

SOCIÉTÉ HYDROTECHNIQUE DE FRANCE -- PRIX HENRI MILON 1971

Contribution à l'étude probabiliste des phénomènes pluvieux

Application aux bassins de la Charente et de la Seudre

par R.-E. Quélenec

Ingénieur E.N.S.A.I.S., ingénieur S.S.H. - E.N.S.E.E.I.H.T.
Docteur Ingénieur, Ex Chef du Service Informatique et Statistique
de l'Agence Financière de Bassin « Adour-Garonne », Toulouse

Cette étude présentera un résumé de la thèse de docteur-ingénieur préparée sous la direction scientifique principale de M. P. Ettinger, professeur de statistique à l'Université Paul-Sabatier de Toulouse, et soutenue devant un jury de cette même université le 25 février 1971.

Elle traitera de l'une des caractéristiques essentielles de l'hydrologie d'un bassin : le régime pluviométrique.

BUT DE L'ÉTUDE :

Les données pluviométriques, relativement abondantes étant disponibles sur ces deux bassins contigus en de nombreuses stations de mesure, il était tentant d'essayer de condenser cette information brute, après l'avoir critiquée, de façon à obtenir un minimum de paramètres fonamen-

taux représentatifs, dans l'espace et dans le temps, du régime pluviométrique de ces bassins.

Les paramètres ainsi obtenus après une analyse statistique des séries chronologiques journalières observées doivent permettre de construire les lois de probabilité des phénomènes pluvieux les plus importants avec une bonne approximation et peu de frais.

Ce rapport s'efforcera donc de décrire, à partir de méthodes d'analyse générales facilement transposables, les lois statistiques auxquelles semblent obéir les phénomènes pluvieux et les liaisons stochastiques existant entre eux.

Les modèles proposés ont donné lieu, pour leur application, à une très large contribution des méthodes de calcul sur ordinateur (IBM 7044, Faculté des Sciences de Toulouse).

Situation géographique des bassins étudiés

Les bassins de la Charente et de la Seudre sont contigus et situés dans des régions de faible altitude comprises entre le massif du Limousin et l'océan Atlantique (fig. 1).

Leur superficie, est, pour la Charente de 10 000 km² et pour la Seudre de 520 km².

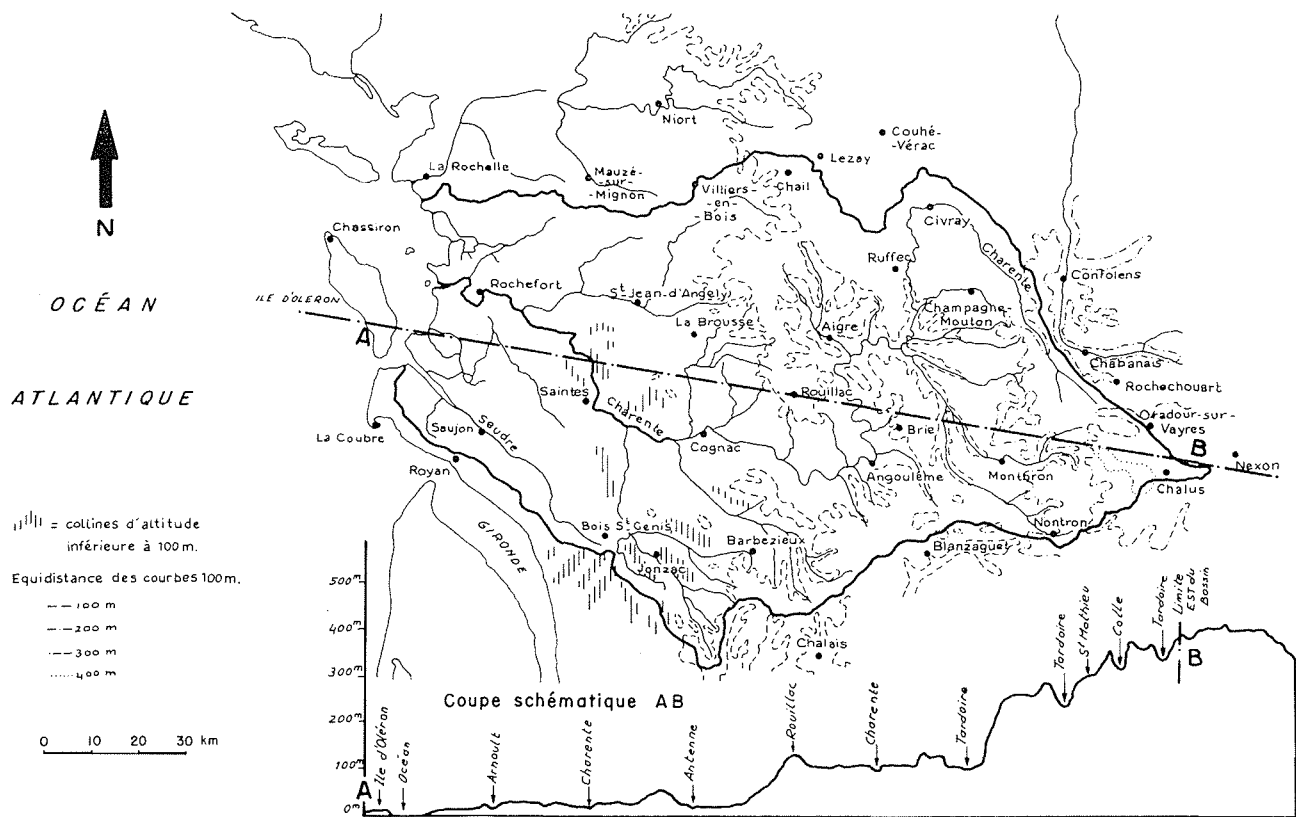
Ils tiennent, réunis, dans un rectangle de 180 km × 110 km et sont très ouverts à l'influence océanique en raison de la douceur du relief qui s'élève, par degrés, de l'Océan aux contreforts du Limousin (300 m à 500 m entre Montbron et Châlus).

Critique de l'information pluviométrique

Elle sera conduite à l'échelon des valeurs annuelles observées, et portera essentiellement sur le contrôle de l'homogénéité de ces séries.

Ce problème est, en résumé, le suivant :

Soit une station pluviométrique A ayant fonctionné N années dans un ensemble de X autres stations. Les N valeurs annuelles de A font-elles partie d'une « même population » de taille N ? Les « distances » entre A (caractérisé par son échantillon de N valeurs observées) et les X autres stations sont-elles demeurées fixes dans le temps ?



1/ Bassin de la Charente et de la Seudre. Courbes hypsométriques.

Une réponse positive à ces questions caractérisera pour nous l'homogénéité de la station étudiée.

Deux hypothèses de départ ont conditionné le choix de certaines méthodes étudiées :

- normalité des observations de pluies annuelles;
- effet multiplicatif constant sur les valeurs annuelles quand intervient une modification dans les conditions de mesure de la pluie à une station.

De nombreuses méthodes numériques ou graphiques, classiques ou inédites, ont été étudiées dans ce chapitre de la Thèse.

- méthode des écarts cumulés (par rapport à la moyenne ou à la droite de régression);
- méthode des doubles cumuls (ou « doubles-masses »);
- test de Student portant sur l'égalité de deux moyennes observées;
- test de « range multiple » de Duncan;
- méthode des « moindres-carrés » et tests de Fisher appliqués à l'ajustement numérique et automatique de p droites au nuage observé sur les graphes des « doubles-masses »;
- tests de Student appliqués aux coefficients caractéristiques des droites de régression.

Nous ne présenterons ici qu'un résumé des observations qui nous apparaissent caractériser les méthodes qui nous

semblent les mieux adaptées, par leur simplicité de mise en œuvre et l'intérêt des conclusions rapides qu'elles permettent de tirer, au problème du contrôle de l'homogénéité des séries pluviométriques annuelles.

Méthode des écarts cumulés.

PAR RAPPORT A LA MOYENNE :

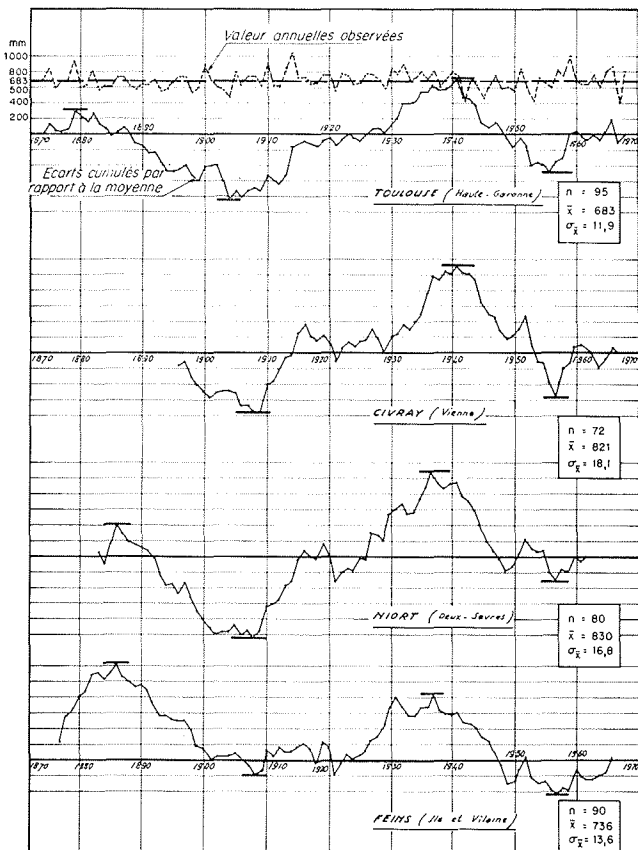
Elle fait intervenir la statistique S_j des écarts à la moyenne cumulés sur la période étudiée :

$$S_j = \sum_{i=1}^j (X_i - \bar{X}) \quad \text{avec} \quad \bar{X} = \sum_{i=1}^N X_i / N$$

