

Etude des effets de dépendance dans une série chronologique. Application à l'étude des séquences de jours de pluies

Ch. BOUVIER
Hydrologue à l'ORSTOM

RÉSUMÉ

Après avoir décrit les divers effets de dépendance pouvant affecter les termes d'une série chronologique, l'auteur propose d'utiliser un schéma de probabilité tenant compte de ces effets, le schéma de POLYA, et l'applique à l'étude des séquences d'un même état hydrologique d'un système.

ABSTRACT : *Dependence effect study in a data time serie — Application to the survey of daily rainfall sequences*

The author first describes dependence effects which can exist in a time serie, and a probability scheme including these effects, POLYA's scheme is proposed to study the chronological followings of an hydrological event.

SOMMAIRE

Introduction

1. Etude des effets de dépendance dans une série chronologique

- 1.1 Position du problème
- 1.2 Représentation analytique des données
- 1.3 Méthode du corrélogramme
 - 1.3.1 Définition et généralités
 - 1.3.2 Interprétation des résultats
 - 1.3.3 Signification des coefficients d'autocorrélation
- 1.4 Méthode du périodogramme
 - 1.4.1 Définitions et interprétation
 - 1.4.2 Comparaison du périodogramme et du corrélogramme-Exemples
- 1.5 Influence de la persistance sur l'exactitude des moyennes

2. Etudes des séquences de jours de pluie

- 2.1 Position du problème
- 2.2 Schéma de Polya
- 2.3 Chaînes de Markov
 - 2.3.1 Généralités
 - 2.3.2 Utilisation des probabilités conditionnelles dans l'étude des séquences de jours de pluie

- 2.4 Résultats et commentaires
 - 2.4.1 Calcul théorique du nombre des séquences
 - 2.4.2 Calcul des récurrences
 - 2.4.3 Influence de l'effet saisonnier
 - 2.4.4 Définition des pluies « utiles »
 - 2.4.5 Conclusion

Annexe A

1. *Calcul des paramètres du schéma de Polya par la méthode des moments*
2. *Valeur modale*
3. *Comportement aux limites*

Annexe B

Application du schéma de Polya à l'étude de la distribution des hauteurs de pluie

Annexe C

Comparaison des ajustements réalisés pour la station de Boyvinière

Bibliographie

INTRODUCTION

Les données représentant les phénomènes hydrologiques et climatologiques sont souvent dépendantes entre elles. La prise en compte de ces interactions revêt un caractère essentiel dans la mesure où, d'une part, leur existence peut mettre en cause la validité de résultats obtenus par des lois statistiques basées sur l'indépendance des événements étudiés et, d'autre part, leur connaissance qualitative et quantitative permet d'apporter une information supplémentaire qui réduit le caractère aléatoire du phénomène.

Pour illustrer ceci, nous nous proposons d'étudier statistiquement les séquences de jours de pluie au cours du temps ; toute autre étude concernant les séquences d'un état qualitatif d'une variable peut, évidemment, être menée de la même façon.

1. ÉTUDE DES EFFETS DE DÉPENDANCE DANS UNE SÉRIE CHRONOLOGIQUE

1.1 POSITION DU PROBLÈME

Les valeurs ou les états que prennent à un instant donné les variables climatologiques ou hydrologiques ne sont pas sans rapport avec le passé. Les divers effets de dépendance affectant les termes d'une série chronologique (1) peuvent être représentés par :

- un effet de tendance : l'espérance mathématique (moyenne) des valeurs observées croît (ou décroît) avec le temps ;
- des effets cycliques ou pseudo-cycliques : l'espérance mathématique d'une valeur observée est fonction de la chronologie, mais la valeur moyenne de séries suffisamment longues peut être considérée comme stationnaire ;

(1) On suppose que les données étudiées sont corrigées des erreurs systématiques.

— un effet de persistance : une valeur n'est pas indépendante de la ou des valeurs précédentes (processus markoviens ou autres), mais la suite est stationnaire.

1.2 REPRÉSENTATION ANALYTIQUE DES DONNÉES

Si $F(t)$ représente la valeur de la variable étudiée à l'instant t , nous admettrons schématiquement que $F(t)$ peut s'exprimer sous la forme :

$$F(t) = T(t) + C(t) + S(t) + R(t)$$

où :

T représente la tendance générale

C représente le mouvement cyclique

S représente la variation saisonnière

et R tient compte à la fois du phénomène de persistance et des variations résiduelles purement aléatoires.

Ainsi, la figure 1 représente graphiquement la composition des trois premiers mouvements.

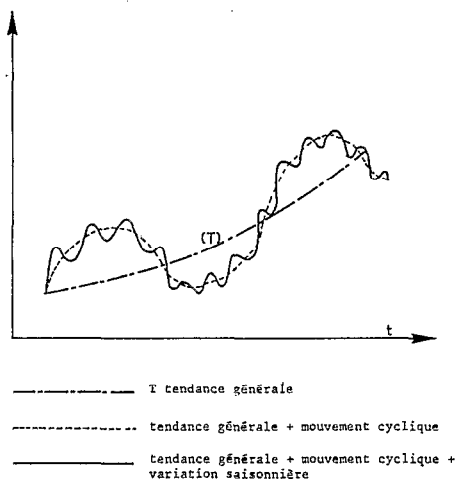


Fig. 1

Signalons que l'on peut être amené à représenter la variable $F(t)$ sous la forme :

$$F(t) = T(t) \times C(t) \times S(t) \times R(t)$$

1.3 MÉTHODE DU CORRÉLOGRAMME

1.3.1 Définition et généralités

Le coefficient de corrélation ρ , calculé entre deux variables X et Y de la manière suivante :

$$\rho = \frac{E((X - \bar{X})(Y - \bar{Y}))}{(\text{var } X \cdot \text{var } Y)^{1/2}}$$

met en évidence le degré de dépendance entre ces deux variables.

L'analyse des effets de dépendance dans une série chronologique (X_i) s'effectue de la même façon en calculant le coefficient de corrélation des variables représentant 2 valeurs successives (resp. 2 valeurs séparées par k intervalles de temps). Ce coefficient s'appelle le coefficient d'autocorrélation d'ordre 1 (resp. d'ordre k) de la variable X .

$$\rho_k = \frac{E (X_i - \bar{X}) (X_{i+k} - \bar{X})}{\text{var } \bar{X}}$$

En pratique, ce coefficient ρ_k se calcule d'après l'échantillon :

$$\rho_k = \frac{\frac{1}{n-k} \sum_{i=1}^{n-k} (X_i - \bar{X}) (X_{i+k} - \bar{X})}{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2}$$

Quel que soit l'ordre k d'un coefficient d'autocorrélation :

$$-1 \leq \rho_k \leq 1$$

la valeur absolue $|\rho_k|$ exprime le degré de dépendance affectant les observations distantes de k intervalle de temps : plus $|\rho_k|$ est proche de 1, plus les écarts $|X_i - \bar{X}|$ et $|X_{i+k} - \bar{X}|$ sont voisins ; les valeurs de la série chronologique sont indépendantes si et seulement si $\rho_1 = \rho_2 = \dots = \rho_k = \dots = 0$.

Le signe de ρ_k indique la répartition des valeurs X_i et X_{i+k} par rapport à la moyenne (fig. 2).

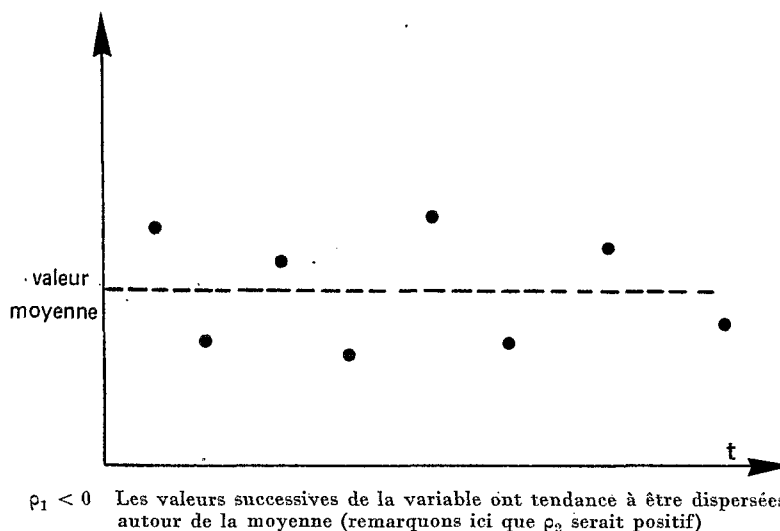
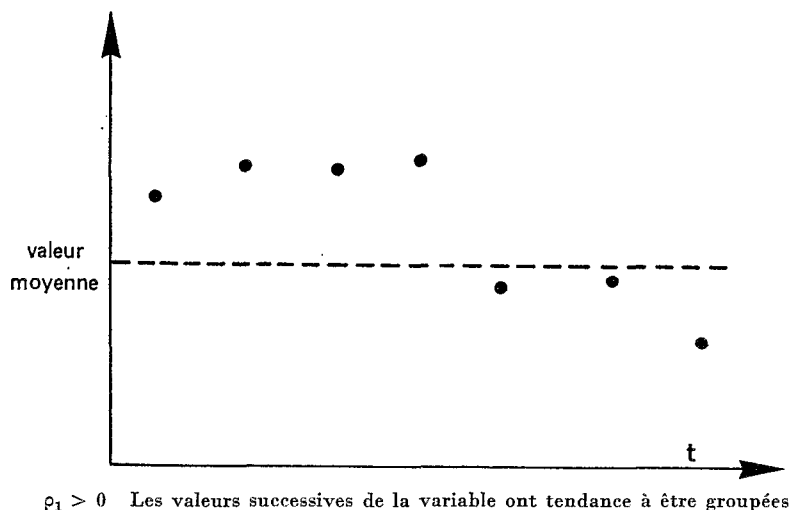


Fig. 2

Le corrélogramme de la série chronologique représente la courbe obtenue en reportant la valeur du coefficient d'autocorrélation d'ordre k en fonction de la valeur k .

1.3.2 Interprétation du corrélogramme

Le corrélogramme d'une série chronologique peut présenter diverses allures caractéristiques :

— Le corrélogramme (tabl. I) calculé à partir des hauteurs d'eau mensuelles à la station Le Raizet (1951-1978) met en évidence l'existence d'un élément cyclique dans l'organisation de la série. Disons en général que, lorsque le corrélogramme d'une série chronologique est périodique de période T , les termes de la série sont eux-mêmes affectés d'un mouvement cyclique, dont la période est également T .

Dans le tableau I, la période est égale à 12 mois, ce qui correspond évidemment au cycle annuel des saisons.

— Si la persistance est le facteur dominant dans l'organisation de la série, les coefficients d'autocorrélation vont décroître en fonction de leur ordre : en effet, la dépendance qui peut exister entre valeurs successives décroît lorsqu'augmente l'intervalle de temps séparant les observations considérées.

TABLEAU I
CORRÉLOGRAMME DES PLUIES MENSUELLES AU RAIZET

1	0.3796	*****
2	0.2345	*****
3	-0.0572	****
4	-0.2049	*****
5	-0.2411	*****
6	-0.3129	*****
7	-0.2555	*****
8	-0.1832	****
9	-0.0486	*****
10	0.2201	*****
11	0.2331	*****
12	0.3655	*****
13	0.3104	*****
14	0.1295	*****
15	-0.0718	*****
16	-0.2545	*****
17	-0.2707	*****
18	-0.3351	*****
19	-0.2903	*****
20	-0.1800	****
21	-0.0574	*****
22	0.2195	*****
23	0.3716	*****
24	0.4174	*****
25	0.3226	*****
26	0.0923	**
27	-0.0125	*****
28	-0.2378	*****
29	-0.2563	*****
30	-0.2818	*****
31	-0.3030	*****
32	-0.1204	***
33	-0.0274	*****
34	0.1097	*****
35	0.2585	*****
36	0.3554	*****
37	0.3238	*****
38	0.1541	***
39	-0.0304	*****
40	-0.1954	*****
41	-0.3594	*****
42	-0.3235	*****
43	-0.2714	*****
44	-0.1297	**
45	-0.0062	*****
46	0.1651	*****
47	0.3236	*****
48	0.3212	*****
49	0.3233	*****
50	0.1598	*****

TABLEAU II
CORRÉLOGRAMME DES PLUIES JOURNALIÈRES AU RAIZET

1	0. 1506	*****
2	0. 0677	****
3	0. 0461	***
4	0. 0350	**
5	0. 0502	***
6	0. 0289	**
7	0. 0340	**
8	0. 0419	***
9	0. 0221	**
10	0. 0415	***
11	0. 0376	**
12	0. 0380	**
13	0. 0130	*
14	0. 0553	***
15	0. 0218	**
16	0. 0398	**
17	0. 0164	*
18	0. 0325	**
19	0. 0172	*
20	0. 0238	**
21	0. 0307	**
22	0. 0224	**
23	0. 0352	**
24	0. 0160	*
25	0. 0207	**
26	0. 0334	**
27	0. 0289	**
28	0. 0202	**
29	0. 0090	*
30	0. 0102	*
31	0. 0277	**
32	0. 0263	**
33	0. 0237	**
34	0. 0316	**
35	0. 0355	**
36	0. 0396	**
37	0. 0381	**
38	0. 0549	***
39	0. 0286	**
40	0. 0175	*
41	0. 0267	**
42	0. 0377	**
43	0. 0420	***
44	0. 0152	*
45	0. 0103	*
46	0. 0163	*
47	0. 0092	*
48	0. 0271	**
49	0. 0102	*
50	0. 0383	**

Le corrélogramme (tabl. II) calculé à partir des hauteurs d'eau journalières en est une illustration, et les coefficients d'autocorrélation expriment ici, à la fois la force et l'étalement de la persistance : ainsi à la station Le Raizet, la persistance semble se manifester pendant 2 jours, d'une façon plus marquée le premier jour ($\rho_1 = 0,15$) que le second ($\rho_2 = 0,07$) ; pour constituer une série de précipitations journalières indépendantes entre elles, il serait nécessaire de prendre les observations de 2 en 2 ou de 3 en 3.

— L'effet de tendance n'apparaît généralement pas de façon caractéristique dans le corrélogramme. Toutefois, si l'on désire, pour une raison ou pour une autre, minimiser l'influence de cette tendance dans le calcul des coefficients d'autocorrélation, une méthode consiste à trouver l'expression analytique de la courbe de tendance (voir fig. 1.1) éventuellement au moyen des moindres carrés, et à remplacer dans le calcul des coefficients la valeur moyenne \bar{X} par la valeur $X(t)$ qui traduit l'évolution de cette valeur moyenne en fonction du temps.

Les coefficients d'autocorrélation calculés à partir des données brutes dans la série tiennent compte des effets de dépendance de toute nature. Aussi, les corrélogrammes de séries affectées de plusieurs effets différents peuvent ne correspondre à aucune des courbes caractéristiques que nous avons présentées, dans la mesure ou la combinaison de ces effets crée souvent une sorte de brouillage.

Il n'est pas inintéressant alors, si cela est possible, d'envisager une correction des données, à partir de leur expression analytique, pour supprimer l'influence de tel ou tel effet : l'examen du nouveau corrélogramme obtenu à partir des données corrigées fera peut-être apparaître une courbe caractéristique.

En conclusion, nous dirons que la méthode du corrélogramme présente de nombreux avantages (facilité de la mise en œuvre, multiplicité de l'information recueillie) et semble assez bien convenir à l'étude des effets de dépendance pouvant lier les termes d'une série chronologique. Dans le cas particulier de l'étude des cycles, nous citerons un autre outil, le périodogramme, qui peut dans certaines conditions constituer un apport non négligeable au corrélogramme.

La méthode ne peut évidemment pas s'appliquer si l'on entreprend d'étudier un système ne prenant qu'un nombre fini et limité d'états (état sec — état pluvieux par exemple), le calcul des coefficients d'autocorrélation s'avérant impossible dans ce cas. Nous verrons plus loin qu'il est possible de caractériser la persistance d'un tel système par un autre coefficient, le paramètre de contagion du schéma de Polya.

