

L'homogénéité des données pluviométriques

G. HIEZ

Directeur de recherches ORSTOM
Université de Fortaleza (Brésil)

RÉSUMÉ

Un réseau d'observations pluviométriques peut être comparé à un système d'information et, comme dans celui-ci, les messages délivrés sont susceptibles de subir des altérations de différentes natures ; celles qui peuvent affecter un total pluviométrique sont énumérées brièvement.

L'un des moyens d'analyse les plus éprouvés pour tester l'homogénéité de l'information pluviométrique recueillie repose sur l'utilisation des sommes cumulées (simples et doubles) ; on tente d'en ébaucher une théorie qui permet de souligner le caractère linéaire de la liaison entre deux séries de totaux cumulés, et de montrer que cette liaison est régie par un processus de Markov de premier ordre. L'emploi de ce procédé implique toutefois une lourde manipulation des données et le recours à des choix qui peuvent ne pas être objectifs.

La « pseudo-proportionnalité » des totaux pluviométriques entre postes permet d'envisager, à partir de l'ensemble des données, considéré comme une matrice, l'élaboration d'un vecteur de référence à composantes annuelles — « le vecteur régional » —, auquel on viendra comparer chacun des postes. On montre que cette élaboration ne peut s'appuyer sur des méthodes conventionnelles, comme celle des composantes principales, ni sur des critères utilisant la minimisation en norme euclidienne, mais sur un principe de maximum de vraisemblance basé sur le postulat que l'information la plus vraisemblable est celle qui se répète le plus fréquemment.

La méthode proposée repose donc essentiellement : — sur le choix d'une formule d'estimation de la valeur modale, en l'absence d'hypothèse sur la forme des fonctions de densité ; — sur un procédé original d'extraction du vecteur régional de référence à partir de la matrice (en général rectangulaire) des données, par itération successive sur les lignes et sur les colonnes.

Deux traitements optionnels sont proposés pour répondre aux cas de distributions plurimodales des composantes des vecteurs-lignes ou des vecteurs-colonnes.

La critique des données de chaque poste se fait à partir des écarts relatifs au poste de référence ; pour la détection des séquences hétérogènes, on utilise le simple cumul de ces écarts relatifs.

Le programme est décrit schématiquement et un certain nombre de recommandations définissent les conditions d'application à partir d'exemples choisis au Brésil.

ABSTRACT

A rainfall station network is similar to a data system where the message sent may be altered diversely. The alterations which may affect a rainfall total are listed briefly below.

One of the most tested means of analysis to check the homogeneity of the rainfall data makes use of the cumulated totals i.e. single and double mass curves. An attempt is made at developing a theory which stresses the linearity of the connection between two series of cumulated totals and shows that this connection is governed by a first-order Markov process. The use of that process requires however a substantial handling of data and the use of choices which may not be objective.

The pseudo-proportionality of rainfall amounts between several posts allows for the elaboration of a reference vector with its annual components called "the regional vector" from the available data taken as a matrix. Each station would then be checked up against such a vector. This elaboration cannot be backed up by the conventional methods exclusively (such as the main components) nor may it make use of the criteria based on the minimization in Euclidian norm. Actually at the very basis of this elaboration lies the principle of the maximum likelihood which in turn lies on the postulate that the most likely information is that which repeats itself the most frequently.

The method suggested lies basically (a) on the choice of a formula to estimate the modal value in the lack of

hypothesis on the form of the density functions, (b) on an original process of extraction of the regional reference vector, from the matrix, which is usually rectangular, of data by successive iteration on the lines and columns.

Two optional treatments are suggested for the multimodal distributions of the components of the line-vectors or the column-vectors.

The critical examination of the data from each station operates with the relative deviations at the reference station ; the single mass of the relative deviations is used for the identification of the heterogeneous sequences.

The program is presented in a simplified form ; a set of recommendations determine the conditions of application from examples taken in Brazil.

RESUMO

A rede de observações pluviométricas assemelha-se a um sistema de informações onde as mensagens transmitidas são suscetíveis de sofrer alterações de natureza diversa. Enumeradas abaixo encontram-se as alterações suscetíveis de alterar o total pluviométrico.

Um dos meios de análise mais comprovados para testar a homogeneidade das informações pluviométricas colhidas recorre à utilização das somas cumuladas (simples e duplas) ; tenta-se esboçar uma teoria que permita destacar a linearidade da ligação entre duas séries de totais cumulados e mostrar que esta ligação é regida pelo processo de Markov de primeira ordem. O emprego do referido processo implica contudo uma substancial manipulação dos dados e o recurso a escolhas que podem não ser objetivas.

A « pseudoproporcionalidade » dos totais pluviométricos entre estações permite encarar, a partir do conjunto dos dados considerados como uma matriz, a elaboração de um vetor de referência com componentes anuais denominados « vetor regional » ao qual cada uma das estações seria comparada. Mostra-se que essa elaboração não pode apoiar-se em métodos convencionais — como os componentes principais — bem como não pode recorrer a critérios que utilizem a minimização em norma euclidiana mas num princípio de máximo de probabilidade baseado no postulado de que a informação mais provável é a que mais frequentemente se repete.

O método sugerido é baseado sobretudo : (a) na escolha de uma fórmula de estimação do valor modal, na ausência de hipótese sobre a forma das funções de densidade ; (b) num processo original de extração do vetor regional de referência, a partir da matriz (geralmente retangular) dos dados por iteração sucessiva em linhas e colunas.

Dois tratamentos opcionais são sugeridos para atender o caso de distribuições multimodais dos componentes dos vetores-linhas ou dos vetores-colunas.

A crítica dos dados de cada estação é feita a partir dos afastamentos relativos à estação de referência ; para a identificação das seqüências heterogêneas utiliza-se o simples cúmulo dos afastamentos relativos.

O programa é descrito sumariamente e algumas recomendações são apresentadas definindo as condições de aplicação a partir de exemplos colhidos no Brasil.

SOMMAIRE

0. Introduction

1. L'information pluviométrique

- 1.1. Le signal
- 1.2. Les « bruits »
- 1.3. L'analyse du signal

2. Essai sur la théorie des sommes cumulées

- 2.1. Le simple cumul
- 2.2. Le double cumul
 - 2.2.1. Poste à poste
 - 2.2.2. Entre un poste et un vecteur de base

3. *Elaboration d'un vecteur de référence*
 - 3.1. Les principes directeurs
 - 3.2. Le tableau des données considéré comme une matrice
 - 3.3. Déficience de la moyenne pour caractériser une tendance centrale
 - 3.4. La valeur modale
 - 3.5. Estimation de la valeur modale
 - 3.5.1. Formule classique simple
 - 3.5.2. Par l'utilisation du coefficient de dissymétrie
 - 3.5.3. Par les développements en série limités
 - 3.6. Mise en œuvre de M sur une matrice
 - 3.6.1. Le traitement « L.C. »
 - 3.6.2. Décomposition en sous-matrices
 - 3.6.3. Correction itérative
 - 3.6.4. Sensibilité et fiabilité de l'algorithme
4. *La critique des données*
 - 4.1. Distribution des EPSI
 - 4.2. Les erreurs relatives cumulées
5. *Description sommaire du programme*
6. *Conditions d'application*
7. *Exemples*
8. *Bibliographie*

0. INTRODUCTION

L'exploitation des ressources agricoles ainsi que l'alimentation des agglomérations urbaines ou la satisfaction des besoins en eau, toujours croissants, des concentrations industrielles, d'une manière générale, l'inventaire des ressources hydriques et leur gestion rationnelle sont étroitement subordonnés à une parfaite connaissance des éléments du climat, parmi lesquels la pluviométrie joue un rôle prépondérant.

Les précipitations atmosphériques constituant un phénomène essentiellement variable dans l'espace et dans le temps, cette connaissance — qui doit être avant tout quantitative — ne peut être acquise sans une « observation » permanente et globale, au moins régionale, des événements pluvieux.

En attendant que soient rendus totalement opérationnels des systèmes de surveillance, de mesure et de cartographie par radar, couvrant de vastes zones, cette « vision » spatiale ne peut être appréhendée que par échantillonnage à travers un « réseau » d'observations *ponctuelles*.

Malheureusement, toute acquisition de données par sondage ou échantillonnage est soumise à des erreurs de toute nature, qui tiennent autant au système d'acquisition lui-même qu'au caractère individuel de la mesure ou de l'enquête, individualité éminemment propre à engendrer des hétérogénéités.

Cette constatation impose une critique sévère des données recueillies, suivie d'éviction pure et simple quand elles sont jugées s'éloigner par trop de la réalité, de corrections prudentes quand la nature systématique des erreurs détectées l'autorise, de messages d'alerte quand leur valeur « anormale » peut être considérée comme localement possible.

Nous désignons l'ensemble de ces opérations sous le titre : « Homogénéisation des données pluviométriques ».

1. L'INFORMATION PLUVIOMÉTRIQUE

Pour tenter d'exposer aussi clairement que possible le problème de l'homogénéité des données pluviométriques tel qu'il se présente, et dégager les directions selon lesquelles il pourrait être résolu, il paraît intéressant de comparer

le réseau d'observations pluviométriques à un « système d'information », et la donnée pluviométrique sur laquelle travaille un utilisateur au mot du « message » reçu par un destinataire.

Ce message, à l'état brut, comporte, outre le « signal » proprement dit contenant l'information que l'on a voulu transmettre, toute une série de « bruits », en cascade, issus aussi bien de la (des) source(s) que de la chaîne d'acquisition, à tous ses niveaux.

Si l'on considère une source émettrice isolée j , il est évident que la qualité du message à la réception, que l'on peut représenter par l'expression :

$$SB_j = S_j + B_j$$

dépendra du rapport $\frac{S_j}{B_j}$; plus ce rapport sera élevé, plus le signal S_j sera « reconnaissable », discernable du bruit B_j qui l'accompagne.

Il existe malheureusement une limite supérieure du bruit, difficile à préciser, au-delà de laquelle le signal n'est plus perceptible. Mais on verra que, si l'on a la chance de disposer d'un réseau d'information d'une densité convenable pour constituer un système redondant, on peut repousser assez loin cette limite et obtenir, malgré des lacunes locales, une information suffisamment « représentative ».

Sans décrire dans les détails le cheminement de l'observation pluviométrique, familier à tous les hydrologues et climatologues, nous rappellerons succinctement, en nous aidant de la figure 1, ses éléments constitutifs, les sources d'erreurs possibles rencontrées tout au long du système d'acquisition, les moyens actuellement disponibles pour en analyser la qualité.

1.1. LE SIGNAL

Si l'on admet comme hypothèse simplificatrice l'appartenance du poste d'observation considéré à une *tendance* climatique régionale *unique*, il est raisonnable de concevoir l'information pluviométrique de base — le total annuel des précipitations — comme le résultat de la superposition de deux variables asynchrones :

— l'une, de période longue, représente la fluctuation éventuelle de la moyenne au cours du temps (cyclique ou non) ;

— la seconde, de période courte — l'année — représente la variation annuelle autour de la moyenne.

Les valeurs annuelles de cette information régionale de base ou, pour plus de commodité, leurs écarts à la moyenne rapportés à l'écart-type peuvent être considérés comme les composantes « w_i » d'un vecteur de référence W .

Pour obtenir les hauteurs théoriques annuelles de précipitations « x_i » de la zone immédiatement voisine d'un lieu d'observation, il y aura lieu d'affecter aux coordonnées du vecteur W , moyenne \bar{x} et écart-type σ_x propres au poste considéré, et intégrant les différents facteurs de variation spatiale (altitude, exposition aux vents dominants, environnement, etc.) ; nous les appellerons « facteurs de site ».

L'information idéale recherchée pourra, dans ces conditions, s'écrire :

$$x_i = \bar{x} + \sigma_x \cdot w_i.$$

Nous avons vu qu'elle est, le plus souvent, déformée par des erreurs — ou bruits — issus des différents maillons de la chaîne d'acquisition.

1.2. LES « BRUITS »

Les bruits que nous allons énumérer ont, pour la plupart, un caractère multiplicatif : autrement dit, en vue de les rattacher au système de référence choisi, il y aurait lieu de multiplier les données brutes par un coefficient correcteur.

Quelques autres, cependant, ont un caractère additif : il s'agit, en général, de lacunes dans les observations ou d'erreurs de lecture ou d'écriture ; leur prise en considération peut souvent faciliter la détection des précédents.

1.2.1. Le site

1.2.1.1. Bruit propre du site

Dû à la nature ponctuelle de l'observation, il traduit les variations aléatoires autour de la valeur moyenne hypothétique d'un élément de surface qui aurait pour centre le pluviomètre ; on peut l'assimiler à un effet de turbulence.

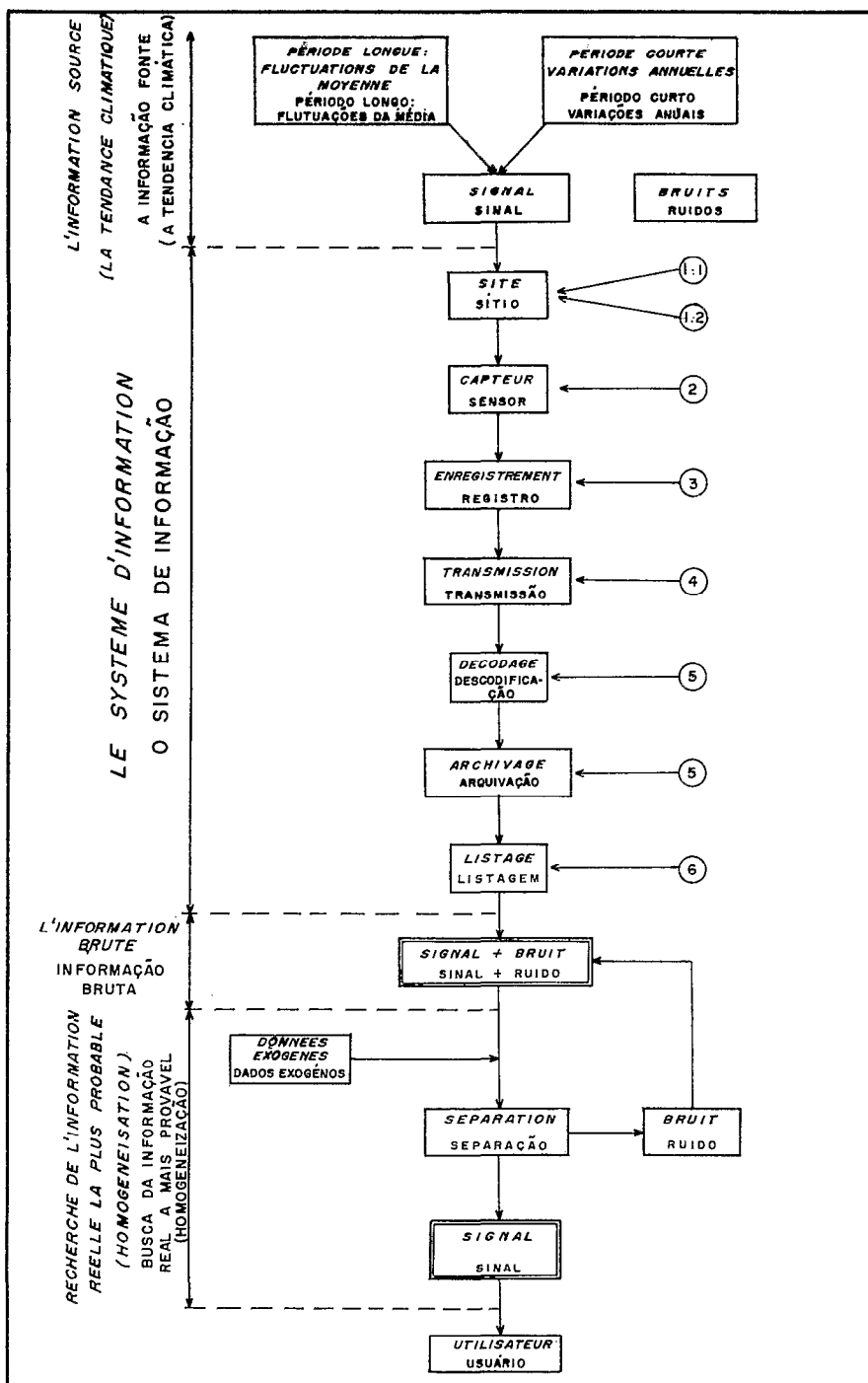


Fig. 1

(L'épicentre d'une micro-tornade a pu passer, par hasard, à l'emplacement précis du pluviomètre, par exemple ; la hauteur enregistrée ne serait alors pas représentative de l'élément de surface qui l'entoure.)

1.2.1.2. Modification de l'environnement du site

- par déplacement du pluviomètre ;
- par modification du paysage limitrophe,
 - brutale (construction),
 - progressive (croissance d'un arbre ou d'une haie) ;
- par changement de la hauteur de l'appareil au-dessus du sol.

1.2.2. *Bruit du capteur*

1.2.2.1. Modification de la surface réceptrice

- par construction (surface imprécise) ;
- par changement (passage de $100 \pi \text{ cm}^2$ à 400 cm^2) ;
- par déformation (bague réceptrice ovalisée).

1.2.2.2. Défaut de calibrage

Dans le cas d'un appareil enregistreur : cela peut être dû à un vice de conception (hauteur enregistrée variable avec l'intensité), ou à un mauvais réglage (contrepois du pluviographe à siphon, etc.).

1.2.3. *Erreurs de mesure et d'enregistrement*

1.2.3.1. Au niveau de l'éprouvette

- précision différente d'une éprouvette à une autre ;
- confusion d'éprouvette ;
- mesure peu soignée ou lecture erronée (additif).

1.2.3.2. A l'enregistrement

- par manque d'encre ;
- mauvais réglage du système d'inscription ;
- erreur de date ;
- écriture erronée de la hauteur mesurée (additif) ou erreur de virgule ;
- erreur de recopie.

1.2.4. *A la transmission*

- Par perte de courrier.
- Pli égaré à la réception.
- Défectuosités du système de transmission (électrique, radio, satellite...).

1.2.5. *Au décodage et à l'archivage*

Erreurs de date, de virgule, de recopie, de perforation, sur le calcul du total, sur l'échelle d'un diagramme, etc. Disparition d'un document par suite d'un mauvais archivage, ou d'une erreur d'adresse (bande magnétique).

1.2.6. *Erreurs de listage*

Elles peuvent se produire à la publication du fichier des données brutes et comporter toutes les formes d'erreurs d'écriture, jusqu'à l'omission de données et à l'interversion de postes.

Nous classerons également sous cette rubrique les erreurs possibles dues à la non utilisation de l'année « hydrologique » pour effectuer les comparaisons entre postes.

1.3. L'ANALYSE DU SIGNAL

Cette énumération rend évidente la nécessité d'effectuer sur les données un rigoureux contrôle de leur qualité.

En poussant plus avant notre comparaison, il faudra en un premier temps « reconnaître » le signal et le « séparer » du bruit : autrement dit rechercher dans l'information brute reçue — les données d'un réseau de postes — les éléments d'une information de référence : avec son aide et (ou) au moyen de tests appropriés, discriminer sur chaque poste entre le « bruit propre » du poste (variation aléatoire ou anomalie locale) et d'éventuelles altérations systématiques (erreur d'éprouvette, changement de site, etc.)

Une fois « séparée » la part de l'information qui relève de déficiences *prouvées* du système d'acquisition, on pourra alors procéder à une opération de « filtrage » qui consiste, en d'autres termes, à corriger les séquences ou valeurs isolées reconnues comme défectueuses.

De nombreux tests ou méthodes ont été expérimentés pour effectuer cette analyse de consistance des données pluviométriques ; signalons parmi d'autres, les travaux récents du Bureau de l'Eau de la Météorologie Nationale [10, 13, 28].

Mais il semble que *seule* la méthode graphique de « double-cumul », complétée au besoin par celle de « simple cumul », a prouvé une réelle efficacité.

Pendant, les conclusions auxquelles elle permet d'aboutir sont souvent subjectives ou incertaines ; elles nécessitent toujours de persévérants et minutieux efforts d'interprétation.

Par ailleurs son application à l'échelle régionale ou d'un pays — échelle qui tend actuellement à se généraliser — exige la manipulation d'une quantité considérable d'informations que seule une automatisation poussée des opérations de traitement permet d'entreprendre.