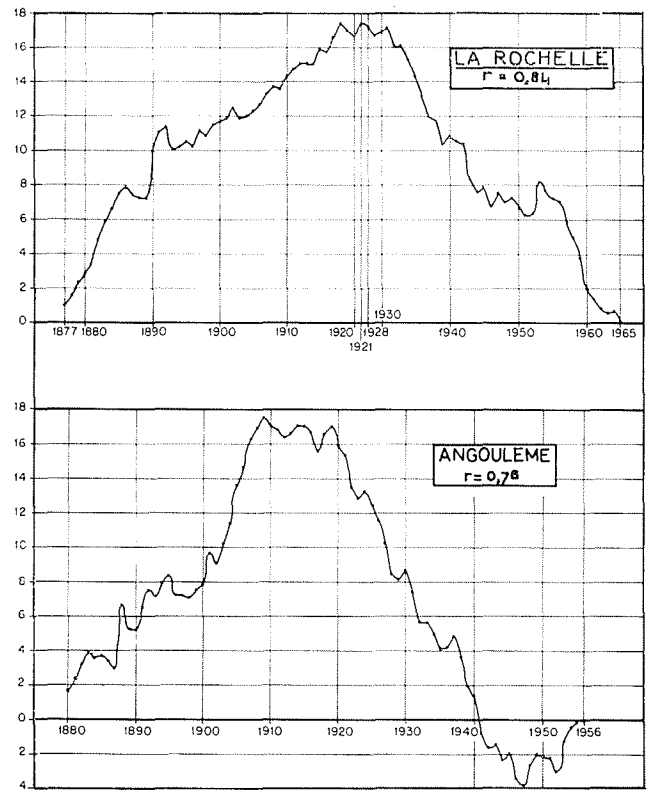
Représentée sur les axes hauteurs-temps, cette statistique définit des courbes des « écarts cumulés » qui nous semblent caractéristiques de l'ensemble de l'évolution de la pluviométrie annuelle d'une région sur une longue période (fig. 2).

Elles indiquent par leurs parties croissantes et décroissantes trois « tendances pluviométriques caractéristiques » des régions soumises au régime océanique d'une :

- tendance déficitaire. . . de 1881-86 à 1906-09 (20 à 27 ans);
- tendance excédentaire. . . de 1906-09 à 1937-41 (28 à 35 ans);
- tendance déficitaire. . . de 1937-41 à 1956-58 (15 à 21 ans).



2/ Graphes des écarts cumulés par rapport à la moyenne.



3/ Lignes des écarts cumulés par rapport à la droite de régression. Station de référence : Niort (*).

Tournée vers le passé, cette simple constatation des faits obtenue au moyen d'une légère transformation des données, doit aider le climatologue et l'hydrologue à interpréter les caractéristiques des séries chronologiques de pluies annuelles.

Cette méthode permet déjà de dégrossir l'étude critique des données. Elle montre également les limites des méthodes de contrôle d'homogénéité qui ne portent que sur la comparaison, à l'aide de tests d'hypothèse, des caractéristiques des séries contiguës dans le temps et appartenant à la même station.

PAR RAPPORT A LA DROITE DE RÉGRESSION :

Elle utilise les propriétés de la régression entre deux variables et complète la méthode précédente.

La référence sera ici, pour chaque valeur X_i , observée à la station A à contrôler, la valeur X'_{yi} de la droite de régression correspondant à la mesure y_i observée à la même date à une station « témoin » homogène T.

La statistique est donc ici :

$$S'j = \sum_{i=1}^j (X_i - X'_{yi}) \quad \text{avec} \quad X'_{yk} \neq X'_{yl}$$

Si la représentation de cette statistique s'enroule apériodiquement autour de l'axe temps, on pourra considérer la station A comme homogène. (Le témoin peut encore être

affiné en le prenant égal à la moyenne de plusieurs stations réputées homogènes d'une région.)

Si cette méthode permet assez facilement de déceler des hétérogénéités, elle ne permet pas toujours, à elle seule, de déterminer avec précision les termes des sous-échantillons homogènes (fig. 3).

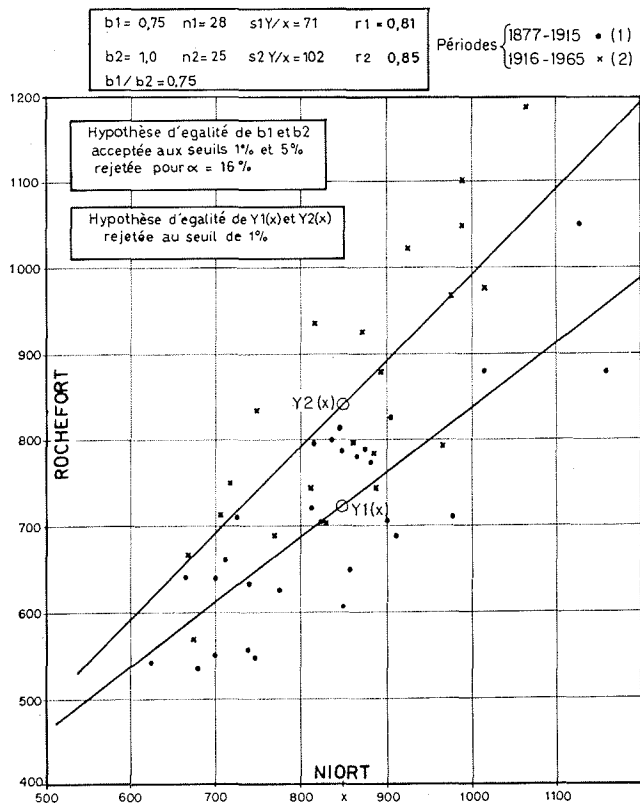
Méthodes des doubles-masses.

Très connue et utilisée, cette méthode consiste à porter sur un graphique les totaux pluviométriques annuels cumulés d'une station A en regard des totaux correspondants d'une station voisine B. Elle représente donc la régression entre les statistiques :

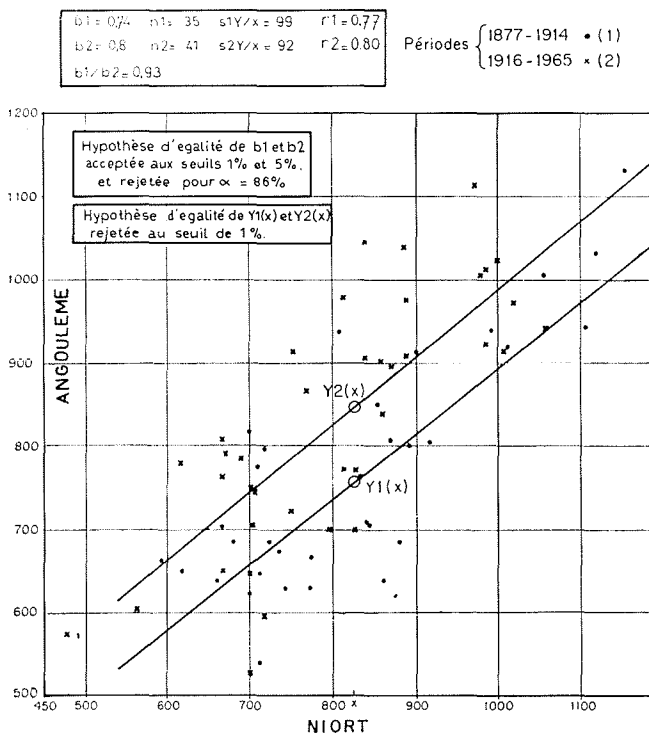
$$\sum_{i=1}^k y_i \quad \text{et} \quad \sum_{i=1}^k x_i - k \in (0, N)$$

Si B est homogène et que A l'est également, ce graphe doit être une droite (constance des coefficients de pente a et b de la régression entre les valeurs de A et B sur la période étudiée).

Cette méthode s'applique d'autant mieux que le coefficient de corrélation entre les valeurs de A et B est grand et que b est différent de l'erreur moyenne de mesure.



4/ Régressions Niort-Rochefort.



4a/ Régressions Niort-Angoulême.

Son inconvénient majeur réside dans le caractère subjectif du choix du tracé des droites quand il y a hétérogénéité (termes des périodes des sous-échantillons homogènes).

Nous n'avons cependant pu obtenir plus de précision par la méthode d'ajustement automatique de p droites au nuage observé sur ces graphes, méthode qui utilise les propriétés des « moindres-carrés » couplées avec un test d'hypothèse de Fisher.

Cet arbitraire peut être considérablement réduit quand on opère sur plusieurs couples de valeurs bien corrélées et que l'on choisit une échelle de représentation qui réduise l'importance de l'erreur de mesure.

