

# REVUE DE STATISTIQUE APPLIQUÉE

H. KHODJA

H. LUBÈS-NIEL

R. SABATIER

J.-M. MASSON

E. SERVAT

J.-E. PATUREL

**Analyse spatio-temporelle de données pluviométriques en  
Afrique de l'ouest. Recherche d'une rupture en moyenne.  
Une alternative intéressante : les tests de permutations**

*Revue de statistique appliquée*, tome 46, n° 1 (1998), p. 95-110

[http://www.numdam.org/item?id=RSA\\_1998\\_\\_46\\_1\\_95\\_0](http://www.numdam.org/item?id=RSA_1998__46_1_95_0)

© Société française de statistique, 1998, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

**ANALYSE SPATIO-TEMPORELLE DE DONNÉES  
PLUVIOMÉTRIQUES EN AFRIQUE DE L'OUEST.  
RECHERCHE D'UNE RUPTURE EN MOYENNE.  
UNE ALTERNATIVE INTÉRESSANTE :  
LES TESTS DE PERMUTATIONS.**

H. Khodja<sup>a</sup>, H. Lubès-Niel<sup>a</sup>, R. Sabatier<sup>b</sup>, J.-M. Masson<sup>c</sup>, E. Servat<sup>d</sup>, J.-E. Paturel<sup>d</sup>

<sup>a</sup> Laboratoire d'Hydrologie, ORSTOM, B.P. 5045, 34032 Montpellier cedex 1, France

<sup>b</sup> Laboratoire de Physique Moléculaire et Structurale, Faculté de Pharmacie, Université Montpellier I, 34060 Montpellier cedex 1, France

<sup>c</sup> Laboratoire Géofluides-Bassins-Eau, URA-CNRS 1767, Université Montpellier II, Place Eugène Bataillon, 34095 Montpellier cedex 5, France

<sup>d</sup> Antenne Hydrologique ORSTOM, 06 BP 1203 cidex 1, Abidjan 06, Côte d'Ivoire

**RÉSUMÉ**

Une étude de détection de ruptures en moyenne a été réalisée sur un ensemble de données pluviométriques de l'Afrique de l'ouest non sahélienne entre 1950 et 1980. L'approche spatio-temporelle généralisée au cas multivarié a été abordée par une analyse exploratoire multidimensionnelle. Des groupes de stations, caractérisés par des comportements pluviométriques similaires, ont ainsi été constitués. Un test de permutation, s'appuyant sur le critère de l'analyse des corrélations canoniques, a confirmé l'existence de différences significatives entre ces groupes et a permis une répartition homogène des stations. Ces études ont par la suite conduit à l'application d'un test multivarié de détection de rupture en moyenne grâce à une statistique bayésienne ainsi qu'à l'estimation des points de ruptures par la méthode du maximum de vraisemblance.

**Mots-clés :** *modèles multilinéaires gaussiens, rupture en moyenne, procédure bayésienne, vraisemblance, analyse exploratoire multidimensionnelle, test de permutation.*

**ABSTRACT**

This papers deals with detecting change in mean of rainfall data of different stations in West Africa from 1950 to 1980. The statistical approach is a spatial-temporal and multivariate study which requires an exploratory multidimensional analysis. Groups of stations with similar characteristics are then constituted. The canonical correlation analysis under the permutation hypothesis confirms a significant difference between these groups and defines an homogeneous distribution of the stations. On each group, a multivariate test for a shift in the mean, with a bayesian argument, and the estimate of a change point, by maximum likelihood, are applied.

**Keywords :** *multilinear gaussian model, change in mean, multivariate analysis, bayes procedure, likelihood, permutation test.*

## Introduction

De nombreux travaux consacrés à l'étude des mécanismes hydrométéorologiques de l'Afrique de l'Ouest et Centrale mettent l'accent sur les variations du régime pluviométrique observées dans ces régions depuis plus de vingt cinq ans et sur leurs conséquences souvent tragiques.

Caractérisées essentiellement par une baisse sensible de la pluviométrie vers la fin des années 60 et une diminution des apports en eau de surface, ces fluctuations déjà très importantes dans les régions sahéliennes, semblent s'étendre plus au sud, vers l'équateur.

C'est dans le cadre du programme ICCARE (Identification et Conséquences de la variabilité Climatique en AfRIque de l'ouest non sahélienne) du projet FRIEND (Flow Regimes for International and Experimental Network Data) AOC (Afrique de l'Ouest et Centrale) du Programme Hydrologique International de l'UNESCO, que plusieurs variables relatives à la pluie ont été prises en compte. On citera par exemple le cumul annuel des pluies, le nombre annuel de jours de pluie, la hauteur annuelle des pluies de la saison sèche ou encore la durée annuelle de la saison des pluies.

Pour caractériser la variabilité du régime pluviométrique de l'Afrique de l'Ouest, des tests univariés ont été mis en œuvre [12]. Ils ont conclu à des ruptures en moyennes sur la plupart des stations de cette région, mais les résultats obtenus sur chaque site de mesure et pour chacune des variables étudiées n'ont pas été toujours faciles à interpréter dans leur ensemble ni à synthétiser. C'est pourquoi, l'objectif de ce travail est de rechercher des ruptures en moyennes à partir d'une analyse spatio-temporelle généralisée au cas multivarié. Cette approche tridimensionnelle permettra alors d'une part de répartir spatialement les stations ayant subi des changements similaires sur une période d'étude commune et d'autre part de détecter des ruptures en moyenne caractérisées par plusieurs variables pluviométriques.