1.3.3 Intervalles de confiance des coefficients d'autocorrélation

La détermination de la distribution des coefficients d'autocorrélation n'est possible théoriquement que lorsque la variable est distribuée normalement.

On peut admettre, en vertu du théorème central limite, que dans le cas d'échantillon de grande taille (au moins plusieurs centaines de valeurs), la distribution des ρ_k est approximativement normale, de paramètres

$$\hat{\rho}_k + \frac{\hat{\rho}_k (1 - \hat{\rho}_k^2)}{2n}$$

et

$$\frac{1 - \hat{\rho}_k^2}{\sqrt{n-1}}$$

où n désigne le nombre de valeurs utilisées pour calculer ($n =$ taille de l'échantillon — ordre du coefficient), ce qui donne pour les extrémités de l'intervalle de confiance $1 - \alpha$

$$\rho_{1,2} \sim \hat{\rho}_k + \frac{\hat{\rho}_k (1 - \hat{\rho}_k^2)}{2n} \pm u_{(\alpha/2)} \cdot \frac{1 - \hat{\rho}_k^2}{\sqrt{n-1}}$$

où $u_{(\alpha/2)}$ représente la variable centrée réduite de la loi $N(0,1)$

Plus précisément, et pour lever les restrictions concernant la taille de l'échantillon, il est souhaitable de calculer les intervalles de confiance par une double anamorphose en considérant la transformation de Fisher

$$\rho \rightarrow z = \text{Argth } \rho = \frac{1}{2} \text{Log } \frac{1 + \rho}{1 - \rho}$$

La distribution de z tend en effet beaucoup plus rapidement vers la normalité et l'approximation peut être raisonnablement validée pour des échantillons petits.

On considère donc que la distribution de z est normale, de paramètres

$$\hat{z} = \frac{\hat{\rho}}{2(n-1)} \quad \text{et} \quad \frac{1}{\sqrt{n-3}}$$

ce qui donne pour les valeurs limites de l'intervalle de confiance $1 - \alpha$

$$z_{1,2} = \hat{z} \pm u_{(\alpha/2)} \cdot \frac{1}{\sqrt{n-3}} = \frac{\hat{\rho}}{2(n-1)} \pm \frac{u_{(\alpha/2)}}{\sqrt{n-3}}$$

d'où il découle que la vraie valeur du coefficient d'autocorrélation est comprise, avec un niveau de confiance $1 - \alpha$, dans les limites $\text{thz}_1 < \rho < \text{thz}_2$

Les calculs effectués à partir des hauteurs de pluie mensuelles de la station Le Raizet (336 valeurs) sont assez semblables avec les deux méthodes (tabl. III).

1.4 MÉTHODE DU PÉRIODOGRAMME

Le périodogramme constitue un outil intéressant pour l'étude de l'élément cyclique d'une série.

1.4.1 Définition et interprétation

Soit X une variable représentant une série chronologique. Pour des valeurs T arbitrairement choisies (1 à 50 par exemple), que nous qualifierons de périodes d'essai, nous allons calculer les quantités suivantes :

$$A = \frac{2}{n} \sum_{t=1}^n X_t \cos \frac{2\pi t}{T}$$

$$B = \frac{2}{n} \sum_{t=1}^n X_t \sin \frac{2\pi t}{T}$$

dans un premier temps, puis :

$$S^2 = A^2 + B^2$$

Le graphique obtenu en reportant ces valeurs S^2 en fonction de T représente le corrélogramme de la série.

TABLEAU III
INTERVALLE DE CONFIANCE DES COEFFICIENTS D'AUTOCORRÉLATION CALCULÉS
A PARTIR DES HAUTEURS DE PLUIES MENSUELLES AU RAIZET

K	VALEUR DE R	VALEUR DE Z	INTERVALLE DE CONFIANCE	VALEURS LIMITES DE R	VALEURS LIMITES DE R(Z)
1	0.3796	0.3996	95%	0.2883/ 0.4719	0.2835/ 0.4673
2	0.2345	0.2390	95%	0.1333/ 0.3363	0.1301/ 0.3331
3	-0.0572	-0.0572	95%	-0.1645/ 0.0500	-0.1636/ 0.0507
4	-0.2049	-0.2078	95%	-0.3084/-0.1020	-0.3055/-0.0991
5	-0.2411	-0.2460	95%	-0.3431/-0.1399	-0.3398/-0.1365
6	-0.3129	-0.3237	95%	-0.4108/-0.2158	-0.4067/-0.2116
7	-0.2555	-0.2613	95%	-0.3571/-0.1547	-0.3535/-0.1512
8	-0.1832	-0.1853	95%	-0.2882/-0.0787	-0.2856/-0.0762
9	-0.0486	-0.0486	95%	-0.1570/ 0.0596	-0.1562/ 0.0603
10	0.2201	0.2237	95%	0.1167/ 0.3238	0.1138/ 0.3207
11	0.2331	0.2374	95%	0.1304/ 0.3364	0.1272/ 0.3331
12	0.3655	0.3832	95%	0.2715/ 0.4604	0.2666/ 0.4558
13	0.3104	0.3210	95%	0.2121/ 0.4095	0.2079/ 0.4054
14	0.1295	0.1302	95%	0.0221/ 0.2372	0.0203/ 0.2353
15	-0.0718	-0.0719	95%	-0.1809/ 0.0371	-0.1798/ 0.0381
16	-0.2545	-0.2602	95%	-0.3575/-0.1522	-0.3539/-0.1486
17	-0.2707	-0.2776	95%	-0.3730/-0.1693	-0.3692/-0.1654
18	-0.3351	-0.3485	95%	-0.4333/-0.2378	-0.4288/-0.2332
19	-0.2903	-0.2989	95%	-0.3917/-0.1898	-0.3877/-0.1857
20	-0.1800	-0.1820	95%	-0.2871/-0.0734	-0.2844/-0.0708
21	-0.0574	-0.0575	95%	-0.1677/ 0.0527	-0.1668/ 0.0535
22	0.2195	0.2231	95%	0.1144/ 0.3253	0.1111/ 0.3220
23	0.3916	0.4137	95%	0.2982/ 0.4861	0.2929/ 0.4811
24	0.4174	0.4446	95%	0.3262/ 0.5098	0.3207/ 0.5046
25	0.3226	0.3345	95%	0.2233/ 0.4228	0.2187/ 0.4183
26	0.0923	0.0925	95%	-0.0181/ 0.2030	-0.0195/ 0.2014
27	-0.0125	-0.0125	95%	-0.1242/ 0.0992	-0.1239/ 0.0992
28	-0.2378	-0.2424	95%	-0.3437/-0.1326	-0.3401/-0.1291
29	-0.2563	-0.2622	95%	-0.3614/-0.1520	-0.3576/-0.1482
30	-0.2818	-0.2897	95%	-0.3856/-0.1789	-0.3815/-0.1748
31	-0.3030	-0.3128	95%	-0.4055/-0.2013	-0.4012/-0.1969
32	-0.1204	-0.1210	95%	-0.2316/-0.0097	-0.2296/-0.0078
33	-0.0274	-0.0274	95%	-0.1401/ 0.0853	-0.1396/ 0.0856
34	0.1097	0.1101	95%	-0.0018/ 0.2215	-0.0034/ 0.2196
35	0.2585	0.2645	95%	0.1533/ 0.3645	0.1494/ 0.3607
36	0.3554	0.3716	95%	0.2568/ 0.4549	0.2517/ 0.4500
37	0.3238	0.3359	95%	0.2227/ 0.4259	0.2179/ 0.4213
38	0.1541	0.1554	95%	0.0434/ 0.2654	0.0410/ 0.2629
39	-0.0304	-0.0304	95%	-0.1443/ 0.0834	-0.1437/ 0.0837
40	-0.1954	-0.1979	95%	-0.3055/-0.0859	-0.3023/-0.0829
41	-0.3594	-0.3762	95%	-0.4595/-0.2604	-0.4545/-0.2552
42	-0.3235	-0.3355	95%	-0.4265/-0.2215	-0.4218/-0.2166
43	-0.2714	-0.2784	95%	-0.3781/-0.1656	-0.3740/-0.1614
44	-0.1297	-0.1305	95%	-0.2429/-0.0170	-0.2407/-0.0149
45	-0.0062	-0.0062	95%	-0.1213/ 0.1089	-0.1211/ 0.1089
46	0.1651	0.1666	95%	0.0532/ 0.2773	0.0506/ 0.2748
47	0.3236	0.3357	95%	0.2207/ 0.4275	0.2158/ 0.4227
48	0.3212	0.3330	95%	0.2180/ 0.4255	0.2130/ 0.4207
49	0.3233	0.3353	95%	0.2200/ 0.4276	0.2150/ 0.4228
50	0.1598	0.1612	95%	0.0470/ 0.2733	0.0444/ 0.2706

Supposons que la variable X_t se décompose sous la forme :

$$X_t = A_0 + A_1 \sin \frac{2\pi t}{T_1} + A_2 \sin \frac{2\pi t}{T_2}$$

où A_0, A_1, A_2 sont des constantes,

nous sommes donc dans le cas où la série chronologique représentée par X_t renferme 2 éléments cycliques, de périodes T_1 et T_2 , d'amplitudes $2A_1$ et $2A_2$.

Dans ce cas, le calcul montre que la quantité S^2 est dominée par un terme en

$$\frac{A_1^2 \sin^2 \left(\frac{1}{2} n \left(\frac{2\pi}{T_1} - \frac{2\pi}{T} \right) \right)}{n^2 \sin^2 \left(\frac{1}{2} \left(\frac{2\pi}{T_1} - \frac{2\pi}{T} \right) \right)}$$

au voisinage de $T = T_1$

et par un terme semblable avec T_2 à la place de T_1 au voisinage de $T = T_2$.

En notant que

$$\sin x \sim x$$

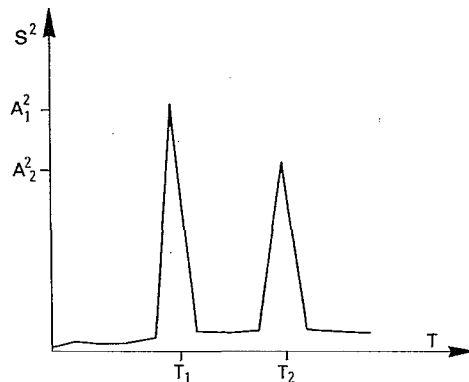
et

au voisinage de 0

$$\frac{\sin x}{x} \sim 1$$

ces termes sont équivalents, dans les voisinages respectifs de T_1 et T_2 , à A_1^2 et A_2^2 .

Le périodogramme de la variable X_t présente alors l'allure suivante (fig. 3).



On observe donc des pics très marqués pour les valeurs de T correspondant aux différentes périodes réelles de la variable X_t et les valeurs de ces pics sont théoriquement égales au carré de la demi-amplitude des divers éléments cycliques correspondants.

1.4.2 Comparaison du périodogramme et du corrélogramme — Exemples

Le périodogramme (tabl. IV) calculé à partir des hauteurs de pluie mensuelles à la station Le Raizet (1951, 1978) montre que cette série renferme effectivement un élément cyclique dont la période est égale à douze mois. Nous retrouvons ainsi le cycle annuel que nous avons caractérisé en utilisant le corrélogramme de cette même série (voir tabl. I).

En admettant que ce mouvement est approximativement sinusoïdal, la demi-amplitude du cycle est de

$$\sqrt{589404} = 768$$

ce qui semble être en accord avec l'observation.

TABLEAU IV
PÉRIODOGRAMME DES HAUTEURS DE PLUIES MENSUELLES AU RAIZET

1	0. 0	*
2	0. 0043	*
3	0. 0038	*
4	0. 0049	*
5	0. 0064	*
6	0. 0625	****
7	0. 0014	*
8	0. 0024	*
9	0. 0055	*
10	0. 0173	*
11	0. 0507	***
12	0. 5621	*****
13	0. 0053	*
14	0. 0032	*
15	0. 0004	*
16	0. 0059	*
17	0. 0078	*
18	0. 0174	*
19	0. 0037	*
20	0. 0004	*
21	0. 0158	*
22	0. 0084	*
23	0. 0177	**
24	0. 0315	**
25	0. 0450	***
26	0. 0070	*
27	0. 0014	*
28	0. 0060	*
29	0. 0280	**
30	0. 0178	*
31	0. 0028	*
32	0. 0129	*
33	0. 0212	**
34	0. 0274	**
35	0. 0311	**
36	0. 0202	**
37	0. 0048	*
38	0. 0020	*
39	0. 0065	*
40	0. 0050	*
41	0. 0005	*
42	0. 0059	*
43	0. 0238	**
44	0. 0418	***
45	0. 0457	***
46	0. 0334	**
47	0. 0153	*
48	0. 0042	*
49	0. 0065	*
50	0. 0189	*

Pour obtenir les valeurs réelles des pics, multiplier les nombres de la colonne par 1 048 576.

Comme nous l'avons déjà remarqué, le périodogramme peut donc fournir des informations non seulement sur la période de l'élément cyclique étudié, mais aussi sur son amplitude. Mais il semble que son emploi doit rester limité à l'étude de mouvements périodiques constitués par la superposition d'un nombre limité de cycles approximativement sinusoïdaux : bien que tout élément périodique soit décomposable en série de Fourier, il est souvent impossible d'interpréter des résultats faisant apparaître des forêts de pics ne se distinguant pas toujours les uns des autres (sans parler des problèmes de convergence qui peuvent se poser en vue d'une extrapolation).

Pour fixer les idées, nous avons tracé les périodogrammes et corrélogrammes relatifs aux séries suivantes :

1. $X_t = 2 + 15 \sin \frac{2\pi t}{17} + 13 \sin \frac{2\pi t}{5} + 17 \sin \frac{2\pi t}{11}$. superposition de mouvements sinusoïdaux
2. $Y_t = t * X_t$ tendance * superposition de mouvements sinusoïdaux
3. $Z_t = t + X_t$ tendance + superposition de mouvements sinusoïdaux

TABLEAU V
CORRÉLOGRAMME ET PÉRIODOGRAMME DE LA VARIABLE X_t

Correlogramme		Periodogramme	
1	0.7414	*****	
2	0.2322	*****	* 0.0006 2
3	-0.0934	****	* 0.0469 3
4	-0.1530	*****	* 0.0035 4
5	-0.2403	*****	***** 168.3715 5
6	-0.5261	*****	* 0.0888 6
7	-0.7578	*****	* 0.0307 7
8	-0.5881	*****	* 0.1240 8
9	-0.0832	****	* 0.0641 9
10	0.3051	*****	* 2.1732 10
11	0.2848	*****	***** 291.7600 11
12	0.0593	**	* 14.1891 12
13	0.0088	*	* 1.9772 13
14	0.1678	*****	* 2.4530 14
15	0.2162	*****	* 6.7588 15
16	-0.0194	**	* 19.1745 16
17	-0.2698	*****	***** 236.6829 17
18	-0.1620	*****	*** 43.2183 18
19	0.2624	*****	* 10.7766 19
20	0.5661	*****	* 3.5701 20
21	0.4590	*****	* 4.4885 21
22	0.1363	****	* 2.9077 22
23	-0.0332	**	* 1.2027 23
24	-0.0198	**	* 3.4402 24
25	-0.1343	*****	* 0.9656 25
26	-0.3255	*****	* 1.0151 26
27	-0.8867	*****	* 2.4754 27
28	-0.8072	*****	* 1.5123 28
29	-0.3015	*****	* 0.3909 29
30	0.1984	*****	* 1.0143 30
31	0.3800	*****	* 1.9173 31
32	0.3901	*****	* 1.6667 32
33	0.5323	*****	* 0.7560 33
34	0.7696	*****	* 0.3284 34
35	0.7400	*****	* 0.7118 35
36	0.2736	*****	* 1.3747 36
37	-0.3104	*****	* 1.6826 37
38	-0.5553	*****	* 1.4403 38
39	-0.4092	*****	* 0.8935 39
40	-0.2321	*****	* 0.4277 40
41	-0.2763	*****	* 0.2922 41
42	-0.3604	*****	* 0.5036 42
43	-0.1778	*****	* 0.7052 43
44	0.2085	*****	* 1.2898 44
45	0.3921	*****	* 1.5011 45
46	0.1539	****	* 1.4809 46
47	-0.2280	*****	* 1.2633 47
48	-0.3214	*****	* 0.9389 48
49	-0.0796	***	* 0.6125 49
50	0.1476	****	* 0.3698 50

Le périodogramme donne donc réellement plus d'information que le correlogramme lorsqu'on étudie une fonction constituée par la superposition d'un nombre limité de cycles approximativement sinusoidaux, intervenant seule ou en facteur multiplicatif, du type :

$$X_t = A_0 + A_1 \sin \frac{2\pi t}{T_1} + A_2 \sin \frac{2\pi t}{T_2} \quad \text{ou} \quad W_t = \varphi(t) * X_t$$

Remarquons que dans cette deuxième éventualité, l'amplitude de l'élément cyclique reste indéterminée.