Méthode de test des caractéristiques de la régression linéaire entre deux stations.

Considérons que nous voulions vérifier, par un test statistique, la présence de deux sous-échantillons distincts Y_1 et Y_2 dans la série des valeurs annuelles relevée à la station A, cette hétérogénéité ayant, par exemple, été remarquée grâce à une des méthodes précédentes (fig. 4 et 4 a).

Couplons la station à tester Z, avec une station témoin homogène T. L'estimation du coefficient de pente b caractéristique de la régression linéaire existant entre les valeurs correspondantes de A et T, s'écrit pour chacun des deux échantillons où :

$$y'_x = a + bx, \quad \hat{b} = r S_y / S_x$$

L'hypothèse H_0 à soumettre au test sera donc :

$$E(b_1) = E(b_2)$$

— Si les effectifs des sous-échantillons sont grands : n_1 et $n_2 > 30$ ou 50 alors :

$$R = \frac{b_1 - b_2}{S(b_1 - b_2)}$$

suit une loi $N(0,1)$.

Pour accepter H_0 , il faudra que R calculé $< R_\alpha$ avec $R_\alpha = 1,96$ ou $2,58$, pour des seuils $\alpha = 5\%$ ou 1% .

— Si n_1 et n_2 sont petits (< 30), alors la statistique :

$$t = \frac{b_1 - b_2}{S} \left[\frac{(n_1 - 1)(n_2 - 1) S^2_{1x} S^2_{2x}}{(n_2 - 1) S^2_{2x} + (n_1 - 1) S^2_{1x}} \right]^{1/2}$$

suit une loi de Student-Fisher à $n_1 + n_2 - 4$ d.d.l.

avec :

$$S^2 = [(n_1 - 1)(1 - r_1^2) S^2_{y1} + (n_2 - 1)(1 - r_2^2) S^2_{y2}] / (n_1 + n_2 - 4)$$

Pour accepter H_0 , il faudra que t' calculé $< t_\alpha$ table, au seuil α .

Ces tests devraient donc permettre de confirmer statistiquement les résultats des méthodes précédentes. Il faut cependant les manier avec précaution et ne pas se borner à les utiliser avec les seuils « scolaires » de 1 ou 5 % qui ne sont pas adaptés aux caractéristiques de tous les échantillons. La « puissance » du test t augmente en effet avec α , $(n_1 + n_2)$ et la valeur du coefficient multiplicatif correcteur c .

Si les méthodes de critique de l'information pluviométrique sont nombreuses, elles sont toutes imparfaites et ne permettent jamais, et d'autant moins qu'on ne possède généralement pas les dossiers historiques des stations, de faire un choix définitif sans une relative part de choix personnel. C'est justement cette part d'arbitraire qui peut être réduite grâce à la comparaison des résultats donnés par plusieurs méthodes, qui, seule, peut dégager un « faisceau de présomptions convergentes » permettant de conclure à la présence d'hétérogénéité dans les séries mesurées.

Les méthodes simples et économiques préconisées font appel à des comparaisons de couples de stations dont l'une est un « témoin » homogène. Elles tiennent compte des caractéristiques des liaisons de ces couples et permettent de tester leur évolution.

D'autres méthodes plus évoluées viennent naturellement à l'esprit; elles permettent de passer du couple (deux dimensions) à l'espace (analyse factorielle des correspondances ou en composantes principales par exemple...) et devraient encore enrichir et densifier l'information. En fait, l'intérêt pratique et économique de telles méthodes demeure moins évident quand les diverses séries à contrôler ne coïncident pas dans le temps et quand les hypothèses en fonction desquelles le modèle a été conçu sont quelque peu transgressées.

En fin de compte, il faudra quand même choisir car aucune méthode, aussi sophistiquée ou performante qu'elle puisse ou semble être ne nous dispensera de ce choix...!

Construction d'un modèle d'étude des lois de répartition des phénomènes pluvieux

Les tentatives d'explication, par l'hydrologue, des phénomènes hydrométéorologiques doivent être accompagnées de la faculté d'apprécier des risques (occurrence d'un phénomène, dépassement d'une intensité ou d'une durée...) dont la représentation nécessite la connaissance de la structure probabiliste des phénomènes de base.

Nous avons, dans ce but, considéré que les phénomènes pluviométriques de base étaient constitués par :

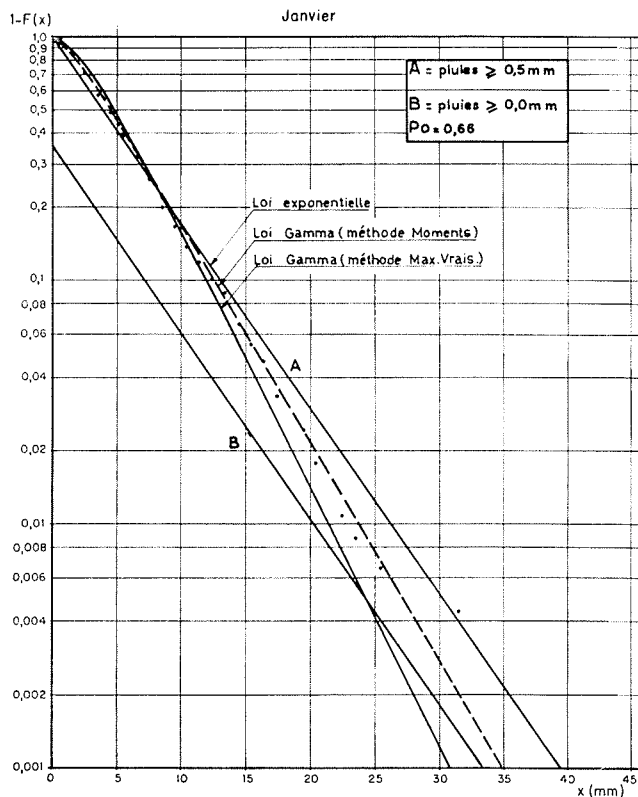
- les hauteurs des précipitations journalières non nulles (Xi);
- les durées des épisodes secs (TS);
- les durées des épisodes pluvieux (TP).

Cela revient à poser que les paramètres des lois de ces phénomènes « élémentaires » résument, quand certaines hypothèses d'indépendance sont vérifiées, la majeure partie de l'information pluviométrique contenue dans les séries mesurées.

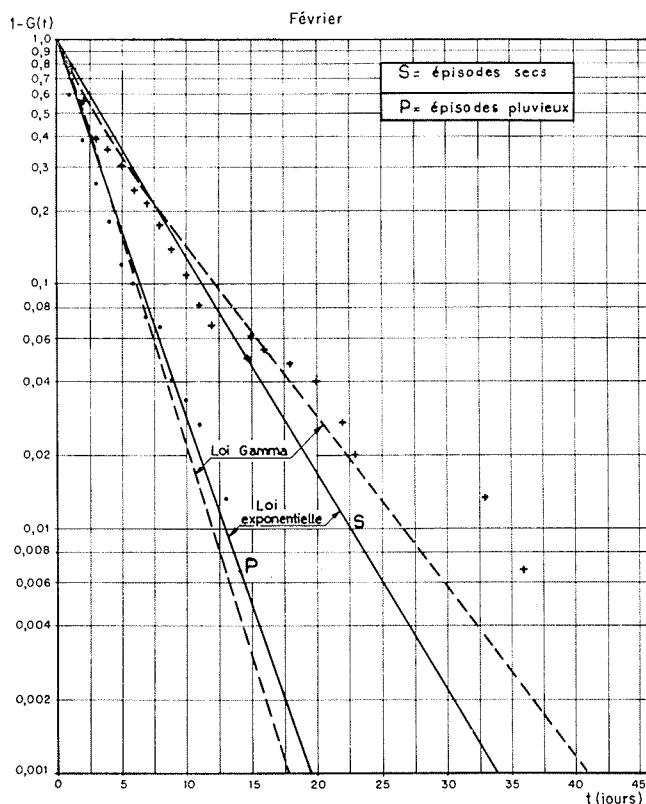
Phénomènes de base.

Nous avons constaté que ces trois phénomènes « élémentaires » obéissaient à des lois générales à deux paramètres : lois gamma incomplète ou de Pearson III, dont la fonction de répartition, qui donne la probabilité de non-dépassement, s'écrit :

$$F_Y(y) = \Pr [Y \leq y] = 1/\Gamma(k) \int_0^y u^{k-1} \cdot e^{-u} du$$



5/ Hauteurs de pluie journalières. Cognac.



6/ Durée des épisodes. Montbron.

avec : $u = a(y - y_0)$ où y_0 est un paramètre de troncature (vaut 0 pour TS et TP), Γ étant la fonction *gamma*, a et k étant les paramètres à valeurs positives de la loi *gamma* (échelle et forme).

Si X est la variable hauteur de pluie journalière ($x \in [0, \infty]$), que X_i est la variable hauteur de pluie journalière non nulle, et que p_0 est tel que $p_0 = \Pr[X = 0]$, alors :

$$F_X(x) = p_0 + (1 - p_0) F_{X_i}(x) = \Pr[X \leq x]$$

Afin de pouvoir admettre la stationnarité de ces lois, nous avons utilisé un découpage mensuel des échantillons de valeurs journalières.

L'adéquation des lois *gamma incomplète* aux valeurs observées de ces trois variables de base s'est avéré en général très satisfaisant.