## 1. Traitement et analyse des données

### 1.1. Description des données

La zone d'étude se situe entre les longitudes 18° ouest et 28° est et les latitudes 2° et 14° nord. Elle recouvre 16 pays en bordure du Golfe de Guinée : Mali, Burkina Faso, Niger, Tchad, Centrafrique, Cameroun, Nigeria, Bénin, Togo, Ghana, Côte d'Ivoire, Liberia, Sierra Leone, Guinée Conakry, Guinée Bissau et Sénégal.

En chacune des 99 stations pluviométriques sont considérées entre 1950 et 1980 et avec un pas de temps annuel, quatre variables quantitatives : cumul des pluies, nombre de jours de pluie, durée de la saison des pluies et hauteur de pluie de la saison sèche, notées par la suite, variables 1, 2, 3 et 4 respectivement. Les mesures, à la base de la définition des variables, sont effectuées à des pas de temps journaliers et décennaires. Cette période d'observations a permis de sélectionner un nombre maximal de stations pour les quatre variables prises en compte.

### 1.2. Analyse du cube des données : STATIS [11]

L'objectif est de résumer, par une méthode d'analyse exploratoire multidimensionnelle, l'information brute décrite ci-dessus en constituant des groupes de stations à l'intérieur desquels des comportements pluviométriques similaires vis à vis de l'ensemble des variables sont mis en évidence.

#### a) Organisation des tableaux et définition des études

Parmi toutes les méthodes d'analyse d'un cube de données [4], celle retenue pour notre étude est STATIS (Structuration des Tableaux A Trois Indices de la Statistique) [11]. Il s'agit d'une analyse conjointe de  $K$  tableaux de données quantitatives, relevées en différentes stations, où  $p$  mêmes variables (colonnes) ont été mesurées sur les mêmes  $N$  individus (lignes), ici  $N$  années.

Les tableaux des données s'organisent alors de la façon suivante (voir figure page suivante).

La démarche que nous allons aborder consiste à comparer les tableaux entre eux et à fournir des graphiques mettant en évidence les ressemblances et dissemblances. C'est donc à la première étape de STATIS, sur les trois qu'elle comporte, que nous allons nous intéresser.

L'approche se limite à la recherche de l'*interstructure*, c'est-à-dire à la définition des relations qui existent entre les tableaux  $X_k$ , mesurés à la station  $k$ , de dimensions  $n \times p$  et à sa représentation graphique.

Les études retenues pour cette analyse sont les triplets  $(X_k, Q, D)$  pour  $1 \leq k \leq K$  où  $Q$  est la métrique des poids des lignes et  $D$  celle des colonnes, c'est-à-dire  $I_p$  et  $\frac{1}{N}I_N$  respectivement, et sont affectées du même poids  $\frac{1}{K}$ . De plus, chaque tableau est bicentré afin de supprimer les effets moyens dûs aux années et aux variables et réduit en colonnes car les variables ont des ordres de grandeur très différents.

Si  $x_{ijk}$  est la valeur de la variable  $i$  à l'instant  $j$  pour la station  $k$ , la transformation effectuée revient à remplacer  $x_{ijk}$  par  $\frac{y_{ijk}}{s_{ik}}$  où :

$$y_{ijk} = x_{ijk} - \overline{x_{jk}} - \overline{x_{ik}} + \overline{x_k}$$

avec

$$\overline{x_{jk}} = \frac{1}{p} \sum_{i=1}^p x_{ijk}, \overline{x_{ik}} = \frac{1}{N} \sum_{j=1}^N x_{ijk}, \overline{x_k} = \frac{1}{Np} \sum_{i=1}^p \sum_{j=1}^N x_{ijk}$$