Une tentative a été faite dans cette voie au Service Hydrologique de l'ORSTOM ; elle a abouti à la rédaction d'un programme de tracé systématique des graphiques de simple et double-cumul avec ou sans correction des données (Programmes POH 102 UNMASS, DBLMSS et UDMCOR).

Son utilisation facilite beaucoup les comparaisons entre postes par la suppression de la plupart des opérations manuelles et apporte une aide précieuse dans le travail d'interprétation visuelle.

Malheureusement, il ne résoud pas le problème fondamental de toute analyse de données : celui de la « décision objective ».

C'est dans ce sens qu'il est apparu opportun de rechercher une voie originale qui, sans supprimer la faculté de décider, permette l'élaboration automatique des facteurs de la décision ; cette élaboration s'appuyant sur les principes mêmes qui sont à la base de la méthode de double-cumul.

2. ESSAI SUR LA THÉORIE DES SOMMES CUMULÉES

2.1. LE GRAPHIQUE DE « SIMPLE-CUMUL »

Reprenons le secteur W défini en 1.1 et soit x_i l'information annuelle idéale observée en un poste ; en l'absence de fluctuations aléatoires ou systématiques, cette information peut s'écrire :

$$x_i = \bar{x} + \sigma_x * w_i$$

Mais nous savons que l'observation est en réalité affectée de fluctuations aléatoires telles que la grandeur x_i ne « suit » pas exactement les variations de w_i , auquel elle est liée par le coefficient de corrélation r_x .

Dans ces conditions x_i s'écrira :

$$x_i = \bar{x} + \sigma_x (r_x w_i + \sqrt{1 - r_x^2} * u_i)$$

dans laquelle u_i représente les fluctuations propres du poste x , indépendantes de w_i .

Effectuons le cumul des données jusqu'à l'année l :

$$\sum_1^l x_i = l \cdot \bar{x} + \sigma_x \cdot (r_x \cdot \sum_1^l w_i + \sqrt{1 - r_x^2} \cdot \sum_1^l u_i)$$

ou en divisant par la moyenne :

$$X_l = \frac{1}{\bar{x}} \sum_1^l x_i = 1 + \frac{\sigma_x}{\bar{x}} \left(r_x \cdot \sum_1^l w_i + \sqrt{1 - r_x^2} \cdot \sum_1^l u_i \right)$$

On note :

a) que $C_x = \frac{\sigma_x}{\bar{x}}$ est le coefficient de variation de la variable x ,

b) que x_l est proportionnel à l , le nombre d'années sur lequel porte la sommation, à un terme aléatoire près que nous pouvons décomposer de la façon suivante :

$$X_l = l + C_x \left(r_x \cdot \sum_1^l w_i + \sqrt{1 - r_x^2} \cdot \sum_1^l u_i \right) + C_x (r_x \cdot w_l + \sqrt{1 - r_x^2} \cdot u_l).$$

Nous appellerons le graphique qui traduit la relation $X_l = f(l)$ un graphique de « simple cumul relatif ».

Cette forme d'écriture permet de mettre en évidence l'existence d'un terme de « persistance » — le deuxième — régi par un processus de Markov, complété par un terme aléatoire pur — le troisième —, de moyenne nulle et d'écart-type « C_x ».

S'il existe une fluctuation à long terme de la moyenne on « verra » le tracé osciller à l'intérieur d'un « couloir » ou « rail » très sensiblement parallèle à la première bissectrice ; si la moyenne subit une déformation systématique pendant une période déterminée, pour une des raisons énumérées en 1.2, le deuxième terme s'écartera significativement de la valeur nulle pendant cette période et le tracé restera à l'intérieur d'un couloir dont la pente ne sera plus celle de la première bissectrice : ce changement de pente, est, en général, aisément détectable à vue.

Tel est le graphique de simple-cumul ; utilisé seul, il ne permet pas en pratique de distinguer avec certitude d'éventuelles déviations systématiques d'un changement de la tendance climatique.

2.2. LE GRAPHIQUE DE « DOUBLE-CUMUL »

2.2.1. Comparaison poste à poste

Considérons maintenant deux postes d'observation pluviométrique situés à une distance suffisamment faible pour que le régime des précipitations enregistrées à chacun d'eux soit régi par une tendance climatique unique, connue, telle qu'elle a été définie précédemment, et soit W de composantes w_i le vecteur des précipitations annuelles représentant cette tendance.

Nous pouvons exprimer la hauteur annuelle des précipitations à chacun des deux postes par les relations :

$$x_i = \bar{x} + \sigma_x (r_x \cdot w_i + \sqrt{1 - r_x^2} \cdot u_i) \quad (1)$$

et

$$y_i = \bar{y} + \sigma_y (r_y \cdot w_i + \sqrt{1 - r_y^2} \cdot v_i) \quad (2)$$

en supposant qu'ils sont liés à la tendance par les coefficients de corrélation linéaires respectifs r_x et r_y et affectés chacun de fluctuations propres indépendantes représentées respectivement par u_i et v_i .

Faisons le cumul des variables jusqu'à l'année l et soit :

$$X_l = \frac{1}{\bar{x}} \sum_1^l x_i = l + \frac{\sigma_x}{\bar{x}} \left(r_x \sum_1^l w_i + \sqrt{1 - r_x^2} \cdot \sum_1^l u_i \right) \quad (3)$$

et

$$Y_l = \frac{1}{\bar{y}} \sum_1^l y_i = l + \frac{\sigma_y}{\bar{y}} \left(r_y \sum_1^l w_i + \sqrt{1 - r_y^2} \cdot \sum_1^l v_i \right) \quad (4)$$

et éliminons l entre les deux relations en notant que $\frac{\sigma_x}{\bar{x}}$ et $\frac{\sigma_y}{\bar{y}}$ représentent les coefficients de variation C_x et C_y de chacun des postes, soit :

$$Y_l = X_l - C_x \left(r_x \sum_1^l w_i + \sqrt{1 - r_x^2} \cdot \sum_1^l u_i \right) + C_y \left(r_y \sum_1^l w_i + \sqrt{1 - r_y^2} \cdot \sum_1^l v_i \right) \quad (5)$$

ou, en groupant différemment les termes :

$$Y_l = X_l + \sum_1^l w_i (r_y \cdot C_y - r_x \cdot C_x) + \left(C_y \sqrt{1 - r_y^2} \cdot \sum_1^l v_i - C_x \sqrt{1 - r_x^2} \cdot \sum_1^l u_i \right) \quad (6)$$

Nous mettons ainsi en évidence l'existence d'une liaison linéaire entre les variables X_l et Y_l à deux termes complémentaires près. En fait, les points de coordonnées (X_l, Y_l) se trouveront à l'intérieur d'un « couloir » rectiligne ou « rail » de pente positive, égale à l'unité ; si les variables sont indépendantes, la largeur du couloir dépendra uniquement des coefficients de variation de chacun des postes.

Dans le cas contraire, la largeur du couloir dépendra à la fois des coefficients de variation et des coefficients de corrélation, et sera d'autant plus étroite que ces derniers seront proches de l'unité.

Toute variation de l'une des moyennes, par modification des conditions de l'observation, se traduira par une modification simultanée de la pente du couloir.

Si la relation (6) justifie l'application de la méthode du double-cumul à des postes quelconques, en pratique la détection des anomalies de persistance ou d'amplitude faible et une précision suffisante des corrections ne seront possibles qu'à l'intérieur de couloirs relativement étroits, autrement dit s'il existe une corrélation suffisante entre postes ; c'est pour cette raison que l'on en restreint l'usage à des postes relativement voisins, cette notion de proximité étant davantage définie par la valeur du coefficient de corrélation que par la distance absolue.

On note également dans la relation (6) que, si les deux postes sont situés suffisamment proches l'un de l'autre pour qu'ils soient soumis à une même tendance climatique, les coefficients de variation C_x et C_y seront sensiblement égaux, ainsi que r_x et r_y : le deuxième terme du second membre, dans ces conditions, tendra vers une valeur nulle.

Cette constatation entraîne une conclusion des plus intéressantes : les « double-cumul » présentent sur les « simple-cumul » un avantage considérable du fait que, si l'on se place sous la stricte condition d'identité de la tendance climatique, la relation à caractère linéaire qui existe entre les sommes cumulées des deux variables pluviométriques est indépendante d'éventuelles fluctuations du climat et n'est modifiée que par des fluctuations aléatoires si les données sont homogènes.

Le troisième terme, en effet, que nous appellerons a_l , est composé de la différence pondérée de deux grandeurs variables aléatoirement ; il constitue donc lui-même une variable aléatoire ; cette variable est, elle aussi, régie par un processus de Markov comme on peut le montrer aisément en la décomposant de la façon suivante :

$$a_l = \left(C_y \sqrt{1 - r_y^2} \cdot \sum_1^{l-1} v_l - C_x \sqrt{1 - r_x^2} \cdot \sum_1^{l-1} u_l \right) + \left(C_y \sqrt{1 - r_y^2} \cdot v_l - C_x \sqrt{1 - r_x^2} \cdot u_l \right) \quad (7)$$

On voit qu'elle est constituée pour le premier terme de la valeur antérieure et pour le second terme d'une variable aléatoire pure, de moyenne nulle dont la variance peut être estimée par :

$$V = (C_x^2 - 2 C_x * C_y * r_x * r_y + C_y^2) \quad (8)$$

L'existence de ce premier terme fait que la variable a_l est assujettie à un effet de persistance qui explique parfaitement les « pseudo-oscillations » entre les « rails » rencontrées dans les applications pratiques [3] ; la figure 2, obtenue à partir de la simulation de 500 couples de valeurs aléatoires corrélées par un coefficient de 0,8 montre très bien que a_l peut rester, de façon « persistante » (plus de 50 ans), très éloignée de la valeur 0 : en conséquence, il y aura lieu de manipuler les rails avec beaucoup de circonspection, car en toute rigueur, ils n'ont de signification que sur des intervalles de temps limités.

L'écartement des rails pourrait être défini, à un coefficient près, dépendant de l'intervalle de confiance choisi (ou du taux d'homogénéité recherché) par :

$$d = \sqrt{\frac{V}{2}} \quad (9)$$

Malheureusement V ne peut être estimé correctement, du moins en début d'analyse, car le calcul des coefficients de variation et de corrélation est en général faussé par la présence d'erreurs ou de valeurs anormalement fortes ou faibles et par d'éventuelles déviations systématiques.

2.2.2. Comparaison entre un poste et son vecteur régional de référence

On peut, par contre, tenter de resserrer l'espacement entre les rails en comparant les postes non plus deux à deux mais chacun des postes à un vecteur de référence représentant la « tendance » climatique régionale des précipitations.

Soit Z ce vecteur de référence tel que :

$$z_i = \bar{z} + \sigma_z * w_i$$

De la même manière que précédemment, nous aurons :

$$Z_l = \frac{1}{l} * \sum_1^l z_i = \bar{z} + C_z \sum_1^l w_i \quad (10)$$

et

$$X_l = \frac{1}{l} * \sum_1^l x_i = \bar{x} + C_x \left(r_x \cdot \sum_1^l w_i + \sqrt{1 - r_x^2} * \sum_1^l u_i \right) \quad (11)$$

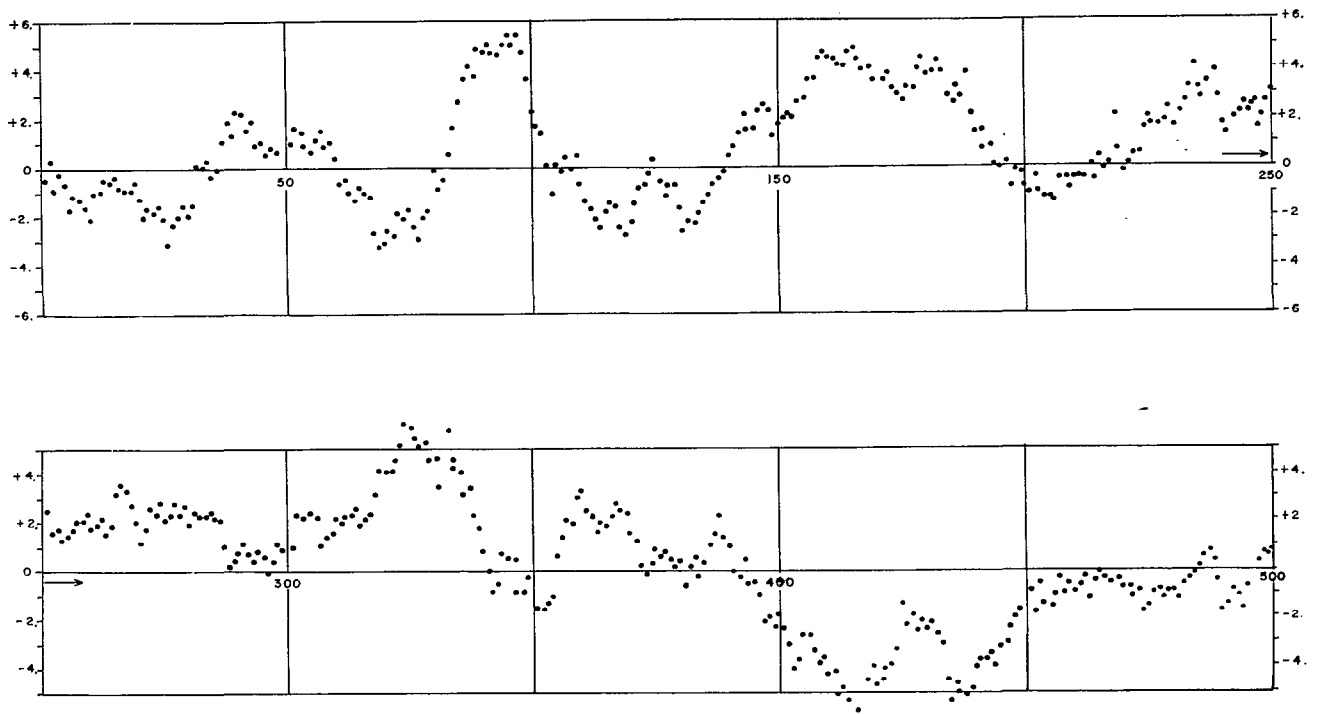


Fig. 2
 Différence cumulée de deux variables normales réduites corrélées ($r = 0,8$)
 Diferença acumulada de duas variáveis normais reduzidas correlatas ($r = 0,8$)

et en éliminant l :

$$X_l = Z_l + \sum_1^l w_1 (r_x * C_x - C_z) + \left(C_x \sqrt{1 - r_x^2} * \sum_1^l u_i \right) \quad (12)$$

Dans cette nouvelle expression, le deuxième terme du second membre tend également vers 0 et ce d'autant mieux que le vecteur Z sera représentatif du poste considéré, ou, ce qui revient au même, que le poste considéré sera corrélé avec le vecteur Z .

En procédant à la décomposition déjà utilisée, nous obtenons :

$$X_l = Z_l + \sum_1^{l-1} w_1 (C_x * r_x - C_z) + w_1 (C_x * r_x - C_z) + C_x \sqrt{1 - r_x^2} * \sum_1^{l-1} u_i + C_x \sqrt{1 - r_x^2} * u_l \quad (13)$$

dans laquelle l'élément aléatoire pur aura une moyenne nulle et une variance qu'il est possible d'estimer par :

$$(C_x^2 - 2 C_x * C_z * r_x + C_z^2) \quad (14)$$

A coefficients de variations semblables on peut vérifier que cette variance est plus faible que celle obtenue à partir de la comparaison des postes pris deux à deux ; l'écartement entre les « rails » sera par conséquent plus réduit, et la détection d'éventuelles cassures s'en trouvera affinée.

Dans toutes les applications des double-cumuls effectuées jusqu'à ce jour, les deux méthodes sont utilisées, soit séparément, soit conjointement.

Y. BRUNET-MORET [3] semble préférer la comparaison poste à poste mais il met l'accent sur l'ampleur et la complexité de la tâche qui peut conduire parfois à l'inextricable (examen de 250 graphiques pour homogénéiser 33 postes) ; il envisage le recours à la deuxième méthode.

P. DUBREUIL [11] préfère élaborer une variable de référence, à partir de la moyenne d'un groupe réputé de « base » de 3 à 5 stations, choisies pour la qualité de leurs données et la longueur de la période d'observation

commune ; l'homogénéité de chacun des postes du groupe de base est soigneusement vérifiée en recourant aussi bien aux simples cumuls qu'à la comparaison des postes deux à deux.

En fait, aucune des deux méthodes n'est totalement satisfaisante quand elle est pratiquée manuellement ; la première en raison du haut degré de patience et de perspicacité nécessaire pour venir à bout d'un nombre de postes même restreint ; la seconde parce que la constitution d'un groupe de base est délicate et il n'est pas toujours possible de trouver un nombre suffisant de postes présentant toutes les garanties de qualité requises, observés durant une période commune suffisamment longue et groupés sous une densité assez forte pour permettre l'estimation d'une valeur moyenne sans introduire de biais systématique.

Enfin, ni l'une, ni l'autre de ces méthodes, utilisées avec des moyens conventionnels, ne s'est prêtée jusqu'à ce jour, à une approche globale et « mécanisée » du problème de l'homogénéisation.

On conçoit tout l'intérêt qu'il y aurait à élaborer à partir d'un *ensemble de données* de précipitations fournies par un certain nombre de postes, situés dans une même région, et *sans aucune intervention manuelle*, une *variable de référence* couvrant toute la période observée, à laquelle on viendrait comparer chacun des postes.

Cet intérêt réside non seulement dans la réduction possible de l'écartement des « rails », comme on l'a montré plus haut, mais aussi, à un plus haut niveau, dans la réduction du nombre de comparaisons à effectuer qui passe, si m est le nombre de postes du groupe à homogénéiser, de $m(m-1)/2$ à m soit un facteur de réduction de $(m-1)/2$.

Enfin — et surtout — le nombre de critères de « décisions » est réduit dans les mêmes proportions : si la variable de référence a été élaborée dans des conditions optimales, la simple comparaison du poste dont il faut vérifier l'homogénéité avec cette variable de référence fournit la décision de corriger ou non les données incriminées sans qu'il soit nécessaire de faire appel à une série de déduction en cascade dont la pratique conduit rapidement à la confusion pour peu que le nombre de postes à traiter augmente au-delà de 5 ou 6.

3. ÉLABORATION D'UN VECTEUR DE RÉFÉRENCE

3.1. LES PRINCIPES DIRECTEURS

Nous avons vu dans les expressions (6) ou (12) que, à un terme correctif aléatoire près, les sommes cumulées des totaux annuels de précipitations d'un poste et celles d'un autre poste voisin ou d'une variable de référence suivent une loi linéaire simple autour de la bissectrice, si les données sont rapportées à leurs moyennes respectives ; on sait que toute perturbation dans les conditions de la mesure, au cours de la période d'observation, entraînera sur toute la durée de la perturbation une modification du coefficient angulaire de la droite autour de laquelle se situent successivement les points.

Sur la figure 3, nous avons représenté schématiquement un graphique de double-cumul ; il est constitué de deux périodes distinctes séparées par une cassure nette ; projetons chacune des deux directions ainsi suivies sur l'origine (translatée pour la clarté de la figure) et soit U et V les vecteurs définis par l'intersection des directions d_1 et d_2 avec le cercle de rayon unitaire ; nous pourrions dire que chaque vecteur élémentaire P_i de coordonnées $(x_i = X_i - X_{i-1}, y_i = Y_i - Y_{i-1})$ est, à tous les « bruits » dont il a été fait mention près, colinéaire respectivement du vecteur U jusqu'au temps 5 et du vecteur V jusqu'au temps 13.

Ceci revient à admettre que, tant que les vecteurs élémentaires P_i restent sensiblement parallèles à la même direction, leurs coordonnées x_i et y_i sont dans un rapport quasi constant et nous pouvons écrire :

$$\frac{x_i}{y_i} \cong \frac{x_{i-1}}{y_{i-1}} \text{ ou ce qui revient au même } \frac{x_i}{x_{i-1}} \cong \frac{y_i}{y_{i-1}}$$

c'est ce que nous appellerons le principe de « pseudo-proportionnalité » des totaux pluviométriques annuels entre postes « voisins » ou entre un poste et le vecteur auquel il se réfère.