Des exemples d'ajustement de ces lois aux variables X_i , TS et TP sont donnés figures 5 et 6.

On remarquera que la loi *exponentielle*, dérivée de la loi *gamma* (paramètre de forme = 1), s'adapte également bien aux échantillons des valeurs observées pour X_i , TS et TP. Notons cependant que la loi exponentielle a quelquefois tendance à sur-estimer les durées de retour des fortes hauteurs journalières précipitées X_i , et à sous-estimer celles des épisodes secs de grande durée TS.

Elle s'ajuste bien, par contre, à la distribution empirique des longueurs des épisodes pluvieux TP.

ESTIMATION DES PARAMÈTRES DES LOIS DE BASE :

Lois gamma :

— la méthode des moments donne :

$$\tilde{k} = \bar{y}^2 / S^2 \quad \text{et} \quad \tilde{a} = \bar{y} / S^2$$

avec :

\bar{y} = moyenne de l'échantillon;

S^2 = variance empirique;

— la méthode du maximum de vraisemblance conduit au système :

$$\hat{a} = \hat{k} / \bar{y}$$

et :

$$\text{Log } \hat{k} - \psi(\hat{k}) = \text{Log } \bar{y} - \sum_{i=1}^n \text{Log } y(i) / n$$

avec $\psi(k)$ = fonction digamma = $d \text{Log } \Gamma(k) / dk$

Lois exponentielles :

Dans ce cas $k = 1$ et le paramètre d'échelle s'estime par $a = 1/\bar{y}$ pour une fonction de répartition de la forme :

$$F_Y(y) = 1 - e^{-ay} = \Pr[Y \leq y]$$

VÉRIFICATION DES HYPOTHÈSES D'INDÉPENDANCE DES PHÉNOMÈNES ÉLÉMENTAIRES :

Nous calculerons les coefficients de corrélation suivants entre :

r_1 = les longueurs des épisodes secs successifs;

r_2 = les longueurs des épisodes pluvieux successifs;

r_3 = les longueurs des épisodes pluvieux et secs successifs;

r_4 = la hauteur d'un épisode pluvieux et la longueur de l'épisode sec précédent;

r_5 = la hauteur d'un épisode pluvieux et sa longueur;

r_6 = la hauteur d'un épisode pluvieux et celle de l'épisode pluvieux suivant... pour chacun des mois de l'année et les trois stations suivantes...

	MONTBRON			COGNAC			LA ROCHELLE		
	moy./an	maxi	mini	moy./an	maxi	mini	moy./an	maxi	mini
r1	-0,013	0,14 Oct.	-0,14 Janv.	-0,035	0,05 Oct.	-0,15 Août	+0,004	0,14 Nov.	-0,10 Juin
r2	+0,050	0,20 Mars	-0,13 Sept.	+0,020	0,17 Août	-0,07 Sept.	+0,122	0,32 Juill.	-0,03 Nov.
r3	-0,068	0,03 Déc.	-0,18 Août	-0,096	0,03 Déc.	-0,15 Août	-0,110	-0,03 Déc.	-0,19 Août
r4	+0,006	0,16 Juill.	-0,13 Oct.	-0,027	0,11 Juill.	-0,17 Mai	+0,049	0,16 Oct.	-0,05 Févr.
r5	+0,770	0,83 Déc.	0,57 Août	+0,792	0,89 Déc.	0,56 Août	+0,731	0,88 Déc.	0,48 Mai
r6	+0,050	0,21 Nov.	-0,10 Sept.	+0,000	0,11 Févr.	-0,10 Juin	+0,020	0,13 Oct.	-0,15 Mai

Nous pouvons donc conclure à l'indépendance des variables TS et TP et à la structure non autocorrélée des échantillons de ces deux variables élémentaires.

On peut également accepter comme hypothèse que la hauteur d'un épisode pluvieux ne dépend pas de la durée de l'épisode sec précédent ni de la hauteur de l'épisode pluvieux suivant. Elle est par contre très bien corrélée, et d'autant mieux que les mois sont à forte et homogène pluviosité, avec la longueur de cet épisode pluvieux.

D'autre part, en ce qui concerne les variables hauteurs de pluie journalières X et Xi, dont nous avons calculé les coefficients d'autocorrélation sériale rΔ qui sont représentés sur la figure 7, on peut conclure :

- à l'indépendance des valeurs successives de Xi, hauteur de pluie journalière non nulle;
- à une structure d'autocorrélation faible, mais réelle, d'allure exponentielle pour la variable hauteur de pluie journalière X (coefficient d'autocorrélation moyen de 0,29 et 0,18 pour les saisons humides et sèches et un décalage Δ = 1 jour).

Principales lois déduites des phénomènes de base.

Nous ne présenterons ici que les lois les plus intéressantes en partant du modèle le plus général qui utilise des lois gamma pour les trois phénomènes de base.

Les paramètres d'échelle et de forme de ces lois sont, successivement :

- Pour la v.a. Xi : a et k;
- Pour la v.a. TS : α et q;
- Pour la v.a. TP : β et r.

LOI DE LA DURÉE D'UN CYCLE TC = TS + TP :

Par convolution on obtient, à partir des lois de TS et TP, la fonction de répartition :

$$GTC(t) = \Pr [TC \leq t]$$

$$GTC(t) = \left(\frac{\beta}{\alpha}\right)^r \frac{1}{\Gamma(r)} \sum_{n=0}^{\infty} \left[\left(1 - \frac{\beta}{\alpha}\right)^n \frac{\Gamma(n+r)}{\Gamma(n+1)} \text{Gam}(\alpha, n+q+r) \right]$$

où Gam(α, n+q+r) est la f.d.r. d'une loi gamma incomplète et β < α (dans le cas inverse, changer α en β et q en r).

Cette formule donne lieu à des simplifications intéressantes au niveau de la sommation car si q < 1, elle devient :

$$\sum_{n=0}^N [...] + \varepsilon$$

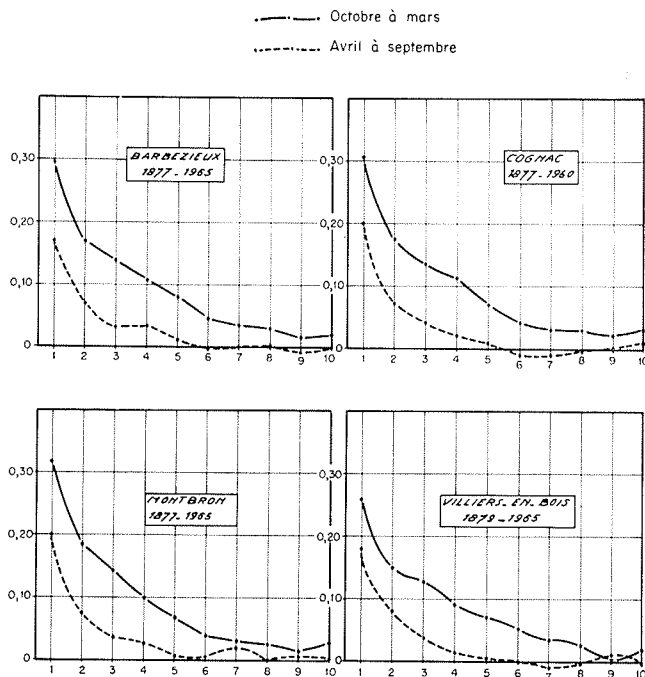
avec :

$$\varepsilon \leq (1+x)^{N+1} - x$$

LOI DE LA DURÉE D'ATTEINTE DU n^{ème} CYCLE TnC :

$$TnC = [TS(1) + \dots + TS(n)] + [TP(1) + \dots + TP(n)] \\ = TnS + TnP$$

Alors TnS suit une loi Gam(α, nq) et TnP suit une loi Gam(β, nr).



7/ Hauteurs de précipitation journalières X. Corrélogrammes moyens

Et pour obtenir la f.d.r. de TnC, on est ramené au schéma précédent et à une formulation identique obtenue en remplaçant q par qn et r par rn.

LOI DU NOMBRE DE CYCLES Nt DANS UN INTERVALLE (0, t) :

$$\Pr [Nt = n] = \Pr [TnC > t] = 1 - GTC(t) \\ = GTC(t) - GT(n+1)c(t) \\ \text{avec : } \Pr [Nt = 0] = 1 - GTC(t)$$

LOI DES VALEURS EXTRÊMES DES DURÉES DES ÉPISODES :

Episodes secs :

Si MTS, t = durée maximum des épisodes secs survenus sur (0, t), sa fonction de répartition s'écrira :

$$GMTS, t(u) = \sum_{n=0}^{n \max} \Pr [Nt = n] [GTS(u)]^n$$

avec n max = nombre maximum d'épisodes susceptibles d'intervenir sur (0, t).

Episodes pluvieux :

Remplacer TS par TP.

