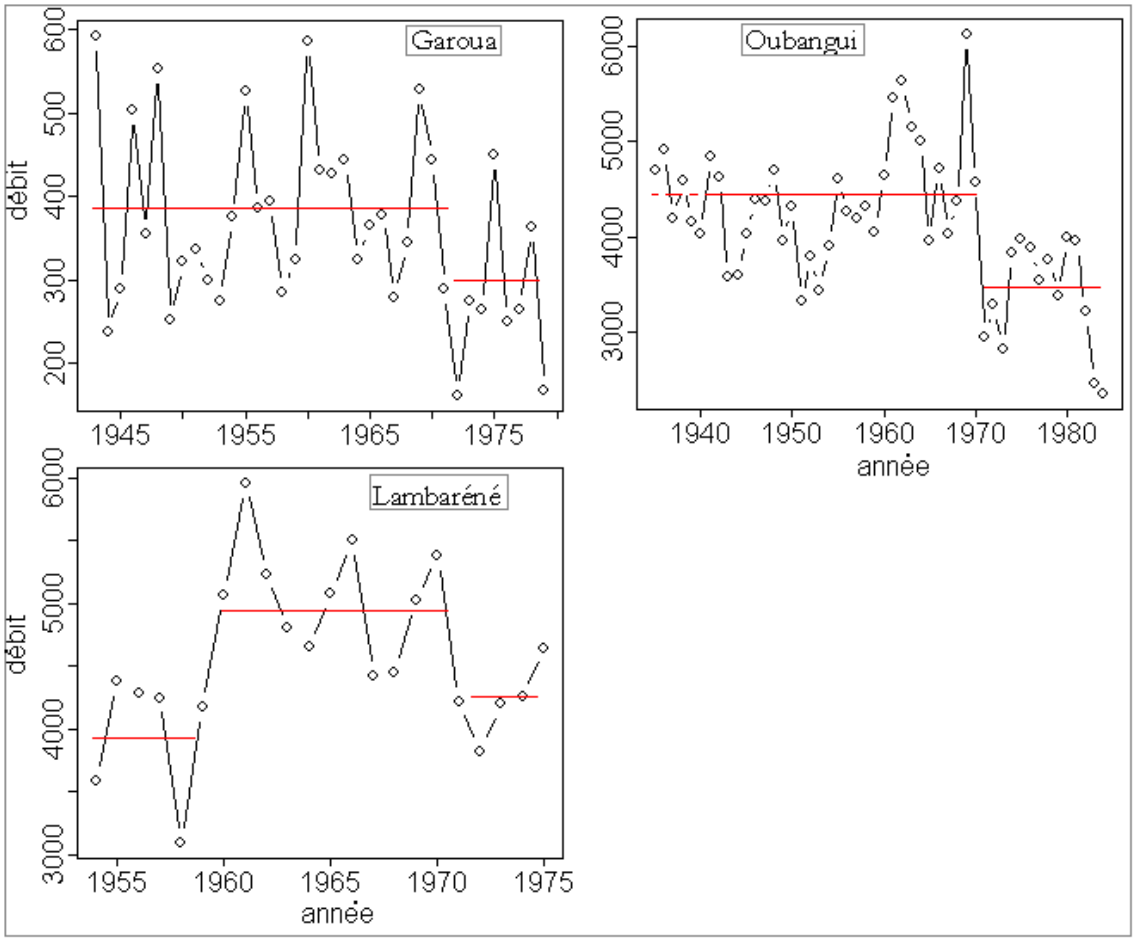

Figure 5: Évolution des débits annuels aux stations de Garoua, Oubangui et Lambaréné.

4.2 Applications

Nous nous limitons dans le cadre de ce travail aux procédures de Monte Carlo de détection d'une(des) date(s) de rupture de la moyenne dans une chronique sans persistance. Nous retiendrons le niveau de signification α=1% pour analyser nos résultats. Ce niveau a été recommandé par Hubert et al. (1989)[26] pour exécuter le test de Schéffé. Pour les échantillons simulés nous retenons $N = 999$. Le tableau 5 présente les résultats obtenus.