La pratique des graphiques de double-cumul implique l'adhésion à ce principe ; c'est également sur lui que nous ferons reposer l'élaboration du vecteur de référence. Mais il convient de lui adjoindre quelques hypothèses de travail complémentaires :

a) cette élaboration ne doit exiger aucune hypothèse particulière sur la distribution des totaux pluviométriques annuels ;

b) on admet qu'au sein du groupement régional de postes que l'on se propose d'homogénéiser, il n'existe (au moins en première analyse) aucune variation sensible de la tendance climatique ;

c) une mesure ponctuelle de la pluviométrie — dans le temps et dans l'espace — est considérée comme entachée

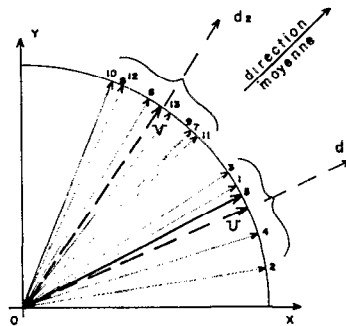
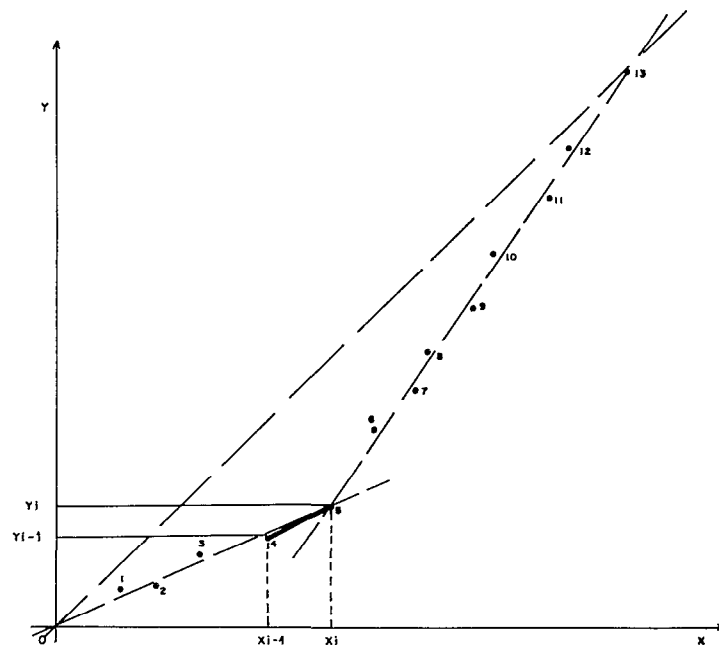


Fig. 3

d'erreur, quelle que soit sa qualité : le total pluviométrique d'un poste pour une année déterminée ne peut donc en aucune façon servir « a priori » de valeur de référence pour combler la lacune d'un autre poste, d'autant moins qu'il est susceptible de représenter une anomalie locale ;

d) corollairement, nous devons supposer qu'au sein de l'information pluviométrique globalement fournie par un « ensemble » de stations, il existe une valeur « estimative » des fluctuations de la pluie dans le temps, plus significative au sens de la représentativité de l'ensemble (régionale), que l'information partielle fournie par chacun des postes, considéré isolément, composant le groupe.

Cela n'implique absolument pas que soit exclue la possibilité d'anomalies purement locales, cela exprime que de telles anomalies ne doivent pas être prises en considération pour l'élaboration d'un vecteur de référence auquel seront comparés, en vue de l'homogénéisation, les données de postes où ces anomalies ne se seraient *en général* pas produites.

Si l'on n'admet pas cette hypothèse de travail, on ne peut envisager la recherche d'une information « représentative » sous la forme d'un vecteur unique ;

e) mais par contre, pour être objectif, le procédé d'homogénéisation doit être conçu de telle manière que toute l'information contenue dans *chacun* des postes puisse participer à l'élaboration de l'information « idéale », la plus probable possible, sans que l'information erronée, localement anormale ou provisoirement déviée, puisse avoir une influence sensible sur le résultat de l'élaboration.

On sera malgré tout contraint d'enfreindre cette règle dans les applications pratiques, lorsque la qualité des observations laisse trop à désirer.

En revanche, on n'éliminera pas a priori un poste observé sur un nombre d'années trop faible (mais supérieur ou égal à 3) dont le hasard peut faire que les données soient déterminantes pour préciser telle ou telle composante annuelle du vecteur.

3.2. LE TABLEAU DES DONNÉES CONSIDÉRÉ COMME UNE MATRICE

Les principes directeurs précédemment définis nous conduisent :

- 3.2.1. A traiter simultanément l'ensemble des postes et non plus deux à deux comme dans la méthode conventionnelle ; ceci nous évitera, entre autre, de recourir à de multiples et laborieuses confrontations.
- 3.2.2. A rechercher, à partir de cet ensemble, un vecteur représentatif unique de référence, auquel chacun des postes sera comparé (ultérieurement, on envisagera la possibilité pour un poste d'appartenir à deux, ou plus, composantes climatiques).
- 3.2.3. Donc à considérer le tableau des données (totaux pluviométriques annuels ou mensuels) comme une matrice munie de ses deux espaces :
 - celui des postes, en colonne ;
 - celui des années en ligne.

Soit A cette matrice, m et n ses dimensions respectives :

$$A = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & x_{ij} & \dots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix}$$

Les « m » données recueillies pendant une année i représentent alors les coordonnées d'un vecteur — pluie — annuel, dans l'espace des postes.

Puisque nous avons retenu comme principe de base celui de la « pseudo-proportionnalité », nous devons admettre que l'ensemble des n vecteurs i constitue un ensemble de vecteurs *presque* colinéaires. Dans ces conditions, et si l'hypothèse est vérifiée, la matrice A pourrait être représentée par une matrice B dégénérée, ou en d'autres termes une matrice de rang 1, entièrement définie si l'on connaît un vecteur ligne et un vecteur colonne, soient L et C tels que :

$$B = \begin{bmatrix} L_1 \\ L_2 \\ \dots \\ L_i \\ \dots \\ L_n \end{bmatrix} * [C_1 \ C_2 \ \dots \ C_j \ \dots \ C_m]$$

où $b_{ij} = L_i * C_j$.

En fait, la matrice A des données réelles est plus ou moins proche de ce modèle théorique, puisque même dans le cas d'observations effectuées dans des conditions optimales, des fluctuations aléatoires propres à chaque poste subsistent ; on doit donc le compléter par une matrice d'erreur E, telle que :

$$A = B + E \quad (15)$$

où B représenterait l'information théorique recherchée, c'est-à-dire toute l'information à caractère linéaire qu'il est possible d'extraire de A, tout en satisfaisant le critère retenu pour caractériser E.

La tentation est grande d'imaginer que les éléments de $E = A - B$ devraient autant que faire se peut être linéairement indépendants entre eux, autrement dit, que la matrice E^TE devrait être une matrice diagonale, ce qui reviendrait à minimiser $\|A - B\|_2$ c'est-à-dire à utiliser un procédé classique de régression.

En ce qui concerne la résolution de :

$$A = B + E$$

sous les contraintes : $B = L * C$ et : $E^TE =$ matrice diagonale,

cela équivaudrait à rechercher la direction principale de la matrice des données.

Or on sait que la minimisation de la norme euclidienne implique la condition d'homoscédasticité des résidus, ce qui ne paraît guère compatible avec la nature a priori hétérogène des données, mais aussi l'appartenance des données à une distribution gaussienne, puisque l'on a admis le principe de pseudo-proportionnalité (ou linéarité) ce qui serait contraire à l'une des hypothèses de travail que nous nous sommes imposées.

De plus cette démarche conduit à accorder un poids plus important aux données anormalement faibles ou fortes et de ce fait laisse dans le résidu un contenu « linéaire » très loin d'être négligeable : si la tendance des données à la « proportionnalité » est forte, on n'est pas certain par ce procédé d'extraire toute l'information à caractère proportionnel.

Enfin le problème que nous cherchons à résoudre ne consiste pas à « minimiser » les « erreurs », et encore moins les erreurs fortes, au profit d'une réduction — somme toute assez arbitraire — de la variance expliquée, mais bien plutôt à les détecter telles qu'elles existent.

Un exemple très simple permettra de mieux saisir où peut conduire l'abus des régressions en minimisation de norme, principalement de normes euclidiennes :

Soit la matrice :

$$A = \begin{bmatrix} 2 & 3,6 & 4 & 6 \\ 4 & 6 & 8 & 12 \\ 6 & 9 & 9,6 & 18 \end{bmatrix}$$

dont les éléments à l'exception de a_{12} et a_{33} sont ceux d'une matrice dégénérée, comme on le remarque immédiatement. Ses valeurs propres sont celles de la matrice (AA^T) , soit par le calcul matriciel :

$$\lambda_1 = 860,29 \quad \lambda_2 = 1,58 \quad \lambda_3 = 0,24$$

alors qu'en réalité ces valeurs devraient être :

$$\lambda_1 = 910 \quad \lambda_2 = 5,76 \quad \lambda_3 = 0,36$$

montrant que la part de variance expliquée a été sous-estimée.

Le vecteur propre principal associé à cette matrice aurait pour coordonnées (non normées) :

<i>En calcul matriciel</i>	<i>En réalité</i>
1,029	1
2,000	2
2,867	3

soit une erreur d'environ 5% sur l'estimation de la troisième composante.

On notera que si les colonnes de A avaient été normées, la direction principale serait devenue

1,049
2,000
2,829

encore plus éloignée de la direction réelle.

Enfin, la matrice des résidus :

$$E = A - B = \begin{bmatrix} -0,112 & +0,384 & +0,308 & -0,337 \\ -0,105 & -0,251 & +0,824 & -0,317 \\ +0,114 & +0,038 & -0,686 & +0,342 \end{bmatrix}$$

permet de mettre en évidence une très forte corrélation entre les lignes, aussi bien qu'entre les colonnes (colonnes 1 et 4) :

$$r = \begin{bmatrix} 1 & 0,493 & -0,691 \\ 0,493 & 1 & -0,970 \\ -0,691 & -0,970 & 1 \end{bmatrix}$$

c'est ce que nous entendons par contenu linéaire du résidu.

Si l'on retient les valeurs (1, 2, 3) comme composantes du vecteur-colonne, et (2, 3, 4, 6), comme composantes du vecteur-ligne, le résidu réel se réduit à :

$$E = \begin{bmatrix} 0 & +0,6 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -2,4 & 0 \end{bmatrix}$$

considérablement différent du précédent et l'on voit qu'il n'est maintenant plus possible d'en extraire une information à caractère linéaire.

Il s'agit, sans doute, d'un exemple particulier dans lequel les données ne ressemblent en rien à des données pluviométriques ; mais on imagine mal ce que ce procédé, si largement utilisé dans tous les domaines, peut produire quand il s'agit de traiter des données réelles affectées d'erreurs grossières ou systématiques dont les distributions, dans ces conditions, ne peuvent être considérées comme normales.

3.3. DÉFICIENCE ÉVENTUELLE DE LA MOYENNE POUR L'ESTIMATION D'UN PARAMÈTRE DE TENDANCE CENTRALE

On sait que la valeur moyenne d'une variable peut être considérée comme la valeur du paramètre a qui rend minimale la quantité $\sum (x_i - a)^2$; elle repose donc sur la notion de distance et résulte d'une minimisation de la norme euclidienne, ainsi que nous venons de la pratiquer dans l'exemple précédent, pour obtenir la direction principale d'une matrice $— \| A - B \|_2 —$. Le piètre résultat fourni ainsi que ceux obtenus lors de nombreuses expérimentations, effectuées tant à partir de données simulées que de données réelles, nous amènent à émettre de sérieux doutes sur la valeur de la moyenne arithmétique en tant que paramètre significatif de tendance centrale, dès que les variables s'écartent sensiblement d'une distribution rigoureusement normale.

Or les cas de non normalité sont extrêmement fréquents : ils peuvent être le fait d'un nombre trop faible de données, ou de données obéissant à d'autres lois (fig. 4 a) (cas de toutes les distributions asymétriques) ou régies par une distribution plurimodale (fig. 4 b) (cas des séries chronologiques hétérogènes avec cassure dans les graphiques de double cumul) ; soit enfin qu'il existe des valeurs extrêmes fortes qui viennent altérer le calcul de la moyenne (fig. 4 c) ce que certains auteurs d'avant-garde appellent la « contamination » de la moyenne par les valeurs extrêmes [2] ; ces mêmes auteurs n'hésitent pas à qualifier cette dernière d'estimateur « désastreux » si les données s'écartent un tant soit peu de la distribution gaussienne.

Il en est de même de l'écart-type, issu du même critère, qui peut être considéré comme impropre pour rendre compte de la dispersion d'une variable autour d'une valeur centrale, si ce n'est dans une distribution gaussienne, autour de la moyenne ; comme on peut le voir sur la figure 4, il peut certes donner une idée de l'étalement des données, mais selon les cas, il peut surestimer ou sous-estimer leur concentration autour de valeurs centrales.

Ces considérations et les échecs subis dans les tentatives d'homogénéisation des données pluviométriques à partir de méthodes conventionnelles conduisent à écarter définitivement la distance euclidienne comme critère de minimisation de $E = A - B$.

Par contre, il est possible d'obtenir des résultats encourageants à partir de la distance en valeur absolue (minimisation de $\sum |x_i - a|$) [8], l'estimation de $B = L.C.$ se faisant par la recherche des valeurs médianes ; mais il apparaît vite que cette recherche pose des problèmes délicats lorsque les séries comportent de nombreuses données manquantes.

3.4. LA VALEUR MODALE

Nous avons vu que la moyenne arithmétique peut être définie comme étant la valeur de a qui rend minimum la quantité $\sum (x_i - a)^2$, de même que la médiane est celle qui minimise $\sum |x_i - a|$; ces deux options pour caractériser une tendance centrale équivalent à réduire *quantitativement* la masse des erreurs, en introduisant le critère de *distance* (euclidienne ou autre).

Ne serait-il pas plus satisfaisant, à de nombreux points de vue, de réduire *numériquement* les erreurs, en astreignant les données à répondre, dans leur majorité, à une norme de qualité ? Le critère utilisé serait alors la fréquence ?

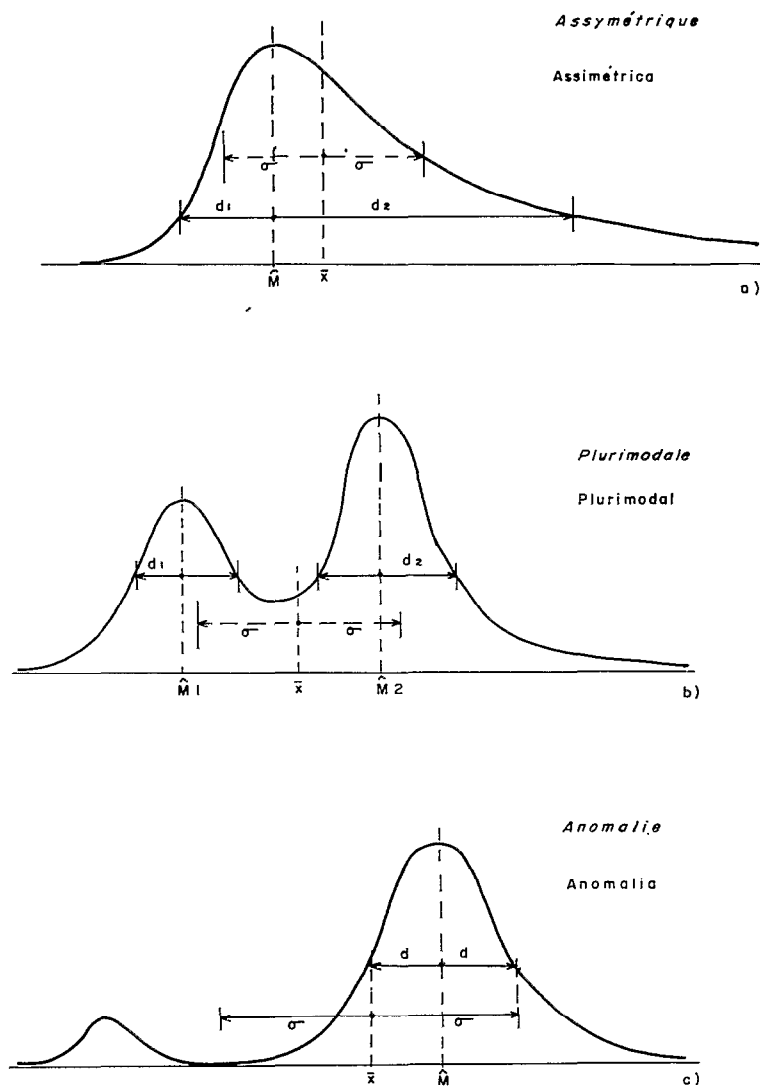


Fig. 4

Cela équivaudrait à utiliser une méthode de maximum de vraisemblance dont la solution réside dans la recherche des valeurs modales.

Si nous revenons à l'équation

$$E = [A - B]$$

définissant la matrice des erreurs, que nous pouvons encore écrire :

$$E = [A - L.C.] \tag{16}$$

avec $e_{ij} = a_{ij} - L_i \cdot C_j$

nous dirons que L_i et C_j seront correctement estimés quand la valeur de e_{ij} la plus fréquente — donc la plus probable, ou la plus vraisemblable — sera la valeur zéro.

En fait, il sera plus intéressant d'utiliser les erreurs relatives définies par :

$$E = L^{-1} EC^{-1} = [L^{-1} AC^{-1} - [1]] \quad (17)$$

avec $\varepsilon_{i,j} = \frac{a_{ij}}{L_i * C_j} - 1$.

Alors, la valeur la plus fréquente de ε_{ij} sera la valeur zéro, si la valeur modale de $(a_{ij}|L_i \cdot C_j)$ est l'unité, ce qui pourra s'écrire

$$\widehat{M}(a_{ij}/L_i * C_j) = 1 \quad (18)$$

Cette forme d'écriture permettra d'attribuer a priori le même poids à toutes les données, chacune se trouve normée par le produit de la ligne par la colonne auxquelles elle appartient. Si toutes les données étaient rigoureusement proportionnelles, la matrice $L^{-1} AC^{-1}$ serait une matrice de [1].

3.5. ESTIMATION DE LA VALEUR MODALE

Il n'existe pas, à notre connaissance, de méthode usuelle permettant le calcul précis du mode. Si les données étaient très nombreuses, on pourrait, à la rigueur, les grouper par classe et prendre pour valeur modale la valeur du point médian de la classe correspondant à la fréquence observée la plus élevée ; dans le cas contraire, il ne paraît pas possible de chercher les valeurs modales composant le vecteur ligne et le vecteur colonne tels qu'ils ont été définis en 3.2.3. sans choisir une formulation mathématique de la fonction de densité « ligne » et de la fonction de densité « colonne », ce que l'on ne saurait faire a priori. Dans ces conditions, il semble préférable de s'affranchir de toutes hypothèses sur la forme des fonctions de densité et de rechercher des méthodes donnant des valeurs suffisamment approchées des modes.

Nous passons donc rapidement en revue dans ce qui suit, une série de formules dont aucune ne prétend répondre avec pleine satisfaction au problème posé.

Si toutes sont fondées sur l'hypothèse que la distribution n'est pas connue a priori, elles ont toutes l'inconvénient majeur de supposer que cette distribution est unimodale ; seules des formules faisant intervenir des moments d'ordre supérieur à 4 permettraient — peut-être — d'envisager des distributions bi ou pluri-modales, mais on connaît la difficulté et le peu de précision du calcul des moments, à partir du cinquième ordre.

3.5.1. FORMULE CLASSIQUE

On peut mentionner tout d'abord une formule empirique qui est donnée dans divers manuels pour calculer une valeur du mode dans les distributions représentées par des courbes unimodales à faible asymétrie (23,5)

$$\widehat{M} = 3 \overline{\overline{M}} - 2 M \quad (19)$$

ou \widehat{M} est la valeur du mode, $\overline{\overline{M}}$ celle de la médiane et M celle de la moyenne arithmétique. Cette formule ne doit être utilisée qu'avec beaucoup de circonspection et uniquement pour les courbes réellement peu dissymétriques ; or nous ignorons, a priori, la forme des distributions dont nous devons extraire le mode, à moins d'en calculer les premiers moments.