et

$$s_{ik}^2 = \frac{1}{N} \sum_{j=1}^N y_{ijk}^2$$

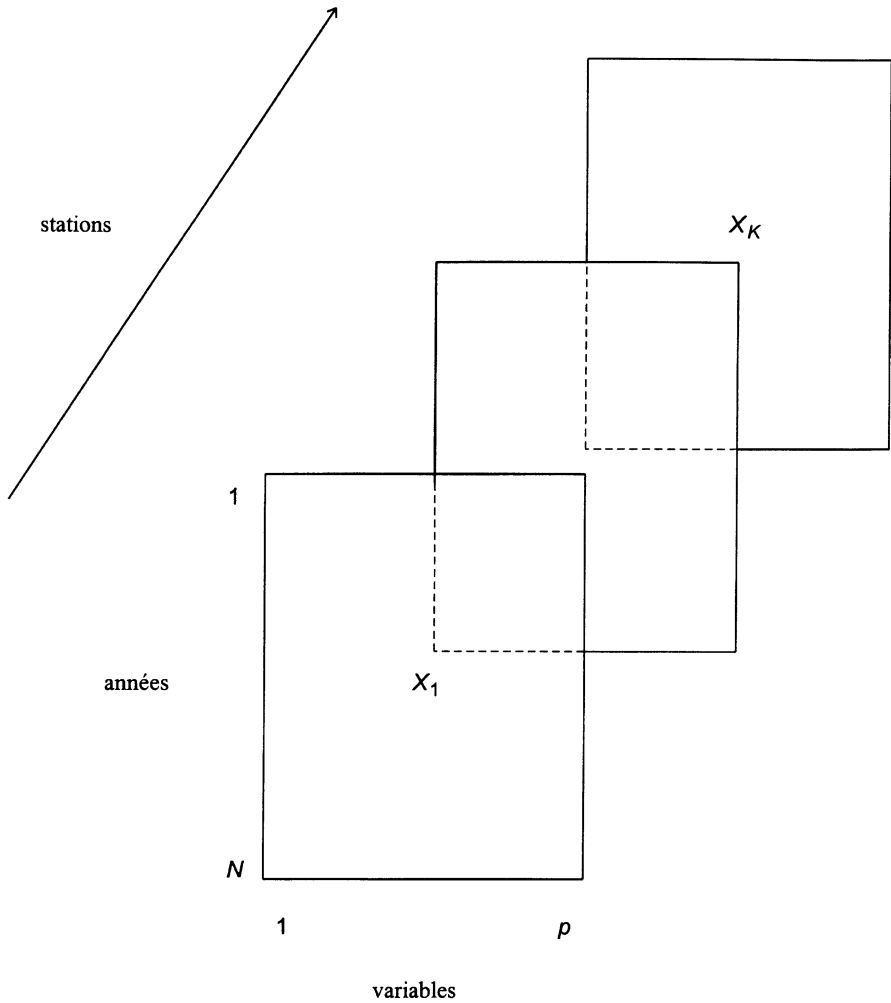


FIGURE 1  
Présentation des tableaux de données

Cette démarche qui consiste à prendre en compte un effet variable définit un profil moyen annuel. Conjugué avec l'effet année, celui ci oriente la recherche du compromis.

*b) Définition du coefficient d'association entre études*

Les relations qui existent entre les tableaux  $X_k$  et  $X_{k'}$  (ou entre les études  $(X_k, Q, D)$  et  $(X_{k'}, Q, D)$ ) sont décrites par les distances entre les  $W_k$  qui sont les matrices des produits scalaires entre les  $N$  observations calculés par :  $W_k = X_k Q X_k'$ .

La matrice des produits scalaires normés entre les objets  $W_k$  est celle des *coefficients de corrélations vectorielles* (ou *RV* d'Escoufier) dont le terme général s'écrit :

$$RV(X_k, X_{k'}) = \text{trace}(W_k W_{k'}) / \sqrt{\text{trace}(W_k^2) \text{trace}(W_{k'}^2)}.$$

C'est sur cette matrice des coefficients de corrélations vectorielles qu'un positionnement multidimensionnel classique est réalisé en retenant les  $r$  vecteurs propres correspondant aux  $r$  plus grandes valeurs propres afin d'obtenir le meilleur résumé possible de l'information détenue par cette matrice.

### 1.3. Résultats

La représentation des stations dans le premier plan principal permet de former trois groupes à l'intérieur desquels les stations ont des comportements similaires au regard des variables étudiées (voir figure 2).

Globalement, et pour simplifier, on peut résumer les groupes autour des noyaux suivants :

groupe I : Togo, Bénin et sud de la Côte d'Ivoire.

groupe II : Cameroun, Centrafrique et Burkina Faso.

groupe III : Guinée Conakry, Guinée Bissau, Mali et Sénégal.

La carte obtenue résume donc les comportements pluviométriques de l'ensemble des stations de la zone étudiée entre 1950 et 1980, tels qu'ils ont été décrits par quatre variables relatives à la pluie (voir figure 3).

Ainsi, il est possible d'associer à chaque groupe de postes pluviométriques des caractéristiques communes qui sont analysées par la suite vis-à-vis de ruptures en moyenne potentielles. L'organisation spatiale des stations fait apparaître assez nettement une répartition nord-sud d'une part et est-ouest d'autre part.

Il est à déplorer cependant l'absence de données au Ghana et au Nigeria qui marquent une discontinuité certaine dans la répartition des stations. Quoique cette contrainte soit pénalisante au niveau des interprétations spatiales, elle n'enlève pas l'intérêt à porter sur le reste de la zone d'étude.

## 2. Validation statistique des groupes : Les tests de permutations

Le pourcentage d'inertie expliquée par les deux premiers axes étant assez faible (20.13 %), on désire appliquer un test de permutation permettant de vérifier si les groupes obtenus par STATIS sont significativement différents vis-à-vis des quatre variables étudiées.

### 2.1. Les tests de permutations : principe général [1] et [6]

La majorité des tests non paramétriques supposent la connaissance approximative, généralement asymptotique, ou même exacte des distributions qui interviennent