Au regard de l'évidence apportée par les données des chroniques étudiées, et au niveau de signification α=1% , le test de MC met en évidence un changement de la moyenne en 1971 dans la station *Oubangui* en RCA. Celle-ci se traduit par une diminution du débit moyen annuel. En adoptant toujours un niveau de signification α=1% , nous avons également appliqué le test séquentiel de détection de plusieurs dates de rupture dans ces mêmes chroniques. La conclusion ne fait aucun doute : au niveau de signification α=1% , aucune des trois stations retenues ne présente plus d'une date rupture en son sein.

Il est intéressant de noter que, d'après la figure 5 et les résultats du tableau 5, les changements « importants » de la moyenne les plus probables sont survenues en 1959 à *Lambaréné* et en 1970 dans l'ensemble des trois stations. Cela cadre bien avec les résultats des travaux de Paturel et

al. (2010)[36] qui ont analysés les principaux traits des changements climatiques survenus au 20ème siècle en Afrique de l'Ouest et Centrale.

Table 5: Résultats avec le test MC de détection d'une rupture brusque de la moyenne

Station	**Date probable de modification**	**p-valeur**
Garoua	1970	0.174
Bangui	1970	0.001*
Lambaréné	1959	0.084

$\varepsilon \overset{\text{iid}}{\sim} N(0,1)$, $N = 999$

*p-valeur significative au niveau $\alpha = 1\%$

V CONCLUSION

Dans ce papier, nous avons proposé un test de MC de détection d'une date de rupture et deux tests séquentiels de détection de plusieurs dates de rupture dans les modèles de régression linéaire. Les résultats de simulation suggèrent que le test de MC présente des avantages par rapport aux tests habituellement utilisés dans la littérature hydrométéorologique pour détecter une modification climatique. Ces résultats confirment également que le test séquentiel 1 de détection de plusieurs dates de rupture contrôle le niveau et a une bonne puissance.

Les tests proposés ont été appliqués aux données de débits moyens annuels des stations retenues en Afrique Centrale. Leur application a permis, pour chaque station, de calculer la probabilité qu'un changement de moyenne de la chronique soit survenu une année donnée et d'identifier la date pour laquelle la probabilité de réalisation du changement est maximale. Pour les trois stations retenues, la date la plus probable d'occurrence du changement se situe en 1970. Cependant, il n'est pas possible de conclure à une modification climatique sur la seule base de ces trois chroniques. La validation des résultats avec d'autres chroniques hydrométéorologiques collectées dans la même région est nécessaire. Même si ce changement est significatif à la station de l'*Oubangui*, il est aussi nécessaire d'expliquer les causes de ce phénomène observé (le changement est-il dû à une modification climatique ou à une cause anthropique ?).

En somme, les tests que nous proposons peuvent s'étendre à tout type de modèle de changement de la moyenne et/ou de la variance d'une chronique lorsque la statistique de test est simulable sous l'hypothèse principale d'absence de rupture. Par conséquent, une généralisation naturelle de ces tests à des modèles considérant la dynamique des chroniques et des extensions semblables aux problèmes des tests multidimensionnels de détection de rupture pourraient aussi s'avérer fort intéressant.

RÉFÉRENCES BIBLIOGRAPHIQUES

[1] D. W. K. Andrews. "Tests for Parameter Instability and Structural Change With Unknown Change Point". In: *Econometrica* 61.4 (1993), pages 821–856.

[2] D. W. K. Andrews. "Optimal Tests when a Nuisance Parameter is Present Only Under the Alternative". In: *Econometrica* 62.6 (1994), pages 1383–1414.