3.5.2. PAR LE FACTEUR γ_1 DE DISSYMETRIE

Nous avons tenté d'en tenir compte, en utilisant la valeur x_0 qui annule le moment du 3^e ordre, soit :

$$\frac{1}{N} \sum (x_i - x_0)^3 = 0 \quad (20)$$

Soient K_2 et K_3 les cumulents du 2^e et du 3^e ordre, et $\gamma_1 = \frac{K_3}{K_2^{3/2}} = 2 \Gamma$ le premier coefficient de Pearson ; effectuons le changement de variable

$$Z_0 = \frac{x_0 - \bar{x}}{K_2^{1/2}}$$

l'expression (20) devient

$$z_0^3 + 3 z_0 - 2 \Gamma = 0$$

et

$$z_0 = \Gamma + (\Gamma^2 + 1)^{1/2}^{1/3} + (\Gamma - (\Gamma^2 + 1)^{1/2})^{1/3}$$

On peut montrer que le mode est lié à la médiane et à la moyenne par la relation approximative (avec les mêmes notations que dans la formule précédente) :

$$\widehat{M} + z_0 = M + \overline{M} \quad (21)$$

cette formule présente un net progrès sur la précédente car elle tient compte de la dissymétrie de la distribution (γ_1), mais elle nécessite outre celle des moments du 2^e et 3^e ordre, la connaissance de la médiane, qui n'est pas forcément estimée avec une bonne précision.

3.5.3. PAR LES DÉVELOPPEMENTS EN SÉRIE LIMITÉS

En définitive, il semble préférable d'utiliser, comme le fait Y. BRUNET-MORET, le développement en série limité aux premiers termes, d'une fonction de distribution quelconque $f(u)$ telle qu'elle est définie par MAC-CORNISH et FISHER (1).

$$f(u) \sim = z^{(0)}(u) - \left[+ \frac{\gamma_1}{6} z^{(3)}(u) \right] + \left[\frac{\gamma_2}{24} z^{(4)}(u) + \frac{\gamma_1^2}{72} z^{(6)}(u) \right] - \left[\frac{\gamma_3}{120} z^{(5)}(u) + \frac{\gamma_1 \gamma_2}{144} z^{(7)}(u) + \frac{\gamma_1^3}{1296} z^{(9)}(u) \right] + \dots \quad (22)$$

avec $u = \frac{x - \bar{x}}{K_2^{1/2}}$ et $z(u) = \frac{1}{\sqrt{2} \pi} \exp\left(-\frac{u^2}{2}\right)$, $z^{(n)}(u)$ étant la dérivée n-ième de $z(u)$.

$$\frac{dz^{(n)}}{du} = (-1)^{n+1} * z(u) * H_{n+1}(u)$$

$H_n(u)$ étant le polynome d'Hermite de degré n obtenu par la récurrence :

$$H_{n+1}(u) = u \cdot H_n(u) - n \cdot H_{n-1}(u)$$

avec $H_0 = 1$, $H_1 = u$.

La valeur modale réduite sera la valeur de u pour laquelle $f(u)$ est maximal, soit pour $f'(u) = 0$.

Ici,

$$f'(u) = \left[-z^{(0)} * H_1 \right] - \left[\frac{\gamma_1}{6} z^{(0)} * H_4 \right] + \left[\frac{\gamma_2}{24} z^{(0)} * H_5 - \frac{\gamma_1^2}{72} z^{(0)} * H_7 \right] - \dots = 0 \quad (23)$$

ou, si l'on se limite au deux premiers termes pour :

$$-z^{(0)} \cdot \left[H_1 + \frac{\gamma_1}{6} H_4 \right] = 0$$

ou

$$u^4 - 6u^2 + \frac{6u}{\gamma_1} + 3 = 0 \quad (24)$$

Les autres termes du développement font intervenir des moments d'ordre plus élevé et de ce fait permettent de tenir compte de la forme de la distribution mais le degré de l'équation à résoudre augmente très rapidement et les temps de calcul deviennent prohibitifs ; de plus, il devient de plus en plus difficile de choisir, parmi les racines, celle qui correspond au maximum principal de $f(u)$ alors que dans l'équation (24) on sait que la plus petite fournit la solution cherchée.

Y. BRUNET-MORET propose si $|\gamma_1| < 0,5$, l'approximation :

$$u = \frac{1 - \sqrt{2\gamma_1^2 + 1}}{2\gamma_1} \quad (25)$$

et pour les très faibles valeurs de $|\gamma_1|$

$$u = -\frac{\gamma_1}{2} \quad (26)$$

Nous préférons utiliser une formule du type :

$$u = -D \gamma_1 \sqrt{\frac{\gamma_1^2 + A}{\gamma_1^2 (\gamma_1^2 + B) + C}} \quad (27)$$

avec $A = 4,297\ 404\ 6$ $C = 9,463\ 056\ 7$
 $B = 11,327\ 238\ 6$ $D = 0,741\ 963\ 8.$

Elle est légèrement plus précise que (25) et respecte les conditions limites ($|u|$ tend vers $\sqrt{3} - \sqrt{6} = 0,742$ quand $|\gamma_1|$ tend vers l'infini et vers $\left|\frac{\gamma_1}{2}\right|$ quand $|\gamma_1|$ tend vers zéro) sans nécessiter un temps de calcul sensiblement plus long, tout en présentant l'avantage d'avoir un dénominateur toujours différent de 0, quelles que soient les valeurs de γ_1 .

Il n'en reste pas moins que cette formule, comme les précédentes, ne constitue qu'une approximation de la valeur modale d'une distribution à un seul maximum, ce qui est très restrictif ; elle aussi peut être, par ailleurs, assez sensible aux valeurs extrêmes — bien que $|u|$ ait une limite supérieure — du fait que le calcul du mode fait intervenir la dispersion de la distribution :

$$\widehat{M} = \bar{x} + \sigma * u \quad (\sigma = K_2^{1/2})$$

Dans une phase ultérieure, on envisage de recourir à une nouvelle méthode, non paramétrique, de détermination de la valeur modale.

3.6. MISE EN ŒUVRE DE \widehat{M} SUR UNE MATRICE

3.6.1. Le traitement « L.C. »

Sous ses formes (16) ou (17) présentées aux pages 144 et 145, le problème de l'élaboration d'un vecteur de référence paraît a priori difficilement soluble, du moins globalement.

Mais on peut remarquer que pour toute colonne j donnée de la matrice des données, on peut écrire quel que soit i :

$$\widehat{M} \left(\frac{a_{ij}}{L_i C_j} \right) = \frac{1}{C_j} \widehat{M} \left(\frac{a_{ij}}{L_i} \right) = 1,$$

de même pour tout i donné

$$\widehat{M} \left(\frac{a_{ij}}{L_i C_j} \right) = \frac{1}{L_i} \widehat{M} \left(\frac{a_{ij}}{C_j} \right) = 1.$$

En supposant le problème résolu, nous en déduisons que :

$$\boxed{\begin{aligned} C_j &= \widehat{M}_{vj} \left(\frac{a_{ij}}{L_i} \right) \\ L_i &= \widehat{M}_{vi} \left(\frac{a_{ij}}{C_j} \right) \end{aligned}} \quad (28)$$

$$L = \widehat{M}(AC^{-1})$$

$$C = \widehat{M}(L^{-1}A)$$

en fait, la forme des expressions (28) est valable quelle que soit la caractéristique de tendance centrale calculée ; on pourra par exemple utiliser la moyenne et l'on écrira :

$$\bar{C} = \frac{1}{N} \sum \frac{a_{ij}}{L_i} \quad \text{ou} \quad \bar{C} = \text{moy}(L^{-1}A).$$

Ces mêmes expressions (28) nous fournissent tous les éléments d'un processus itératif : en appelant p l'ordre des itérations, il suffira d'écrire :

$$\begin{aligned} C_j^p &= \widehat{M}\left(\frac{a_{ij}}{L_i^{p-1}}\right) \\ L_i^p &= \widehat{M}\left(\frac{a_{ij}}{C_j^p}\right) \end{aligned} \quad (29)$$

en initialisant par $L_i^1 = 1$.

Nous appellerons ce procédé le « Traitement Ligne-Colonne » ou « L.C. ».

Il est toujours convergent, mais plus ou moins rapidement suivant la qualité des données et la formule d'estimation de la valeur modale.

Afin de réduire le nombre d'itérations, on introduit un seuil de convergence s limitant la précision du calcul du produit $L_i * C_j$, pour un i et un j donnés ; on arrête le traitement « L.C. » pour

$$\gamma_{ij} = \frac{L_i^p * C_j^p}{L_i^{p-1} * C_j^{p-1}} - 1 \leq s.$$

Mais le choix de i et de j reste assez arbitraire et il serait peu économique d'explorer la totalité du champ des γ_{ij} .

On peut alors, avantagusement, ne considérer que le vecteur

$$PR_i = \frac{L_i^p}{L_i^{p-1}}$$

et faire agir le seuil quand

$$\frac{PR_{\max} - PR_{\min}}{PR_{\text{moy}}} \leq 2s.$$

Par suite de l'utilisation de la valeur modale, le traitement n'est pas altéré si des données manquent sur une ligne ou une colonne à condition toutefois qu'elles soient en nombre suffisant (au moins 3) ; mais bien entendu l'estimation sera d'autant plus précise que les données sont nombreuses.

Notons enfin que le traitement « L.C. », en utilisant un autre centrage que celui de la valeur modale, fournit un algorithme extrêmement performant pour la détermination, au sens classique, de la composante principale d'une matrice.

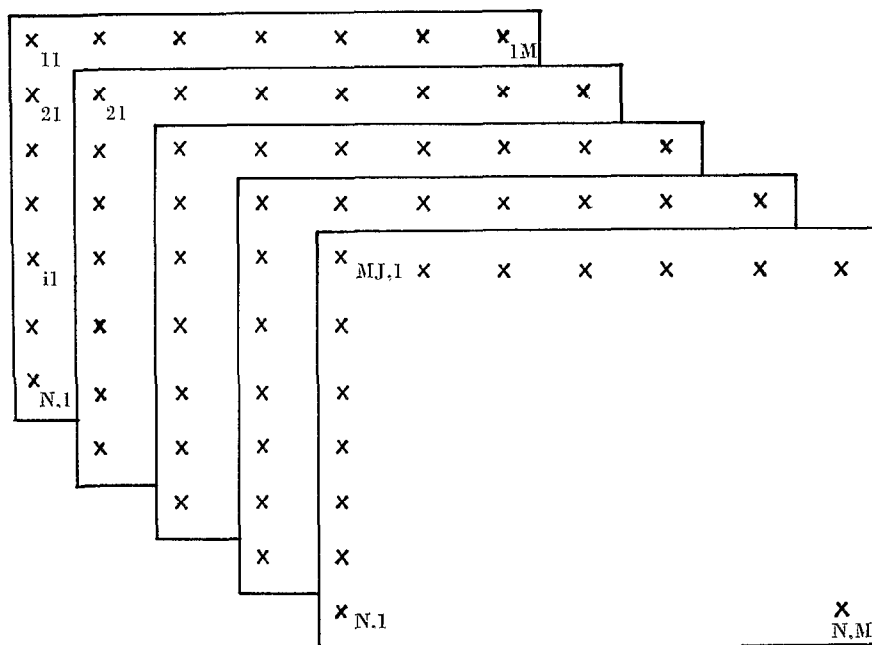
3.6.2. DÉCOMPOSITION EN SOUS-MATRICES

Pour toutes les raisons déjà énumérées, mais principalement parce que l'on ne peut jamais être certain d'être en présence d'une distribution unimodale sans recourir au calcul long et imprécis de moments d'ordre élevé, il s'est avéré nécessaire de rechercher des artifices complémentaires permettant une détermination plus fiable de la composante dominante de la matrice des données.

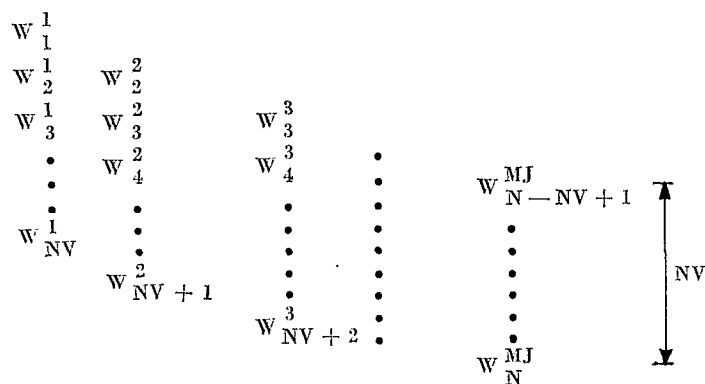
Partant du principe que plus la période d'observation est courte, plus on a de chance de se trouver à l'intérieur d'une séquence homogène unique, dont les données sont à distribution plus ou moins unimodale, on peut penser

qu'en réduisant la hauteur de la matrice, un plus grand nombre de postes sont susceptibles d'être homogènes durant la même période ; et qu'ainsi la formule d'estimation serait plus valide (γ_1 plus faible) et permettrait une meilleure approche des valeurs annuelles du vecteur de référence (tout au moins en valeur relative).

Soit NV , la hauteur de ces matrices réduites ; nous devons donc décomposer la matrice de hauteur N des données en $(N - NV + 1) = MJ$ sous-matrices imbriquées.



Le traitement L.C. appliqué à chacune d'elles (en utilisant par exemple, la formule (27)) fournira un vecteur partiel de référence W_k , dont, comme nous allons le voir, il ne sera pas nécessaire de connaître la base : si nous repérons les sous-matrices par les indices supérieurs 1, 2, ..., nous pouvons représenter les vecteurs partiels par le tableau suivant de dimensions : N, MJ .



Si ces vecteurs partiels avaient une base unique et que leurs composantes soient connues avec précision il y aurait redondance de l'information ($w_3^1 = w_3^2 = w_3^3$) ; il nous suffirait donc de diviser chaque vecteur par sa base respective (C_j) pour obtenir des lignes de valeurs à peu près semblables dont on extrairait la valeur modale (L_i). Nous voyons ainsi réapparaître la possibilité d'utiliser le traitement « L.C. » pour la détermination d'un vecteur unique de référence \bar{W} , de longueur N .

3.6.3. Corrections successives

Une autre approche consiste, après la sortie de chaque traitement L.C., à établir la matrice des erreurs relatives :

$$\varepsilon_{i,j} = a_{ij}/L_i * C_j - 1$$

comme elle a été établie en 3.4, et à rechercher parmi toutes les lignes et toutes les colonnes, la valeur de $x_{i,j}$ pour laquelle $\varepsilon_{i,j}$ est maximum.

On corrige alors cette donnée $x_{i,j}$ en la multipliant par le coefficient :

$$K_{i,j} = \frac{1}{1 + \varepsilon_{ij}} \quad \text{ou par} \quad K_{ij} = \frac{1}{1 + \alpha * \varepsilon_{ij}}$$

dans laquelle α est un facteur d'accélération ou de ralentissement de convergence, selon qu'il est supérieur ou inférieur à l'unité.

On procède alors à une nouvelle détermination des vecteurs L_i et C_j .

On arrête les corrections quand tous les ε_{ij} sont inférieurs à un certain seuil ; l'expérience a montré que ce seuil pouvait être choisi assez bas (de l'ordre de 1×10^{-2}) sans altérer notablement le résultat : on réduit ainsi le temps de calcul.

Sous certaines conditions l'application à cette approche de la décomposition en sous-matrices décrite précédemment permettra également de réduire très substantiellement les temps de calcul (1).

3.6.4. Sensibilité et fiabilité de l'algorithme

L'élaboration du vecteur de référence est soumise, nous l'avons vu, à des contraintes très sévères si l'on veut respecter les principes jugés indispensables à toute recherche d'une information pluviométrique « représentative » (voir par. 3.1).

Nous avons tenu à nous assurer qu'ils sont bien respectés ; pour cela deux tests expérimentaux ont été utilisés : ils permettent d'éprouver la sensibilité et la fiabilité de l'algorithme.

Le premier a consisté à construire une matrice dégénérée purement artificielle, ne comportant pas de composante aléatoire, de dimension 9×5 , sur laquelle on a appliqué des déviations systématiques et introduit des anomalies locales de telle façon qu'aucune ligne ni colonne ne reste sans altérations.

<i>Vecteur- ligne origine</i>	<table style="border-collapse: collapse; width: 100%; text-align: center;"> <tr><td style="padding: 5px;">5</td><td style="padding: 5px;">5</td><td style="padding: 5px;">10</td><td style="padding: 5px;">13</td><td style="padding: 5px;">20</td><td style="padding: 5px;">25</td></tr> <tr><td style="padding: 5px;">2</td><td style="padding: 5px;">2</td><td style="padding: 5px;">4</td><td style="padding: 5px;"><u>3,6</u></td><td style="padding: 5px;">8</td><td style="padding: 5px;">10</td></tr> <tr><td style="padding: 5px;">3</td><td style="padding: 5px;">3</td><td style="padding: 5px;">6</td><td style="padding: 5px;"><u>6,4</u></td><td style="padding: 5px;">12</td><td style="padding: 5px;">15</td></tr> <tr><td style="padding: 5px;">4</td><td style="padding: 5px;">4</td><td style="padding: 5px;">8</td><td style="padding: 5px;"><u>7,2</u></td><td style="padding: 5px;">16</td><td style="padding: 5px;"><u>24</u></td></tr> <tr><td style="padding: 5px;">2</td><td style="padding: 5px;"><u>1,5</u></td><td style="padding: 5px;">4</td><td style="padding: 5px;">6</td><td style="padding: 5px;"><u>12</u></td><td style="padding: 5px;"><u>10</u></td></tr> <tr><td style="padding: 5px;">3</td><td style="padding: 5px;"><u>3,6</u></td><td style="padding: 5px;"><u>4,8</u></td><td style="padding: 5px;">9</td><td style="padding: 5px;"><u>18</u></td><td style="padding: 5px;">15</td></tr> <tr><td style="padding: 5px;">5</td><td style="padding: 5px;"><u>6,0</u></td><td style="padding: 5px;"><u>8,0</u></td><td style="padding: 5px;">15</td><td style="padding: 5px;"><u>20</u></td><td style="padding: 5px;"><u>22</u></td></tr> <tr><td style="padding: 5px;">4</td><td style="padding: 5px;"><u>4,8</u></td><td style="padding: 5px;"><u>6,4</u></td><td style="padding: 5px;">12</td><td style="padding: 5px;">16</td><td style="padding: 5px;"><u>20</u></td></tr> <tr><td style="padding: 5px;">6</td><td style="padding: 5px;"><u>7,2</u></td><td style="padding: 5px;"><u>9,6</u></td><td style="padding: 5px;">18</td><td style="padding: 5px;">24</td><td style="padding: 5px;">30</td></tr> </table>	5	5	10	13	20	25	2	2	4	<u>3,6</u>	8	10	3	3	6	<u>6,4</u>	12	15	4	4	8	<u>7,2</u>	16	<u>24</u>	2	<u>1,5</u>	4	6	<u>12</u>	<u>10</u>	3	<u>3,6</u>	<u>4,8</u>	9	<u>18</u>	15	5	<u>6,0</u>	<u>8,0</u>	15	<u>20</u>	<u>22</u>	4	<u>4,8</u>	<u>6,4</u>	12	16	<u>20</u>	6	<u>7,2</u>	<u>9,6</u>	18	24	30
5	5	10	13	20	25																																																		
2	2	4	<u>3,6</u>	8	10																																																		
3	3	6	<u>6,4</u>	12	15																																																		
4	4	8	<u>7,2</u>	16	<u>24</u>																																																		
2	<u>1,5</u>	4	6	<u>12</u>	<u>10</u>																																																		
3	<u>3,6</u>	<u>4,8</u>	9	<u>18</u>	15																																																		
5	<u>6,0</u>	<u>8,0</u>	15	<u>20</u>	<u>22</u>																																																		
4	<u>4,8</u>	<u>6,4</u>	12	16	<u>20</u>																																																		
6	<u>7,2</u>	<u>9,6</u>	18	24	30																																																		
	<table style="border-collapse: collapse; width: 100%; text-align: center;"> <tr><td style="padding: 5px;">1</td><td style="padding: 5px;">2</td><td style="padding: 5px;">3</td><td style="padding: 5px;">4</td><td style="padding: 5px;">5</td></tr> </table>	1	2	3	4	5																																																	
1	2	3	4	5																																																			
	<i>Vecteur-colonne origine</i>																																																						

(Les valeurs modifiées ont été soulignées.)