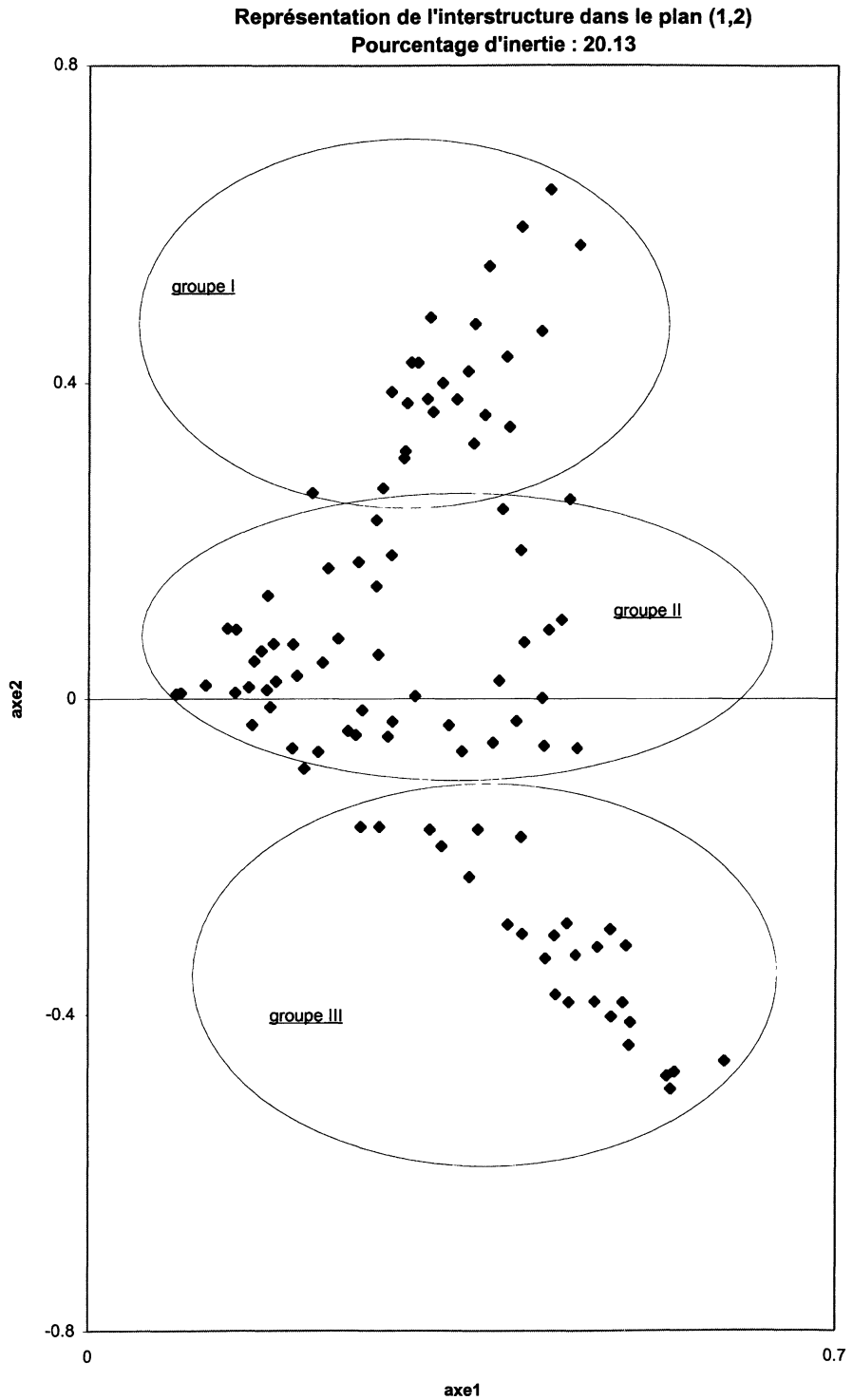
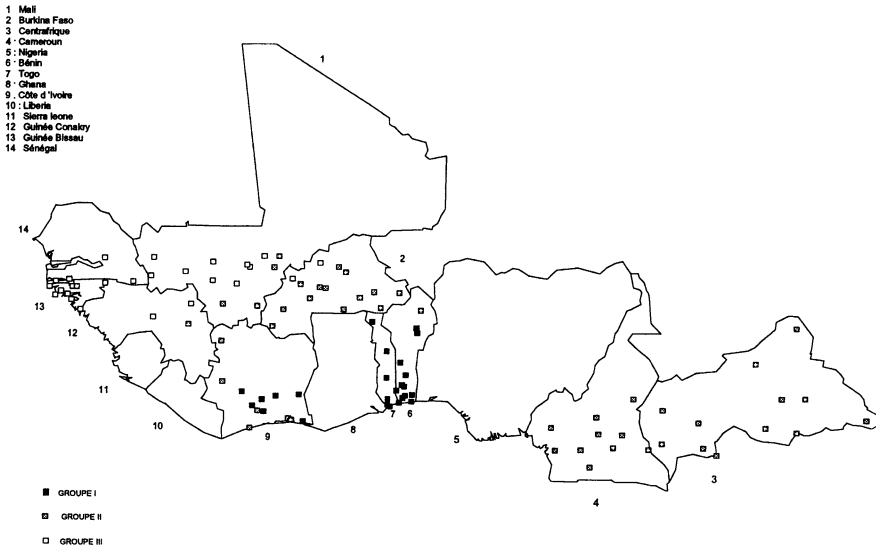


FIGURE 2

*Représentation de l'interstructure dans le premier plan principal*



**FIGURE 3**  
*Répartition spatiale des stations et identification des trois groupes  
 obtenus par l'analyse exploratoire multidimensionnelle*

dans leur réalisation. Souvent, celles-ci sont difficiles à obtenir même sous l'hypothèse gaussienne. Sous l'hypothèse de permutabilité des observations, la statistique observée à partir des données initiales est comparée à toutes celles obtenues après permutations.

Si l'on dispose de  $N$  observations iid alors les  $N!$  permutations possibles de l'ensemble des observations sont équiprobables. Cependant, toutes les permutations étant en pratique irréalisables, on procède au tirage aléatoire d'un certain nombre d'entre elles.

La probabilité de dépassement  $\hat{\alpha}$  est alors la proportion des permutations pour lesquelles les valeurs calculées de la statistique sont supérieures à la valeur observée initialement<sup>1</sup>. Elle est d'autant meilleure, (sa précision suit une loi binomiale), que le nombre de permutations est grand. La valeur exacte de la probabilité n'est obtenue que lorsque toutes les permutations sont effectuées.

L'hypothèse nulle de permutabilité est alors rejetée avec un risque  $\alpha$  si :  $\hat{\alpha} \leq \alpha$ .

### 2.2. Le principe de Mardia

Ces tests peuvent être obtenus par la méthode de Monte Carlo qui permet de générer aléatoirement un échantillon de permutations, mais l'intérêt des tests de permutation, tels qu'ils ont été présentés par Mardia, est qu'il est rendu possible, sous l'hypothèse de permutabilité des observations, de calculer les valeurs exactes des

<sup>1</sup> ou inférieures si la valeur observée est proche de la borne inférieure des valeurs obtenues après permutations



premiers moments de la statistique de test choisi ([13] pour les deux premiers moments et [9] pour le moment d'ordre 3) *en ne réalisant explicitement aucune permutation*. Initialement, ce test était appliqué pour étudier les relations entre tableaux de données à partir de la statistique de l'Analyse des Corrélations Canoniques [13], mais cette approche a été généralisée et trois autres statistiques ont été exhibées pour tester l'indépendance entre tableaux [9].