Dans l'exemple 1, le périodogramme nous donne toutes les caractéristiques (période et amplitude) des éléments cycliques que renferme la série. A partir d'un échantillon de 200 valeurs, nous pouvons savoir que l'organisation de la série consiste en la superposition de 3 cycles sinusoidaux de périodes 5, 11 et 17, d'amplitudes respectives $2 * 13$, $2 * 17$, $2 * 15$. Par contre, le correlogramme ne pourrait caractériser que la période du mouvement qui affecte les valeurs de la série (ici $5 * 11 * 17 = 935$), période qui ne peut apparaître puisque le calcul des coefficients d'auto-corrélation s'arrête ici à 50 pour des raisons de fiabilité des résultats (de toute manière, le calcul ne pourrait se poursuivre au-delà de 200).

TABLEAU VI
CORRÉLOGRAMME ET PÉRIODOGRAMME DE LA VARIABLE Y_t

Corrélogramme		Periodogramme
1 0.7445	*****	
2 0.2402	*****	* 204.6133 2
3 -0.0888	****	* 445.9534 3
4 -0.1617	*****	* 2277.8457 4
5 -0.2607	*****	***** 1669812.00 5
6 -0.5469	*****	* 1895.6611 6
7 -0.7730	*****	* 653.7280 7
8 -0.6040	*****	* 1662.6943 8
9 -0.1054	****	* 6918.2773 9
10 0.2871	*****	* 69578.1250 10
11 0.2912	*****	***** 2910116.00 11
12 0.0968	***	** 210313.125 12
13 0.0606	**	* 112135.937 13
14 0.2121	*****	* 110232.500 14
15 0.2488	*****	** 189698.500 15
16 0.0108	*	***** 806012.000 16
17 -0.2448	*****	***** 2420610.00 17
18 -0.1651	*****	***** 763350.500 18
19 0.2137	*****	* 84444.5625 19
20 0.4872	*****	* 66485.3750 20
21 0.3885	*****	* 27412.6094 21
22 0.0977	***	* 26261.1406 22
23 -0.0481	***	* 15468.9453 23
24 -0.0233	**	* 13963.6484 24
25 -0.1207	****	* 12827.2500 25
26 -0.4697	*****	* 9390.6094 26
27 -0.7935	*****	* 8810.9141 27
28 -0.7201	*****	* 9032.1406 28
29 -0.2632	*****	* 7918.0664 29
30 0.1886	*****	* 6719.0859 30
31 0.3560	*****	* 6508.0273 31
32 0.3699	*****	* 6755.4570 32
33 0.4983	*****	* 6666.6836 33
34 0.7060	*****	* 6156.7109 34
35 0.6727	*****	* 5619.3320 35
36 0.2509	*****	* 5370.1914 36
37 -0.2763	*****	* 5417.0977 37
38 -0.5076	*****	* 5573.0625 38
39 -0.3954	*****	* 5650.1445 39
40 -0.2490	*****	* 5571.9453 40
41 -0.2837	*****	* 5370.5703 41
42 -0.3430	*****	* 5133.1055 42
43 -0.1677	*****	* 4942.7656 43
44 0.1835	*****	* 4847.1445 44
45 0.3608	*****	* 4851.2852 45
46 0.1759	*****	* 4929.6445 46
47 -0.1384	*****	* 5042.4687 47
48 -0.2211	*****	* 5151.2617 48
49 -0.0332	**	* 5227.0000 49
50 0.1371	****	* 5255.4727 50

Dans le cas 3, on retrouve les 3 pics correspondant aux périodes des éléments cycliques de la série, mais à partir de $T = 18$, le périodogramme présente une nette tendance uniforme. Le corrélogramme n'indique rien.

Le calcul montre que les quantités S^2 calculées pour le périodogramme sont dominées à partir d'un certain rang par un terme en $\cotg \frac{2\pi}{T}$ et la fonction $T \rightarrow S^2(T)$ est alors strictement croissante.

Un autre essai effectué avec la série $t + \sin \frac{2\pi t}{10}$ n'a pas permis de faire apparaître le pic caractéristique en $T = 10$. En d'autres termes lorsque l'élément cyclique d'une série se combine additivement avec la tendance, la lisibilité du périodogramme dépend de leur importance relative.

1.5 INFLUENCE DE LA PERSISTANCE SUR L'EXACTITUDE DES MOYENNES

Soit X une variable représentant les valeurs d'une série chronologique, dont la moyenne \bar{X} et l'écart-type σ_x de la population mère sont parfaitement connus.

TABLEAU VII
CORRÉLOGRAMME ET PÉRIODOGRAMME DE LA VARIABLE Z_t

Correlogramme		Periodogramme	
1	0.9648	*****	
2	0.9134	*****	* 0.9512 2
3	0.8785	*****	* 1.8566 3
4	0.8668	*****	* 1.9059 4
5	0.8525	*****	***** 135.1564 5
6	0.8202	*****	* 4.5221 6
7	0.7912	*****	* 6.1681 7
8	0.7949	*****	* 7.0660 8
9	0.8253	*****	* 7.3299 9
10	0.8443	*****	* 17.0180 10
11	0.8277	*****	***** 283.5164 11
12	0.7940	*****	* 0.7191 12
13	0.7766	*****	* 8.6108 13
14	0.7788	*****	* 8.5865 14
15	0.7735	*****	* 6.0117 15
16	0.7456	*****	* 15.0863 16
17	0.7176	*****	***** 210.2917 17
18	0.7203	*****	***** 147.3995 18
19	0.7492	*****	*** 55.4655 19
20	0.7663	*****	*** 59.4092 20
21	0.7470	*****	*** 52.6773 21
22	0.7080	*****	*** 71.7275 22
23	0.6809	*****	*** 44.2107 23
24	0.6688	*****	**** 82.0087 24
25	0.6456	*****	**** 71.6271 25
26	0.5992	*****	*** 59.7958 26
27	0.5557	*****	**** 89.9777 27
28	0.5496	*****	**** 99.3885 28
29	0.5793	*****	**** 82.6507 29
30	0.6082	*****	**** 84.9146 30
31	0.6100	*****	**** 109.7805 31
32	0.5976	*****	**** 126.5967 32
33	0.5969	*****	**** 122.9947 33
34	0.6051	*****	**** 114.4461 34
35	0.5920	*****	**** 118.4149 35
36	0.5435	*****	**** 136.1752 36
37	0.4862	*****	**** 156.6017 37
38	0.4575	*****	**** 169.1965 38
39	0.4607	*****	**** 171.8392 39
40	0.4646	*****	**** 169.7676 40
41	0.4477	*****	**** 170.0625 41
42	0.4252	*****	**** 176.8639 42
43	0.4229	*****	**** 190.3979 43
44	0.4363	*****	**** 207.2281 44
45	0.4327	*****	**** 223.5834 45
46	0.3951	*****	**** 236.7256 46
47	0.3477	*****	**** 245.8689 47
48	0.3265	*****	**** 251.9521 48
49	0.3352	*****	**** 256.8770 49
50	0.3440	*****	**** 262.6208 50

Si les valeurs X_i de cette variable sont indépendantes, l'espérance mathématique (valeur moyenne) de m termes consécutifs en chronologie est \bar{X} et la variance de cette moyenne est :

$$V_0 = \frac{\sigma_x^2}{m}$$

Il n'en est pas de même si les valeurs X_i ne sont pas indépendantes : définissons la variable Y_i comme la moyenne des m valeurs consécutives de la variable X , X_{i+1}, \dots, X_{i+m} , et calculons la moyenne et la variance de cette variable

$$Y_i = \frac{X_{i+1} + \dots + X_{i+m}}{m} \quad \text{moyenne de } m \text{ termes consécutifs}$$

$$E(Y) = E\left(\sum_{b=1}^m X_{i+b}/m\right) = \sum_{b=1}^m E(X_{i+b})/m = \frac{mE(X)}{m} = E(X)$$

$$\text{var}(Y) = \text{var}(g(X_{i+1}, \dots, X_{i+m})) \text{ avec } g(X_{i+1}, \dots, X_{i+m}) = \frac{1}{m} \sum_{b=1}^m X_{i+b}$$

$$\begin{aligned} \text{donc var}(Y) &= \sum_{i+1 \leq p, q \leq i+m} \frac{\delta g}{\delta x_p}(\bar{X}_p) \cdot \frac{\delta g}{\delta x_q}(\bar{X}_q) \text{cov}(X_p, X_q) \\ &= \sum_{i+1 \leq p, q \leq i+m} \frac{1}{m} \cdot \frac{1}{m} \text{cov}(X_p, X_q) \\ &= \frac{1}{m^2} \sum_{i+1 \leq p \leq i+m} \text{var}(X_p) + \frac{1}{m^2} \sum_{i+1 \leq p \neq q \leq i+m} \text{cov}(X_p, X_q) \\ \text{or cov}(X_p, X_q) &= \rho_{p-q} \cdot (\text{var}(X_p) \cdot \text{var}(X_q))^{1/2} = \rho_{q-p} \cdot \sigma_x^2 \end{aligned}$$

et donc

$$\text{var}(Y) = \frac{1}{m^2} \cdot m \sigma_x^2 + \frac{2\sigma_x^2}{m^2} (\rho_{m-1} + 2\rho_{m-2} + \dots + (m-1)\rho_1)$$

$$\text{i.e. var}(Y) = \frac{\sigma_x^2}{m^2} (m + 2((m-1)\rho_1 + \dots + 2\rho_{m-2} + \rho_{m-1}))$$

où $\rho_1, \dots, \rho_{m-1}$ sont les coefficients d'autocorrélation d'ordre 1, ..., $m-1$ de la variable X.

Si les valeurs de cette variable sont indépendantes

$$\rho_1 = \rho_2 = \dots = \rho_{m-1} = 0$$

et on retrouve bien

$$V_0 = \sigma_x^2/m$$

Dans le cas où les coefficients d'autocorrélation sont liés par la relation

$$\rho_k = \rho_1^k, k \in |1 \dots| \quad (\text{processus markovien})$$

on trouve alors

$$V = \frac{\sigma_x^2}{m} \left(\frac{1 + \rho_1}{1 - \rho_1} - \frac{2\rho_1}{m} \cdot \frac{1 - \rho_1^m}{(1 - \rho_1)^2} \right)$$

Pour m assez grand, la variance de la valeur moyenne de m termes consécutifs en chronologie est donc :

$$V = \frac{\sigma_x^2}{m} \cdot \frac{1 + \rho_1}{1 - \rho_1}$$

2. ÉTUDE DES SÉQUENCES DES JOURS DE PLUIES

2.1 POSITION DU PROBLÈME

La répartition des jours de pluie au cours d'une période donnée n'est pas purement aléatoire. Le phénomène de persistance se manifeste en disant que la probabilité d'observer un jour de pluie dépend des états (secs ou pluvieux) des jours précédents. Ainsi, en Guadeloupe, cette probabilité augmente jusqu'à atteindre un maximum, enregistré lorsqu'on a observé précédemment 5 ou 6 jours de pluie, puis décroît ensuite.

Pour étudier la succession chronologique des jours de pluie, il est donc souhaitable d'employer des schémas tenant compte de cet effet de persistance : le schéma de Polya et les chaînes de Markov.

Nous définirons ainsi deux états : un état sec et un état pluvieux au moyen de seuils (inférieur et éventuellement supérieur) ; à chaque jour sera attribué l'un de ces deux états en fonction de la précipitation observée ce jour et des valeurs choisies pour les seuils.

2.2 SCHÉMA DE POLYA

Le schéma de Polya considère qu'entre les tirages au sort successifs et effectués dans un certain ordre dans N urnes contenant toutes des boules blanches et des boules noires en proportion différente, il y a interdépendance, c'est-à-dire qu'après chaque tirage, non seulement on remet la boule tirée dans l'urne, mais en plus b boules de la même couleur, ce qui caractérise la contagion ou la persistance du phénomène (1).

Dans le schéma de Polya, la probabilité P_k de k occurrences d'un événement dans une série de n observations est :

$$P_0 = \frac{1}{(1+d)^{m/d}}$$

et pour $k \geq 1$

$$P_k = \frac{m(m+d) \dots (m+(k-1)d)}{k!(1+d)^{m/d+k}}$$

ou encore

$$P_k = P_{k-1} \cdot \frac{m+(k-1)d}{k(1+d)}$$

La loi est donc entièrement définie par deux paramètres, m et d , qui représentent respectivement (voir annexe A) :

m = moyenne des occurrences de l'événement considéré.

$d = \frac{\sigma^2}{m} - 1$ où σ^2 est la variance des occurrences de l'événement considéré.

d caractérise la persistance du phénomène de la manière suivante :

- si $d = -1$, c'est-à-dire si $\sigma^2 = 0$, on obtient toujours le même nombre n d'occurrences, ce qui signifie en particulier que le fait d'avoir observé n occurrences rend impossible une occurrence supplémentaire du même événement ;
- si $d = 0$, c'est-à-dire si $\sigma^2 = m$, on retrouve la loi de Poisson (voir annexe A) et l'occurrence d'un événement est indépendante des précédentes ;
- si $-1 < d < 0$, l'occurrence d'un événement rend plus probable l'occurrence de l'événement contraire ;
- si $d > 0$, l'occurrence d'un événement rend plus probable une nouvelle occurrence du même événement.

Dans le cas de notre étude des séquences de jours de pluie, l'événement considéré est : « 1 jour de pluie suit un jour de pluie ». Notons d'abord que, dans ces conditions, une séquence de k jours de pluie se traduit par $(k-1)$ occurrences de l'événement cité : par exemple, 0 occurrence de l'événement correspond à un jour de pluie isolé, autrement dit à une séquence de longueur 1. Si on appelle S_k le nombre de séquences observées de longueur k et \bar{S} le nombre total des séquences observées, les probabilités expérimentales de k occurrences $P_{\text{exp}}(k)$ où k varie de 0 à l'infini sont donc égales $\frac{S_{k+1}}{\bar{S}}$.

Les paramètres m et d sont calculés d'après l'échantillon :

$$m = \sum_0 k \cdot P_{\text{exp}}(k) = \sum_0 k \frac{S_{k+1}}{\bar{S}} = \sum_1 \frac{(k-1)S_k}{\bar{S}} = \frac{J - \bar{S}}{\bar{S}}$$

où $J = \sum_1 k S_k$ représente le nombre de jours de pluie

$$\sigma^2 = \sum_0 (k-m)^2 \cdot P_{\text{exp}}(k) = \sum_1 (k-m-1)^2 \cdot \frac{S_k}{\bar{S}}$$

d'où l'on déduit $d = \frac{\sigma^2}{m} - 1$; « paramètre de contagion »

(1) Une variante consiste à soustraire, et non pas ajouter, b boules de la couleur tirée.

Les nombres de séquences de longueur k ($= k - 1$ occurrences) théoriquement calculés par le schéma de Polya sont obtenus en multipliant \bar{S} par P_{k-1} .