(1) Le procédé des corrections successives est équivalent à une contraction progressive du rang de la matrice des données.

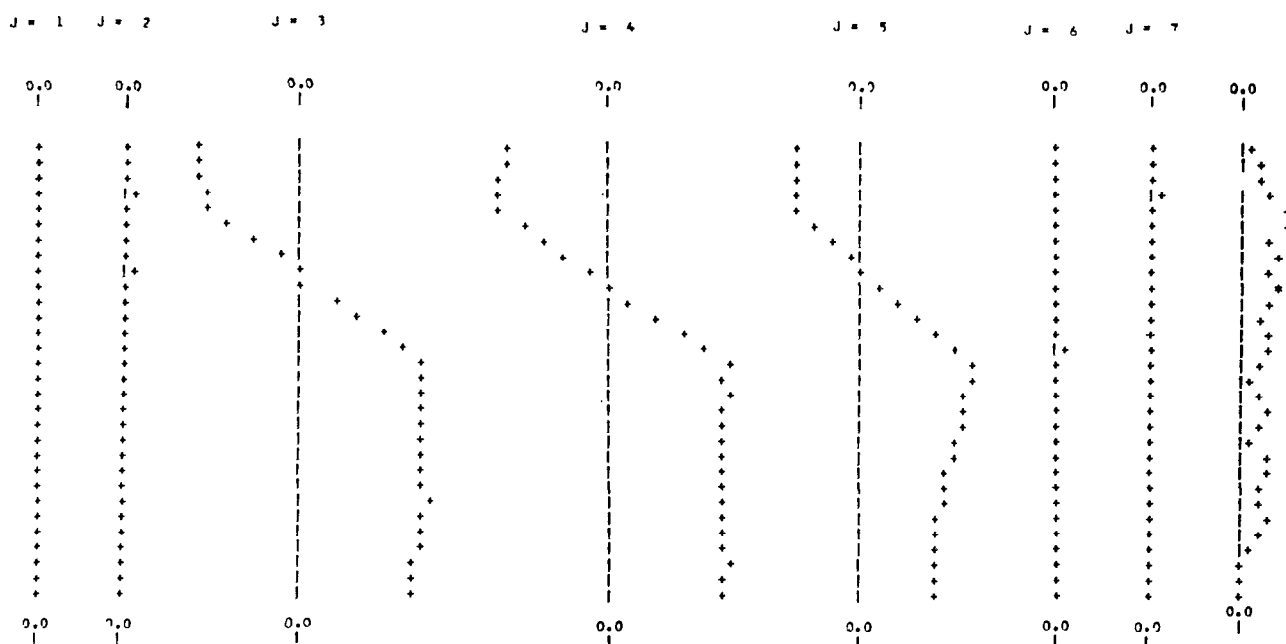


Fig. 5
 Test de sensibilité (Teste de sensibilidade)
 Données des postes 3-4-5 \times 1,25 pendant 10 ans
 Dados dos postos 3-4-5 \times 1,25 durante 10 anos

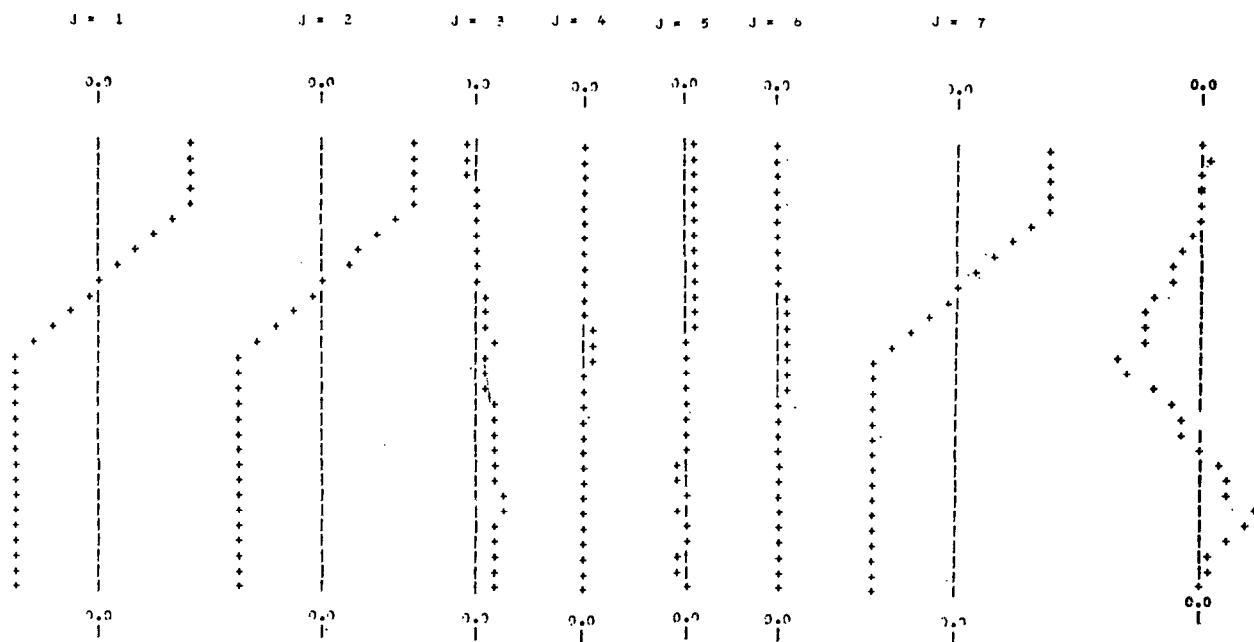


Fig. 6
 Test de sensibilité (Teste de sensibilidade)
 Données des postes 3-4-5-6 \times 1,25 de 1945 à 1954
 Dados dos postos 3-4-5-6 \times 1,25 de 1945 à 1954

Cette matrice, présentée ci-dessus, sera reprise en exemple en annexe 1 et l'on pourra constater que le procédé a permis de retrouver sans aucune erreur les vecteurs-ligne et colonne d'origine.

Il est important de noter que de nombreuses autres méthodes, plus conventionnelles, ont été expérimentées sur cette même matrice sans aucun succès (y compris la méthode des composantes principales).

Un deuxième test porte sur une autre série simulée (par F. MONIOD), dont les données ressemblent davantage à des totaux pluviométriques annuels (composante aléatoire) et l'on pourra remarquer (fig. 5) que le « simple-cumul » du vecteur représentatif n'est pas des plus réguliers ; cette série comporte 7 postes observés fictivement de 1940 à 1969, soit pendant 30 ans.

En un premier temps on a multiplié les postes 3, 4 et 5, de l'année 1945 à l'année 1954, par le coefficient 1,25 : sur la figure 5, on notera que le graphique de double-cumul entre le vecteur obtenu par l'algorithme et chacun des postes fait ressortir une cassure sur ces 3 mêmes postes alors que le simple cumul du vecteur (à droite de la figure) est normal : on en conclut que l'élaboration du vecteur n'a pas été influencée par les déviations introduites.

En un deuxième temps, on a multiplié un quatrième poste, le 6, par le même coefficient 1,25 : on a alors la surprise de constater que les cassures n'apparaissent plus sur les postes 3, 4, 5 et 6 mais sur les postes 1, 2 et 7 et le simple cumul du vecteur permet de mettre en évidence une cassure pendant la même période ; ceci illustre tout simplement le fait que l'algorithme a obéi à la loi du plus grand nombre (4 postes déviés sur 7 observés).

Cette constatation est très satisfaisante sur le plan de la sensibilité de l'algorithme, mais elle entraîne des conséquences importantes quant à la fiabilité des résultats qu'il fournit.

Le vecteur de référence obtenu, réputé représentatif de l'ensemble des postes considérés, pourra être systématiquement faussé, sans que l'on puisse incriminer le procédé de traitement :

- soit par le fait qu'une majorité des postes de l'ensemble a été observée dans des conditions particulières : par exemple, distribution à la même date d'éprouvettes inadéquates ;
- soit par le fait que les postes n'obéissent pas tous à la même tendance climatique, impliquant un mauvais choix du groupement régional (par exemple dans les zones de transition de régime littoral-intérieur).

On devra donc *toujours* vérifier la validité du vecteur de référence par simple cumul et par double cumul avec des vecteurs de références voisins et, éventuellement, procéder à une modification du groupement des postes.

4. LA CRITIQUE DES DONNÉES

Si nous retenons le principe du double cumul comme méthode d'analyse des totaux pluviométriques annuels, il suffira maintenant de comparer les données de chaque poste au vecteur de référence qui vient d'être établi, sous réserve que toutes les conditions favorables aient été réunies pour son obtention.

Les modalités différeront toutefois profondément de la méthode classique (voir 2.2.2), en ce sens que l'on ne rapportera plus les données à leur valeur moyenne, mais à leur valeur la plus vraisemblable.

Les équations (16) ou (17) définissent en effet parfaitement les écarts à cette valeur la plus probable, soit en valeur absolue soit en valeur relative, dont nous rappelons l'expression :

$$\varepsilon_{ij} = a_{ij}/L_i * C_j - 1.$$

Or le procédé qui nous permet d'obtenir le vecteur de référence (L), nous permet, par la même occasion, de connaître la norme en valeur modale de chacun des postes (C_j) ; nous disposons donc d'une mesure a priori des écarts des données à leur valeur la plus probable.

C'est cette mesure ε_{ij} , que nous appellerons par la suite « EPSI », que nous utiliserons pour l'analyse des données ; l'utilisation de l'écart ou erreur relative étant, à l'évidence, préférable par sa nature adimensionnelle.

4.1. DISTRIBUTION DES EPSI

Avant d'étudier l'évolution des EPSI en fonction du temps, spécifique de chaque poste, il paraît intéressant d'observer leur comportement global par ensemble de postes. Les histogrammes de leur distribution présentent, en effet une allure tout à fait remarquable et des caractéristiques communes semblent se dégager, quel que soit le groupe de postes considéré.

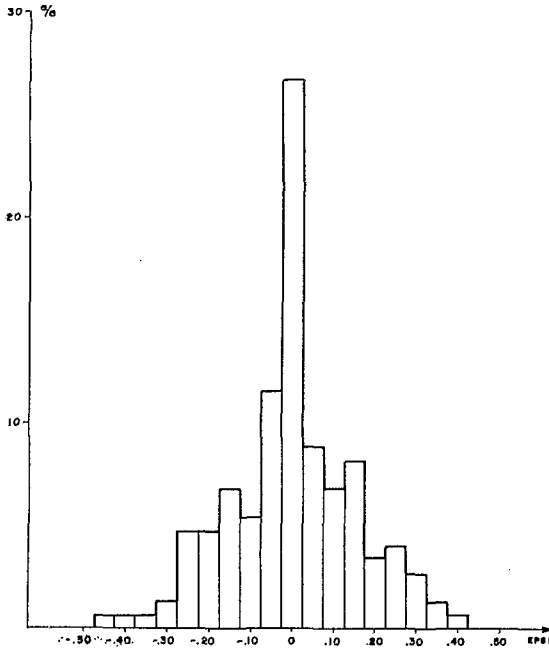


Fig. 7
 Histogramme de la distribution de « epsi »
 Données du Sénégal sur 146 postes-années
 Histograma da distribuição de « epsi »
 ados do Senegal sobre 146 postos-anos

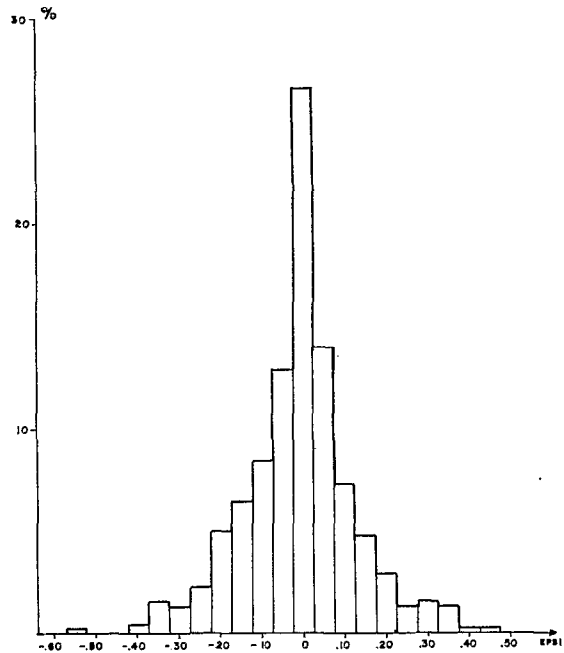


Fig. 8
 Histogramme de la distribution de « epsi »
 Données de Côte-d'Ivoire sur 434 postes-années
 Histograma da distribuição de « epsi »
 Dados da costa do marfim sobre 434 postos-anos

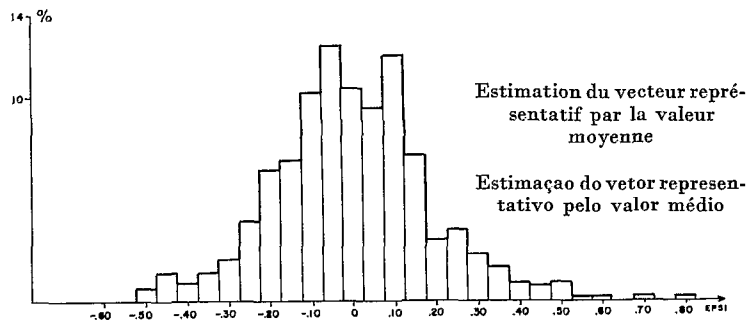


Fig. 9 a

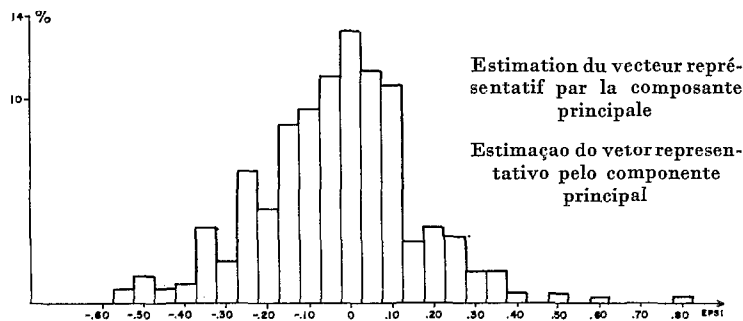


Fig. 9 b

Ensemble de 8 postes de l'Etat de Pernambuco (Brésil)
 Conjunto de 8 postos do estado de Pernambuco (Brasil)

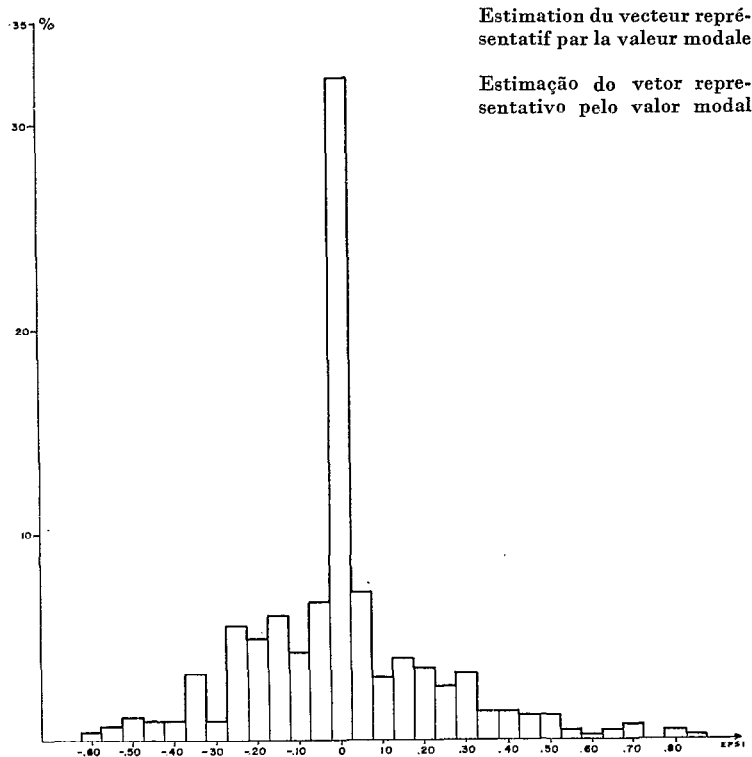


Fig. 9 c

Ensemble de 8 postes de l'Etat de Pernambuco (Brésil)
 Conjunto de 8 postos do estado de Pernambuco (Brasil)

On constate (fig. 7 et 8) que :

— l'intervalle de classe de EPSI qui permet de décrire au mieux les distributions se situe autour de 0,05 ;
 — la forme de la distribution est très éloignée de toutes les lois paramétriques connues et singulièrement de la loi normale ;

— la classe centrale de $-0,025$ à $+0,025$ a toujours la fréquence la plus élevée et cette fréquence dépasse de loin celle de toutes les autres classes ;

— ces distributions pourraient être considérées comme composées d'une loi sensiblement normale groupant les valeurs centrales et représentant en quelque sorte le « bruit propre » du système, et d'une superposition de lois quelconques figuratives des erreurs systématiques ou des anomalies. S'il en était ainsi la mesure du « bruit propre » permettrait de disposer d'un critère objectif pour détecter les ruptures d'homogénéité.

On constate enfin que seule l'utilisation de la valeur modale permet d'obtenir des fréquences aussi élevées pour la classe centrale de EPSI ; l'estimation du vecteur à partir de toute autre caractéristique (moyenne, composante principale...) conduit à des distributions considérablement plus aplaties. On s'en fera une idée en consultant les figures 9 où sont représentés les histogrammes correspondant à chaque type d'estimation et relatifs à un même ensemble de données.

Ces histogrammes de EPSI, permettent, en outre :

- 1) De vérifier que l'hypothèse initiale de linéarité est justifiée.
- 2) De démontrer la supériorité de la valeur modale comme paramètre de tendance centrale, en particulier dans un système linéaire.
- 3) De conclure à l'efficacité du traitement « L.C. ».

Ils montrent aussi la tendance du procédé à limiter le nombre des corrections et la nécessité d'effectuer celles qui ont été nettement détectées, surtout s'il s'agit d'établir des « normales » pluviométriques.

4.2. LES ERREURS RELATIVES CUMULÉES

Rejoignant ici seulement la méthode classique des double-cumul, la détection des anomalies ou des déviations systématiques pourra se faire visuellement sur le graphique des valeurs cumulées de EPSI, appelées SEPSI, telles que pour chaque poste j :

$$\text{SEPSI}(j) = \sum_1^j \text{EPSI}(i) = \frac{1}{C_j} * \sum_1^j \frac{a_{ij}}{L_i} - j$$

où $\frac{a_{ij}}{L_i}$ tend (le plus souvent) vers C_j en absence de déviations.

Les graphiques ne se referment pas sur l'origine comme dans le cas d'un double cumul classique (fig. 10), à moins que le mode trouvé se situe très près de la moyenne arithmétique ; le tracé reste parallèle à l'axe lorsque les données sont homogènes par rapport à la période de référence sur laquelle a été calé le mode ; cette particularité facilite très généralement l'interprétation des graphiques ; il peut cependant arriver que la période de référence retenue ne corresponde pas à celle où se situe la valeur modale de L_i et dans ce cas l'alignement des SEPSI peut fort bien ne plus être parallèle à l'axe ; cela peut facilement se produire lorsque la fonction de densité-colonne est plurimodale, c'est-à-dire lorsqu'un poste est affecté de plusieurs cassures.

L'unique façon de lever les doutes consistera à établir le tableau des coefficients de correction correspondant à chaque période et éventuellement à changer de période de référence.

Bien que présentant un avantage considérable sur la méthode classique des double-cumuls poste à poste, puisque l'on dispose immédiatement des éléments de décision, sans qu'il soit nécessaire de recourir à des comparaisons avec d'autres postes du groupe, le cumul des EPSI soulève encore quelques difficultés d'interprétation.

