

Test d'homogénéité

Y. BRUNET-MORET
Ingénieur hydrologue,
Bureau Central Hydrologique Paris

RÉSUMÉ

Présentation d'un test d'homogénéité spécialement conçu pour vérifier l'homogénéité des suites chronologiques de précipitations annuelles dans une zone climatique, en utilisant, si possible, le vecteur régional Hiez.

ABSTRACT

Exposing a test of homogeneity of chronological series of annual rainfall in a climatic area, with using, if possible, the regional vector Hiez.

1. INTRODUCTION

1.1. GÉNÉRALITÉS

Une suite chronologique de valeurs observées est de caractère aléatoire simple si toutes les valeurs proviennent d'une même population mère par tirages au hasard et indépendants.

Le caractère aléatoire simple peut être détruit :

— Par un effet de persistance : une valeur n'est pas indépendante de la ou des valeurs précédentes (processus markovien ou autre), mais la suite est stationnaire et les paramètres de la distribution des valeurs observées ne varient pas dans le temps : entre autres l'espérance mathématique.

— Par un effet de tendance : l'espérance mathématique des valeurs observées croît (ou décroît) avec le temps.

— Par 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 et les paramètres de la distribution de cette valeur moyenne ne varient pas dans le temps : entre autres l'espérance mathématique.

— Par des erreurs systématiques d'observation affectant plusieurs termes consécutifs de la suite chronologique, et qui peuvent se surajouter aux trois effets ci-dessus. Ce sont ces erreurs qui détruisent l'homogénéité de la suite.

1.2. TESTS EXISTANTS

Parmi les tests du caractère aléatoire simple d'une suite chronologique, les principaux sont :

Le test de corrélation sériale (III p. 360, p. 437) (1) de Wald, Wolfowitz, Anderson qui a été proposé comme test de détection d'une tendance et qui est en fait un test de persistance.

Le test des pics et celui des phases (III, p. 351) qui sont surtout utilisables dans le cas d'effets cycliques ou pseudo-cycliques.

Le test de corrélation de rang de Mann (II p. 478, III p. 357) test de tendance.

Le test de corrélation de rang de Spearman (II, p. 476) également test de tendance.

(1) Les références entre parenthèses se rapportent à la publication de M. G. KENDALL et A. STUART intitulée *The advanced theory of statistics*, 3 volumes 1968.

Le premier test demande des suites chronologiques longues pour être significatif. Le test des pics, celui des phases semblent n'être pas assez puissants pour être utilisables. Les deux derniers tests ont une puissance convenable pour être utilisés avec des suites relativement courtes, si la tendance, exprimée en variation de l'espérance mathématique de la population mère, est suffisamment forte par rapport à l'écart type observé dans la suite.

1.3. POSITION DU PROBLÈME

Le véritable problème n'est pas de détecter les effets de persistance, de tendance ou pseudo-cyclique, mais de détecter les erreurs d'observations. Il semble que ce soit impossible, si on ne dispose que d'une seule suite chronologique.

En étude de précipitations annuelles, par exemple, on dispose facilement d'assez nombreuses suites chronologiques provenant de stations situées dans une même zone climatique, avec des coefficients de corrélation linéaire significativement positifs.

Étant dans la même zone climatique, les stations sont soumises aux mêmes effets de persistance, de tendance, de pseudo-cycles. Leurs précipitations moyennes interannuelles sont voisines, et, d'après les observations, leurs coefficients de variations sont encore plus voisins.

1.4.

Soient deux stations de moyennes \bar{x} et \bar{y} , de variance $\text{var } x$ et $\text{var } y$, de coefficient de variation commun :

$$c = \frac{\sqrt{\text{var } x}}{\bar{x}} = \frac{\sqrt{\text{var } y}}{\bar{y}}$$

de coefficient de corrélation linéaire

$$\rho = \frac{\text{covar}(x, y)}{\sqrt{\text{var } x \text{ var } y}}$$

La variate transformée $Y' = Y \frac{\bar{x}}{\bar{y}}$ a même moyenne et même variance que la variate X .

En supposant une relation structurelle et linéaire entre les variates X et Y' (II chap. 29), c'est-à-dire une relation fonctionnelle entre les variates $X + \delta$, $Y' + \varepsilon'$

$$x_i + \delta_i = y'_i + \varepsilon'_i$$

quelle que soit l'année i , c'est-à-dire que les valeurs x_i et y'_i représentent la même chose (précipitation de l'année i) avec des erreurs d'appréciations δ_i et ε_i qui sont des variables aléatoires, indépendantes entre elles, indépendantes des valeurs observées x_i et y_i , de moyennes nulles de mêmes variances $\text{var } x (1 - \rho)$, indépendantes des valeurs δ_i , ε'_i , x_i , y_i (homoscedasticité), on peut écrire $x_i - y'_i = \varepsilon_i$.