2.3 CHAINES DE MARKOV

2.3.1 Généralités

Considérons une variable climatologique X ne pouvant prendre qu'un nombre fini d'états $E_1 \dots E_n$. Si la probabilité que la variable X prenne un état quelconque E_i dans l'intervalle t ne dépend que des états qu'elle a pris dans l'intervalle précédent $t - 1$, on peut représenter la succession des valeurs prises par X par une chaîne de Markov d'ordre 1. Si cette probabilité $\text{Prob}(X(t) = E_i)$ dépend non seulement des états pris par X dans l'intervalle $t - 1$, mais aussi des états pris aux instants $t - 2, t - 3 \dots, t - k$, la chaîne est dite d'ordre k . Les chaînes de Markov permettent donc de représenter la structure de dépendance chronologique de variables ne pouvant prendre qu'un nombre fini d'états n au moyen de matrices de probabilités dites de transition. On peut noter ces probabilités :

$$P(E_1 \dots E_k : E_{k+1}) = \text{Prob}(X(t) = E_{k+1} / X(t-1) = E_1 \dots X(t-k) = E_k) \quad (2)$$

ceci implique que le nombre de probabilités à estimer est directement proportionnel au nombre d'états n d'une part, et à l'ordre de la chaîne d'autre part. Dans la pratique, pour un nombre d'états n fixé, l'information dont on dispose n'étant pas toujours très importante, il convient de réaliser un compromis quant au choix de l'ordre de la chaîne pour obtenir une bonne précision dans les résultats et dans l'estimation des paramètres (ou du moins dans la confiance avec laquelle on peut affirmer que ceux-ci représentent bien les paramètres de la population mère).

2.3.1 Utilisation des probabilités conditionnelles dans l'étude des séquences de jours de pluie

Le nombre des états du système est ici limité à 2 : un état sec et un état pluvieux.

Convenons de noter 1 l'état pluvieux,

Convenons de noter 2 l'état sec.

L'événement « observer une séquence d'exactly k jours pluvieux » est réalisé si l'on observe la succession chronologique suivante :

Jour sec, k jours pluvieux, jour sec, c'est-à-dire :

$$2 \quad \underbrace{1 \dots 1}_k \quad 2$$

Nous appelons P_k la probabilité d'obtenir une séquence de longueur k

$$P_k = \text{prob}(2 \quad \underbrace{1 \dots 1}_k \quad 2)$$

— dans le cas où les états du système sont indépendants entre eux (assimilable à une chaîne d'ordre 0)

$$P_1 = \text{prob}(212) = \text{prob}(2) \cdot \text{prob}(1) \cdot \text{prob}(2) \\ = \text{prob}(2)^2 \cdot \text{prob}(1)$$

et

$$P_k = \text{prob}(2 \quad \underbrace{1 \dots 1}_k \quad 2) = \text{prob}(2) \cdot \underbrace{\text{prob}(1) \dots \text{prob}(1)}_k \cdot \text{prob}(2) \\ = \text{prob}(2)^2 \cdot \text{prob}(1)^k \\ = \text{prob}(1) \cdot P_{k-1}$$

(2) Il est également possible de représenter l'évolution chronologique d'une variable continue X par un processus markovien d'ordre k , sous la forme :

$$X_i = U_i + Y(X_{i-k} \dots X_{i-1})$$

où U désigne une variable dont les valeurs sont indépendantes

Y est une fonction de k variables.

(voir l'étude de MM. Brunet-Moret et Roche. Cahiers Hydro, vol. XII, n° 3, 1975).

Pour calculer ces probabilités, il est donc nécessaire de calculer d'après l'échantillon la valeur du paramètre prob (1)

$$\text{prob (1)} = \frac{\text{nombre de jours de pluie}}{\text{nombre de jours observés}}$$

d'où l'on déduit prob (2) = 1 — prob (1)

— dans le cas où l'état pris à l'instant t dépend de l'état pris à l'instant t — 1 (dépendance d'ordre 1)

$$P_1 = \text{prob (212)} = \text{prob (2/21)} \cdot \text{prob (21)} \quad (3)$$

or prob (2/21) = prob (2/1) et prob (21) = prob (1/2) · prob (2)

$$\text{pour } k \geq 2 \quad P_k = \text{prob (2 } \underbrace{1 \dots 1}_k \text{ 2)}$$

$$= \text{prob (2/1)} \cdot \underbrace{\text{prob (1/1)} \dots \text{prob (1/1)}}_{k-1} \cdot \text{prob (1/2)} \cdot \text{prob (2)}$$

$$P_k = \text{prob (2/1)} \cdot \text{prob (1/1)}^{k-1} \cdot \text{prob (1/2)} \cdot \text{prob (2)}$$

$$= \text{prob (1/1)}^{k-1} \cdot P_1$$

il est alors nécessaire d'estimer 4 paramètres, dont 3 seulement sont indépendants,

$$\begin{aligned} &\text{prob (1/1)} \\ &\text{prob (1/2)} \\ &\text{prob (2)} \\ &\text{prob (2/1)} = 1 - \text{prob (1/1)}, \end{aligned}$$

de la manière suivante :

$$\text{prob (j/i)} = \frac{\text{nombre de couples } ij}{\text{nombre de jours } i} \quad i = 1, 2 \quad j = 1, 2$$

— dans le cas où l'état pris à l'instant t dépend des états pris aux instants t — 1 et t — 2 (dépendance d'ordre 2)

$$P_1 = \text{prob (212)} = \text{prob (2/21)} \cdot \text{prob (21)}$$

$$P_2 = \text{prob (2112)} = \text{prob (2/211)} \cdot \text{prob (211)}$$

$$= \text{prob (2/11)} \cdot \text{prob (1/21)} \cdot \text{prob (21)}$$

pour k ≥ 3

$$P_k = \text{prob (2 } \underbrace{1 \dots 1}_k \text{ 2)} = \text{prob (2/11)} \cdot \underbrace{\text{prob (1/11)} \dots \text{prob (1/11)}}_{k-2} \cdot \text{prob (1/21)} \cdot \text{prob (21)}$$

$$= \text{prob (2/11)} \cdot \text{prob (1/11)}^{k-2} \cdot \text{prob (1/21)} \cdot \text{prob (21)}$$

$$= \text{prob (1/11)}^{k-2} \cdot P_2$$

il faut donc calculer d'après l'échantillon 5 paramètres, dont 3 seulement sont indépendants :

$$\text{prob (21)} = \frac{\text{nombre de couples } 21}{\text{nombre de couples observés}}$$

$$\left. \begin{aligned} &\text{prob (1/11)} \\ &\text{prob (1/21)} \end{aligned} \right| \text{prob (k/ij)} = \frac{\text{nombre de triplets } ijk}{\text{nombre de couples } ij}$$

$$i = 1, 2 ; j = 1, 2 ; k = 1, 2$$

$$\text{prob (2/11)} = 1 - \text{prob (1/11)}$$

$$\text{prob (2/21)} = 1 - \text{prob (1/21)}$$

Quel que soit l'ordre de dépendance retenu (0, 1 ou 2) le nombre théorique de séquences de longueur k s'obtient en multipliant les probabilités P_k par N, nombre de jours observés.

(3) D'après le schéma des probabilités conditionnelles

$$P(A B) = P(B/A) \cdot P(A)$$

2.4 RÉSULTATS ET COMMENTAIRES

Les groupes de stations étudiées ont été choisis en fonction de leur appartenance à des régimes climatiques variés :

- 3 stations de la Guadeloupe :

621445 Le Raizet 1951-1978	}	climat tropical maritime
620810 Boyvinière 1929-1978		
622130 Duclos 1954-1978		

 (2 années manquantes)
- 2 stations de Côte d'Ivoire :

90055 Bouake 1921-1973	climat équatorial de transition
90160 Odienne 1921-1973	climat tropical de transition

2.4.1 Calcul théorique du nombre de séquences

Dans un premier temps, nous étudions les séquences de jours de pluie pour chacune des stations précédentes à l'échelle annuelle, c'est-à-dire en prenant en considération tous les jours de l'année.

Le tableau VIII établit la comparaison des nombres théoriques des séquences *d'exactement* k jours ($k = 1 \dots$) de pluie calculés pour chaque schéma (d'après les considérations faites aux paragraphes 2.2 et 2.3) par rapport aux nombres observés, à la station Le Raizet, pour un seuil inférieur fixé à 0,1 mm.

Dans la colonne « différence observée » sont calculées pour chaque schéma les différences entre les nombres de séquences calculés par ce schéma et le nombre de séquences observées de même longueur.

Le calcul est effectué jusqu'à la plus grande longueur l_{\max} observée, plus 1, soit 35 jours ici. En conséquence, la dernière longueur du tableau représente ici les nombres de séquences observés et calculés *d'au moins* 35 jours.

Dans ce tableau figure également la valeur du paramètre d , dénommé « contagion », intervenant dans le schéma de Polya (ici contagion = 4,089).

Il est souvent plus utile, dans la plupart des problèmes, de connaître le nombre de séquences *d'au moins* k ($= 1 \dots$) jours de pluie. C'est l'étude de ces nombres, observés d'une part et calculés d'autre part pour chaque schéma, que nous développerons par la suite. Ces nombres sont représentés par le tableau IX.

Pour rendre compte clairement de la qualité de l'ajustement fourni par chaque schéma, en ce qui concerne les séquences de longueur supérieure à k jours de pluie, nous traçons, à partir des résultats du tableau IX, le graphique 4 de la manière suivante : nous reportons l'erreur relative commise sur le calcul du nombre théorique de séquences, pour chaque schéma, en fonction de la longueur k de ces séquences.

$$\text{En abscisse : erreur relative} = \frac{\text{différence observée}}{\text{nombre de séquences observé}}$$

En ordonnée : longueur des séquences

donc, pour chaque longueur k ($k = 1 \dots, l_{\max} + 1$), nous obtenons 4 points :

- les points + représentent l'ajustement de Polya
- les points * représentent l'ajustement Indépendance
- les points . représentent l'ajustement Markov 1
- les points o représentent l'ajustement Markov 2

Il est évident que la meilleure qualité d'un ajustement est obtenue par un maximum de points dont l'abscisse (erreur relative) est proche de la valeur nulle. Toutefois, il convient de ne pas accorder une trop grande importance aux points représentant les séquences de longueurs extrêmes, compte tenu de la faiblesse de l'information fournie par l'échantillon à ce sujet. Pratiquement, les points d'abscisse voisins de 0 sont ceux qui correspondent aux séquences relativement courtes, donc les plus significatives du point de vue de l'information, et ce n'est pas trahir l'interprétation du résultat que de comparer la qualité de l'ajustement fourni par un schéma au nombre de points obtenus par le schéma correspondant situés dans une bande encadrant la valeur d'abscisse 0.

Par exemple, si nous retenons comme critère de qualité une erreur relative comprise entre $\pm 10\%$, on compte d'après le graphique 4 :

- 19 points pour le schéma de Polya
- 13 points pour le schéma Markov 2
- 8 points pour le schéma Markov 1
- 2 points pour le schéma Indépendance

répondant à ce critère, pour un maximum de 34 points; les résultats obtenus pour l'ensemble des stations étudiées en Guadeloupe et en Côte d'Ivoire et exprimés sous cette forme sont consignés dans le tableau X.

LE RAIZET Station no 621445 Seuil inférieur = 0.1 mm

Période d'observation : 1951-1978

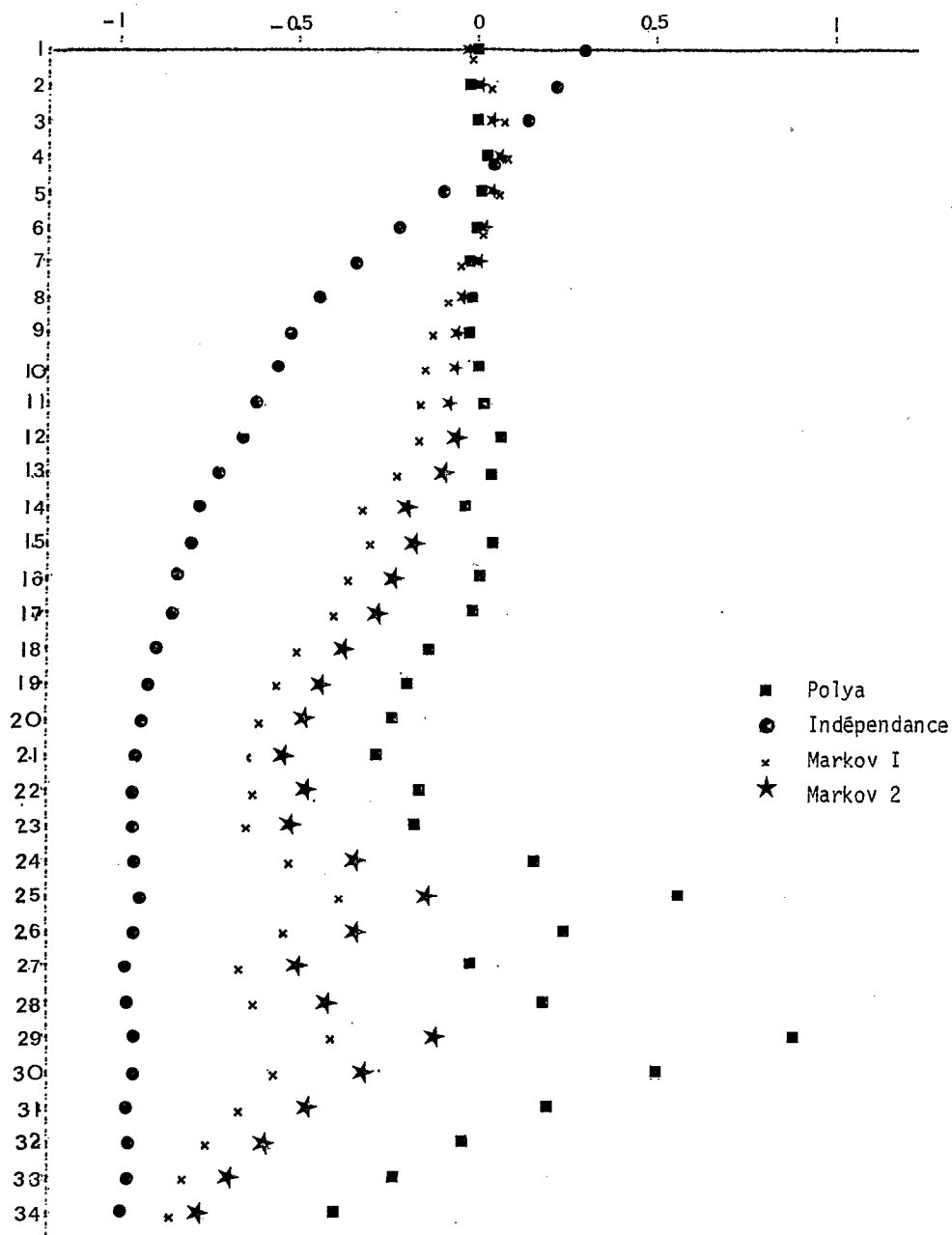


Fig. 4. — Comparaison des ajustements fournis par les schémas de Polya, Indépendance, Markov 1 et 2