L'une des principales est due aux ruptures de séquence provoquées par la présence de valeurs isolées anormalement élevées ou faibles — anomalie locale possible, à prendre en considération, ou erreur grossière à éliminer — ; les valeurs trop faibles se rencontrent d'ailleurs plus fréquemment et correspondent bien souvent à des lacunes dans les observations.

Cette difficulté peut être levée en éliminant *momentanément* les valeurs « anormales » des *graphiques de cumul* : on voit alors réapparaître des alignements qu'il était difficile de déceler auparavant.

Une autre difficulté provient de l'imprécision des transitions de séquence à séquence ; elle peut heureusement être considérablement atténuée en utilisant une nouvelle variable « DÉPSI », constituée de la différence première des variables EPSI :

$$\text{DEPSI}_i = \text{EPSI}_i - \text{EPSI}_{i-1}.$$

Cette variable offre, à notre sens, autant si non plus d'intérêt que la propre variable SEPSI, dont pourtant on ne peut nier l'efficacité quand on limite l'enquête critique à un examen visuel (1).

5. DESCRIPTION DU PROGRAMME

Outre le premier programme d'études, élaboré en 1971 au Brésil (Campina Grande), dans le cadre des accords de coopération internationale entre l'OMN-UNESCO et la Sudene, il existe actuellement plusieurs versions du programme de traitement utilisant les nouveaux principes précédemment définis :

— Un premier programme — expérimental — a été mis au point au Bureau Central à Paris en 1973 sur ordinateur IBM 370-168, avec la collaboration de G. VUILLAUME, Maître de Recherches à l'ORSTOM.

— En vue de l'application de cette méthode aux données du Nordeste du Brésil, deux versions ont été simultanément adaptées, à Recife et à Fortaleza, au cours du premier trimestre de 1976 :

- l'une sur l'ordinateur IBM 360-30 de la Sudene par G. JACCON, Chef de la Mission hydrologique française,
- l'autre, par l'auteur, sur l'ordinateur IBM 1130 de l'Université fédérale de Fortaleza (État du Ceara).

Pour tenir compte à la fois des capacités de mémoire réduites de ces deux unités de calcul et des temps de passage considérablement plus élevés que ceux d'un 370, ces deux versions ont exigé de profondes modifications

(1) La détection automatique des ruptures de séquence sera exposée dans une prochaine note.

CHUVA - DOMINICA				ALFARO - 197314				REGIÃO MATRIZ MONTE						
N	ANO	OBS.	CALC.	EPST	SEPSI	-4.0	I	-2.0	I	0.0	I	+2.0	I	+4.0
1	1940	85.8	157.2	-0.4287	1.304									
2	1941	57.8	160.0	-0.6760	0.620									
3	1942	50.6	158.0	-0.4264	0.199									
4	1943	147.9	148.0	-0.0005	0.193									
5	1944	162.5	162.0	0.0001	0.193									
6	1945	150.0	168.8	0.0003	0.199									
7	1946	162.6	164.4	-0.0050	0.194									
8	1947	124.4	184.0	-0.1570	-0.004									
9	1948	128.2	184.0	-0.2202	-0.324									
10	1949	98.3	164.0	-0.4667	-0.023									
11	1950	164.2	162.0	0.0103	-0.015									
12	1951	152.2	152.0	-0.0030	-0.017									
13	1952	189.1	185.0	0.0210	-0.790									
14	1953	152.2	152.0	-0.0002	-0.001									
15	1954	171.0	170.0	0.0000	-0.797									
16	1955	132.9	130.0	-0.0003	-0.033									
17	1956	150.0	160.0	0.0000	-0.037									
18	1957	172.8	170.0	-0.0000	-0.070									
19	1958	127.6	162.0	-0.2216	-1.097									
20	1959	120.3	170.0	-0.2310	-1.303									
21	1960	144.8	170.0	-0.1500	-1.510									
22	1961	128.6	181.0	-0.2210	-1.740									
23	1962	163.6	184.0	-0.0000	-1.744									
24	1963	152.1	152.0	-0.0000	-1.740									
25	1964	171.0	171.0	-0.0000	-1.747									
26	1965	150.7	160.0	0.0000	-1.730									
27	1966	167.5	167.1	0.0003	-1.730									
28	1967	178.2	175.0	0.0140	-1.701									
29	1968	200.3	200.0	-0.0000	-1.704									
30	1969	154.0	160.0	0.0000	-1.703									

N	ANO	OBS.	CALC.	EPST	SEPSI	-4.0	I	-2.0	I	0.0	I	+2.0	I	+4.0	
VALOR MOIAL											173.2				
CHUVA MEDIA NO PERIODO DE OBSERVACOES						1940-1969	30 ANOS	OBSERVADA	153.6	CALCULADA	173.2				
CHUVA MEDIA ESTIMADA NO PERIODO DE VALOR						1940-1969	30 ANOS		173.2						

Fig. 10

du logiciel ; celle de Recife est davantage orientée vers la production de documents définitifs ; celle de Fortaleza, dont les sorties sont actuellement simplifiées, a servi à la mise au point d'algorithmes plus efficaces.

Nous ne décrivons ici qu'un schéma général de traitement commun à toutes les versions : il est relativement simple puisqu'on peut le résumer en trois grands sous-titres dont l'organisation ne présente aucune complexité (fig. 11) :

- 5.1. Préparation des données.
- 5.2. Traitement de l'information.
- 5.3. Résultats.

La partie qui se réfère au traitement aurait pu être encore simplifiée — et elle l'est dans le programme opérationnel — mais nous avons cru devoir y faire figurer la combinaison des différentes options proposées en 3.6.2. et 3.6.3., options commandées par les variables NV et KOR.

Le choix définitif de l'une d'elles permet évidemment d'éliminer certaines parties de l'organigramme.

5.1. PRÉPARATION DES DONNÉES

On s'est inspiré pour une grande part des programmes existant à l'ORSTOM : LMPDON, LECCOR et UNMASV.

On y trouve :

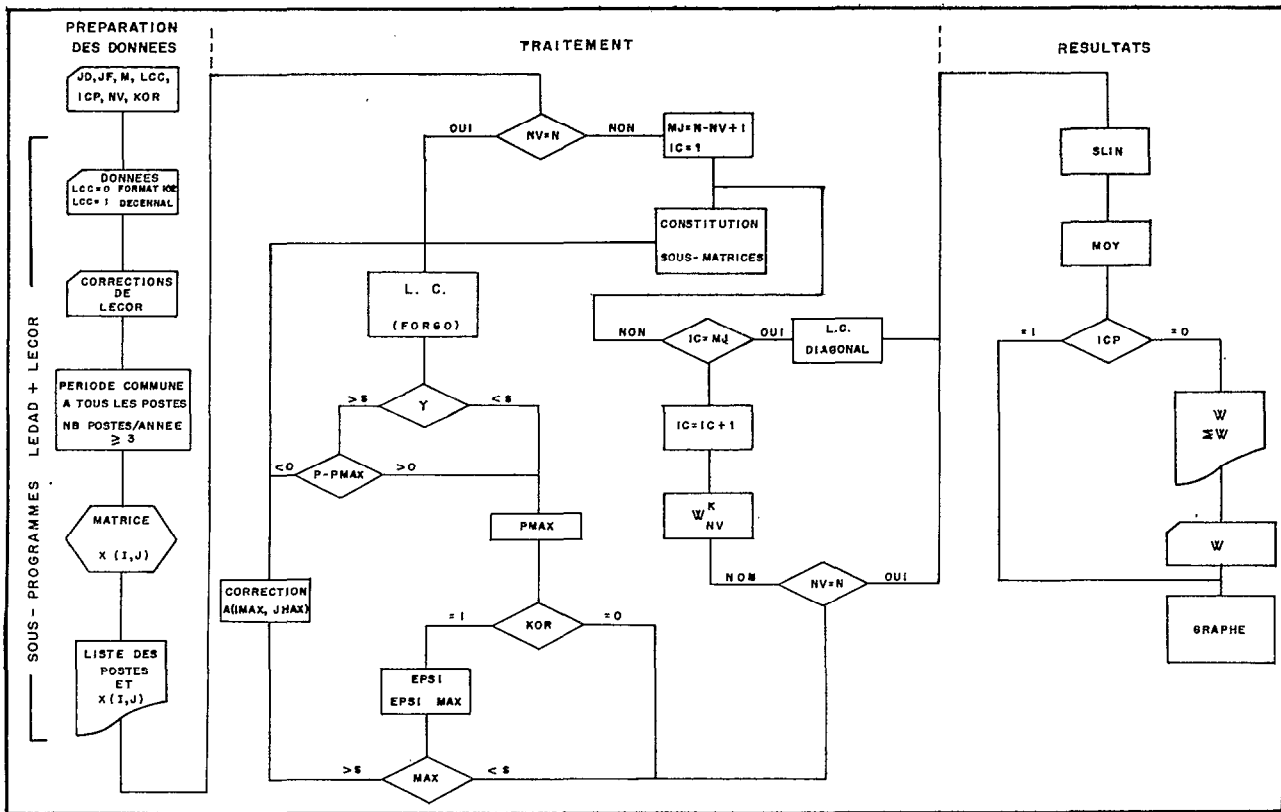


Fig. 11
Homogénéisation des données — Organigramme de traitement —

5.1.1. La lecture des paramètres d'entrée

- JD** année de début de traitement
JF année fin de traitement
M le nombre de postes à traiter
LCC = 0 si les données sont en format 102
 = 1 si les données sont en format décennal
ICP = 0 si toutes les sorties sont réclamées
 = 1 si seule la sortie graphique est demandée
NV = la hauteur des sous-matrices, le cas échéant
KOR = 0 pas de corrections itératives
 = 1 avec corrections itératives.

5.1.2. La lecture des données

Elle se fait par :

- 1 carte d'identification du poste,
- 1 carte de correction,
- x cartes au format COH 102 ou au format décennal,
- 1 carte blanche de fin de poste,

et par groupe de postes :

- 1 carte blanche de fin de groupe.

5.1.3. La constitution de la matrice des données

Le programme établit la période d'observation commune à tous les postes dans l'intervalle JD-JF, en éliminant les années durant lesquelles le nombre de postes ayant des observations serait inférieur à 3 (ou 5).

Si le nombre de postes à traiter est faible, il sera préférable de n'en conserver qu'un nombre impair.

5.1.4. L'impression de la liste des postes à traiter et de la matrice des données

5.2. LE TRAITEMENT PROPREMENT DIT

Comme nous l'avons vu les variables NV et KOR déterminent le type de traitement à appliquer ; les options sont au nombre de quatre :

5.2.1. $NV = N$ et $KOR = 0$

La matrice est traitée globalement et le programme de corrections itératives (3.6.3) n'est pas appliqué, ce qui implique une bonne précision de la formule utilisée pour l'estimation de la valeur modale, soit que la distribution soit unimodale, soit que l'on sache estimer le mode principal d'une distribution plurimodale.

C'est le mode de traitement le plus économique en temps machine, puisqu'il ne nécessite qu'une seule série d'itérations par « L.C. ».

5.2.2. $NV = N$ et $KOR = 1$

La matrice est traitée globalement et son rang est progressivement contracté par l'application de corrections successives aux données brutes. C'est *certainement* le traitement le plus fiable, même si la formule d'estimation du mode est imprécise. Il coûte malheureusement très cher, le temps de calcul croissant sensiblement comme le carré du produit des dimensions de la matrice. Son utilisation sur des matrices importantes n'est concevable qu'à partir d'ordinateurs de très forte puissance. (Une matrice de 18 postes sur 80 années a exigé 7,5 minutes de temps CPU sur un ordinateur IBM 370-168.)

5.2.3. $NV \neq N$ et $KOR = 0$

Ce traitement n'est concevable que si NV est suffisamment petit pour que la période observée corresponde à une distribution unimodale, pour laquelle la formule d'estimation (27) est jugée satisfaisante. Il est, lui aussi, très économique en temps machine.

5.2.4. $NV \neq N$ et $KOR = 1$

Cette option constitue un compromis pas trop onéreux, si l'on veut tout de même utiliser les corrections successives par souci de fiabilité. Il est souhaitable toutefois que, par poste, le nombre de séquences homogènes reste limité.

Le choix des options 3 et 4 est commandée par un premier branchement du programme qui permet la constitution éventuelle des sous-matrices de hauteur NV ; le programme de traitement « L.C. » est commun à toutes les options.

On voit qu'il peut être interrompu par un seuil « s » ou pour un nombre maximal prédéterminé d'itérations PMAX ; à sa sortie, un branchement permet d'appliquer ou non les corrections successives (options 1 ou 2).

Dans le cas où l'option $NV \neq N$ a été choisie, et quand la décomposition arrive à la dernière ligne de la matrice des données, le dernier vecteur partiel W^{M_j} de hauteur NV est calculé et l'ensemble des vecteurs partiels W^K ($\forall K : 1, M_j$) est lui-même traité, comme une matrice, par un programme L.C. spécial, tenant compte de la forme particulière de la matrice (voir 3.6.2).

Toutes les options débouchent alors sur le calcul de SLIN, valeur modale du vecteur L, et les L_i sont normées par $L_i = L_i/SLIN$.

L'option 1 ($NV = N$, $KOR = 0$) est actuellement utilisée avec la nouvelle méthode de détermination du mode dont il a été fait mention en 3.5. Encore en cours de test, elle permet déjà d'atteindre les meilleures performances temps-précision, elle fera prochainement l'objet d'un second article.

5.3. LES RÉSULTATS

Avec l'option $ICP = 0$, toutes les sorties sont commandées, soit :

5.3.1. L'impression du vecteur normé ainsi que son simple cumul.

5.3.2. La perforation du vecteur en format décennal (si nécessaire).

5.3.3. La sortie GRAPHE où sont résumés les résultats poste par poste.

Pour chacun, nous trouvons le numéro de code, le nom du poste et la région à laquelle il appartient ; puis successivement de gauche à droite et année par année :

- le numéro de l'année observée,
- le millésime de l'année observée,
- la hauteur de pluie observée,
- la hauteur théorique calculée,
- EPSI,
- SEPSI et le graphique de SEPSI,
- la moyenne observée et calculée sur la période observée du poste et la moyenne sur la période totale couverte par le vecteur.

Si le sous-programme de détection des « cassures » de la courbe des SEPSI a été introduit, un tableau est imprimé sur lequel figurent, par période homogène ou par groupe de périodes, les coefficients multiplicateurs de correction recommandés.

6. CONDITIONS D'APPLICATION

L'application de la méthode d'homogénéisation qui vient d'être présentée requiert le respect d'un certain nombre de règles assez strictes ; il est cependant difficile d'en dresser une liste rigoureuse et précise, chaque cas traité pouvant se révéler un cas particulier.

On peut cependant formuler les recommandations suivantes que l'on s'efforcera d'adapter au mieux, en fonction de l'ensemble traité :

6.1. Utilisation de l'année hydrologique comme base de calcul de la hauteur pluviométrique annuelle, surtout si la répartition mensuelle fait ressortir une saison sèche bien marquée.

6.2. Le nombre de postes servant à l'élaboration d'un vecteur de référence doit obligatoirement être supérieur à trois, mais le minimum souhaitable est de sept ; il n'y a pas de limite supérieure si tous les postes appartiennent à une même tendance climatique et si les plus éloignés sont à des distances n'excédant pas certaines limites d'ailleurs variables avec les zones traitées en fonction de l'homogénéité spatiale des précipitations.

Pour le Nordeste brésilien, on a admis que le nombre optimal de postes par groupe se situe entre neuf et quinze et les distances maximales entre postes à 120 km (100-150) ; mais ces indications dépendent naturellement de la densité du réseau d'observations : plus cette dernière sera élevée, meilleure pourra être l'estimation du vecteur régional.

8.3. Pour la constitution des groupes de postes, on s'efforcera dans toute la mesure du possible de respecter les limites des régions climatiques homogènes (même orientation aux vents dominants, proximité du littoral, chaînes de montagne, etc.).

Le découpage en régions devra être orienté en fonction d'éventuelles variations progressives du climat.

6.4. On pourra être amené à effectuer des tests préliminaires par simple et double cumul, pour éliminer des postes dont les données sont jugées par trop douteuses, mais on courra ainsi le risque de perdre de l'information utile pour définir ne serait-ce qu'une composante annuelle.

Par contre on aura souvent intérêt, à procéder à un regroupement des postes après un premier passage par le programme de traitement.

8.5. Lorsqu'on a obtenu un vecteur régional, on procédera toujours à une comparaison par double cumul avec le vecteur le plus voisin et si des « cassures » apparaissent, avec le vecteur de la deuxième région la plus proche, ce sera en général suffisant pour établir s'il y a lieu d'apporter des corrections au vecteur considéré.

8. BIBLIOGRAPHIE

- [1] ABRAMOWITZ (Milton) and SEGUN IRENE (A.) – *Handbook of mathematical functions.*
- [2] ANDREWS (D. F.), BICKEL (P. J.), HAMPEL (Fr.), HUBER (P. J.), ROGERS (W. H.), TUKEY (J. W.) – 1972 – *Robust Estimates of Location. Survey and advances.* Princeton University Press, New Jersey.
- [3] BRUNET-MORET (Y.) – 1971 – Étude de l'homogénéité de séries chronologiques de précipitations annuelles par la méthode des doubles masses. *Cahier ORSTOM*, vol. VIII, n° 4, p. 3-31.
- [4] CALOT (G.) – 1967 – *Cours de calcul des probabilités.* Éd. Dunod.
- [5] CALOT (G.) – 1945 – *Cours de statistique descriptive.* Éd. Dunod.
- [6] CANCEILL (M.) – Mars 1971 – *Optimisation des réseaux pluviométriques et hydrométriques.* BRGM, Département Hydrogéologie, 20 p.
- [7] CHILOV (G.) – 1973 – *Analyse mathématique.* Traduit du russe, Éditions de Moscou.
- [8] CLAERROUT (John F.) and MUIR (Francis) – Octobre 1973 – Robust modeling with erratic data. *Geophysics*, vol. 8, n° 5, p. 826-844.
- [9] CORMARY (M. Y.) – Étude des ressources en eau à l'échelle régionale. Comptes rendus Symposium scientifique international « L'informatique et les problèmes de l'eau », p. 119-128.
- [10] DONGUY (P.) et LASNIER (J.) – Septembre 1972 – *Application des composantes principales à un champ pluviométrique.* Mémoire de fin de stage, Bureau de l'Eau.
- [11] DUBREUIL (P.) – Avril 1965 – *Étude des précipitations appliquée à la mise en valeur hydro-agricole.* Groupe d'Étude du Val de Jaguaribe, ORSTOM.
- [12] DUBREUIL (P.) et GUISCAFRE (J.) – Juillet 1976 – *Étude des précipitations dans le massif forestier landais.* ORSTOM, ministère de l'Agriculture, Paris.
- [13] GRISOLLET (H.), GUILMET (B.) et ARLERY (R.) – 1962 – *Climatologie. Méthodes et pratiques.* Monographie de météorologie. Gauthier-Villars et Cie, Éd.
- [14] HANNAN (Ed. J.) – 1970 – *Multiple time series.* John Wiley and Sons Ed.
- [15] HIEZ (G.) – Mai 1969 = *Homogénéisation des données pluviométriques. Note provisoire.* Sudene-UNESCO, 25 p.
- [16] HIEZ (G.) – Novembre 1969 – *Homogénéisation des données pluviométriques. Instructions de calculs pour la programmation.* Sudene-UNESCO, 19 p. + annexes.
- [17] HIEZ (G.) – Décembre 1971 – *Homogénéisation des données pluviométriques* (texte présenté à l'occasion du Séminaire organisé conjointement par l'OMM et la Sudene sur « Les sécheresses et les crues du Nordeste du Brésil », Sudene-OMM, 19 p. (ronéo), 6 graph.
- [18] KARLIN (S.) – 1969 – *Initiation aux processus aléatoires.* Éd. Dunod.
- [19] KORGANOFF (A.) – 1961 – *Méthodes de calcul numérique.* Éd. Dunod.
- [20] KORGANOFF (A.) – 1967 – *Éléments de théorie des matrices carrées et rectangulaires en analyse numérique.* Éd. Dunod.
- [21] LABOUREUR (M.), CHOSSAT (M.), CARDOT (C.) – 1968 – *Cours de calcul mathématique moderne.* Paris, Éd. Dunod.
- [22] LENTIN et RIBAUD (J.) – 1961 – *Leçons d'algèbre moderne.* Vuibert Éditeur.
- [23] MORICE (E.) et CHARTIER (F.) – 1954 – *Méthode statistique.* INSEE, Paris, Imprimerie Nationale.
- [24] ORGANISATION MÉTÉOROLOGIQUE MONDIALE – 1970 – *Guide des pratiques hydrométéorologiques.* OMM, n° 168, TP 82, Genève, annexe A, p. 28-31.
- [25] ROCHE (M.) – 1963 – *Hydrologie de surface.* Paris, Gauthier-Villars Éd., p. 62-66.
- [26] SOKOLNIKOFF (Ivan S.) – *Advanced calculus.* International Student Edition.
- [27] TORRENS-IBERN (J.) – 1972 – *Modèles et méthodes de l'analyse factorielle.* Dunod.
- [28] TRENDEL (R.), MEGREDITCHAN (G. DER), RULLIERE (Mme M. C.) – 1973 – *Sur le traitement opérationnel des données pluviométriques entachées d'erreurs ou insuffisantes.* Météorologie Nationale.
- [29] VIALAR (J.) – 1956 – *Calcul des probabilités et statistiques.* Direction de la Météorologie Nationale.