LOI DES HAUTEURS CUMULÉES Hi RECUEILLIES AU COURS DES ÉPISODES PLUVIEUX DE i JOURS :

Hi est donc une v.a. gamma, Gam(a, ik), dont la fonction de répartition s'écrit, si ik est entier (Xi exponentielle) :

$$FHi(x) = 1 - e^{-ax} \left[1 + \sum_{l=1}^{ik-1} \frac{(ax)^l}{l!} \right]$$

LOI DES VALEURS EXTRÊMES DES HAUTEURS JOURNALIÈRES SURVENUES SUR (0, t) :

Soit Mx , t cette variable. Sa fonction de répartition s'écrira :

$$F_{MX, t}(x) = [p0 + (1 - p0) F_{Xi}(xi)]^t$$

avec :

$$F_{Xi}(xi) = 1/\Gamma(k) \int_0^u u^{k-1} e^{-u} du$$

et

$$u = a(xi - x0)$$

Xi étant une variable du type $\text{Gam}(a, k)$ et $x0$ = seuil de troncature au-dessous duquel une pluie journalière est considérée comme « nulle » (pris égal à 0,5 mm) pour notre modèle.

LOI DES HAUTEURS H DE PLUIE RECUEILLIES AU COURS DES DIVERS ÉPISODES PLUVIEUX :

$$\Pr [H \leq x] = \sum_i \Pr [\text{épisode dure } i \text{ jours} \\ \text{et } Hi \leq x]$$

$$F_H(x) = \sum_i [G_{TP}(i) - G_{TP}(i-1)] \cdot \Pr [Hi \leq x]$$

LOI DE LA HAUTEUR CUMULÉE TOTALE Ht RECUEILLIE SUR UNE PÉRIODE QUELCONQUE (0, t) :

$$F_{Ht}(x) = \Pr [Ht \leq x]$$

$$= \sum_{i=0}^t \left[C_t^i p0^i (1 - p0)^{t-i} \cdot F_H(t-i)(x) \right]$$

avec $H(t-i)$ = v.a. qui suit une loi $\text{Gam}(a, (t-i)k)$.

Commentaires.

Nous donnerons ici quelques remarques relatives à l'adéquation des lois précédemment obtenues pour la station de Cognac située au centre du bassin de la Charente (voir fig. 8, 9, 10, 11, 12).

Cette adéquation, que l'on parte de lois de base du type *gamma* ou *exponentielle*, est généralement très satisfaisante, ce qui est intéressant car la simplification apportée dans les calculs successifs par cette dernière hypothèse est importante (calculs plus simples donc moins coûteux).

Les valeurs extrêmes (durées et intensités) sont, dans la majorité des cas étudiés (mois et station), bien représentées (fig. 9 et 11).

L'adéquation des échantillons des pluies totales Ht recueillies sur un intervalle quelconque (0, t) est bonne jusqu'à $t = 15$ jours. La loi théorique obtenue a tendance à diverger pour des valeurs de t supérieures (sous-estimation des probabilités de dépassement) (fig. 12).

Ceci est sans doute dû au fait que nous avons négligé l'autocorrélation faible mais réelle existant entre les valeurs des pluies journalières.

Les lois déduites du modèle de base représentent donc assez fidèlement les échantillons observés dans la plupart des exemples cités.

La procédure utilisée va donc dans le sens d'une *valorisation* de l'utilisation des observations les plus nombreuses (journalières) sans grande perte d'information quant aux lois dérivées, et d'une *condensation* de cette information en quelques paramètres simples représentatifs des diverses lois les plus utilisées (fig. 13).

Le modèle présenté, qui peut paraître quelque peu « sophistiqué », est néanmoins intéressant quant à sa démarche, grâce au retour aux sources qu'il permet par sa vision en « raccourci » de la structure probabiliste des phénomènes pluvieux.

Essais d'interprétation des facteurs conditionnels principaux des régimes pluviométriques mensuels

Méthode. Analyse factorielle des correspondances.

Notre but sera ici, à partir de la répartition, à l'échelle mensuelle, des précipitations journalières en une vingtaine de stations pluviométriques, de tenter de mettre en évidence quelques facteurs principaux qui conditionnent la pluviométrie des bassins de la Charente et de la Seudre.

Nous allons pour cela utiliser la technique de *l'analyse factorielle des correspondances* qui a donné lieu à des applications intéressantes depuis quelques années dans des domaines très divers (linguistique, économétrie, sociologie, psychologie, ...).

Notre *tableau de correspondances* sera constitué ici par ni lignes et nj colonnes représentant respectivement les ni stations pluviométriques, et les nj classes de hauteurs journalières précipitées (un tableau par mois de l'année).

Nous posons donc, comme hypothèse de travail, que les facteurs conditionnels du régime pluviométrique peuvent être appréhendés à partir du caractère probabiliste du tableau de correspondances ainsi défini.

C'est encore dire que les *caractères fondamentaux du régime pluviométrique d'un bassin doivent modéliser la répartition en classes des hauteurs de pluie journalières* observées aux diverses stations pluviométriques de ce bassin.

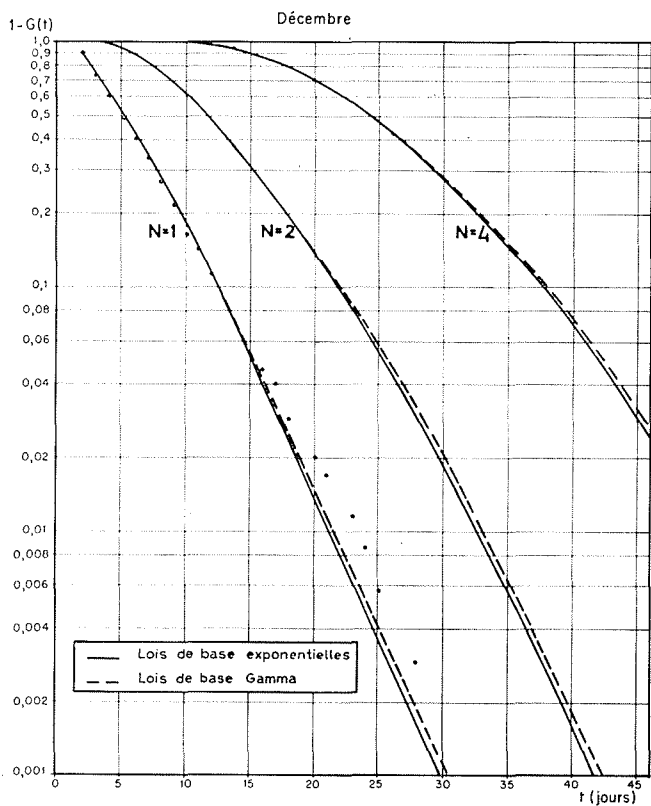
A l'intersection de la colonne j et de la ligne i on trouvera donc l'effectif de la j° classe de hauteurs à la i° station considérée, $k(i, j)$, (donc $k(i, j)$ = nombre de fois que l'on a observé des précipitations journalières comprises dans les limites de la classe j , à la station i , au cours des n mois « M » de la période de référence considérée).

La méthode suivante permettra de représenter les données de ce tableau, et les liaisons existant entre elles, dans un espace à deux ou trois dimensions, le nombre de dimensions étant égal au nombre de facteurs principaux conditionnant la structure et la proximité des ensembles (I) ou (J).

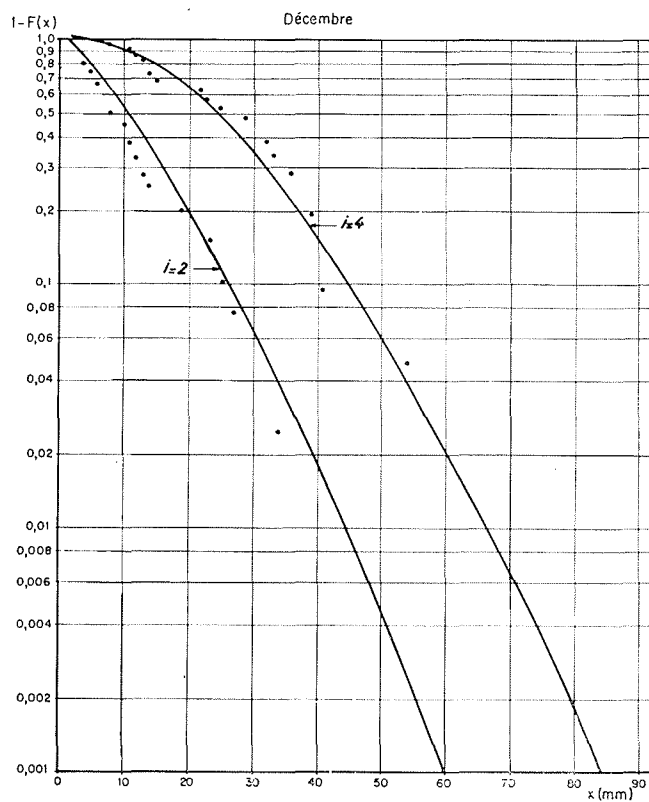
Ainsi chaque élément de (I) ou (J) pourra être représenté dans cet espace par ses coordonnées qui seront fonction de la contribution de ces facteurs à sa structure.

Cette analyse tentera donc de décrire les *liaisons, au point de vue forme ou structure*, existant entre ces divers éléments.