La statistique qui intervient dans le test peut alors être approximée par une distribution connue.

En notant  $S_{obs}$ , la valeur de la statistique calculée à partir des données initiales,  $E_p(S)$  et  $V_p(S)$  respectivement les valeurs exactes de la moyenne et de la variance de la statistique sous l'hypothèse de permutabilité, l'écart à la moyenne réduit :

$$z = \frac{S_{obs} - E_p(S)}{\sqrt{V_p(S)}}$$

peut être ajusté à une loi normale centrée-réduite, à une loi Beta ou encore à une distribution de Pearson de type III si le moment d'ordre 3 le suggère.

L'hypothèse nulle de permutabilité est alors rejetée avec un risque  $\alpha$  si :  $\hat{\alpha} \leq \alpha$  où  $\hat{\alpha}$  est l'estimation de la probabilité de dépassement de  $z$ .

### 2.3. Application aux données

Le test de permutation, tel qu'il a été décrit ci-dessus, va être utilisé pour étudier les relations qui existent entre deux tableaux de données et plus précisément leur indépendance [7].

#### a) Organisation des tableaux de données (cf. figure 4)

Les tableaux sur lesquels le test de permutation s'applique sont les suivants :

- $Y$  est le tableau croisant en lignes les stations classées par groupes et en colonnes les variables pluviométriques et dans lequel  $Y_{ij}$  est la moyenne, calculée sur l'ensemble des années, de la variable  $j$  pour la station  $i$ .

- $X$  est le tableau croisant en lignes les stations classées par groupes et en colonnes les indicatrices des groupes, où

$$\begin{cases} X_{ij} = 0 & \text{si la } i^{\text{ème}} \text{ station n'appartient pas au groupe } j. \\ X_{ij} = 1 & \text{si la } i^{\text{ème}} \text{ station appartient au groupe } j. \end{cases}$$

Il s'agit donc de tester l'hypothèse nulle  $H_0$  : les tableaux sont indépendants contre l'hypothèse alternative  $H_1$  : les tableaux sont liés.

L'hypothèse  $H_0$  se traduit encore en termes de permutabilité des lignes de l'un des deux tableaux.

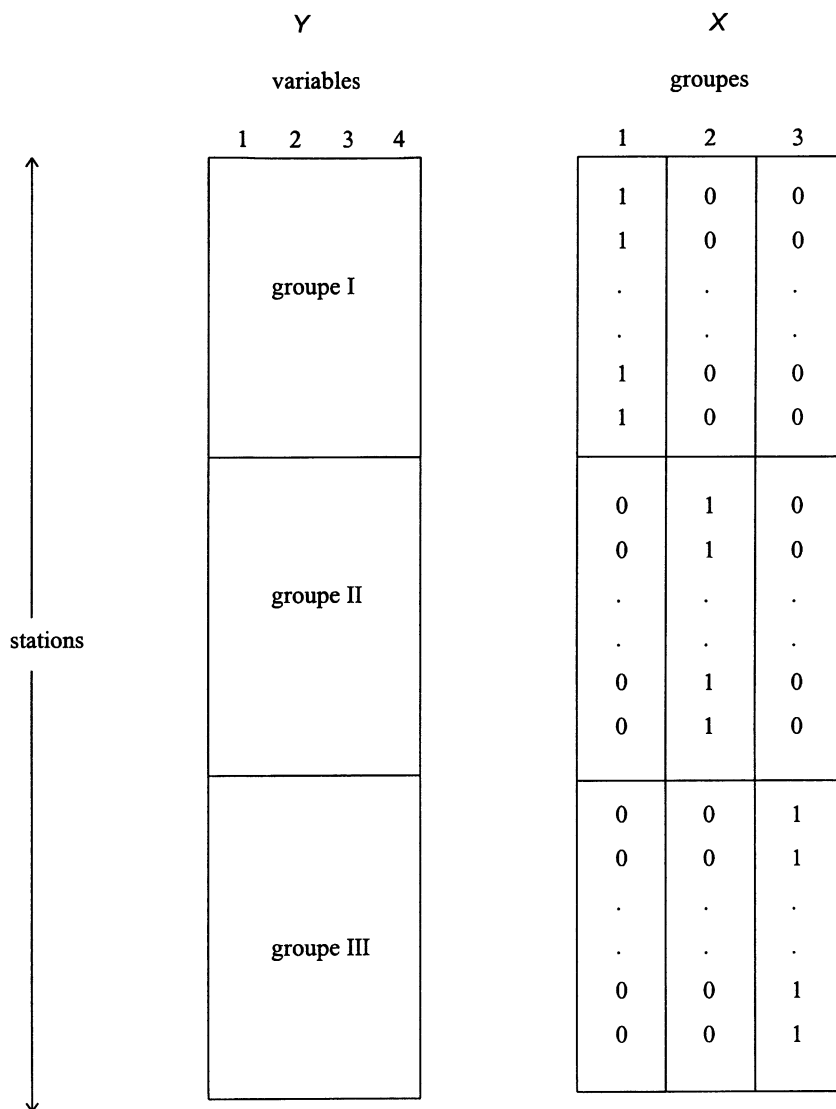


FIGURE 4  
 Présentation des tableaux sur lesquels est effectué  
 le test de permutation

*b) Le critère de l'Analyse Canonique des Corrélations*

On dispose ainsi de deux tableaux de données où deux groupes de variables différents (variables pluviométriques et indicatrices des groupes) ont été mesurés sur les mêmes unités statistiques (stations classées par groupes).