[3] J. Bai and P. Perron. “Estimating and Testing Linear Models with Multiple Structural Changes”. In: *Econometrica* 66.1 (1998), pages 47–78.
[4] A. Banerjee, S. Lazarova, and G. Urga. *Bootstrapping Sequential Tests for Multiple Structural Breaks*. Technical report. Discussion Paper No 17-98, 1998.
[5] A. P. Barten and L. S. Bronsard. “Two-Stage Least-Squares Estimation with Shifts in the Structural Form”. In: *Econometrica* 38.6 (1970), pages 938–941.
[6] M. I. Bhatti and J. Wang. “Power Comparaison of some Tests for Testing Change Point”. In: *Statistica* 58.1 (1998), pages 127–137.
[7] L. D. Broemling and J. H. Choy. “Detecting Structural Change in Linear Models”. In: *Communications in Statistics - Theory and Methods* 10.24 (1981), pages 2551–2561.
[8] R. L. Brown, J. Durbin, and J. M. Evans. “Techniques for testing the constancy of regression relationships over time”. In: *Journal of the Royal Statistical Society, Series B* 37 (1975), pages 149–192.
[9] T. A. Buishand. “Some Methods for Testing the Homogeneity of Rainfall Records”. In: *Journal of Hydrology* 58.1-2 (1982), pages 11–27.
[10] T. A. Buishand. “Tests for Detecting a Shift in the Mean of Hydrological Time Series”. In: *Journal of Hydrology* 73.1-2 (1984), pages 51–69.
[11] G. S. Cavadas. *A Survey of Current Approaches to Modelling of Hydrological Time Series with Respect to Climate Variability and Change*. Technical report. WMO Workshop on WCP Water, Geneva, 1992, pages 3–13.
[12] G. S. Cavadas. *Detection and Modelling of the Impact of Climatic Change on River Flows*. Technical report. NATO Advanced Study Institute, Deauville (France), 1993.
[13] G. C. Chow. “Tests of Equality Between Sets of Coefficients in Two Linear Regressions”. In: *Econometrica* 28.3 (1960), pages 591–605.
[14] L. J. Christiano. “Searching for a Break in GNP”. In: *Journal of Business Economic Statistics* 10.3 (1992), pages 237–250.
[15] F. X. Diebold and C. Chen. “Testing Structural Stability with Endogenous Breakpoint A Size Comparaison of Analytic and Bootstrap Procedures”. In: *Journal of Econometrics* 70.1 (1996), pages 221–241.
[16] J. M. Dufour. *Monte Carlo Tests with Nuisance Parameters: A General Approach to Finite-Sample Inference and Nonstandard Asymptotics in Econometrics*. Technical report. C.R.D.E., University of Montréal, 1995.
[17] J. M. Dufour. “Exact tests for structural change in first-order dynamic models”. In: *Journal of Econometrics* 70.1 (1996), pages 39–68.
[18] J. M. Dufour and E. Ghysels. “Recent Developments in the Econometrics of Structural Change”. In: *Journal of Econometrics* 70 (1996), pages 1–8.
[19] J. M. Dufour and J. F. Kiviet. “Exact tests for structural change in first-order dynamic models”. In: *Journal of Econometrics, Elsevier* 70.1 (1996), pages 39–68.
[20] M. Dwass. “Modified Randomization Tests for Nonparametric Hypotheses”. In: *Annals of Mathematical Statistics* 28.1 (1957), pages 181–187.
[21] R. A. Fisher. *Statistical Methods for Research Workers*. 4th. Oliver and Boyd, 1932.
[22] Z. Fu, Y. Hong, and X. Wang. “Testing for Structural Changes in Large Dimensional Factor Models via Discrete Fourier Transform”. In: *Journal of Econometrics* 233.1 (2023), pages 302–331.
[23] B. E. Hansen. “Tests for Parameter Instability in Regressions with I(1) Processes”. In: *Journal of Business Economic Statistics* 10.3 (1992), pages 321–335.
[24] D. V. Hinkley. “Inference about the Intersection in Two-Phase Regression”. In: *Biometrika* 56.3 (1969), pages 495–504.