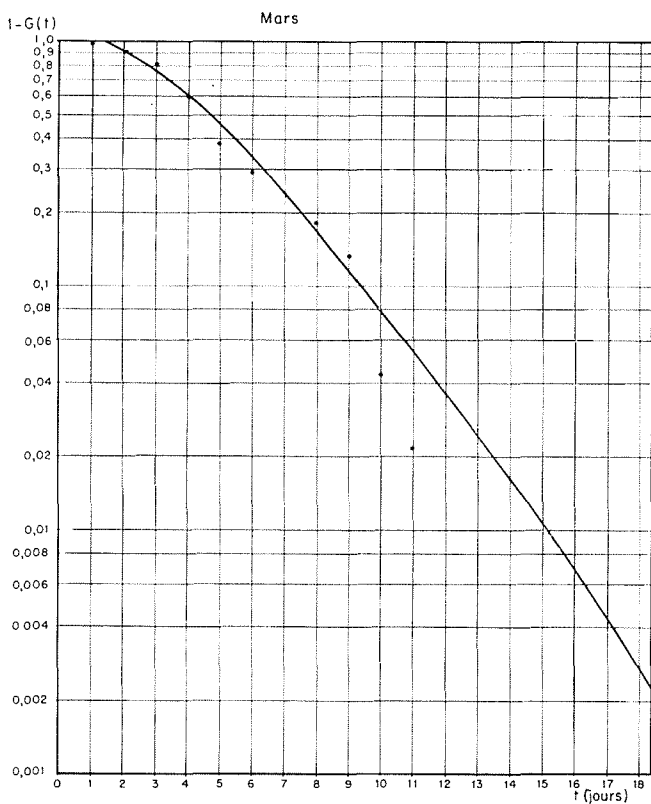
Son avantage essentiel par rapport à la méthode d'analyse en composantes principales est de faire jouer un rôle *symétrique* aux ensembles (I) et (J).



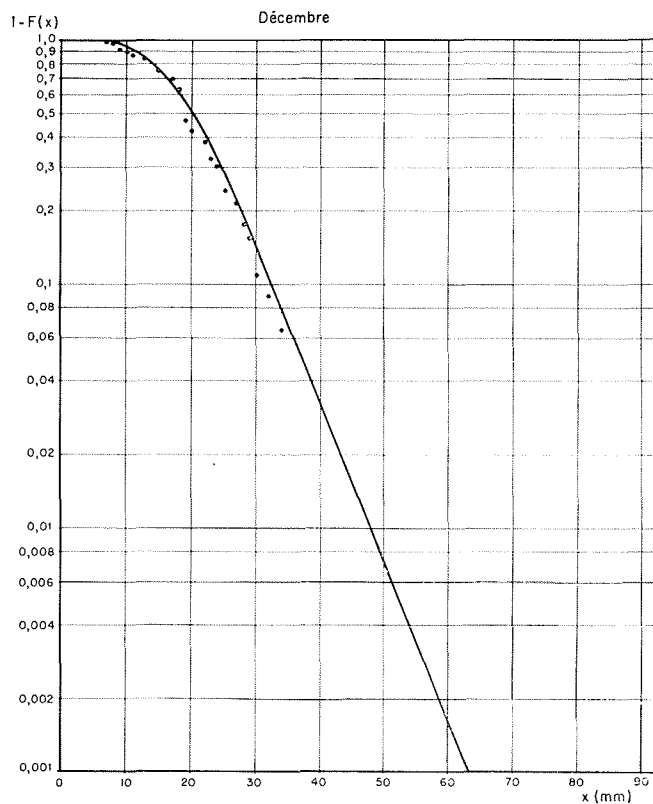
8/ Durée de N cycles. Cognac.



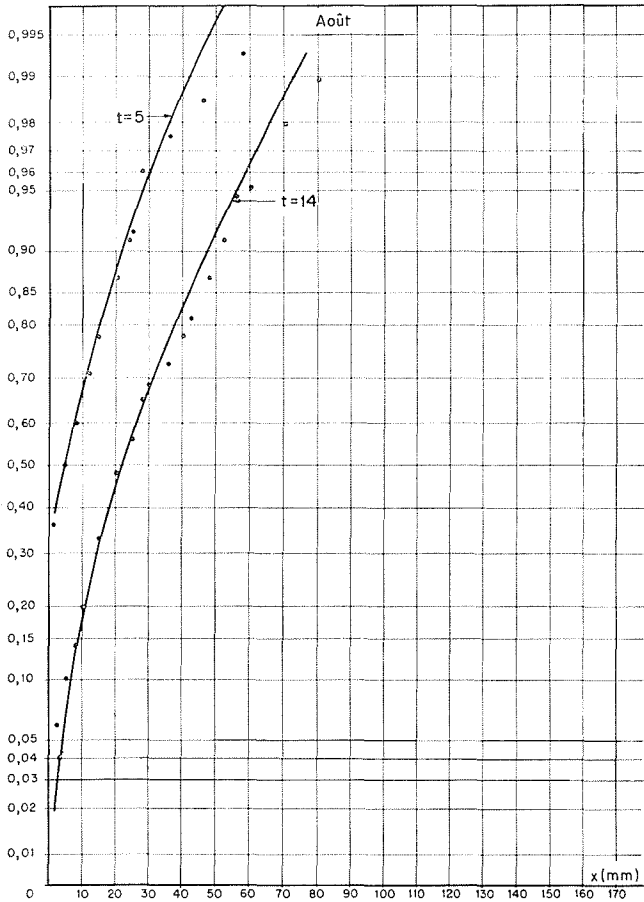
10/ Hauteurs de pluie des épisodes de i jours. Cognac.



9/ Durées extrêmes des épisodes pluvieux. Cognac.



11/ Hauteurs de pluie journalières maximales. Cognac.



12/ Pluies totales d'une période de N jours. Cognac.

Afin de présenter quelques étapes de calcul, plaçons-nous dans l'espace R_j des (j) colonnes où nous construirons un nuage de (i) points (les stations).

Précisons d'abord que les coordonnées du tableau de correspondances permettent d'établir certaines probabilités d'occurrence :

$$p(i) = \sum_j p(i, j) \quad p(j) = \sum_i p(i, j)$$

avec :

$$k = \sum_{i,j} k(i, j) \quad \text{et} \quad p(i, j) = \frac{k(i, j)}{k}$$

donc :

$$\sum_{i,j} p(i, j) = 1$$

et $p(i, j)$ = probabilité d'occurrence du couple (i, j) , ainsi $p(i)$, probabilité qu'une des k observations ait été réalisée à la station i , q.q.s. la classe j (caractérise l'importance de la station i).

De la même façon :

$p(j/i) = p(i, j)/p(i)$ = probabilité qu'une observation appartienne à la classe j , pour la station i [affecté du poids $p(i)$].

Alors la « distance » de deux points stations i et i' , représentés par leurs coordonnées $p(j/i)$ dans R_j , pourra s'écrire :

$$d^2(i, i') = \sum_j \frac{1}{p(i)} \left[\frac{p(i, j)}{p(i)} - \frac{p(i', j)}{p(i')} \right]^2 = \sum_j (y_i - y_{i'}) (y_i - y_{i'})$$

avec $y_i = C x_i$ si $x_i = p(i, j)/p(i)$ et C matrice diagonale telle que $C_{jj} = 1/p(j)$.

Notons dès à présent que cette distance « pondérée » par $1/p(j)$ avec $j \in (J)$ permet d'harmoniser les distributions en atténuant les différences entre les effectifs totaux des différentes colonnes (classes).

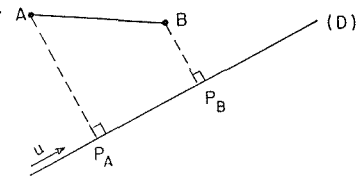
Ceci signifie donc que l'on s'intéresse plus à la répartition des distributions qu'à la taille des échantillons observés, ce qui veut encore dire, par exemple, que l'on peut comparer des distributions en classes de hauteurs observées à des stations ayant des périodes de fonctionnement différentes au point de vue longueur, ce qui est intéressant.

Nous voulons maintenant déterminer les axes principaux d'inertie du nuage, dans l'ordre décroissant des moments d'inertie, et le projeter sur ces axes de façon à obtenir des déformations minima.

— Obtenir des « déformations minima » veut dire que nous désirons que les distances du type $d^2(i, i')$ soient le mieux possible conservées après projection des points i et i' sur les axes factoriels choisis.

— Si l'on cherche tout d'abord le sous-espace à une dimension (la droite premier axe factoriel) on voit qu'il faut, pour réussir ces conditions (en moyenne) maximiser les projections du type $\overline{PA PB}$ pour les n_i points A et B appartenant à l'ensemble (i) ;

— Si \vec{u} est un vecteur unitaire porté par le premier axe factoriel (n_j composantes).



On doit donc avoir :

$$S^2 = \sum_{A, B} \overline{PA PB}^2 \text{ max.}$$

— Les composantes d'un vecteur unitaire porté par (D) sont données par la résolution de :

$$u' \cdot \mathcal{Q} \cdot u \text{ maximum avec } u \cdot u' = 1$$

avec \mathcal{Q} = matrice des covariances des y .

Comme \mathcal{Q} est symétrique définie positive et que l'on peut la diagonaliser ($\lambda = P' \cdot \mathcal{Q} \cdot P$), on voit que le vecteur u cherché sera le vecteur propre correspondant à la plus grande valeur propre placée sur la diagonale de la matrice λ .

— Si l'on généralise et que l'on désire le sous-espace à n dimensions ayant les propriétés de déformation minimale, on démontre qu'il contient (D) , premier axe factoriel, et qu'une de ses bases est constituée des n premiers vecteurs propres de \mathcal{Q} .

Les axes factoriels ainsi déterminés sont orthogonaux et ont la propriété d'extraire, de façon progressive, le plus « d'information » possible du nuage initial.