Parmi toutes les statistiques utilisées pour étudier l'indépendance de deux tableaux de données, c'est celle de l'Analyse Canonique des Corrélations (ACC) qui a été retenue ici. Elle semble en effet être la plus appropriée à nos données car son approche symétrique permet l'étude des relations linéaires entre les groupes de variables observés sur les mêmes individus en accordant aux deux tableaux des considérations égales.

Elle est donnée par :

$$\begin{aligned} S_{ACC} &= \text{trace}(X(X'X)^{-1}X'Y(Y'Y)^{-1}Y') \\ &= \text{trace}(P_X P_Y), \end{aligned}$$

$Y$  et  $X$  étant préalablement centrés en colonnes.

$P_X$  (resp  $P_Y$ ) représente le projecteur orthogonal sur  $Im(X)$  (resp  $Im(Y)$ ), espace vectoriel engendré par les colonnes de  $X$  (resp  $Y$ ).

Les résultats du test d'indépendance des tableaux  $Y$  et  $X$  sont résumés ci-dessous.

TABLEAU 1  
*Résultats du test de permutation*

Valeur de la statistique $S_{ACC}$	0.4564
Moyenne	0.0816
Ecart type	0.0389
Coefficient de dissymétrie	0.8938
Ecart réduit à la moyenne	9.6142
Probabilité de dépassement associée à une distribution Gaussienne	$< 10^{-6}$
Probabilité de dépassement associée à une distribution Bêta	$< 10^{-6}$
Probabilité de dépassement associée à une distribution de Pearson de type III	$< 10^{-6}$

Ainsi, l'hypothèse  $H_0$  est rejetée quelle que soit la distribution d'ajustement choisie et on conclut donc à l'existence de différences significatives entre les trois groupes.

### 3. Etude multivariée de détection d'une rupture en moyenne

#### 3.1. Tests univariés : possibilités et limites

Ce paragraphe reprend – très succinctement et n'est en aucun cas une étude bibliographique des tests univariés de détection de rupture en moyenne – les principaux tests univariés de rupture en moyenne retenus dans le programme ICCARE [12] et qui ont déjà été appliqués aux analyses de séries de précipitations et de débits de l'Afrique de l'Ouest.

– Statistique de Buishand [2] et [3]. Son approche Bayésienne est performante pour détecter une rupture en moyenne en milieu de série, mais il ne permet pas de donner la date de la rupture.

– Procédure Bayésienne [10]. La supposition *a priori* d'un changement en moyenne permet de donner la probabilité que celui-ci se produise à un moment donné ainsi que son amplitude.

– Test de Pettitt [5] et [16]. Réputé pour sa robustesse, ce test s'appuie sur le découpage de la série en deux sous échantillons.

– Test de segmentations de Hubert [8]. Il est plus approprié à la détection de plusieurs ruptures en moyenne car découpe la série d'observations en plusieurs sous échantillons.

Contrairement aux deux premiers tests, ces derniers ne sont pas paramétriques et aucun d'entre eux n'impose une longueur minimale à la série.

Appliqués à chaque site de mesure et variable par variable, ces tests donnent des résultats souvent concordants quant à l'existence de ruptures en moyennes dans les années 1960 et 1970. Ils restent cependant difficiles à synthétiser dans leur ensemble du point de vue de l'organisation spatiale des réponses aux variations pluviométriques vis-à-vis de l'ensemble des variables [15].

#### 3.2. Test multivarié de rupture en moyenne : fondement

Soient  $x_1, \dots, x_N$ ,  $N$  vecteurs aléatoires obéissant au modèle multilinéaire suivant :  $x_i = \mu_i + \varepsilon_i, i = 1, \dots, N$ , où

$$x_i = \begin{pmatrix} x_i^1 \\ \vdots \\ x_i^p \end{pmatrix}, \mu_i = \begin{pmatrix} \mu_i^1 \\ \vdots \\ \mu_i^p \end{pmatrix}, \varepsilon_i = \begin{pmatrix} \varepsilon_i^1 \\ \vdots \\ \varepsilon_i^p \end{pmatrix}, \varepsilon_i \approx N_p(0, \Sigma), \varepsilon_i \text{ iid}, p \text{ étant le nombre de variables.}$$

Il s'agit de tester l'hypothèse d'égalité des moyennes,

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_N,$$

contre l'hypothèse alternative de rupture en moyenne en  $r$ ,

$$H_1 : \mu_1 = \mu_2 = \dots = \mu_r \neq \mu_{r+1} = \mu_{r+2} = \dots = \mu_N$$

dans le cas où  $r$  et  $\Sigma$  sont inconnus.

Srivastava et Carter [18] proposent la statistique bayésienne suivante :

$$S = u'V^{-1}u$$

avec

$$u' = \frac{1}{\sqrt{a'a}} \sum_{i=1}^{N-1} i(x_{i+1} - \bar{x}), a' = \frac{1}{2}(- (N-1), - (N-3), - (N-5), \dots, (N-5), (N-3), (N-1))$$

$$V = \sum_{i=1}^N (x_i - \bar{x})(x_i - \bar{x})' \text{ et } \bar{x} = N^{-1} \sum_{i=1}^N x_i.$$

L'hypothèse d'égalité des moyennes  $H_0$  est ainsi rejetée au niveau  $\alpha$  si :

$$S \geq c_\alpha \text{ où } c_\alpha = 1 - [1 + p(N-p-1)^{-1} F_{p, N-p-1}^\alpha]^{-1}$$

avec

$$\alpha = pr[F_{p, N-p-1} \geq F_{p, N-p-1}^\alpha], F_{p, N-p-1}$$

étant la distribution d'une loi de Fisher à  $p$  et  $N-p-1$  degrés de liberté.