TABLEAU VIII
CALCUL DES SÉQUENCES D'EXACTEMENT k JOURS DE PLUIE

LE RAIZET

NO DE LA STATION 621445
PERIODE D'OBSERVATION=1951-1978

SÉQUENCES JOURNALIÈRES

NOMBRE TOTAL DE JOURS 10227
DE PLUIE... 6922
SÉQUENCES..... 1715

SEUIL INFÉRIEUR= 0.1 SEUIL SUPÉRIEUR=***** CONTAGION= 4.089

LONGUEUR DES SÉQUENCES	NOMBRE DE SÉQUENCES OBS.	POLYA		INDEPENDANCE		CHAÎNE D'ORDRE 1		CHAÎNE D'ORDRE 2	
		SÉQUENCES CALCULÉES	DIFFÉRENCE OBSERVÉE	SÉQUENCES CALCULÉES	DIFFÉRENCE OBSERVÉE	SÉQUENCES CALCULÉES	DIFFÉRENCE OBSERVÉE	SÉQUENCES CALCULÉES	DIFFÉRENCE OBSERVÉE
1	492	512.4	20.4	722.9	230.9	424.7	-67.3	491.0	-1.0
2	327	305.7	-21.3	489.3	162.3	319.5	-7.5	287.3	-39.7
3	229	214.0	-13.0	331.2	102.2	240.4	11.4	219.8	-9.2
4	150	157.2	7.2	224.1	74.1	180.9	30.9	168.2	18.2
5	110	118.2	8.2	151.7	41.7	136.1	26.1	128.7	18.7
6	87	90.1	3.1	102.7	15.7	102.4	-15.4	98.5	11.5
7	69	69.3	0.3	69.5	0.5	77.0	8.0	75.3	6.3
8	33	33.6	0.6	47.0	-6.0	57.9	4.9	57.6	4.6
9	47	41.7	-5.3	31.8	-15.2	43.6	-3.4	44.1	-2.9
10	34	32.5	-1.5	21.5	-12.5	32.8	-1.2	33.7	-0.3
11	28	25.5	-2.5	14.6	-13.4	24.7	-3.3	25.8	-2.2
12	18	20.0	2.0	9.9	-8.1	18.6	0.6	19.8	1.8
13	10	15.7	5.7	6.7	-3.3	14.0	4.0	15.1	5.1
14	16	12.4	-3.6	4.5	-11.5	10.5	-3.5	11.6	-4.4
15	8	9.8	1.8	3.1	-4.9	7.9	-0.1	8.8	-0.8
16	7	7.7	0.7	2.1	-4.9	5.9	-1.1	6.8	-0.2
17	3	6.1	3.1	1.4	-1.6	4.5	1.5	5.2	2.2
18	4	4.8	0.8	0.9	-3.1	3.4	-0.6	4.0	-0.0
19	4	3.8	-0.2	0.6	-3.4	2.5	-1.5	3.0	-1.0
20	3	3.0	0.0	0.4	-2.6	1.9	-1.1	2.3	-0.7
21	5	2.4	-2.6	0.3	-4.7	1.4	-3.6	1.8	-3.2
22	2	1.9	-0.1	0.2	-1.8	1.1	-0.9	1.4	-0.6
23	4	1.5	-2.5	0.1	-3.9	0.8	-3.2	1.0	-3.0
24	2	1.2	-0.8	0.1	-1.9	0.6	-1.4	0.8	-1.2
25	0	1.0	1.0	0.1	0.1	0.5	0.5	0.6	0.6
26	0	0.8	0.8	0.0	0.0	0.3	0.3	0.5	0.5
27	1	0.6	-0.4	0.0	-1.0	0.3	-0.7	0.4	-0.6
28	1	0.5	-0.5	0.0	-1.0	0.2	-0.8	0.3	-0.7
29	0	0.4	0.4	0.0	0.0	0.1	0.1	0.2	0.2
30	0	0.3	0.3	0.0	0.0	0.1	0.1	0.2	0.2
31	0	0.2	0.2	0.0	0.0	0.1	0.1	0.1	0.1
32	0	0.2	0.2	0.0	0.0	0.1	0.1	0.1	0.1
33	0	0.2	0.2	0.0	0.0	0.0	0.0	0.1	0.1
34	1	0.1	-0.9	0.0	-1.0	0.0	1.0	0.1	-0.9
35	0	0.5	0.5	0.0	0.0	0.1	0.1	0.2	0.2

TABLEAU IX

CALCUL DES SÉQUENCES D'AU MOINS k JOURS DE PLUIE

LE RAIZET

NQ DE LA STATION 621445
PERIODE D'OBSERVATION=1951-1978

SÉQUENCES JOURNALIÈRES

NOMBRE TOTAL DE JRS 10227
DE PLUIE... 4922
SÉQUENCES..... 1715

SEUIL INFÉRIEUR= 0.1 SEUIL SUPÉRIEUR=***** CONTAGION= 4 089

!LONGUEUR DES! ! SÉQUENCES	! NOMBRE DE ! SÉQUENCES OBS.	POLYA		INDEPENDANCE		CHAÎNE D ORDRE1		CHAÎNE D ORDRE 2	
		! SÉQUENCES ! CALCULÉES	! DIFFERENCE! ! OBSERVÉE	! SÉQUENCES ! CALCULÉES	! DIFFERENCE! ! OBSERVÉE	! SÉQUENCES ! CALCULÉES	! DIFFERENCE! ! OBSERVÉE	! SÉQUENCES ! CALCULÉES	! DIFFERENCE! ! OBSERVÉE !
> 1	1715	1715.0	0.0	2237.0	522.0	1715.0	0.0	1714.2	-0.8
> 2	1223	1202.6	-20.4	1514.1	291.1	1290.3	67.3	1223.1	0.1
> 3	896	876.9	0.9	1024.8	128.8	970.7	74.7	935.8	39.8
> 4	667	683.0	16.0	693.6	26.6	730.3	63.3	716.0	49.0
> 5	517	523.8	8.8	469.5	-47.3	549.3	32.5	547.9	30.9
> 6	407	407.6	0.6	317.8	-89.2	413.4	6.4	419.2	12.2
> 7	320	317.6	-2.4	215.1	-104.9	311.0	-9.0	320.7	0.7
> 8	251	248.3	-2.7	145.6	-105.4	234.0	-17.0	245.4	-5.6
> 9	198	194.7	-3.3	98.3	-99.3	176.0	-22.0	187.8	-10.2
> 10	151	153.0	2.0	66.7	-84.3	132.4	-18.6	143.7	-7.3
> 11	117	120.5	3.5	45.1	-71.9	99.6	-17.4	109.9	-7.1
> 12	89	93.0	6.0	30.6	-58.4	75.0	-14.0	84.1	-4.9
> 13	71	75.0	4.0	20.7	-50.3	56.4	-14.6	64.3	-6.7
> 14	61	59.3	-1.7	14.0	-47.0	42.4	-18.6	49.2	-11.8
> 15	45	46.9	1.9	9.5	-35.5	31.9	-13.1	37.7	-7.3
> 16	37	37.2	0.2	6.4	-30.6	24.0	-13.0	28.8	-8.2
> 17	30	29.3	-0.3	4.4	-25.6	18.1	-11.9	22.1	-7.9
> 18	27	23.4	-3.6	2.9	-24.1	13.6	-13.4	16.9	-10.1
> 19	23	18.6	-4.4	2.0	-21.0	10.2	-12.8	12.9	-10.1
> 20	19	14.7	-4.3	1.4	-17.6	7.7	-11.3	9.9	-9.1
> 21	16	11.7	-4.3	0.9	-15.1	5.8	-10.2	7.6	-8.4
> 22	11	9.3	-1.7	0.6	-10.4	4.4	-6.6	5.8	-5.2
> 23	9	7.4	-1.6	0.4	-8.6	3.3	-5.7	4.4	-4.6
> 24	5	5.9	0.9	0.3	-4.7	2.5	-2.3	3.4	-1.6
> 25	3	4.7	1.7	0.2	-2.8	1.9	-1.1	2.4	-0.4
> 26	3	3.7	0.7	0.1	-2.9	1.4	-1.6	2.0	-1.0
> 27	3	3.0	-0.0	0.1	-2.9	1.1	-1.9	1.5	-1.5
> 28	2	2.4	0.4	0.1	-1.9	0.8	-1.2	1.2	-0.8
> 29	1	1.9	0.9	0.1	-0.9	0.6	-0.4	0.9	-0.1
> 30	1	1.5	0.3	0.0	-1.0	0.5	-0.3	0.7	-0.3
> 31	1	1.2	0.2	0.0	-1.0	0.3	-0.7	0.5	-0.5
> 32	1	1.0	-0.0	0.0	-1.0	0.3	-0.7	0.4	-0.6
> 33	1	0.8	-0.2	0.0	-1.0	0.2	-0.8	0.3	-0.7
> 34	1	0.6	-0.4	0.0	-1.0	0.1	-0.9	0.2	-0.8
> 35	0	0.3	0.5	0.0	0.0	0.1	0.1	0.2	0.2

TABLEAU X

Nom de la station	Polya	Markov 2	Markov 1	Indép.	Longueur maximale observée
621445 Le Raizet	19	13	5	2	34
620810 Boyvinière	11	7	5	1	23
622130 Duclos	19	15	12	2	40
90055 Bouaké	5	3	2	1	27
90160 Odienné	9	4	4	1	31

De la même façon, le tableau XI indique, pour chaque schéma, le nombre de points affectés d'une erreur relative comprise entre $\pm 25\%$.

TABLEAU XI

Nom de la station	Polya	Markov 2	Markov 1	Indép.	Longueur maximale observée
621445 Le Raizet	29	18	13	5	34
620180 Boyvinière	13	8	7	2	23
622130 Duclos	22	22	18	5	40
90055 Bouaké	11	6	5	1	27
90160 Odienné	14	7	5	1	31

Les résultats détaillés sont reportés sous la forme de graphique du type fig. 4 en annexe C.

L'utilisation du test du χ^2 confirme ce que nous avons pressenti quant à la qualité respective des ajustements.

Voici, par exemple, les résultats de ce test, à partir des données du tableau IX, en ce qui concerne les ajustements de Polya, Markov 1 et Markov 2. Nous considérons les séquences de jours de pluie dont les longueurs varient entre 1 et 34.

Limites des classes	Effectif observé	Effectif théorique Polya	Effectif théorique Markov 1	Effectif théorique Markov 2
< 2	6	6	10	8
2-10	6	7	5	7
10-50	8	7	6	6
50-200	6	6	5	5
200-1715	8	8	8	8
TOTAL	34	34	34	34

Soit :

$$\begin{aligned} \chi^2 \text{ Polya} &= 0,285 & \text{d}^\circ \text{ de liberté} &= 2 & 0,100 < P(\chi^2) < 0,900 \\ \chi^2 \text{ Markov 1} &= 3,333 & \text{d}^\circ \text{ de liberté} &= 1 & 0,050 < P(\chi^2) < 0,100 \\ \chi^2 \text{ Markov 2} &= 1,509 & \text{d}^\circ \text{ de liberté} &= 1 & 0,100 < P(\chi^2) < 0,900 \end{aligned}$$

Si on choisit 0,05 comme valeur du seuil du risque d'adéquation, aucun des 3 ajustements n'est à rejeter : la meilleure probabilité du χ^2 est obtenue par l'ajustement de Polya et ensuite par l'ajustement de Markov 2, ces 2 probabilités étant nettement supérieures à la probabilité du χ^2 fournie par l'ajustement de Markov 1.

2.4.2. Calcul des récurrences

L'estimation des périodes de retour d'un événement est indispensable à la réalisation de tout équipement et répond souvent à des impératifs économiques ou humains, tout en garantissant le bon fonctionnement de l'ouvrage en accord avec le phénomène étudié : à cet égard, le calcul des récurrences revêt une très grande importance.

Les données observées ne permettent pas, du fait de leur nombre limité, d'établir avec précision les récurrences de certains événements, particulièrement dans le cas des événements qui se produisent rarement. C'est pourquoi

on cherche à adapter les données à un modèle mathématique, la loi statistique qui conduit à extrapoler les résultats dans les zones où l'information est insuffisante.

Notre étude consiste ici à déterminer les périodes de retour des séquences de jours de pluie, tous les jours de l'année étant comptabilisés. Remarquons qu'il est également possible de n'étudier le phénomène qu'au cours d'une certaine période de l'année (mois, saison ... voir 2.4.3.).

Quelle que soit la période de l'année considérée, nous définirons les probabilités expérimentales et théoriques suivant 2 échelles de temps distinctes :

● La récurrence journalière R_j telle qu'une séquence de longueur supérieure à k débute en moyenne tous les R_j jours de la période considérée

$$R_j(k) = \frac{\text{nombre de jours d'observations}}{\text{nombre de séquences de longueur } \geq k} = \frac{N}{\text{PCVM}(K, k)}$$

● La récurrence annuelle R_a telle qu'une séquence de longueur supérieure à k débute en moyenne tous les R_a ans

$$R_a(k) = \frac{\text{nombre d'années d'observations}}{\text{nombre de séquences de longueur } \geq k} = \frac{\text{NAN}}{\text{PCVM}(K, k)}$$

$$\text{i.e. } R_a(k) = R_j(k) \times \frac{\text{NAN}}{N}$$

Ainsi, si notre étude s'effectue à l'échelle annuelle, c'est-à-dire si tous les jours de l'année sont comptabilisés, alors :

$$R_a(k) = R_j(k)/365$$

par contre si nous ne considérons que les jours appartenant à la période mai-octobre

$$R_a(k) = R_j(k)/184$$

Nous calculons ainsi, dans le cadre de l'étude à l'échelle annuelle des séquences de jours de pluie à la station Le Raizet, les récurrences qui figurent dans le tableau XII.

L'interprétation du graphique 4, qui montre que le meilleur ajustement est fourni par le schéma de Polya, nous conduit à penser que la récurrence d'une séquence d'au moins 34 jours de pluie est proche de 45 ans.

2.4.3. Influence de l'effet saisonnier

L'étude de certains problèmes n'implique pas la prise en compte de tous les jours de l'année. C'est le cas par exemple lorsque l'agronome doit calculer la récurrence d'une séquence d'au moins k jours secs au cours de la saison des pluies, en vue de procéder éventuellement à une irrigation de complément ou de définir le volume d'une retenue.

Quelles que soient la nature des états (secs ou pluvieux) et la période de l'année considérées, le problème peut se poser en ces termes :

« Sachant que la récurrence annuelle d'un événement est R si tous les jours de l'année sont comptabilisés, et que $a\%$ de ces événements sont relevés pendant une certaine période de l'année, quelle est la récurrence annuelle R_p de cet événement au cours de la période considérée ? ».

La réponse est $R_p = \frac{1}{a} \times R$ ($\frac{1}{a}$ étant compris entre 1 et ∞ , R_p est toujours supérieur à R quelle que soit la période de l'année considérée).

Ainsi, si par exemple, à l'échelle annuelle, la récurrence d'une séquence d'au moins 10 jours de pluie est égale à 100 ans et que 80% de ces séquences sont relevées lors de la saison des pluies, une séquence d'au moins 10 jours de pluie se produit en moyenne :

$$1 \text{ fois tous les } \frac{100}{0,8} = 125 \text{ ans au cours de la saison des pluies,}$$

$$1 \text{ fois tous les } \frac{100}{0,2} = 500 \text{ ans au cours de la saison sèche.}$$

Pour calculer ces récurrences saisonnières, nous préférons néanmoins étudier directement la série chronologique formée uniquement des jours appartenant à la période de l'année considérée.