Pour revenir aux données issues de notre tableau de correspondances, on démontre qu'il suffit de chercher les vecteurs propres de la matrice symétrique, définie et positive de terme général :

$$S_{jk} = \sum_i \frac{p(i, j) p(i, k)}{p(i) [p(j) p(k)]^{1/2}}$$

En général, deux, trois ou quatre facteurs suffisent à représenter 85 % à 95 % de l'« inertie » du nuage de points (« information » ou part expliquée de la variance des variables de départ).

Ces axes factoriels permettent ainsi d'extraire la plus grande partie de l'information contenue dans le tableau de correspondances initial, d'une façon progressive (le pourcentage d'inertie du nuage expliqué par les facteurs décroît avec le rang des facteurs puisque ceux-ci correspondent à des valeurs propres décroissantes).

Application de cette méthode. Interprétation des résultats.

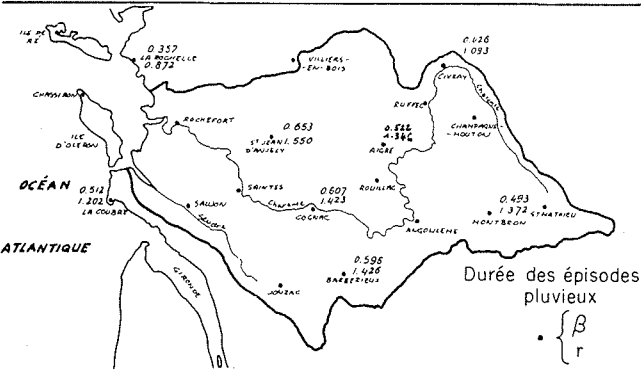
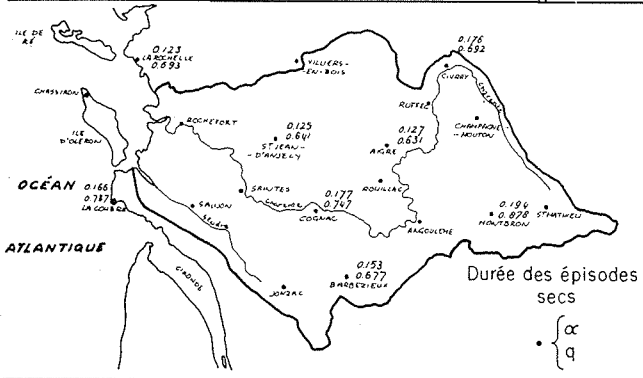
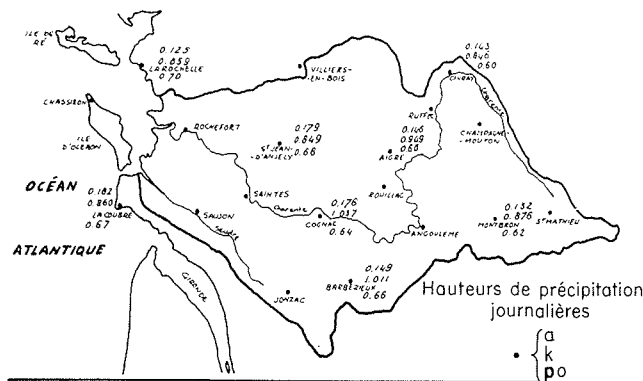
Nous avons choisi, pour tester cette méthode, de prendre pour éléments de l'ensemble (J), dix-neuf stations pluviométriques réparties de manière homogène sur les bassins étudiés.

Les éléments de (I) sont représentés par six classes de hauteurs journalières de précipitation à effectifs décroissants : CL 1 (< 0,5 mm), CL 2 (0,5-5 mm), CL 3 (6-10 mm), CL 4 (11-15 mm), CL 5 (16-20 mm), CL 6 (> 20 mm).

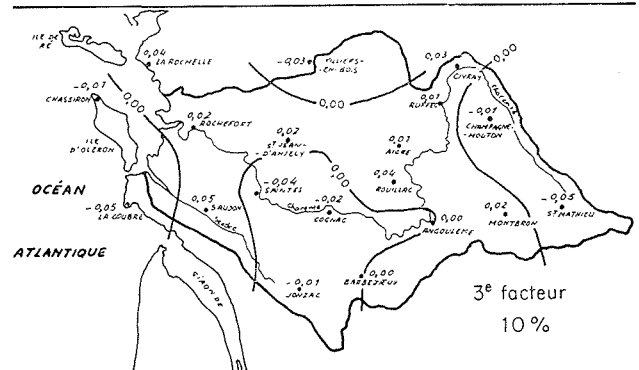
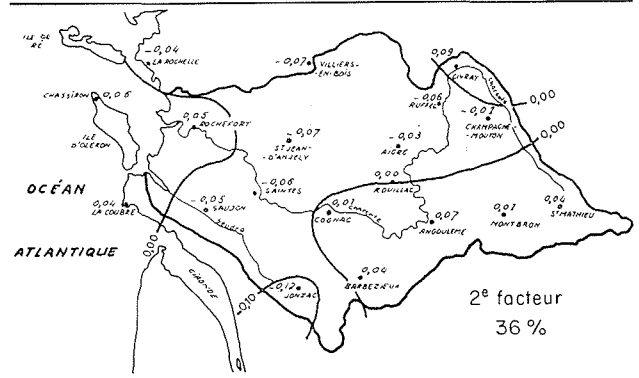
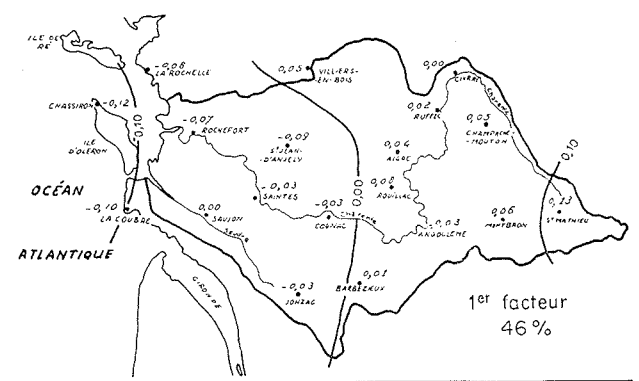
Ces six classes résultent de la subdivision de quatre classes de hauteurs choisies dans un précédent exemple : CL'1 (< 0,5 mm), CL'2 (0,5-10 mm), CL'3 (11-20 mm), CL'4 (> 20 mm).

Les pourcentages d'inertie du nuage expliqué par les trois premiers facteurs pour ces classes CL' sont donnés sur la figure A (en pointillés).

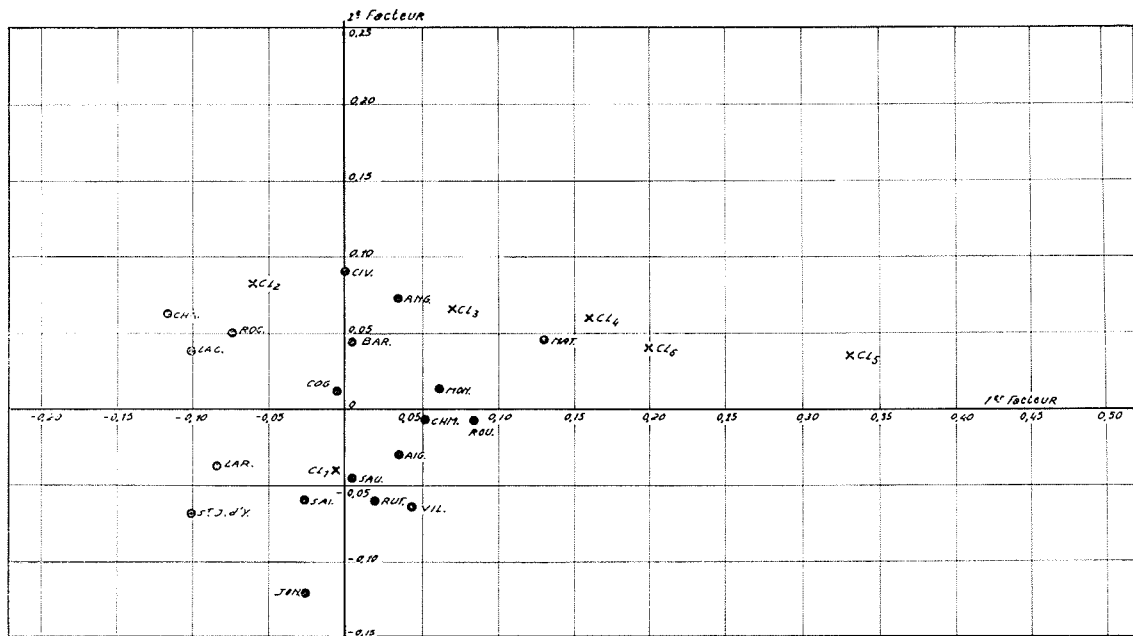
On constate que l'explication de l'information donnée par les deux premiers facteurs (issus de ce découpage en quatre classes) est supérieure à celle donnée pour ces mêmes facteurs pour l'exemple suivant (six classes) et que la part du quatrième facteur est nulle.



13/ Bassins de la Charente et de la Seudre. Paramètres des lois de base - Mars.



14/ Bassins de la Charente et de la Seudre. Variation spatiale des composantes factorielles - Février.