### 3.3. Estimation du point de rupture

Si l'hypothèse  $H_0$  est rejetée on va chercher à estimer par la méthode du maximum de vraisemblance la position de la rupture, en se plaçant toujours dans le cadre du modèle multilinéaire gaussien.

Soient  $\bar{x}_r, \bar{x}_{N-r}$  et  $\hat{\Sigma}$  les estimateurs du maximum de vraisemblance, sous l'hypothèse  $H_1$ , de  $\mu_1, \mu_N$  et  $\Sigma$  respectivement.

En faisant varier  $r$  entre 1 et  $N-1$ , on considère les couples d'échantillons  $(x_1, x_2, \dots, x_r)$  et  $(x_{r+1}, x_{r+2}, \dots, x_N)$ , pour lesquels on calcule les estimations des distances de Mahalanobis données par :

$$D_r^2 = (\bar{x}_r - \bar{x}_{N-r})' \hat{\Sigma}^{-1} (\bar{x}_r - \bar{x}_{N-r}).$$

Alors, le point de rupture estimé  $\hat{r}$  est tel que  $D_r^2$  soit maximale.

En notant :

$$\tilde{\Sigma} = \sum_{i=1}^r (x_i - \bar{x}_r)(x_i - \bar{x}_r)' + \sum_{i=r+1}^N (x_i - \bar{x}_{N-r})(x_i - \bar{x}_{N-r})'$$

«la matrice de variance covariance intra-groupes»,  $\hat{\tau}$  est encore celui qui maximise la statistique  $T_r^2 = N^{-1}r(N-r)(\bar{x}_r - \bar{x}_{N-r})'\tilde{\Sigma}^{-1}(\bar{x}_r - \bar{x}_{N-r})$ .

#### 4. Résultats de l'étude multivariée de détection de ruptures en moyennes

L'étude de détection de ruptures en moyennes s'effectue sur chacun des trois groupes, obtenus à l'issue de l'analyse exploratoire multidimensionnelle et validés par les résultats du test de permutation.

Le test multivarié de changement en moyenne et l'estimation du point de rupture s'appliquent aux séries des moyennes intergroupes pour chaque variable entre 1950 et 1980.

D'après les applications du test citées dans la littérature [18], cette courte période d'observations – 31 ans – ne doit pas mettre en défaut les performances du test.

Les résultats obtenus figurent ci-dessous.

TABLEAU 2  
Résultats du test de détection de rupture en moyenne  
et de l'estimation du point de rupture

valeur critique : 0.2967		groupe I	groupe II	groupe III
var 1, 2, 3 et 4	valeur de S	0.305	0.4463	0.6946
	année de rupture	1962	1969	1967

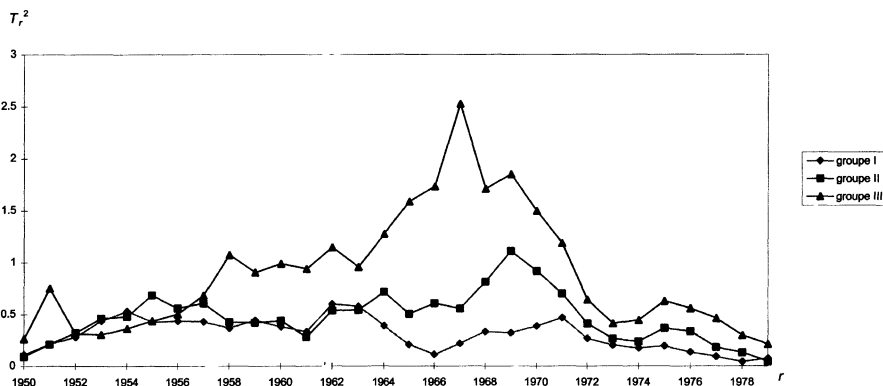


FIGURE 5  
Graphes de  $T_r^2$  en fonction de  $r$

La valeur critique  $c_\alpha$  au-delà de laquelle l'hypothèse  $H_0$  d'égalité des moyennes est rejetée avec un risque de première espèce  $\alpha = 5\%$  étant égale à 0.2967, on conclut donc à l'existence d'une rupture en moyenne sur les groupes I, II et III respectivement en 1962, 1969 et 1967 et c'est sur le groupe III qu'elle est la plus significative.

**TABLEAU 3**  
*Résultats du test de détection de rupture en moyenne  
 et de l'estimation du point de rupture relatifs aux différentes études*

valeur critique : 0.2475		groupe I	groupe II	groupe III
var 2, 3 et 4	valeur de S	0.2903	0.4161	0.6764
	année de rupture	1954	1969	1967
var 1, 3 et 4	valeur de S	0.2959	0.4386	0.6552
	année de rupture	1962	1969	1967
var 1, 2 et 4	valeur de S	0.1226	0.4071	0.6893
	année de rupture	pas de rupture	1969	1967
var 1, 2 et 3	valeur de S	0.2906	0.428	0.6881
	année de rupture	1954	1969	1967
valeur critique : 0.1926		groupe I	groupe II	groupe III
var 3 et 4	valeur de S	0.288	0.415	0.564
	année de rupture	1962	1969	1967
var 2 et 4	valeur de S	0.034	0.171	0.628
	année de rupture	pas de rupture	pas de rupture	1967
var 2 et 3	valeur de S	0.279	0.399	0.667
	année de rupture	1954	1969	1967
var 1 et 4	valeur de S	0.08	0.389	0.647
	année de rupture	pas de rupture	1969	1967
var 1 et 3	valeur de S	0.281	0.427	0.655
	année de rupture	1954	1969	1967
var 1 et 2	valeur de S	0.1	0.388	0.683
	année de rupture	pas de rupture	1969	1967

Les ruptures détectées en 1967 et 1969 confirmeraient le début d'une période déficitaire, celle du groupe I en 1962, traduirait un déficit très peu significatif.