Cette série, que nous qualifierons de « tronquée » pour la différencier de la série chronologique initiale tenant compte de tous les jours de l'année, s'obtient par le découpage de l'année en périodes distinctes : nous raccordons artifi-

TABLEAU XII
CALCUL DES RÉCURRENCES JOURNALIÈRES ET ANNUELLES
DANS LE CADRE D'UNE ÉTUDE A L'ÉCHELLE ANNUELLE

LE RAIZET NO DE LA STATION 421445
PERIODE D'OBSERVATION=1951-1970

SÉQUENCES JOURNALIÈRES

SEUIL INFÉRIEUR= 0 1 SEUIL SUPÉRIEUR=***** CONTAGION= 4.089

CALCUL DES RÉCURRENCES

LONGUEUR DES SÉQUENCES OBSERVÉES		SÉQUENCES POLYA		SÉQUENCES INDEP.		SÉQUENCES D'ORDRE 1		SÉQUENCES D'ORDRE 2		
SÉQUENCES	JOURS	ANNÉES	JOURS	ANNÉES	JOURS	ANNÉES	JOURS	ANNÉES	JOURS	ANNÉES
> 1	3.96	0.02	3.96	0.02	4.37	0.01	3.96	0.02	3.97	0.02
> 2	8.36	0.02	8.36	0.02	6.75	0.02	7.93	0.02	8.36	0.02
> 3	11.41	0.03	11.40	0.03	9.98	0.03	10.54	0.03	10.93	0.03
> 4	15.33	0.04	14.97	0.04	14.74	0.04	14.00	0.04	14.28	0.04
> 5	19.78	0.05	19.43	0.05	21.78	0.06	18.61	0.05	18.67	0.05
> 6	25.13	0.07	25.09	0.07	32.19	0.09	24.74	0.07	24.40	0.07
> 7	31.96	0.09	32.21	0.09	47.53	0.13	32.88	0.09	31.89	0.09
> 8	40.74	0.11	41.19	0.11	70.25	0.19	43.71	0.12	41.68	0.11
> 9	51.65	0.14	52.53	0.14	103.79	0.28	58.09	0.16	54.47	0.15
> 10	67.73	0.19	66.83	0.18	153.34	0.42	77.21	0.21	71.19	0.20
> 11	87.41	0.24	84.88	0.23	226.54	0.62	102.63	0.28	93.04	0.25
> 12	114.91	0.31	107.63	0.29	334.63	0.92	136.41	0.37	121.60	0.33
> 13	144.04	0.39	136.30	0.37	494.34	1.35	181.31	0.50	158.93	0.44
> 14	187.64	0.46	172.39	0.47	730.14	2.00	240.98	0.66	207.71	0.57
> 15	227.27	0.62	217.84	0.60	1078.28	2.95	320.29	0.88	271.47	0.74
> 16	276.41	0.76	275.01	0.75	1592.08	4.36	425.69	1.17	354.79	0.97
> 17	340.90	0.93	346.92	0.95	2349.98	6.44	565.77	1.55	463.68	1.27
> 18	378.78	1.04	437.32	1.20	3467.06	9.50	751.93	2.04	605.98	1.64
> 19	444.65	1.22	550.92	1.51	5111.71	14.00	999.30	2.74	791.93	2.17
> 20	538.24	1.47	693.62	1.90	7529.03	20.63	1327.98	3.64	1034.93	2.84
> 21	639.19	1.75	872.82	2.39	11073.30	30.34	1764.67	4.83	1352.44	3.71
> 22	929.73	2.55	1097.77	3.01	16251.28	44.52	2344.75	6.42	1747.28	4.84
> 23	1134.33	3.11	1380.05	3.78	23776.30	65.14	3115.16	8.53	2309.24	6.33
> 24	1645.40	5.00	1734.17	4.75	34629.26	94.87	4138.10	11.34	3017.18	8.27
> 25	3409.00	9.34	2178.24	5.97	50111.11	137.29	5495.05	15.06	3941.77	10.80
> 26	3409.00	9.34	2734.93	7.49	71853.62	196.86	7297.19	19.99	5149.04	14.11
> 27	3409.00	9.34	3432.55	9.40	101720.19	278.71	9685.58	26.54	6724.97	18.42
> 28	5113.50	14.01	4306.43	11.80	141565.75	387.85	12849.77	35.20	8781.34	24.06
> 29	10227.00	28.02	5400.71	14.80	192620.56	527.73	17037.28	46.68	11463.30	31.41
> 30	10227.00	28.02	6770.29	18.55	254622.12	698.14	22571.23	61.84	14958.93	40.98
> 31	10227.00	28.02	8483.57	23.24	326095.44	893.41	29870.87	81.84	19511.24	53.46
> 32	10227.00	28.02	10625.56	29.11	402244.00	1102.04	39475.82	108.15	25433.18	69.60
> 33	10227.00	28.02	13301.76	36.44	477793.94	1308.91	52073.22	142.67	33125.86	90.76
> 34	10227.00	28.02	16642.72	45.60	547291.00	1499.43	68525.19	187.74	43100.31	118.08

ciellement le dernier jour de la période étudiée au premier jour de la même période de l'année suivante. Si la période considérée est assez longue et si nous disposons d'un grand nombre d'années d'observations, ce découpage n'introduit pas d'erreurs d'échantillonnage notables.

L'avantage de cette méthode par rapport à la précédente réside dans le fait de pouvoir estimer directement les paramètres en fonction de la période considérée. Cette estimation s'avère nécessairement plus adéquate qu'une estimation globale à l'échelle annuelle, particulièrement dans le cas où les différentes saisons sont très marquées, et fournit par conséquent des résultats plus fiables.

Le tableau XIII répertorie les résultats obtenus lors de l'étude des séquences de jours pluvieux dans le cadre de la période novembre-avril à la station Le Raizet. Si nous nous référons au schéma de Polya, nous constatons ainsi qu'une séquence de longueur supérieure à 34 jours, qui se produit en moyenne :

1 fois tous les 46 ans pendant l'année (voir tableau XII) se produit :

1 fois tous les 58 ans au cours de la période novembre-avril.

TABLEAU XIII

CALCUL DES RÉCURRENCES JOURNALIÈRES ET ANNUELLES DES SÉQUENCES DE LONGUEUR SUPÉRIEURE A k JOURS DE PLUIE PENDANT LA PÉRIODE NOVEMBRE-AVRIL

LE RAIZET

NO DE LA STATION 621445
PÉRIODE D'OBSERVATION 1951-1970

SÉQUENCES JOURNALIÈRES

SAISON SECHE Novembre-Avril

SEUIL INFÉRIEUR 0.1 SEUIL SUPÉRIEUR ***** CONTAGION 1.677

CALCUL DES RÉCURRENCES

LONGUEUR DES SÉQUENCES	SÉQUENCES OBSERVÉES	SÉQUENCES POLYA	SÉQUENCES INDEP.	SÉQUENCES D'ORDRE	SÉQUENCES D'ORDRE 2
	JOURS ANNÉES	JOURS ANNÉES	JOURS ANNÉES	JOURS ANNÉES	JOURS ANNÉES
> 1	6.08 0.03	6.08 0.03	4.35 0.02	6.08 0.03	6.08 0.03
> 2	8.73 0.03	7.24 0.03	4.78 0.04	8.17 0.05	8.73 0.05
> 3	12.47 0.07	12.35 0.07	10.57 0.06	10.97 0.06	11.67 0.06
> 4	17.03 0.09	16.50 0.09	14.47 0.09	14.78 0.08	15.26 0.08
> 5	21.50 0.12	21.33 0.12	23.72 0.14	19.88 0.11	19.94 0.11
> 6	27.43 0.13	27.27 0.13	40.11 0.22	26.73 0.13	26.06 0.14
> 7	34.52 0.19	34.62 0.19	62.53 0.35	33.94 0.20	34.05 0.19
> 8	43.73 0.24	43.71 0.24	97.55 0.54	48.34 0.27	44.50 0.25
> 9	57.02 0.31	54.95 0.30	132.13 0.84	65.00 0.36	58.16 0.32
> 10	73.53 0.41	68.86 0.30	237.24 1.31	87.41 0.40	75.00 0.42
> 11	95.73 0.53	86.04 0.47	369.93 2.04	117.54 0.65	99.32 0.55
> 12	120.83 0.67	107.26 0.59	576.03 3.18	158.06 0.87	129.79 0.72
> 13	145.00 0.80	133.46 0.74	899.30 4.96	212.55 1.17	167.62 0.94
> 14	198.59 0.88	165.77 0.91	1401.72 7.73	285.81 1.58	221.66 1.22
> 15	220.65 1.22	205.60 1.13	2184.12 12.05	384.31 2.12	289.66 1.60
> 16	253.73 1.40	254.66 1.41	3401.46 18.77	516.76 2.85	378.52 2.09
> 17	298.53 1.65	315.08 1.74	5293.01 29.20	694.01 3.83	494.63 2.73
> 18	338.33 1.87	387.41 2.13	8226.14 43.39	934.18 5.15	646.33 3.57
> 19	390.30 2.13	480.84 2.61	12737.87 70.40	1255.92 6.93	844.54 4.66
> 20	461.36 2.55	393.22 3.27	19733.10 108.87	1680.31 9.31	1103.47 6.09
> 21	507.50 2.80	731.27 4.03	30377.79 167.60	2367.28 12.52	1441.70 7.93
> 22	843.83 4.67	900.85 4.97	46440.04 256.23	3049.63 16.83	1883.46 10.39
> 23	845.83 4.67	1108.90 6.12	70239.94 387.64	4077.47 22.61	2460.31 13.57
> 24	1268.73 7.00	1364.34 7.53	104689.44 577.60	5503.62 30.36	3213.37 17.73
> 25	1691.67 9.33	1677.30 9.26	132632.07 842.22	7389.34 40.77	4196.23 23.13
> 26	1691.67 9.33	2061.36 11.37	216148.00 1192.54	9913.76 54.71	5478.38 30.23
> 27	1691.67 9.33	2531.68 13.97	274739.00 1626.26	13296.26 73.36	7130.12 39.45
> 28	2537.50 14.00	3107.64 17.13	384397.37 2120.81	17811.87 98.27	9328.27 51.47
> 29	5075.00 28.00	3812.61 21.04	477505.31 2634.31	23030.01 131.48	12163.62 67.11
> 30	5075.00 28.00	4679.03 25.79	565299.56 3118.89	31826.34 175.59	15850.05 87.45
> 31	5075.00 28.00	5729.41 31.61	640047.81 3535.71	42408.40 233.98	20635.58 113.83
> 32	5075.00 28.00	7017.66 38.72	700908.25 3867.08	56337.87 310.83	26835.38 148.06
> 33	5075.00 28.00	8590.50 47.40	745717.49 4114.31	74545.73 411.29	34846.45 192.26
> 34	5075.00 28.00	10507.16 57.98	777993.44 4290.18	98129.12 541.40	45163.02 249.10

La même étude effectuée dans le cadre de la période mai-octobre montre que cette séquence se produit :
1 fois tous les 150 ans au cours de cette période.

Il est facile d'en déduire que d'après Polya :

- 72% des séquences de plus de 34 jours de pluie sont relevées au cours de la période novembre-avril.
- 28% des séquences de plus de 34 jours de pluie sont relevées au cours de la période mai-octobre.

Nous pouvons ainsi calculer la récurrence à l'échelle annuelle R :

$$R = 58 \times \frac{72}{100} \left(\text{ou } R = 150 \times \frac{28}{100} \right) = 42 \text{ ans}$$

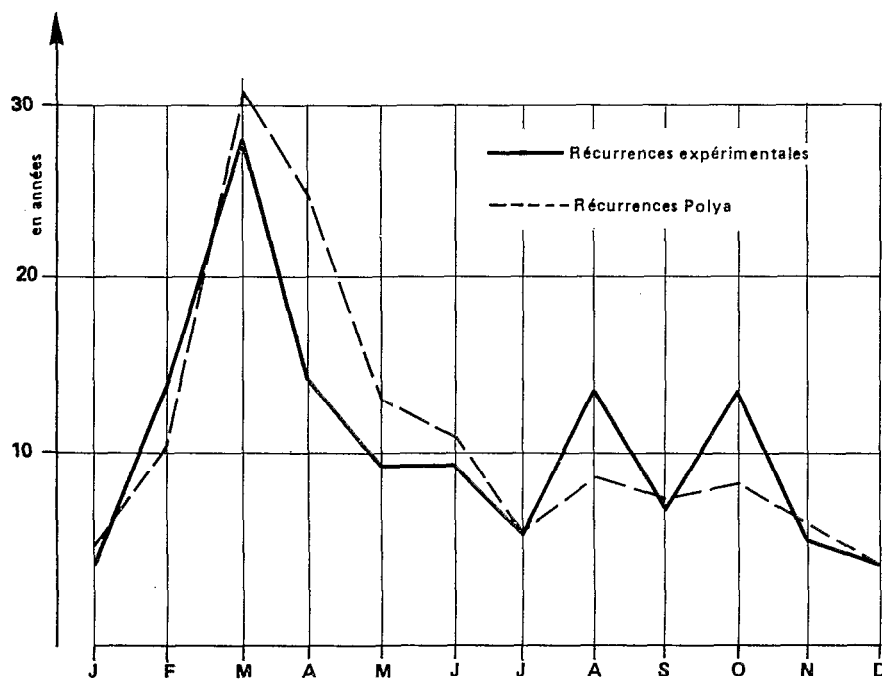


Fig. 5. — Evolution des récurrences des séquences de plus de 15 jours de pluie au cours de l'année

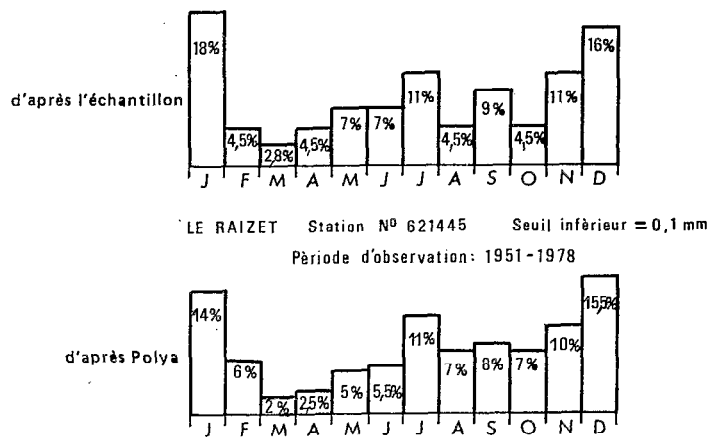


Fig. 6. — Répartition des séquences de plus de 15 jours de pluie au cours de l'année

Ce résultat, calculé à partir des récurrences saisonnières, s'accorde assez bien avec la valeur trouvée lors de l'étude à l'échelle annuelle, qui était de 46 ans.

D'éventuelles différences entre les résultats obtenus par les 2 méthodes évoquées résultent des interférences que peut produire un effet saisonnier trop marqué dans l'estimation des paramètres des schémas de probabilité. Il n'est donc pas très étonnant d'enregistrer des résultats très cohérents pour les stations de la Guadeloupe, dont le régime des précipitations est assez régulier. Il n'en est pas toujours de même en ce qui concerne les stations de la Côte d'Ivoire par exemple, à propos desquelles une étude saisonnière est vivement conseillée, même si la nature du problème ne l'exige pas : cet examen approfondi peut conduire en particulier, en cas de différences notables, à rejeter les résultats obtenus lors de l'étude globale à l'échelle annuelle. Notons que les récurrences expérimentales, dans

lesquelles n'intervient aucun paramètre, ne sauraient être soumises à ces interférences et que, si des différences apparaissent lors du calcul de la récurrence à l'échelle annuelle R, soit par une étude globale, soit à partir des récurrences saisonnières expérimentales, ces différences ne traduisent que les erreurs d'échantillonnage introduites par le découpage de l'année en plusieurs périodes.

Une étude à l'échelle mensuelle (découpage de l'année en 12 périodes mensuelles) peut être également envisagée : la figure 5 décrit l'évolution des récurrences annuelles des séquences de plus de 15 jours de pluie à la station Le Raizet :

- la courbe en trait plein représente les récurrences expérimentales;
- la courbe en pointillé représente les récurrences Polya.

Nous joignons également la figure 6 montrant la répartition de ces séquences au cours de l'année, déterminée d'après les nombres expérimentaux et d'après les nombres théoriques du schéma de Polya.

2.4.4. Définition des pluies « utiles »

Nous avons considéré jusqu'à maintenant l'état pluvieux comme un état de pluie effective (hauteur d'eau journalière supérieure à 0,1 mm). Il est bien sûr possible de choisir d'autres valeurs pour les seuils, en fonction du problème étudié. Ainsi, si nous voulons connaître la répartition et la fréquence de jours de pluie « utiles » au terroir, nous sommes amenés à fixer les valeurs des seuils en fonction de la région considérée : la pluie utile peut, par exemple, être définie par une hauteur d'eau journalière comprise entre 5 et 20 mm, ou encore supérieure au seuil d'évapotranspiration (variable dans le temps et dans l'espace).

Dans ces conditions, l'état pluvieux doit être considéré comme l'occurrence d'une certaine pluie « utile » à l'opérateur. L'état sec ne représente pas alors l'absence de précipitations, mais simplement l'événement contraire de l'état pluvieux précédemment évoqué.

Pour juger des comportements de nos schémas de probabilité, en fonction de la valeur des seuils, nous avons étudié les séquences d'états pluvieux définis successivement par :

- Une hauteur d'eau journalière > 1 mm
seuil inférieur = 1 mm; seuil supérieur = $+\infty$
- Une hauteur d'eau journalière > 5 mm
seuil inférieur = 5 mm; seuil supérieur = $+\infty$
- Une hauteur d'eau journalière > 10 mm
seuil inférieur = 10 mm; seuil supérieur = $+\infty$

Les tableaux suivants, du type X et XI, comparent qualitativement les résultats obtenus pour chaque schéma, le critère de qualité retenu étant une erreur relative comprise entre $\pm 25\%$.