ANNEXE I

TRAITEMENT DE LA MATRICE DE DIMENSION 9×5
(voir 3.6.4., p. 150)
par le programme PLUV 75 (Sudene IBM 360/30)

MATRIZ DOS DADOS						

	1	2	3	4	5	
1 18C0	5.0	10.0	15.0	20.0	25.0	
2 18C1	2.0	4.0	6.0	8.0	10.0	
3 18C2	3.0	6.0	9.0	12.0	15.0	
4 18C3	4.0	8.0	12.0	16.0	20.0	
5 18C4	1.5	4.0	6.0	12.0	10.0	
6 18C5	2.0	4.0	5.0	18.0	15.0	
7 18C6	4.0	8.0	15.0	20.0	22.0	
8 18C7	4.0	6.0	12.0	16.0	20.0	
9 18C8	1.2	5.0	16.0	24.0	30.0	

NUMERO DO VETOR	SECCAO	PERICDO	1977 - 1978	I	ANO	VETOR	POSICAO	
*****	*****	*****						
					1	1870	1.25010	5
					2	1801	0.49995	5
					3	1872	0.74992	5
					4	1873	1.00000	5
					5	1874	0.50000	5
					6	1875	0.74992	5
					7	1876	1.24992	5
					8	1877	0.99993	5
					9	1878	1.49989	5

C	4.0	8.0	12.0	16.0	20.0	
---	-----	-----	------	------	------	--

I.1 Matrice des données et vecteur de référence

POSTO - POSTO 1		NUMERO -		1		REGIÃO		MATRIZ EXPERIMENTAL		E S M	
*****		*****		*****		*****		*****		*****	
N	ANO	UES	CALC.	EFSE	SEFJA	-4.0	I	-2.0	I	0.0	+2.0
1	1800	5.0	5.0	0.0000	-0.270						
2	1801	2.0	2.0	0.0000	-0.270						
3	1802	3.0	3.0	0.0000	-0.270						
4	1803	4.0	4.0	0.0000	-0.270						
5	1804	1.5	2.0	0.2500	-0.250						
6	1805	3.0	3.0	0.2000	-0.200						
7	1806	6.0	5.0	0.2000	-0.100						
8	1807	4.0	4.0	0.2000	0.000						
9	1808	7.0	6.0	0.2000	0.270						

N	ANO	UES	CALC.	EFSE	SEFJA	4.0	I	-2.0	I	0.0	+2.0
						1		1		1	1

VALOR MODAL										3.7	
CHUVA MEDIA NO PERICDO DE OBSERVACOES						1800- 1808	9 ANOS	OBSERVADA		4.1	CALCULADA
CHUVA MEDIA ESTIMADA NO PERICDO DE VETOR						1800- 1808	9 ANOS			3.4	

I.2 Graphique de double-masse du poste 1 avec le vecteur (voir 5.3., p. 159)

POSTO - PUSTO 2				ALPERC - 2				REGIÃO MATRIZ EXPERIMENTAL 2 x 4					
*****				*****				*****					
N	ANO	OBS.	CHUVA CALC.	EPSI	SEFAL	-4.0	1	-2.0	1	0.0	1	+2.0	1
1	1800	10.0	10.0	0.0000	0.400								
2	1801	4.0	4.0	0.0000	0.400								
3	1802	6.0	6.0	0.0000	0.400								
4	1803	6.0	6.0	0.0000	0.400								
5	1804	4.0	4.0	0.0000	0.400								
6	1805	4.0	4.0	-0.2000	0.200								
7	1806	6.0	10.0	-0.2000	0.600								
8	1807	6.0	6.0	-0.2000	0.200								
9	1808	6.0	10.0	-0.2000	0.400								
N	ANO	OBS.	CHUVA CALC.	EPSI	SEFAL	-4.0	1	-2.0	1	0.0	1	+2.0	1
VALOR MODAL											7.4		
CHUVA MEDIA NO PERIODO DE OBSERVACAO						1800- 1808	9 ANOS	OBSERVADA	6.6	CALCULADA	7.6		
CHUVA MEDIA ESTIMADA NO PERIODO DE VERDADE						1800- 1808	9 ANOS		7.6				

I.3 Graphique de double-masse du poste 2 avec le vecteur (voir 5.3., p. 159)

POSTO - PUSTO 3				ALPERC - 3				REGIÃO MATRIZ EXPERIMENTAL 3 x 4					
*****				*****				*****					
N	ANO	OBS.	CHUVA CALC.	EPSI	SEFAL	-4.0	1	-2.0	1	0.0	1	+2.0	1
1	1800	10.0	10.0	-0.1000	0.400								
2	1801	3.0	3.0	-0.1000	0.300								
3	1802	4.0	4.0	-0.2000	0.200								
4	1803	7.0	10.0	-0.1000	0.400								
5	1804	6.0	6.0	0.0000	0.600								
6	1805	6.0	6.0	0.0000	0.600								
7	1806	10.0	10.0	0.0000	0.600								
8	1807	10.0	10.0	0.0000	0.600								
9	1808	10.0	10.0	0.0000	0.600								
N	ANO	OBS.	CHUVA CALC.	EPSI	SEFAL	-4.0	1	-2.0	1	0.0	1	+2.0	1
VALOR MODAL											11.1		
CHUVA MEDIA NO PERIODO DE OBSERVACAO						1800- 1808	9 ANOS	OBSERVADA	10.0	CALCULADA	11.1		
CHUVA MEDIA ESTIMADA NO PERIODO DE VERDADE						1800- 1808	9 ANOS		11.1				
TABELA DAS CORRECOES													

ANOS	N	POSTO	CHUVA OBS.	CHUVA CALC.	EPSI	CORRECCOES DE CASACAO							
1	1	1800	10.0	10.0	0.0000	1.00000							
1	1	1801	3.0	3.0	0.0000	1.00000							
1	1	1802	4.0	4.0	0.0000	1.00000							
1	1	1803	7.0	10.0	-0.1000	1.00000							
1	1	1804	6.0	6.0	0.0000	1.00000							
1	1	1805	6.0	6.0	0.0000	1.00000							
1	1	1806	10.0	10.0	0.0000	1.00000							
1	1	1807	10.0	10.0	0.0000	1.00000							
1	1	1808	10.0	10.0	0.0000	1.00000							

I.4 Graphique de double-masse du poste 3 avec le vecteur (voir 5.3., p. 159) (noter la table des corrections fournie par le sous-programme de détection des cassures)

POSTO - POSTO 4		ALPERC -		4		REGIAD		MATRIZ EXPERIMENTAL		5* 6 M		
*****		*****		*****		*****		*****		*****		
N	ANO	DES.	CALC.	EPSI	SEFJA	-4.0	1	-2.0	1	0.0	1	+2.0
1	1800	20.0	20.0	C.C	-C.500					*		
2	1801	18.0	18.0	-C.CCCC	-C.500					+		
3	1802	12.0	12.0	-C.CCCC	-C.500					+		
4	1803	16.0	16.0	-C.CCCC	-C.500					+		
5	1804	12.0	12.0	C.5CCC	-C.500					+		
6	1805	18.0	18.0	C.5CCC	C.500					+		
7	1806	20.0	20.0	C.CCCC	C.500					+		
8	1807	16.0	16.0	C.C	C.500					+		
9	1808	24.0	24.0	-C.CCCC	C.500					+		
N	ANO	DES.	CALC.	EPSI	SEFJA	-4.0	1	-2.0	1	0.0	1	+2.0
VALOR MODAL										14.6		
CHUVA MEDIA NO PERICCO DE OBSERVACÇES						1800- 1808	9 ANOS	OBSERVADA	16.2	CALCULADA	15.1	
CHUVA MEDIA ESTIMADA NO PERICCO DE VETUR						1800- 1808	9 ANOS		15.1			

I.5 Graphique de double-masse du poste 4 avec le vecteur (voir 5.3., p. 159)

POSTO - POSTO 5		ALPERC -		5		REGIAD		MATRIZ EXPERIMENTAL		5* 6 M		
*****		*****		*****		*****		*****		*****		
N	ANO	DES.	CALC.	EPSI	SEFJA	-4.0	1	-2.0	1	0.0	1	+2.0
1	1800	25.0	25.0	-C.CCCC	-C.600					*		
2	1801	10.0	10.0	-C.CCCC	-C.600					+		
3	1802	15.0	15.0	-C.CCCC	-C.600					+		
4	1803	24.0	24.0	C.5CCC	C.100					+		
5	1804	10.0	10.0	-C.CCCC	C.100					+		
6	1805	15.0	15.0	-C.CCCC	C.100					+		
7	1806	22.0	22.0	-C.12CC	C.600					+		
8	1807	20.0	20.0	-C.CCCC	C.600					+		
9	1808	20.0	20.0	-C.CCCC	C.600					+		
N	ANO	DES.	CALC.	EPSI	SEFJA	-4.0	1	-2.0	1	0.0	1	+2.0
VALOR MODAL										18.5		
CHUVA MEDIA NO PERICCO DE OBSERVACÇES						1800- 1808	9 ANOS	OBSERVADA	19.0	CALCULADA	14.9	
CHUVA MEDIA ESTIMADA NO PERICCO DE VETUR						1800- 1808	9 ANOS		18.9			

I.6 Graphique de double-masse du poste 5 avec le vecteur (voir 5.3., p. 159)

ANNEXE II

TRAITEMENT DE LA MATRICE SIMULÉE PAR F. MONIOD (voir p. 152)
PAR LE PROGRAMME PLUV 75 (Sudene IBM 360/30)

M A T R I X D O S D A D O S

	1	2	3	4	5	6	7
1 1940	184.0	134.1	127.8	93.9	167.5	171.4	180.2
2 1941	120.5	155.2	149.5	57.8	183.5	199.6	205.3
3 1942	144.5	124.0	127.5	90.6	169.1	179.1	177.7
4 1943	138.7	127.4	116.0	147.0	153.5	145.5	176.0
5 1944	153.2	140.0	103.5	165.5	176.1	124.2	185.8
6 1945	176.0	162.4	91.5	130.0	191.7	189.4	214.6
7 1946	104.0	143.2	95.5	153.6	154.9	182.3	184.5
8 1947	153.0	131.6	54.4	124.4	120.1	171.0	178.5
9 1948	125.0	180.4	40.7	125.2	132.9	201.1	205.7
10 1949	175.5	244.1	163.5	33.3	170.1	216.3	227.0
11 1950	150.4	185.3	149.0	154.2	164.8	179.9	209.4
12 1951	141.0	107.7	135.5	152.2	159.7	167.7	154.2
13 1952	168.5	157.0	157.2	139.1	137.9	205.0	152.7
14 1953	178.0	162.7	171.1	192.2	152.2	218.1	265.8
15 1954	157.5	141.5	141.2	171.0	137.0	184.7	197.3
16 1955	125.0	115.6	121.2	133.9	113.4	155.3	156.8
17 1956	176.0	204.8	163.4	130.0	130.5	204.5	214.6
18 1957	210.0	155.8	158.9	172.8	101.7	193.6	209.2
19 1958	154.7	182.5	141.5	127.6	135.8	190.2	181.8
20 1959	180.7	167.5	135.5	120.3	119.5	169.8	176.4
21 1960	215.6	156.7	141.0	144.5	142.7	199.0	206.6
22 1961	151.0	125.8	133.1	125.9	161.0	176.0	147.2
23 1962	172.5	157.8	150.0	133.6	192.3	202.5	165.8
24 1963	140.0	128.8	128.1	152.1	158.6	130.9	135.1
25 1964	144.5	145.4	130.7	171.0	174.6	145.0	154.8
26 1965	168.4	163.8	152.9	196.7	232.4	217.7	220.5
27 1966	158.2	158.0	146.9	197.5	217.7	206.4	215.1
28 1967	147.2	145.2	132.9	178.2	191.0	194.0	198.6
29 1968	173.0	174.5	165.5	200.3	238.6	227.3	232.1
30 1969	166.5	165.6	154.7	194.5	230.0	213.5	226.2

II.1 Matrice des données

NUMERO DO VETOR *****	77330	Região 1940 - 1969 *****					MATRIZ MONITOR *****					DATA 01/02/77 ****
		I	ANO	N. POSICAO	VETOR	SOMA	PIF.	-2.0	-1.0	0.0	+1.0	
1	1940	7	0.90499	0.30489	-0.04774	*						
2	1941	7	1.03638	1.24127	-0.01136	*						
3	1942	7	0.90930	2.85056	-0.10207	* I						
4	1943	7	0.95692	3.70755	-0.24508	* I						
5	1944	7	0.95263	4.56017	-0.29246	* I						
6	1945	7	1.08697	5.74714	-0.20540	* I						
7	1946	7	0.94651	6.69364	-0.25800	* I						
8	1947	7	0.89245	7.58905	-0.36654	* I						
9	1948	7	1.06026	8.54434	-0.30620	* I						
10	1949	7	1.13099	9.77732	-0.17531	* I						
11	1950	7	0.93566	10.71207	-0.23064	* I						
12	1951	7	0.87957	11.50254	-0.34000	* I						
13	1952	7	1.06529	12.65791	-0.29407	* I						
14	1953	7	1.11333	13.77113	-0.18150	* I						
15	1954	7	0.99047	14.75160	-0.20103	* I						
16	1955	7	0.80323	15.55482	-0.32791	* I						
17	1956	7	1.09337	16.44919	-0.30446	* I						
18	1957	7	1.03411	17.43227	-0.27034	* I						
19	1958	7	0.94371	18.52594	-0.32647	* I						
20	1959	7	0.90051	19.52544	-0.42617	* I						
21	1960	7	1.02737	20.55402	-0.30951	* I						
22	1961	7	0.93022	21.43422	-0.44947	* I						
23	1962	7	1.05106	22.54527	-0.40735	* I						
24	1963	7	0.87794	23.42314	-0.52040	* I						
25	1964	7	0.99533	24.40845	-0.54415	* I						
26	1965	7	1.12140	25.53904	-0.42259	* I						
27	1966	7	1.07684	26.50799	-0.34574	* I						
28	1967	7	1.01091	27.51749	-0.33474	* I						
29	1968	7	1.15837	28.79854	-0.16409	* I						
30	1969	7	1.11870	29.30524	-0.24739	* I						
							-2.0	-1.0	0.0	+1.0	+2.0	
							I	I	I	I	I	
0	160.066	145.433	143.672	175.744	179.591	171.762	192.277					

II.2 Vecteur de référence et son « simple-cumul »

POSTO - COMUS INCI				NUMERO - 137311		REGION		MATRICE MON IND				J 1		
N	ANO	ORS.	CHUVA CALC.	EPSI	SEPSI	-4.0	I	-2.0	I	0.0	I	+2.0	I	+4.0
1	1940	164.0	144.5	-0.2152	0.020									
2	1941	120.5	145.5	-0.2137	-0.290									
3	1942	144.5	145.5	-0.0073	-0.290									
4	1943	126.7	127.2	0.0110	-0.264									
5	1944	153.3	152.5	0.0052	-0.237									
6	1945	116.0	116.0	0.0114	-0.223									
7	1946	164.0	151.5	-0.3136	-0.554									
8	1947	193.0	142.5	-0.3500	-0.110									
9	1948	124.0	165.5	-0.2400	-0.460									
10	1949	175.5	181.1	-0.0055	-0.497									
11	1950	150.4	165.0	0.0041	-0.490									
12	1951	141.0	140.0	0.0014	-0.490									
13	1952	166.0	170.5	-0.0115	-0.443									
14	1953	178.0	178.2	-0.0013	-0.444									
15	1954	171.0	175.0	0.0034	-0.444									
16	1955	125.0	125.0	0.0032	-0.443									
17	1956	176.0	171.0	0.0077	-0.404									
18	1957	210.0	165.5	-0.2488	-0.104									
19	1958	154.7	151.1	0.2855	0.100									
20	1959	160.7	160.2	0.2930	0.302									
21	1960	213.8	164.5	-0.2007	0.000									
22	1961	151.0	149.5	0.0100	0.790									
23	1962	172.5	168.0	0.0155	0.790									
24	1963	140.0	150.5	-0.0019	0.790									
25	1964	144.0	157.1	-0.0014	0.200									
26	1965	188.4	175.0	-0.0071	0.300									
27	1966	156.2	172.4	-0.0003	0.470									
28	1967	147.5	181.1	-0.0003	0.330									
29	1968	171.0	157.1	-0.0088	0.300									
30	1969	160.5	175.1	-0.0002	0.260									
N	ANO	ORS.	CHUVA CALC.	EPSI	SEPSI	-4.0	I	-2.0	I	0.0	I	+2.0	I	+4.0
VALOR MODAL										140.1				
CHUVA MEDIA NO PERIODO DE OBSERVACAO						1940- 1969	30 ANOS	OBSERVADA	161.6	CALCULADA	149.4			
CHUVA MEDIA ESTIMADA NO PERIODO DE V.L.O.						1940- 1969	30 ANOS		160.6					

II.3 Graphe de la colonne 1 de la matrice

POSTO - COMUS INCI				NUMERO - 197315		REGION		MATRICE MON IND				J 1		
N	ANO	ORS.	CHUVA CALC.	EPSI	SEPSI	-4.0	I	-2.0	I	0.0	I	+2.0	I	+4.0
1	1940	167.5	162.5	-0.0218	1.230									
2	1941	162.5	161.5	-0.0019	1.230									
3	1942	166.1	165.1	0.0072	1.230									
4	1943	153.5	154.2	-0.0045	1.230									
5	1944	176.1	171.5	-0.0271	1.230									
6	1945	151.3	150.0	-0.0201	1.230									
7	1946	154.5	110.0	-0.0513	1.180									
8	1947	120.1	160.0	-0.2523	0.990									
9	1948	122.0	150.0	-0.3038	0.990									
10	1949	170.1	203.4	-0.1644	0.420									
11	1950	164.5	161.0	-0.0214	0.420									
12	1951	155.7	155.2	0.0055	0.420									
13	1952	187.8	161.0	-0.0200	0.390									
14	1953	152.2	200.0	-0.2400	0.100									
15	1954	137.0	170.0	-0.2218	-0.050									
16	1955	132.0	160.0	-0.0715	-0.090									
17	1956	130.0	150.0	-0.3348	-0.050									
18	1957	101.7	110.1	-0.4438	-1.070									
19	1958	125.5	160.0	-0.2000	-1.270									
20	1959	138.5	162.1	-0.2055	-1.380									
21	1960	152.1	160.0	-0.0286	-1.300									
22	1961	161.0	167.0	-0.0330	-1.300									
23	1962	102.2	151.0	-0.0070	-1.790									
24	1963	188.4	150.0	0.0077	-1.790									
25	1964	174.0	173.2	-0.0056	-1.830									
26	1965	232.0	161.0	0.1510	-1.850									
27	1966	217.7	150.0	0.1232	-1.850									
28	1967	161.0	180.0	-0.0000	-1.970									
29	1968	220.0	210.0	0.1302	-1.990									
30	1969	220.0	201.2	0.1473	-1.990									
N	ANO	ORS.	CHUVA CALC.	EPSI	SEPSI	-4.0	I	-2.0	I	0.0	I	+2.0	I	+4.0
VALOR MODAL										140.0				
CHUVA MEDIA NO PERIODO DE OBSERVACAO						1940- 1969	30 ANOS	OBSERVADA	161.6	CALCULADA	149.4			
CHUVA MEDIA ESTIMADA NO PERIODO DE V.L.O.						1940- 1969	30 ANOS		160.6					