15/ Représentation des variables sur les deux premiers axes factoriels - Février.

On pourrait donc penser qu'il était inutile de tester un autre découpage... Mais ces résultats n'ont de valeurs que si on parvient à donner un nom aux influences pluviométriques qui se traduisent par ces facteurs. Or l'essai de définition de ces influences s'est avéré négatif, sauf en ce qui concerne la première qui est prédominante et que l'on retrouve dans le deuxième essai.

Nous pensons que ceci est dû au fait qu'un découpage assez grossier en quatre classes atténue les disparités entre les stations pluviométriques.

Un découpage plus fin doit donc mieux rendre compte des « nuances » de la répartition de la pluie journalière sur un bassin (ceci a été confirmé depuis sur le bassin de l'Agôut, sous-bassin du Tarn).

Pour le deuxième essai (six classes), on constate que les deux premiers facteurs rendent compte ensemble de 72 à 87 % (juin et janvier) de cette inertie, que le troisième facteur en explique de 6 à 15 % (janvier et juin) et que la part du quatrième est très faible (3 à 4 % en moyenne).

L'évolution de ces pourcentages par ces trois premiers facteurs au cours des mois est représentée sur la figure A.

On peut donc déjà conclure que les deux premiers facteurs résument la majeure partie de l'information contenue dans le nuage.

Une représentation des éléments de (I) et (J) par rapport aux deux premiers axes factoriels est donc avantageuse (fig. 15).

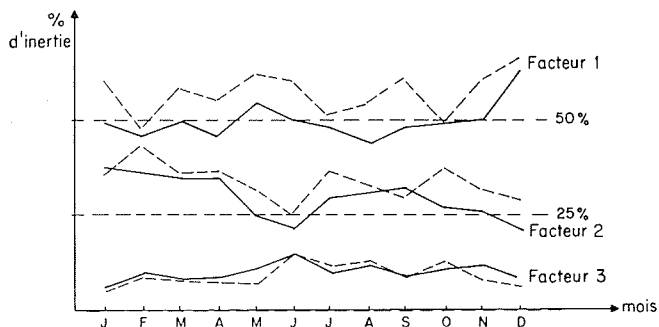
On peut également représenter ces coordonnées (par rapport aux axes factoriels) aux différentes stations sur des cartes géographiques et tracer ainsi des isolignes qui permettront d'aider, par leurs directions, à interpréter ces facteurs (fig. 14).

Les représentations mensuelles, par rapport aux deux premiers axes factoriels permettent de remarquer que le nuage des points-stations (ensemble (j)) est relativement bien groupé autour de l'origine, ce qui laisse présager quelques difficultés d'interprétation.

On notera cependant que le premier axe factoriel « oppose » les stations côtières aux stations situées à l'est des bassins.

On pourrait donc interpréter le premier facteur comme étant lié à l'influence océanique, cette influence conduisant à une augmentation des précipitations au fur et à mesure que l'on s'éloigne de la côte atlantique, c'est-à-dire que les reliefs apparaissent (on pourrait encore l'appeler influence océano-orographique, qualitatif quelque peu barbare qui montre l'influence du relief sur des précipitations d'origine océanique).

Signalons encore que la proximité des points CL 1, CL 2 et de ceux représentant les stations côtières montre bien, sur la représentation par rapport aux deux premiers axes factoriels, le rôle important joué par les effectifs de ces classes dans la répartition globale de la pluviométrie journalière à ces stations (rôle important des faibles précipitations).



A/

On démontre en effet que ces *proximités* peuvent s'interpréter en termes de *corrélation*.

Un rôle identique est joué par les autres classes de hauteurs dans la distribution des précipitations pour les stations d'altitudes plus élevées.

Cette dernière remarque paraît intéressante car elle pourrait aider à la résolution du problème de la *rationalisation des réseaux pluviométriques*.

En effet on peut essayer de définir, à partir de la représentation dans le plan des deux premiers axes factoriels, des *groupes* de stations. Ces groupes ont donc un comportement voisin vis-à-vis des deux premières composantes pluviométriques qui sont fondamentales. On peut donc choisir à l'intérieur de ces groupes de stations (les nuages constituant ces groupes doivent se conserver au cours des mois de l'année) une ou deux stations qui seraient dites « principales », constitueraient le réseau *pluviométrique de base* et devraient être préservées.

Les groupes caractéristiques que nous avons pu ainsi repérer sont les suivants :

- Saint-Mathieu, Montbron, Champagne-Mouton, Rouillac;
- Barbezieux, Cognac, Angoulême, Civray;
- Aigre, Ruffec, Villiers-en-Bois;
- Jonzac, Saintes.
- La Rochelle, La Coubre, Chassiron, Rochefort, Saint-Jean-d'Angély.

Une telle analyse donnerait évidemment des résultats plus nets sur un bassin moins homogène que le nôtre.

Quant au deuxième axe factoriel, les douze représentations du type de celles représentées sur la planche 14 indiqueraient, en moyenne sur les douze mois, une similitude de comportements, vis-à-vis du deuxième facteur d'un groupe, à priori, assez disparate de stations :

- Saintes, Jonzac, Saint-Jean-d'Angély, Villiers-en-Bois, Rouillac, Aigre, Ruffec, Champagne-Mouton, Saint-Mathieu.

Si l'on ne tient pas compte de Jonzac, on constate que ces stations semblent influencées différemment de celles qui sont situées dans la moitié sud-ouest des bassins étudiés.

La carte orographique (fig. 1) indique, de façon évidemment simplifiée, que le groupe de stations noté plus haut serait, grâce à l'ouverture de certaines vallées, soumis à l'influence de ces vents alors que le second groupe pourrait être considéré comme étant dans « l'ombre » des reliefs.

Le facteur 2 pourrait alors représenter *l'influence des caractéristiques de l'exposition des stations aux vents continentaux dominants*.

Ceci demeure encore trop subjectif pour que nous n'attirions pas l'attention du lecteur sur le caractère « flou » donné à l'interprétation de cette seconde composante pluviométrique.

Disons enfin que nous n'avons pu interpréter le troisième facteur qui ne représente, en moyenne, que 9 à 10 % de l'inertie du nuage, en raison de la trop faible dispersion des coordonnées des points du nuage par rapport au troisième axe factoriel.

Ces difficultés d'interprétation nous amènent à faire remarquer que :

a) Le choix que nous avons fait des éléments représentatifs des ensembles (*i*) et (*j*) (classes et stations) ne semble pas conduire à une réduction de l'efficacité de la méthode d'analyse utilisée.

Il serait sans doute intéressant de refaire une expérience semblable en introduisant dans le modèle proposé une répartition des hauteurs de pluie selon la longueur des épisodes pluvieux puisqu'il semble en effet que ces deux facteurs intensité et durée « résument » l'information pluviométrique.

b) Bien que dans le calcul des « distances » entre deux points d'un ensemble intervienne une pondération qui tienne compte des effectifs totaux de chaque station, il vaut mieux travailler sur des stations ayant fonctionné sur la même période (voir en effet le rôle quelque peu « excentrique » de Saujon, station courte période, sur les différents graphiques).

c) La méthode d'analyse factorielle des correspondances devrait donner lieu à une interprétation plus immédiate des facteurs pluviométriques principaux pour des études semblables appliquées à des bassins plus *étendus et moins homogènes*. Les bassins de la Charente et de la Seudre n'offrent en effet que peu de variation dans leurs caractéristiques en raison de leurs faibles étendues, reliefs, et proximité de l'Océan atlantique.

d) Ces difficultés d'interprétation semblent être courantes dans ce genre d'analyse quand les premiers facteurs se traduisent par des influences voisines en intensité.

Conclusion

Les modèles étudiés sont évidemment perfectibles. On leur reprochera déjà d'être compliqués, bien que les idées de base sur lesquels ils s'appuient soient simples, car ils nécessitent l'utilisation de calculateurs assez puissants.

Cette nécessité va cependant dans le sens de l'histoire et l'on peut dire que cette utilisation n'a été faite que dans le souci de condenser au maximum l'information spatiale.

Tous nos remerciements vont à nos collègues de l'« Agence Financière de Bassin Adour-Garonne » qui a financé cette étude, et plus particulièrement à M. A. Dubœ qui nous a guidé dans nos recherches, ainsi qu'à tout le personnel du « Laboratoire de Statistique de l'Université Paul Sabatier de Toulouse ».

Bibliographie sommaire

- BENZECRI (J. P.). — « Leçons sur l'analyse factorielle des correspondances ». Faculté des Sciences, Paris.
- BERNIER (J.). — « Sur la théorie du renouvellement et son application en Hydrologie ». *E.D.F.* (1967).
- LOBERT (A.). — « Modèle probabiliste de base pour les pluies dans le bassin de l'Allier ». *E.D.F.*, 45/67.
- DUBAND (M.). — « Composition des lois de probabilité des précipitations journalières ». *S.H.F.* (mai 1967).
- REMENIERAS (G.). — « L'Hydrologie de l'ingénieur ». *Eyrolles* (1965).
- ROCHE (M.). — « Hydrologie de surface ». *Gauthier-Villars* (1963).
- QUELENNEC (R. E.). — « Sur quelques méthodes générales d'analyse des phénomènes hydrométéorologiques ». *Agence de bassin Adour-Garonne* (avril 1970).