[25] D. V. Hinkley. “Inference about the Change-Point from Cumulative Sum Tests”. In: *Biometrika* 58.3 (1971), pages 509–523.
[26] P. Hubert, J. Carbonnel, and A. Chaouche. “Segmentation des Séries Hydro-météorologiques - Application à des Séries de Précipitation et de Débits de l’Afrique de l’Ouest”. In: *Journal of Hydrology* 110.3 (1989), pages 349–367.
[27] Y. Karavias, P. K. Narayan, and J. Westerlund. “Structural Breaks in Interactive Effects Panels and the Stock Market Reaction to COVID-19”. In: *Journal of Business and Economic Statistics* 41.3 (2023), pages 653–666.
[28] C. M. Kuan and K. Hornik. “The Generalized Fluctuation Test: A Unifying View”. In: *Econometric Reviews* 14 (1995), pages 135–161.
[29] A. F. S. Lee and S. M. Hegnhinian. “A Shift Of The Mean Level In A Sequence Of Independent Normal Random Variables - A Bayesian Approach”. In: *Technometrics* 19.4 (1977), pages 503–506.
[30] H. Lubes-Niel, J. M. Masson, J. E. Paturel, and E. Servat. “Variabilité Climatique et Statistiques. Étude par Simulation de la Puissance et de la Robustesse de quelques Tests Utilisés pour Vérifier l’Homogénéité de Chroniques”. In: *Revue des Sciences de l’Eau* 11.3 (1998), pages 383–408.
[31] R. L. Lumsdaine, R. Okui, and W. Wang. “Estimation of Panel Group Structure Models with Structural Breaks in Group Memberships and Coefficients”. In: *Journal of Econometrics* 233.1 (2023), pages 45–65.
[32] R. L. Lumsdaine and D. H. Papell. “Multiple Trend Breaks And The Unit-Root Hypothesis”. In: *Review of Economics and Statistics* 79.2 (1997), pages 212–218.
[33] E. S. Page. “Continuous Inspection Schemes”. In: *Biometrika* 41 (1954), pages 100–155.
[34] E. S. Page. “A Test for a Change in a Parameter Occurring at an Unknown Point”. In: *Biometrika* 42.3-4 (1955), pages 523–527.
[35] E. S. Page. “On Problems in which a Change of Parameters Occurs at an Unknown Point”. In: *Biometrika* 44.1-2 (1957), pages 248–252.
[36] J. E. Paturel, A. Boubacar A. L’Aour, and G. Mahé. “Analyse de Grilles Pluviométriques et Principaux Traits des Changements Survenus au 20ème Siècle en Afrique de l’Ouest et Centrale”. In: *Journal des Sciences Hydrologiques* 55.8 (2010), pages 1281–1288.
[37] K. Pearson. “On a Method of Determining whether a Sample of Size n Supposed to have been drawn from a Parent Population having a Known Probability Integral has Probably been Drawn at Random”. In: *Biometrika* 25.3-4 (1933), pages 379–410.
[38] P. Perron. “Dealing with Structural Breaks”. In: *Palgrave Handbook of Econometrics, Vol. 1*. Edited by K. Patterson and T. C. Mills. Palgrave, London, 2006, pages 278–352.
[39] A. N. Pettitt. “A Non-Parametric Approach to the Change-Point Problem”. In: *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28.2 (1979), pages 126–135.
[40] Z. Qu and P. Perron. “Estimating and Testing Structural Changes in Multivariate Regressions”. In: *Econometrica* 75.2 (2007), pages 459–502.
[41] R. E. Quandt. “The Estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes”. In: *Journal of the American Statistical Association* 53.284 (1958), pages 873–880.
[42] K. R. Ryberg, G. A. Hodgkins, and R. W. Dudley. “Change Points in Annual Peak Streamflows: Method Comparisons and Historical Change Points in the United States”. In: *Journal of Hydrology* 583 (2020), 124307pp.
[43] X. Shao and Y. Zhang. “Testing for Change Points in Time Series”. In: *Journal of the American Statistical Association* 105.491 (2010), pages 1228–1240.
[44] L. H. C. Tippett. *The Methods of Statistics*. Williams and Norgate, 1931.

[45] C. Truong, L. Oudre, and N. Voyatis. “Selective review of offline change point detection methods”. In: *Signal Processing* 167 (2020), page 107299.
[46] UNESCO. *Débits de certains Cours d’Eau du Monde*. Volume II (partie II). Etudes et rapports d’hydrologie. 1993. ISBN: 92-3-002940-8.
[47] T. Vogelsang. “Wald-Type Tests for Detecting Breaks in the Trend Function of a Dynamic Time Series”. In: *Econometric Theory* 13.6 (1997), pages 818–848.
[48] L. J. Vostrikova. “Detecting "Disorder" in Multidimensional Random Processes”. In: *Soviet Mathematics Doklady* 24 (1981), pages 55–59.
[49] WMO. *Analyzing long time series of hydrological data with respect to climate variability and change*. World Meteorological Organization, World Climate Program, WCAP-3, WMO/TD, no 224. 1988.
[50] K. J. Worsley. “Confidence Regions and Tests for a Change-Point in a Sequence of Exponential Family Random Variables”. In: *Biometrika* 73.1 (1986), pages 91–104.