Tout en conservant une approche multivariée d'étude de détection de rupture en moyenne et afin de caractériser plus précisément les changements pluviométriques de ces régions, les mêmes analyses statistiques ont été effectuées sur trois puis sur deux variables parmi les quatre retenues initialement.

Les conclusions de ces différentes études se rapprochent des résultats obtenus précédemment.

Le groupe I ne semble pas avoir subi de modification du régime pluviométrique de façon très notable. En effet, lorsque le test détecte une rupture en moyenne, elle reste très peu significative quel que soit le nombre de variables retenues. Par contre, les changements pluviométriques des groupes II et III se confirment et ce sont les années 1969 et 1967 qui caractérisent l'apparition du phénomène sensible sur chaque groupe de variables.

## 5. Conclusion

S'il est vrai que la littérature consacrée aux séries chronologiques et plus particulièrement à l'étude de détection de rupture en moyenne est abondante [12], il n'en reste pas moins que les approches multivariées sont plus rarement abordées.

Associée à un travail préliminaire d'analyse exploratoire multidimensionnelle, l'étude multivariée de détection de rupture en moyenne présentée dans cet article a permis de caractériser, dans leur globalité, les modifications pluviométriques survenues en Afrique de l'Ouest francophone non sahélienne.

C'est ainsi que se sont dégagés trois groupes de régimes pluviométriques caractérisés par des comportements spécifiques vis-à-vis de l'ensemble des variables étudiées et par des réponses communes au test de rupture en moyenne associées aux dates respectives d'apparition du phénomène.

Au-delà de la mise en évidence de fluctuations importantes dans ces régions, viennent s'ajouter des difficultés d'acheminement, de distribution et d'approvisionnement en eau. L'urgence réside donc certainement dans une optimisation de la gestion de l'eau, source de vie de l'économie africaine.

A défaut de prévoir et de prévenir l'apparition de tels changements pluviométriques, il semble donc incontestable de tout mettre en œuvre pour une meilleure caractérisation des changements passés pouvant conduire à la fois à une plus grande compréhension des mécanismes pluviométriques et à une démarche plus fiable de définition de scénarii de gestion. Une meilleure maîtrise des ressources en eau en dépend.

## Références bibliographiques

- [1] BIONDINI M.E., MIELKE P.W., REDENTE E.F., Permutation techniques based on euclidian analysis : a new and powerful statistical method for ecological research. *Coenoses*, 3, 155-174, 1988.
- [2] BUISHAND T.A., Some methods for testing the homogeneity of rainfall records. *Journal of Hydrology*, vol. 58, 11-27, 1982.
- [3] BUISHARD T.A., Tests for detecting a shift in the mean of hydrological time series. *Journal of Hydrology*, vol. 58, 51-69, 1984.



- [4] CARLIER A., LAVIT C., PAGES M., PERNIN M.O., TURLLOT J.-C., A comparative review of methods which handle a set of indexed data tables. *Multway data Analysis*, 85-101, 1989.
- [5] CERESTA (Centre d'Enseignement et de Recherche Statistique Appliquée), Aide-mémoire pratique des techniques statistiques pour ingénieurs et techniciens supérieurs. *Revue de Statistique Appliquée*, vol. XXXIV, numéro spécial, 1986.
- [6] EDGINGTON E.S., *randomization tests*. Marcel Dekker, New York, 1987.
- [7] HITIER S., *Quelques éléments pour tester l'indépendance de deux tableaux de données*. Rapport de DEA. Ecole Nationale Supérieure d'Agronomie de Montpellier, 1993.
- [8] HUBERT P., CARBONNEL J.-P., CHAUCHE A., Segmentation des séries hydrométéorologiques. Application à des séries de précipitations et de débits de l'Afrique de l'Ouest. *Journal of Hydrology*, vol. 110, 349-367, 1989.
- [9] KAZI-AOUAL F., HITIER S., SABATIER R., LEBRETON J.-D., Refined approximations to permutation tests for multivariate inference. *Computational statistics & Data Analysis*, 20, 643-656, 1994.
- [10] KOTZ S., JOHNSON N.L., READ C.B., *Encyclopedia of statistical sciences*, New York, Wiley, vol. 1, 197-205, vol. 8, 157-163, vol. 9, 244-255, 1981.
- [11] LAVIT C., *Analyse conjointe de tableaux quantitatifs*. Editions MASSON, 1988.
- [12] LUBES H., MASSON J.-M., SERVAT E., PATUREL J.-E., KOUAME B., BOYER J.-F., *Caractérisation de fluctuations dans une série chronologique par applications de tests statistiques. Etude bibliographique*. Programme ICCARE, Rapport 3. ORSTOM-DEC, UD2, 1994.
- [13] MARDIA K.V., The effect of nonnormality on some multivariate tests and robustness to nonnormality in the linear model. *Biometrika*, 58, n°1, 105-122, 1971.
- [14] MUIRHEAD R.J., *Aspects of multivariate statistical theory*. Wiley series in probability and mathematical statistics, 1982.
- [15] PATUREL J.E., SERVAT E., KOUAME B., LUBES H., OUEDRAOGO M., MASSON J.-M., Climatic variability in humid Africa along the Gulf of Guinea. Part two : An integrated regional approach. *Journal of Hydrology*, Vol. 191, 16-36, 1997.
- [16] PETTITT A.N., A non parametric approach to the change point problem. *Applied Statistics*, 28, n°2, 126-135, 1979.
- [17] SEBER G.A.F., *Multivariate observations*, Wiley series in probability and mathematical statistics, 1984.
- [18] SRIVASTAVA M.S., CARTER M., *An Introduction to Applied Multivariate Statistics*. Elsevier Science Publishing, 1983.