Les résultats détaillés sont reportés sous la forme de graphiques de type figure 4 en annexe C.

L'examen de ces tableaux montre que la qualité des ajustements n'est pas moindre lorsqu'augmente la valeur du seuil inférieur. Remarquons également qu'à partir d'une valeur du seuil inférieur égale à 5 mm, le processus markovien d'ordre 2 présente des résultats sensiblement équivalents à ceux du schéma de Polya.

Seuil inférieur = 1 mm

Nom de la station	Polya	Markov 2	Markov 1	Indép.	Longueur maximale observée
621445 Le Raizet	16	10	8	3	21
620810 Boyvinière	13	8	7	1	23
622130 Duclos	18	18	14	4	27
90055 Bouaké	8	5	4	2	14
90160 Odienné	9	7	5	1	14

Seuil inférieur = 5 mm

Nom de la station	Polya	Markov 2	Markov 1	Indép.	Longueur maximale observée
621445 Le Raizet	8	7	7	2	8
620810 Boyvinière	10	10	6	2	11
622130 Duclos	9	8	7	2	11
90055 Bouaké	5	4	3	1	9
90160 Odienné	7	6	4	2	13

Seuil inférieur = 10 mm

Nom de la station	Polya	Markov 2	Markov 1	Indép.	Longueur maximale observée
621445 Le Raizet	6	6	5	1	6
620810 Boyvinière	6	7	3	1	11
622130 Duclos	6	6	6	1	6
90055 Bouaké	3	5	3	1	9
90160 Odienné	7	6	3	1	9

2.4.5. Conclusion

Les résultats obtenus tout au long de ce paragraphe montrent que, quelles que soient la nature de l'étude envisagée (découpage de l'année, valeur des seuils) et les conditions climatiques rencontrées, le schéma de Polya convient davantage à l'étude des séquences de jours de pluie que les processus markoviens. En ce qui concerne ces derniers, les calculs effectués sous l'hypothèse d'indépendance conduisent systématiquement à surestimer les séquences courtes au détriment des séquences plus longues, ce qui n'est pas étonnant compte tenu que le paramètre de persistance d reste toujours supérieur à 2, quelle que soit la station étudiée : il convient de rejeter ce schéma. D'autre part, si l'ajustement fourni par le processus markovien d'ordre 2 se montre supérieur en qualité à celui d'ordre 1, n'oublions pas qu'il fait intervenir dans le calcul des paramètres une plus grande liaison vis-à-vis de l'échantillon, ce qui limite la confiance qu'on peut lui accorder dans certaines extrapolations.

Au niveau des zones géographiques, quel que soit l'ajustement considéré, les résultats obtenus aux stations de la Guadeloupe sont bien meilleurs que ceux de la Côte d'Ivoire. Nous avons attribué pour une grande part cette différence de qualité à l'existence de fluctuations saisonnières, beaucoup plus nettes en Côte d'Ivoire qu'en Guadeloupe. Là encore, le schéma de Polya semble se comporter de meilleure façon que les processus markoviens, dont la plus grande rigidité les rend moins aptes à l'étude d'un phénomène trop irrégulier. Rappelons que, pour pallier cet inconvénient, il est possible d'effectuer une étude saisonnière (d'ailleurs souhaitable dans la plupart des cas) destinée à uniformiser les données.

Le tableau XIV illustre la plus grande souplesse que manifeste le schéma de Polya à s'adapter à l'échantillon (les probabilités $p(k)$ calculées sont égales au rapport du nombre de séquences de longueur supérieure à $k + 1$ et du nombre de séquences de longueur supérieure à k).

Signalons que le programme conçu pour étudier les séquences de jours de pluie permet également, entre autres utilisations, d'effectuer la même étude en ce qui concerne les séquences de jours secs. Dans ce cas, les résultats sont très satisfaisants et il apparaît très nettement que, au niveau de la qualité, le schéma de Polya précède encore les processus markoviens (voir graphique 7).

Toutes ces considérations nous amènent naturellement à privilégier le schéma de Polya, dans le cadre de notre étude des séquences d'états pluvieux (ou d'états secs) d'une série d'observations chronologiques.

TABLEAU XIV
CALCUL DES PROBABILITÉS CONDITIONNELLES

BOYVINIERE

NO DE LA STATION 420010
PERIODE D'OBSERVATION-1929-1978

Seuil inférieur=0.1 mm

PROBABILITE D UN JOUR DE PLUIE SACHANT QUE LES K JOURS PRECEDENTS SONT EGALEMENT PLUVIEUX

K	OBSERVEE	POLYA	INDEPENDANCE	MARKOV 1	MARKOV 2
1	0.53	0.51	0.42	0.50	0.53
2	0.58	0.60	0.42	0.50	0.62
3	0.57	0.63	0.42	0.50	0.62
4	0.64	0.65	0.42	0.50	0.62
5	0.65	0.66	0.42	0.50	0.62
6	0.63	0.66	0.42	0.50	0.62

LE RAIZET

NO DE LA STATION 421443
PERIODE D'OBSERVATION-1951-1978

Seuil inférieur =0.1 mm

PROBABILITE D UN JOUR DE PLUIE SACHANT QUE LES K JOURS PRECEDENTS SONT EGALEMENT PLUVIEUX

K	OBSERVEE	POLYA	INDEPENDANCE	MARKOV 1	MARKOV 2
1	0.71	0.70	0.60	0.75	0.71
2	0.73	0.75	0.60	0.75	0.77
3	0.74	0.76	0.60	0.75	0.77
4	0.78	0.77	0.60	0.75	0.77
5	0.77	0.70	0.60	0.75	0.77
6	0.77	0.70	0.60	0.75	0.77

TABLEAU XV
CALCUL DES PROBABILITÉS CONDITIONNELLES

ODIENNE

NO DE LA STATION 70160
PERIODE D'OBSERVATION=1921-1973

Seuil inférieur=0.1 mm

PROBABILITE D UN JOUR DE PLUIE SACHANT QUE LES K JOURS PRECEDENTS SONT EGALLEMENT PLUVIEUX

K	OBSERVEE	POLYA	INDEPENDANCE	MARKOV 1	MARKOV 2
1	0.44	0.40	0.30	0.51	0.44
2	0.52	0.57	0.30	0.51	0.50
3	0.56	0.62	0.30	0.51	0.50
4	0.50	0.64	0.30	0.51	0.50
5	0.64	0.66	0.30	0.51	0.50
6	0.66	0.67	0.30	0.51	0.50

BOUAKE

NO DE LA STATION 70055
PERIODE D'OBSERVATION=1921-1973

Seuil inférieur=0.1 mm

PROBABILITE D UN JOUR DE PLUIE SACHANT QUE LES K JOURS PRECEDENTS SONT EGALLEMENT PLUVIEUX

K	OBSERVEE	POLYA	INDEPENDANCE	MARKOV 1	MARKOV 2
1	0.40	0.35	0.27	0.45	0.40
2	0.44	0.52	0.27	0.45	0.51
3	0.40	0.57	0.27	0.45	0.51
4	0.56	0.57	0.27	0.45	0.51
5	0.63	0.61	0.27	0.45	0.51
6	0.67	0.62	0.27	0.45	0.51

DUCLOS

NO DE LA STATION 622130
PERIODE D'OBSERVATION=1954-1974

Seuil inférieur=0.1 mm

PROBABILITE D UN JOUR DE PLUIE SACHANT QUE LES K JOURS PRECEDENTS SONT EGALLEMENT PLUVIEUX

K	OBSERVEE	POLYA	INDEPENDANCE	MARKOV 1	MARKOV 2
1	0.72	0.71	0.69	0.76	0.72
2	0.76	0.76	0.69	0.76	0.70
3	0.78	0.77	0.69	0.76	0.70
4	0.75	0.78	0.69	0.76	0.70
5	0.77	0.78	0.69	0.76	0.70
6	0.70	0.79	0.69	0.76	0.70

BOUAKE

Station no 90055

Seuil inférieur = 0.1 mm

Période d'observation : 1921-1973

SAISON SÈCHE : Novembre-Avril

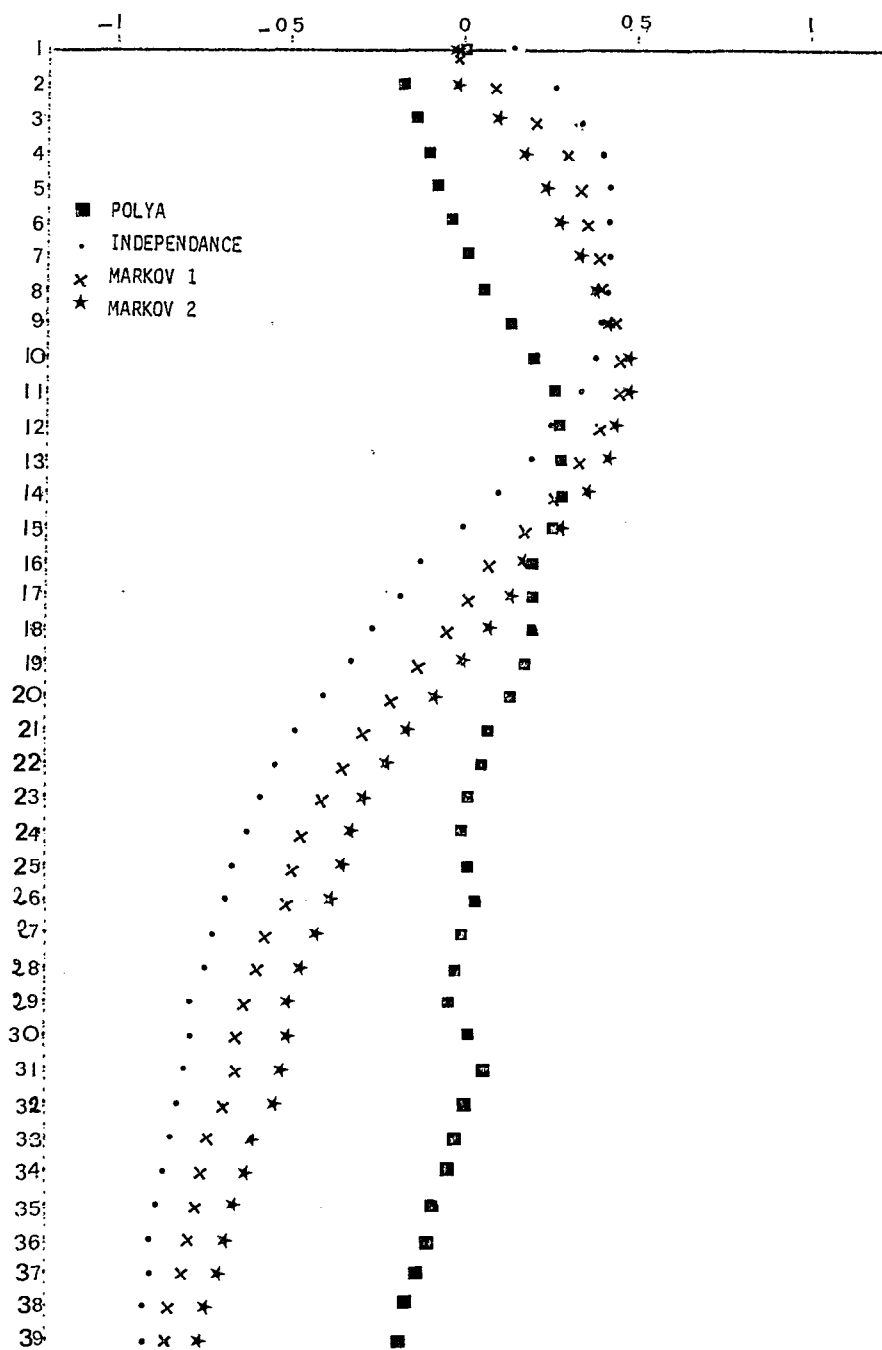


Fig. 7. — Comparaison des ajustements fournis par les schémas de Polya, Indépendance, Markov 1 et 2 relativement aux séquences de jours secs pendant la période novembre-avril à Bouake

ANNEXE A

1. CALCUL DES PARAMÈTRES DU SCHÉMA DE POLYA PAR LA MÉTHODE DES MOMENTS

Les calculs suivants découlent du développement en série entière de :

$$(1+x)^a = 1 + ax + \dots + \frac{a(a-1)\dots(a-n+1)}{n!} x^n + \dots$$

Moment d'ordre 0

$$\begin{aligned} I_0 &= \sum_0^{+\infty} P_i = P_0 + \sum_1^{+\infty} \frac{1}{(1+d)^{m/d}} = \frac{1}{(1+d)^{m/d}} \left(1 + \sum_1^{\infty} \frac{m(m+d)\dots(m+(i-1)d)}{i!(1+d)^i} \right) \\ &= \frac{1}{(1+d)^{m/d}} \left(1 + \sum_1^{\infty} \frac{m/d(m/d+1)\dots(m/d+i-1)}{i!} \left(\frac{d}{1+d} \right)^i \right) \\ &= \frac{1}{(1+d)^{m/d}} \left(1 + \sum_1^{\infty} \frac{(-m/d)(-m/d-1)\dots(-m/d-i+1)}{i!} \left(-\frac{d}{1+d} \right)^i \right) \\ &= \frac{1}{(1+d)^{m/d}} \cdot \left(1 - \frac{d}{1+d} \right)^{-m/d} = \frac{1}{(1+d)^{m/d}} \cdot \frac{1}{(1+d)^{-m/d}} = 1 \end{aligned}$$

Moment d'ordre 1 (espérance mathématique)

$$\begin{aligned} I_1 &= \sum_0^{+\infty} iP_i = \sum_1^{+\infty} \frac{i m (m+d)\dots(m+(i-1)d)}{i!(1+d)^{m/d+i}} \\ &= \frac{1}{(1+d)^{m/d+1}} \left(\sum_1^{\infty} \frac{m(m+d)\dots(m+(i-1)d)}{(i-1)!(1+d)^{i-1}} \right) \\ &= \frac{m}{(1+d)^{m/d+1}} \left(1 + \sum_2^{\infty} \frac{(m+d)\dots(m+d+(i-2)d)}{(i-1)!(1+d)^{i-1}} \right) \\ &= \frac{m}{(1+d)^{m/d+1}} \left(1 + \sum_1^{\infty} \frac{(m+d)\dots(m+d+(i-1)d)}{i!(1+d)^i} \right) \\ &= \frac{m}{(1+d)^{m/d+1}} \cdot \frac{1}{(1+d)^{-m/d-1}} = m \end{aligned}$$

De même, sans entrer dans le détail des calculs, on trouve :

$$I_2 = \text{variance} = m(1+d) \Rightarrow d = \frac{\sigma^2}{m} - 1$$

$$I_3 = \sum_0^{+\infty} (i-m)^3 P_i = m(1+3d+2d^2)$$

$$I_4 = \sum_0^{+\infty} (i-m)^4 P_i = m(3m+3md^2+6md+6d^3+12d^2+7d+1)$$

ce qui permet de calculer les coefficients B_1 et B_2 de Pearson, mesurant respectivement la dissymétrie et l'aplatissement de la distribution :

$$B_1 = \frac{I_3^2}{I_2^3} \quad \text{et} \quad B_2 = \frac{I_4}{I_2^2}$$

2. VALEUR MODALE

La valeur modale correspond à la fréquence la plus grande, c'est-à-dire ici à l'entier positif dont la probabilité P_i est la plus grande.

Etudions la différence $P_i - P_{i-1}$ où $i \in |1...|$

$$P_i - P_{i-1} = \left(\frac{m + (i-1)d}{i(1+d)} - 1 \right) \cdot P_{i-1}$$

Le signe de la différence $P_i - P_{i-1}$ est donc celui de l'expression $f(i) = \frac{m + (i-1)d}{i(1+d)} - 1$, car P_{i-1} est toujours positif.

Considérons f comme une fonction du réel i , variant de 1 à $+\infty$: $f'(i) = \frac{d-m}{i^2(1+d)}$ garde un signe constant sur l'intervalle $]1, +\infty[$.