II.4 Graphe de la colonne 5 de la matrice

ANNEXE III

TRAITEMENT D'UNE MATRICE DE DONNÉES RÉELLES
DE 10 POSTES DE L'ÉTAT DU PERNAMBOUC
PAR LE PROGRAMME VETOR H (IBM 1130)

RELACAO DOS POSTOS			
NUMERO	NOME	PERIODO	CORRECOES
1 3823822.	LUIZ GOMEZ	1912-1971	NENHUMA CORRÉCAP
2 3823258.	PAU DOS FERROS	1912-1971	NENHUMA CORRECAO
3 3833555.	SOUZA	1912-1971	NENHUMA CORRECAO
4 3823898.	ALEXANDRIA	1913-1971	NENHUMA CORRECAO
5 3824119.	MARTINS	1912-1971	NENHUMA CORRECAO
6 3834538.	POMBAL	1912-1971	NENHUMA CORRECAO
7 3824751.	CATOLE DA ROCHA	1912-1971	NENHUMA CORRECAO
8 3823402.	SAO MIGUEL	1912-1971	NENHUMA CORRECAO
9 3833413.	ANTENOR NAVARRO	1912-1971	NENHUMA CORRECAO
10 3824274.	PATU	1912-1971	NENHUMA CORRECAO

III.1 Liste des postes

MATRIZ DOS DADOS

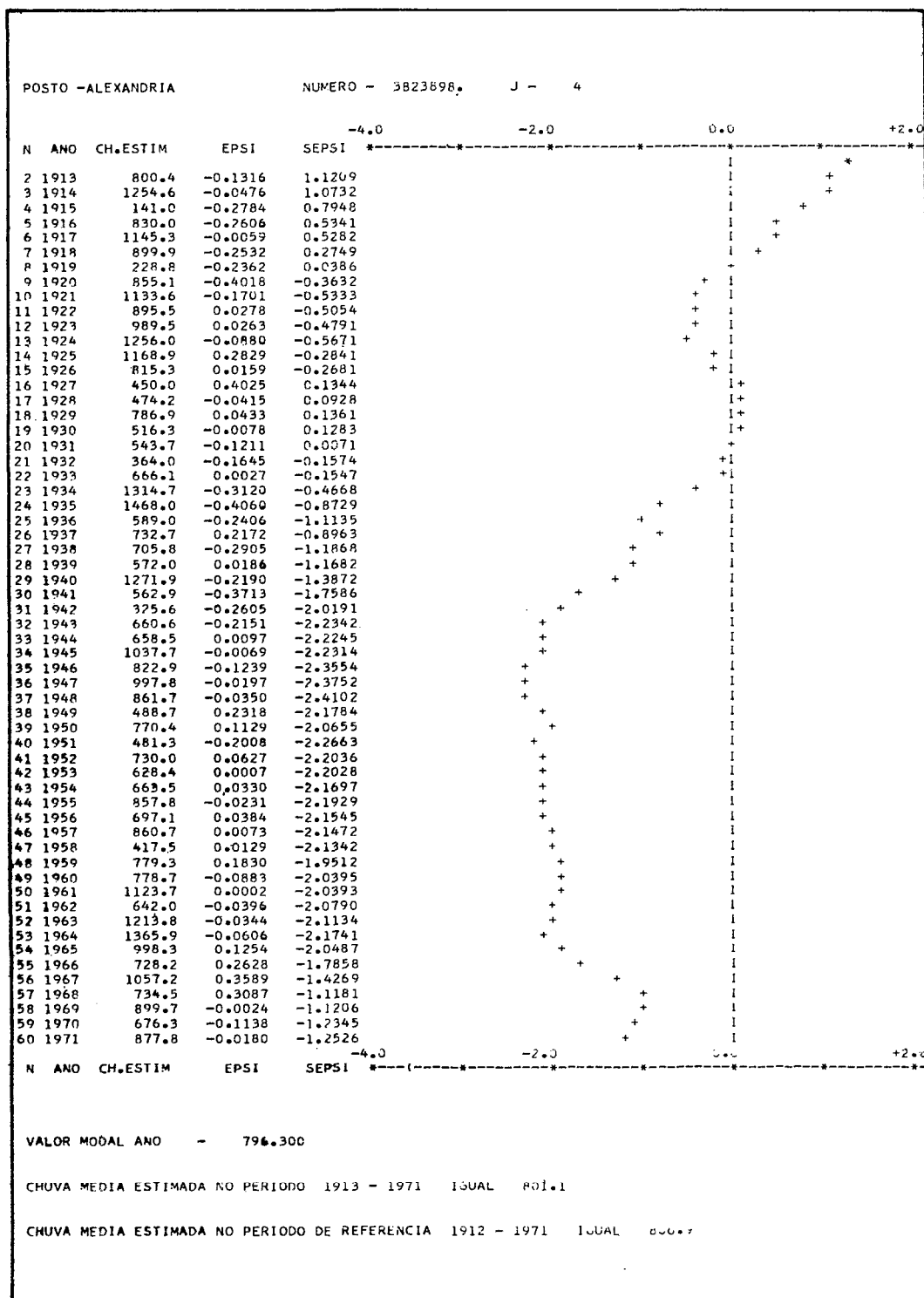
		1	2	3	4	5	6	7	8	9	10
1	1912	1226.3	1285.2	1050.1	-10.0	1586.3	608.8	1155.1	1181.6	954.8	1046.3
2	1913	847.5	706.5	767.3	695.0	1156.6	709.7	983.9	797.1	944.7	781.2
3	1914	1357.5	1173.4	982.2	1194.9	1910.1	1208.5	1636.7	1332.7	1132.7	1386.8
4	1915	149.5	153.6	181.7	101.8	153.4	89.4	154.7	142.4	212.7	146.0
5	1916	741.4	923.3	964.5	613.7	1116.6	884.0	770.6	850.5	822.7	807.3
6	1917	1028.4	1182.9	1323.6	1138.5	1922.4	835.0	1067.0	1202.2	1672.6	1058.6
7	1918	952.9	480.9	709.4	672.0	1224.6	446.4	1341.8	943.7	879.2	891.4
8	1919	329.3	140.1	181.4	174.8	265.7	134.2	208.2	331.0	224.0	229.3
9	1920	676.1	729.9	502.8	511.5	1280.6	504.3	884.3	911.5	881.0	697.7
10	1921	1221.5	1012.7	1041.2	940.8	1483.3	960.3	1207.1	1184.8	1145.4	994.1
11	1922	1194.5	783.6	608.0	920.5	1138.2	988.9	1461.3	992.8	889.1	991.6
12	1923	1064.0	828.6	465.4	1015.6	1234.0	771.8	1372.9	1061.4	803.0	1232.3
13	1924	1316.5	1140.9	1585.5	1145.5	1899.8	370.8	2331.8	1283.9	1799.9	1219.4
14	1925	1315.0	788.8	766.3	1499.7	1495.7	847.4	2112.4	1199.5	1350.8	1116.7
15	1926	972.9	740.7	623.5	828.4	971.1	773.6	513.5	1171.0	1157.6	791.5
16	1927	1008.9	381.0	588.9	631.2	664.7	435.0	338.9	846.7	1085.2	546.7
17	1928	847.0	392.3	382.5	454.5	658.8	493.3	394.0	547.3	1038.8	451.5
18	1929	881.0	440.0	529.8	821.0	1032.3	1800.8	528.0	754.4	1251.8	798.1
19	1930	652.6	335.6	524.3	512.3	593.5	336.8	159.0	542.5	669.7	498.8
20	1931	672.5	396.2	410.9	477.9	729.6	537.9	403.7	543.2	610.9	523.2
21	1932	386.1	253.2	173.0	304.1	519.7	358.4	326.9	522.9	223.3	345.5
22	1933	632.7	462.9	732.0	668.0	903.2	435.8	611.6	668.9	854.7	531.5
23	1934	825.8	1149.7	1187.6	904.4	1428.2	964.2	1147.1	1683.6	1391.9	1120.2
24	1935	849.6	1280.7	1167.6	871.9	1451.6	890.5	1291.1	742.4	1776.2	1442.0
25	1936	617.8	214.5	797.7	447.3	771.9	523.1	536.9	663.0	912.8	593.7
26	1937	820.7	594.4	773.8	891.9	978.8	538.4	1145.6	458.6	762.5	1002.8
27	1938	750.2	490.9	753.5	503.8	754.7	688.4	669.4	477.7	692.0	470.5
28	1939	642.7	501.6	926.7	582.7	813.5	534.5	792.1	556.6	797.4	507.6
29	1940	1498.3	1113.7	1306.0	993.4	1688.7	1256.8	1592.2	1246.0	1478.4	1002.8
30	1941	759.8	311.4	801.8	353.9	661.6	531.6	519.7	449.2	841.2	309.3
31	1942	329.0	421.5	271.3	240.8	427.3	432.8	518.2	238.2	346.4	313.0
32	1943	983.5	499.2	514.9	518.5	879.8	462.5	602.2	586.0	696.3	559.1
33	1944	865.8	525.0	644.7	665.0	951.2	646.1	558.3	512.0	808.2	625.0
34	1945	1120.1	927.5	829.3	1030.5	1280.0	805.5	827.0	420.0	1012.6	569.7
35	1946	1005.5	721.0	652.1	720.9	1102.5	757.9	1169.3	675.5	640.3	809.4
36	1947	1284.8	882.0	1057.5	978.1	1264.6	843.3	1543.9	662.9	1153.0	1003.0
37	1948	1074.6	639.0	672.8	831.5	1011.4	886.4	1435.2	-10.0	1406.3	857.8
38	1949	803.2	398.0	1048.7	602.0	720.1	959.3	1248.1	442.4	1363.4	975.4
39	1950	1007.7	533.7	561.9	857.5	1049.2	737.8	823.7	699.6	812.6	1290.8
40	1951	843.5	459.8	492.4	384.7	657.6	427.8	792.2	-10.0	1206.6	718.7
41	1952	1027.7	591.5	395.2	775.8	978.0	713.5	1463.5	527.0	1360.7	1433.5
42	1953	933.0	867.0	512.9	629.0	814.9	299.0	708.8	505.0	1050.4	924.4
43	1954	1127.9	726.0	721.8	685.5	815.3	691.7	581.6	469.0	1404.5	1089.4
44	1955	1146.9	1186.0	681.0	838.0	1498.2	868.1	1100.9	877.0	1707.1	1016.2
45	1956	707.0	-10.0	848.0	724.0	1011.1	950.5	588.2	1252.4	2202.9	937.9
46	1957	899.0	-10.0	402.4	867.0	1202.8	823.9	723.4	968.4	1249.7	1342.0
47	1958	519.0	-10.0	328.0	423.0	590.6	309.4	375.7	431.0	461.9	520.1
48	1959	824.0	667.6	697.0	922.0	1152.0	753.0	673.2	1318.6	687.3	1732.6
49	1960	854.3	672.6	620.0	710.0	1031.7	464.1	798.4	1311.7	206.2	1266.9
50	1961	1346.0	993.3	874.6	1124.0	1338.0	700.3	839.4	1503.9	967.5	1330.5
51	1962	691.5	636.7	509.8	616.6	894.4	506.2	423.1	663.4	623.3	782.0
52	1963	1150.0	1067.2	951.4	1172.0	1320.0	999.6	580.6	1393.4	1288.4	678.7
53	1964	1425.0	1180.2	1384.4	1283.0	1856.4	1289.4	1258.3	1416.7	1428.1	1111.3
54	1965	1045.7	803.5	927.9	1123.6	1345.2	984.9	968.5	1033.5	714.1	902.5
55	1966	871.6	742.4	573.1	919.7	846.2	789.7	529.0	782.7	742.7	524.0
56	1967	1113.1	1030.8	1283.6	1436.7	1426.9	822.0	996.4	1259.2	1023.0	888.2
57	1968	928.2	638.1	592.1	961.3	984.7	1000.9	983.3	945.8	746.3	831.2
58	1969	862.8	676.7	-10.0	897.5	989.3	433.3	794.4	625.0	1003.2	888.2
59	1970	714.6	325.3	-10.0	599.4	755.7	650.5	636.6	376.3	458.4	515.0
60	1971	1041.2	714.2	1142.9	862.0	1179.9	992.6	839.1	934.7	894.5	1062.1

III.2 Matrice des données

NUMERO DO VETOR		10ç60	PERIODO	1912 - 1971		
I	ANO	VETOR	N.ºPOSTOS			
1	1912	1.43774	9	1.43774	0.43774	
2	1913	1.00515	10	2.44289	0.44289	
3	1914	1.57564	10	4.01853	1.01853	
4	1915	0.17716	10	4.19570	0.19570	
5	1916	1.04242	10	5.23811	0.23811	
6	1917	1.43832	10	6.67643	0.67643	
7	1918	1.13015	10	7.80658	0.80658	
8	1919	0.28742	10	8.09400	0.09400	
9	1920	1.07395	10	9.16795	0.16795	
10	1921	1.42364	10	10.59160	0.59160	
11	1922	1.12460	10	11.71619	0.71619	
12	1923	1.24265	10	12.95884	0.95884	
13	1924	1.57734	10	14.53618	1.53618	
14	1925	1.46795	10	16.00413	2.00413	
15	1926	1.02393	10	17.02807	2.02807	
16	1927	0.56514	10	17.59321	1.59321	
17	1928	0.59553	10	18.18874	1.18874	
18	1929	0.98821	10	19.17695	1.17695	
19	1930	0.64842	10	19.82536	0.82536	
20	1931	0.65289	10	20.50825	0.50825	
21	1932	0.45713	10	20.96538	-0.03461	
22	1933	0.83661	10	21.80199	-0.19801	
23	1934	1.65101	10	23.45300	0.45300	
24	1935	1.84362	10	25.29661	1.29661	
25	1936	0.73975	10	26.03637	1.03637	
26	1937	0.92013	10	26.95650	0.95650	
27	1938	0.88643	10	27.80493	0.80493	
28	1939	0.71838	10	28.56131	0.56131	
29	1940	1.59734	10	30.15864	1.15864	
30	1941	0.70699	10	30.86563	0.86563	
31	1942	0.40893	10	31.27456	0.27456	
32	1943	0.82964	10	32.10420	0.10420	
33	1944	0.82702	10	32.93122	-0.06877	
34	1945	1.30316	10	34.23438	0.23438	
35	1946	1.03344	10	35.26782	0.26782	
36	1947	1.25310	10	36.52092	0.52092	
37	1948	1.08214	9	37.60306	0.60306	
38	1949	0.61372	10	38.21679	0.21679	
39	1950	0.96758	10	39.18437	0.18437	
40	1951	0.60450	9	39.78887	-0.21112	
41	1952	0.91676	10	40.70563	-0.29436	
42	1953	0.78927	10	41.49490	-0.50509	
43	1954	0.83328	10	42.32817	-0.67162	
44	1955	1.07734	10	43.40552	-0.59447	
45	1956	0.87554	9	44.28105	-0.71894	
46	1957	1.08088	9	45.36193	-0.63806	
47	1958	0.52441	9	45.88634	-1.11365	
48	1959	0.97871	10	46.86505	-1.13494	
49	1960	0.97800	10	47.84305	-1.15694	
50	1961	1.41120	10	49.25425	-0.74574	
51	1962	0.80634	10	50.06059	-0.93940	
52	1963	1.52434	10	51.58493	-0.41506	
53	1964	1.71531	10	53.30024	0.30024	
54	1965	1.25370	10	54.55393	0.55393	
55	1966	0.91458	10	55.46852	0.46852	

56	1967	1.32764	10	56.79615	0.79615	
57	1968	0.92243	10	57.71058	0.71058	
58	1969	1.12990	9	58.84648	0.84648	
59	1970	0.84942	9	59.69790	0.69790	
60	1971	1.10245	10	60.80035	0.80035	
851.9 692.3 629.9 796.31066.3 773.2 722.8 826.4 825.1 769.1						

III.3 Vecteur de référence



III.4 Graphe d'un poste

POSTO -WARTINS				NUMERO - 3824119. J - 5				
N	ANO	CH.ESTIM	EPSI	SEPSI	-4.0	-2.0	0.0	+2.0
1	1912	1533.1	0.0346	0.4519				
2	1912	1071.8	0.0790	0.5310				
3	1914	1680.2	0.1365	0.6678				
4	1915	188.9	-0.1880	0.4797				
5	1916	1111.6	0.0044	0.4842				
6	1917	1533.7	0.2533	0.7376				
7	1918	1205.1	0.0161	0.7537				
8	1919	306.4	-0.1331	0.6206				
9	1920	1145.2	0.1182	0.7388				
10	1921	1518.1	-0.0229	0.7159				
11	1922	1199.2	-0.0508	0.6650				
12	1923	1325.1	-0.0687	0.5962				
13	1924	1682.0	0.1294	0.7257				
14	1925	1565.3	-0.0445	0.6812				
15	1926	1091.8	-0.1106	0.5706				
16	1927	602.6	0.1029	0.6735				
17	1928	635.0	0.0373	0.7109				
18	1929	1053.7	-0.0204	0.6905				
19	1930	691.4	-0.1416	0.5488				
20	1931	728.2	0.0019	0.5507				
21	1932	487.4	0.0661	0.6169				
22	1933	892.1	0.0124	0.6293				
23	1934	1760.5	-0.1887	0.4405				
24	1935	1965.9	-0.2616	0.1788				
25	1936	788.8	-0.0214	0.1573				
26	1937	981.2	-0.0024	0.1549				
27	1938	945.2	-0.2015	-0.0466				
28	1939	766.0	0.0619	0.0152				
29	1940	1703.3	-0.0086	0.0066				
30	1941	753.9	-0.1224	-0.1157				
31	1942	436.0	-0.0201	-0.1358				
32	1943	884.7	-0.0055	-0.1414				
33	1944	881.9	0.0785	-0.0628				
34	1945	1389.6	-0.0789	-0.1417				
35	1946	1102.0	0.0004	-0.1413				
36	1947	1336.2	-0.0536	-0.1949				
37	1948	1153.9	-0.1235	-0.3185				
38	1949	654.4	0.1002	-0.2182				
39	1950	1031.7	0.0163	-0.2013				
40	1951	644.6	0.0201	-0.1812				
41	1952	977.6	0.0004	-0.1808				
42	1953	841.6	-0.0317	-0.2126				
43	1954	888.5	-0.0824	-0.2950				
44	1955	1148.8	0.3040	0.0089				
45	1956	933.6	0.0829	0.0919				
46	1957	1152.6	0.0435	0.1354				
47	1958	559.2	0.0561	0.1916				
48	1959	1043.6	0.1037	0.2954				
49	1960	1042.9	-0.0107	0.2846				
50	1961	1504.8	-0.1108	0.1737				
51	1962	859.8	0.0401	0.2139				
52	1963	1625.5	-0.1879	0.0260				
53	1964	1829.1	0.0148	0.0409				
54	1965	1336.9	0.0062	0.0471				
55	1966	975.2	-0.1323	-0.0952				
56	1967	1415.7	0.0078	-0.0773				
57	1968	983.6	0.0010	-0.0763				
58	1969	1204.8	-0.1789	-0.2552				
59	1970	905.7	-0.1657	-0.4209				
60	1971	1175.6	0.0036	-0.4173				

N	ANC	CH.ESTIM	EPSI	SEPSI	-4.0	-2.0	0.0	+2.0
VALOR MODAL ANO - 1066.365								
CHUVA MEDIA ESTIMADA NO PERIODO 1912 - 1971 IGUAL 1088.5								
CHUVA MEDIA ESTIMADA NO PERIODO DE REFERENCIA 1912 - 1971 IGUAL 1088.5								

III.5 Graphe d'un poste