• si $d - m \geq 0$ ou $d \geq m$

$$f \text{ croît de } f(1) = \frac{m-d-1}{1+d} \text{ à } \lim_{+\infty} f = -\frac{1}{1+d}$$

ces 2 valeurs étant strictement négatives, le signe de $f(i)$ reste toujours négatif et $\forall i \in |1...|$.

$$P_{i-1} \geq P_i$$

La valeur modale est donc 0.

• si $d - m \leq 0$ ou $d \leq m$

$$f \text{ décroît de } f(1) = \frac{m-d-1}{(1+d)} \text{ à } \lim_{+\infty} f = -\frac{1}{1+d}$$

Nous devons alors considérer 2 cas :

$$m - d - 1 < 0 \text{ ou } m < 1 + d$$

alors $f(i)$ garde un signe négatif et la valeur modale est 0.

$$m - d - 1 \geq 0 \text{ ou } m \geq 1 + d$$

alors il existe un réel k tel que $k \geq 1$ et $f(k) = 0$.

Ce réel est égal à $m - d$ (solution de l'équation $f(k) = 0$).

Dans ce cas, l'entier $j = E(m - d)$ représente la valeur modale car :

$$\forall i \leq j \quad P_i \geq P_{i-1}$$

$$\forall i > j \quad P_i \leq P_{i-1}$$

En conclusion :

Valeur modale = sup (0, partie entière de $m - d$).

3. COMPORTEMENT AUX LIMITES ($d = -1$ ou $d = 0$)

a) $d = -1$ ou $\sigma^2 = 0$

$$\bullet \lim_{-1^+} P_0 = \lim_{-1^+} e^{-m/d \operatorname{Log}(1+d)} \quad \text{or } \operatorname{Log}(1+d) \rightarrow -\infty$$

$$\text{et si } m \neq 0 \quad -m/d \operatorname{Log}(1+d) \rightarrow +\infty$$

$$m = 0 \quad -m/d \operatorname{Log}(1+d) \rightarrow 0$$

$$\text{donc } \lim_{-1^+} P_0 = \begin{cases} 1 & \text{si } m = 0 \\ 0 & \text{si } m \neq 0 \end{cases}$$

$$\bullet \lim_{-1^+} P_i = \lim_{-1^+} \frac{m(m+d) \dots (m+(i-1)d)}{i! (1+d)^{m/d+1}} i$$

si $i \neq m$ le numérateur tend vers une quantité finie

$$\text{et } \lim_{-1^+} (1+d)^{m/d+1} = \lim_{-1^+} e^{(m/d+1) \operatorname{Log}(1+d)}$$

or $\operatorname{Log}(1+d) \rightarrow -\infty$ et $m/d + 1 \rightarrow i - m (\neq 0)$.

$$\text{donc } \lim_{-1^+} (1+d)^{m/d+1} = \infty \quad \text{et } \lim_{-1^+} P_i = 0$$

si $i = m$

$$\lim_{-1^+} (1+d)^{m/d+1} = \lim_{-1^+} e^{(m/d+1) \operatorname{Log}(1+d)}$$

$$= \lim_{-1^+} e^{m/d (1+d) \operatorname{Log}(1+d)}$$

$$= 1 \quad \text{car } \lim_{-1^+} (1+d) \operatorname{Log}(1+d) = 0$$

$$\text{Donc } \lim_{-1^+} P_m = \frac{m(m-1) \dots (m-(m-1))}{m!} = \frac{m!}{m!} = 1$$

En conclusion :

Quel que soit l'entier $i \in \{0, \dots\}$

$$\lim_{-1^+} P_i = \begin{cases} 1 & \text{si } i = m \\ 0 & \text{si } i \neq m \end{cases}$$

On retrouve bien le cas où si $d = -1$, le résultat de l'épreuve est toujours le même nombre d'occurrences m de l'événement.

b) $d = 0$ ie $\sigma^2 = m$

$$\bullet \lim_{\underset{o}{d} \rightarrow 0} P_o = \lim_{\underset{o}{d} \rightarrow 0} \frac{1}{(1+d)^{m/d}} = \lim_{\underset{o}{d} \rightarrow 0} e^{-m/d \operatorname{Log}(1+d)}$$

or si $d \rightarrow 0$ $\operatorname{Log}(1+d) \sim d$

$$\text{donc } \lim_{\underset{o}{d} \rightarrow 0} P_o = e^{-m}$$

quel que soit l'entier $i \in |1...|$

$$\begin{aligned} \bullet \lim_{\underset{o}{d} \rightarrow 0} P_i &= \lim_{\underset{o}{d} \rightarrow 0} \frac{m(m+d) \dots (m+(i-1)d)}{i!(1+d)^{m/d+i}} \\ &= \lim_{\underset{o}{d} \rightarrow 0} \frac{m(m+d) \dots (m+(i-1)d)}{i!(1+d)^{m/d} \cdot (1+d)^i} = \frac{m^i}{i!} e^{-m} \end{aligned}$$

En conclusion :

Quel que soit l'entier $i \in |0...|$

$$\lim_{\underset{o}{d} \rightarrow 0} P_i = \frac{m^i}{i!} e^{-m}$$

On retrouve les probabilités de la Loi de Poisson.

ANNEXE B

APPLICATION DU SCHÉMA DE POLYA A L'ETUDE DE LA DISTRIBUTION DES HAUTEURS DE PLUIE ANNUELLES

Nous proposons ici de compléter le programme des 9 lois (POH 143) par l'adjonction d'une 10^e loi, la loi de Polya.

Dans le contexte de l'étude des hauteurs de pluie annuelles (mensuelles, journalières), on admet alors qu'une chute de h mm en un an (mois, jour) est équivalente à h occurrences de 1 mm en cette même année (mois, journée).

Les résultats obtenus par la loi de Polya sont assez variables, en fonction des stations étudiées. Toutefois, la méthode proposée permet d'envisager des prolongements intéressants à l'échelle journalière ou mensuelle par exemple, pour lesquelles la loi de Polya s'avère plus adéquate.

Les données représentant les hauteurs d'eau mensuelles (de 1957 à 1970) observées à la station Le Raizet, le tableau XVI établit la comparaison des ajustements fournis par les 10 lois du programme, le meilleur ajustement correspondant à celui dont la valeur du test (proposé par M. Brunet-Moret) est minimale.

TABLEAU XVI

ANALYSE STATISTIQUE DES PLUIES ANNUELLES

STATION NO 621445 LE RAIZET GUADELOUPE

AJUSTEMENTS AUX 10 DISTRIBUTIONS RETENUES
 PAR LA METHODE DU MAXIMUM DE VRAISEMBLANCE
 LES PARAMETRES D ECHELLE ETANT POSITIFS A PRIORI

VALEURS CALCULEES

VALEUR LIMITE INFERIEURE CHOISIE POUR LES BORNES INFERIEURES 0.0

GAUS DISTRIBUTION DE LAPLACE-GAUSS DITE NORMALE
 GUMB DISTRIBUTION DE GUMBEL, DOUBLEMENT EXPONENTIELLE
 GALT DISTRIBUTION DE GALTON, GAUSSO-LOGARITHMIQUE
 PEAR DISTRIBUTION DE PEARSON III, GAMMA INCOMPLETE, EN X
 PEAV DISTRIBUTION DE PEARSON V, GAMMA INCOMPLETE, EN 1/X
 GOOD DISTRIBUTION DE GOODRICH, EXPO GENERALISEE, EN X**A
 FREC DISTRIBUTION DE FRECHET, EXPO GENERALISEE, EN 1/X**A
 LGAM DISTRIBUTION WRC-USA, LOG-GAMMA DE 1ERE ESPECE
 FUIT DISTRIBUTION DE LA LOI DES FUITES
 POLY DISTRIBUTION DE POLYA

DISTRIBUTION PARAMETRES	GAUS	GUMB	GALT	PEAR	PEAV	GOOD	FREC	LGAM	FUIT	POLY
D ECHELLE	967.648	707.957	1211.408	671.387	1771.103	1562.613	838.404	0.028	276.105	623.397
DE POSITION	1508.470	1079.078	0.0	59.699	0.0	98.672	0.0	0.028	0.0	1508.470
DE FORME PREMIER	*****	*****	0.769	2.158	1.935	0.668	-0.784	223.744	5.463	*****
DE FORME SECOND	*****	*****	*****	*****	*****	*****	*****	0.048	*****	*****
BORNE INFERIEURE	*****	*****	0.0	59.699	0.0	98.672	0.0	0.028	0.0	*****
BORNE SUPERIEURE	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
MOYENNE	1508.470	1487.721	1628.769	1508.470	1894.779	1509.693	3551.923	1577.107	1508.469	1508.470
MEDIANE	1508.470	1338.553	1211.408	1291.885	1097.622	1321.878	1117.589	1192.271	1368.248	1300.671
MODE	1508.470	1079.078	670.120	837.083	603.498	846.457	532.427	710.874	1069.325	885.073
VARIANCE	936343.437	824444.750	*****	972685.750	*****	921590.750	*****	*****	832992.937	941884.437
COEF VARIATION	0.641	0.610	0.899	0.654	*****	0.636	*****	0.870	0.605	0.643
COEF ASYMETRIE	0.0	1.139	3.422	1.361	*****	1.076	*****	3.791	0.908	1.653
COEF APLATISSEMENT	0.0	2.400	26.299	2.781	*****	1.403	*****	36.238	1.098	5.480
VALEUR DE TEST	58.261	9.154	17.649	2.092	61.772	3.782	82.458	19.612	9.589	2.113
FREQ AU DEPASSEMENT	0.000	0.097	0.004	0.948	0.000	0.681	0.0	0.002	0.082	0.946

PROBABILITE RECURRENCE

0.0100	.1000E+03	-742.597	-2.099	202.247	184.202	271.749	170.953	253.116	246.572	57.895	156.00
0.0200	.5000E+02	-478.814	113.387	249.447	235.521	309.761	213.917	287.659	293.965	131.710	215.00
0.0300	.3333E+02	-311.465	190.852	284.954	275.870	337.741	250.292	313.432	329.096	189.315	260.00
0.0400	.2500E+02	-185.571	251.454	314.959	310.669	361.164	283.058	335.196	358.509	238.526	298.00
0.0500	.2000E+02	-83.164	302.316	341.680	342.038	381.899	313.448	354.624	384.442	282.091	332.00
0.1000	.1000E+02	268.385	488.619	451.884	472.674	467.013	446.096	435.908	490.056	455.920	472.00
0.2000	.5000E+01	694.079	742.172	633.929	685.224	609.068	672.276	577.263	661.103	716.880	693.00
0.5000	.2000E+01	1508.470	1338.553	1211.408	1291.885	1097.654	1321.878	1117.589	1192.266	1368.051	1306.00
0.8000	.5000E+01	2322.859	2140.970	2314.937	2209.302	2268.593	2246.203	2718.378	2199.273	2217.944	2211.00
0.9000	.1000E+02	2748.552	2672.239	3247.527	2827.763	3553.991	2826.793	4896.590	3056.437	2740.733	2816.00
0.9500	.2000E+02	3100.100	3181.846	4294.949	3415.036	5375.133	3351.220	8611.051	4030.353	3213.151	3394.00
0.9600	.2500E+02	3202.504	3343.500	4659.336	3599.484	6112.980	3511.156	10299.703	4371.812	3357.888	3576.00
0.9700	.3333E+02	3328.398	3550.814	5149.937	3834.652	7195.496	3712.024	12958.652	4833.797	3539.676	3812.00
0.9800	.5000E+02	3495.749	3841.481	5883.020	4162.316	9022.055	3986.087	17881.098	5530.125	3788.674	4144.00
0.9900	.1000E+03	3759.523	4335.781	7255.918	4712.328	13182.629	4433.730	30916.789	6848.750	4198.512	4731.00

ANNEXE C

Comparaison des ajustements obtenus à la station :

620810 BOYVINIÈRE (GUADELOUPE)

Pour les valeurs du seuil inférieur suivantes :

- 1 mm
- 5 mm
- 10 mm

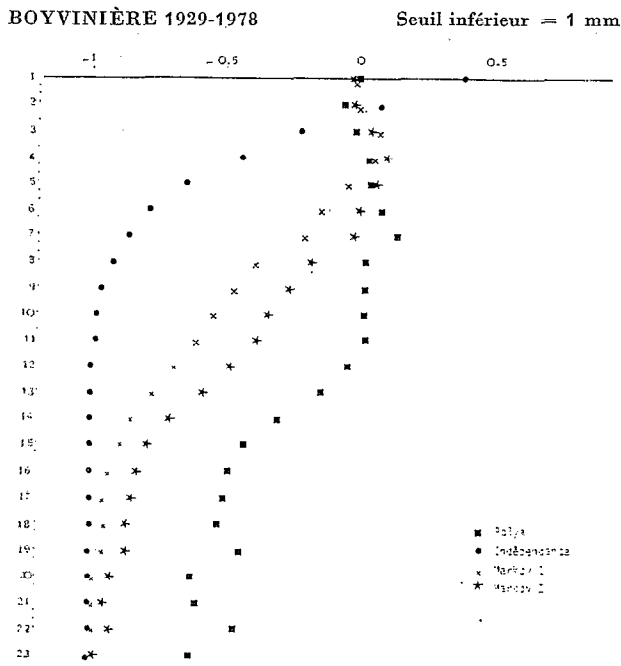


Fig. 8

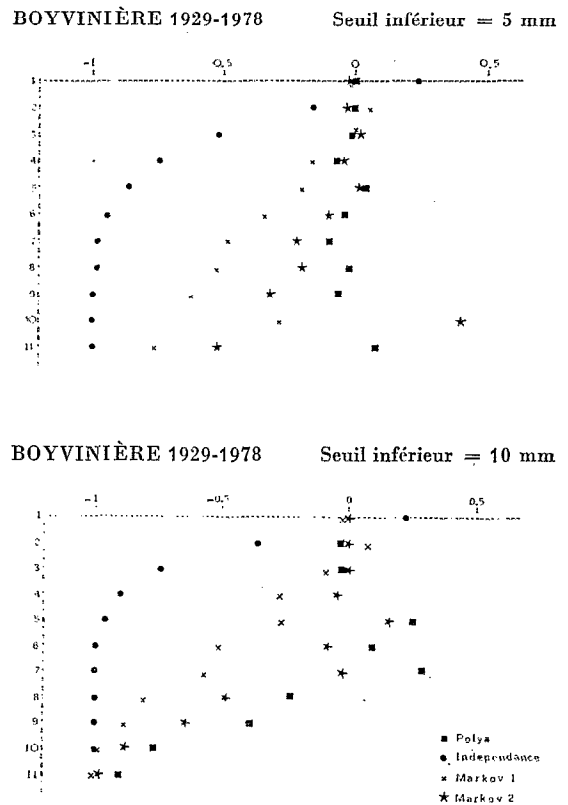


Fig. 9

BIBLIOGRAPHIE

GRISOLLET (M.), GUILMET (B.), ARLERY (R.) – 1962 – *Climatologie, Méthodes et Pratiques*.

KENDALL (M.) et STUART (A.) – *The advanced theory of statistics, volume 3*.

BRUNET-MORET (Y.) – 1977 – Test d'homogénéité, *Cahiers Hydrologie, Vol. XIV, n° 2, 1977*.

BRUNET-MORET (Y.) et ROCHE (M.) – 1975 – Persistance dans les suites chronologiques de précipitations annuelles. *Cahiers Hydrologie, Vol. XII, n° 3, 1975*.

MASSON (J. M.) – 1977 – Persistance des états pluvieux en fonction de leur durée. *Cahiers Hydrologie, Vol. XIV, n° 2, 1977*.

MONIOD (F.) – 1977 – Les précipitations en un et plusieurs jours dans le massif forestier landais. *Cahiers Hydrologie, Vol. XIV, n° 2, 1977*.

DUBREUIL (P.) – 1975 – Etude des précipitations appliquée à la mise en valeur hydro-agricole du bassin du Jaguaribe. Service Hydrologique de l'ORSTOM.

AIVAZIAN (S.) – 1970 – Etude statistique des dépendances.