La variate ε est distribuée avec une moyenne nulle et une variance : $2 \text{ var } x (1 - \rho)$.

Les hypothèses ci-dessus n'ont rien de contraire à ce qui peut être déduit des observations. De plus, même si les distributions de x et de y ne sont pas normales (elles sont bornées inférieurement par une valeur nulle et positive) on peut constater que, lorsque les coefficients de variation et d'asymétrie de ces distributions ne sont pas grands, mettons inférieurs à $1/3$, la distribution de ε est quasi-normale.

1.5. CONCLUSION DE L'INTRODUCTION

La variate ε se trouve pratiquement débarrassée des effets de persistance, de tendance, de pseudo-cycles, mais conserve les erreurs systématiques. La suite chronologique des valeurs ε est de caractère aléatoire simple.

Les seuls tests utilisables sont alors ceux de Mann ou de Spearman qui permettent de détecter, comme tendance, les erreurs systématiques affectant une seule extrémité de la suite chronologique. Ces tests ne détectent rien (vérifié par tirages au hasard) si un erreur systématique se produit au milieu de la suite ou s'il se produit plusieurs erreurs systématiques.

Il manque un véritable test d'homogénéité permettant de détecter les erreurs systématiques par détection des variations de la moyenne.

1.6. REMARQUE

Dans ce qui précède, il n'a été question en fait que de tester la constance de la moyenne. Pour le problème qui nous concerne, il semble difficile de tester la constance de la racine carrée de la variance (écart-type). On pourrait appliquer un des tests de 1.2 à la valeur absolue de ε_1 . Il est probable qu'on n'obtiendrait pas de valeurs significatives d'un des tests.

2. PROPOSITION POUR UN TEST D'HOMOGENÉITÉ

2.1.

Compte tenu de 1.4, il semble suffisant de chercher un test valable en distribution normale.

L'échantillon de taille n « suite chronologique » est constitué de n variables x_i avec les hypothèses :

- les variables sont aléatoires et indépendantes ;
- elles sont tirées d'une population mère distribuée normalement, de moyenne et variance inconnues.

L'échantillon est transformé en échantillon chronologique de variables réduites u_i

$$x_0 = \frac{1}{n} \sum_1^n x_i \quad s^2 = \frac{1}{n-1} \sum_1^n (x_i - x_0)^2 \quad u_i = \frac{x_i - x_0}{s}$$

2.2.

La simple masse des quantités u_i , en posant

$$u_0 = 0, \quad z_k = \sum_{i=0}^k u_i = - \sum_{i=k+1}^n u_i$$

correspond à la figure 1 :

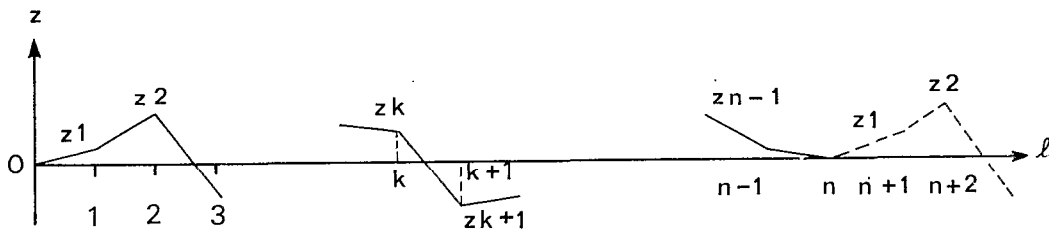


Fig. 1.

Le carré de la surface élémentaire comprise entre l'axe des l , les abscisses k et $k + 1$ et la droite $z_k z_{k+1}$ (surfaces comptées positivement si elles sont au-dessus de l'axe des l , négativement en-dessous) s'écrit :

$$\frac{1}{4} (z_k + z_{k+1})^2$$

dont la sommation étendue à toute la simple masse est :

$$S = \frac{1}{4} \sum_{k=0}^{n-1} (z_k + z_{k+1})^2$$

qui peut s'écrire :

$$S = \frac{1}{4} \sum_{k=0}^{n-1} (2 z_k + u_{k+1})^2 = \sum_{k=0}^{n-1} z_k^2 + \frac{1}{4} \sum_{i=1}^n u_i^2 + \sum_0^{n-1} z_k u_{k+1}$$

d'après la définition des variables réduites :

$$\sum_1^n u_i^2 = n - 1$$

$$\left(\sum_1^n u_i \right)^2 - \sum_1^n u_i^2 = - (n - 1) = 2 \sum_{j=2}^n \sum_{i=1}^{j-1} u_i u_j = 2 \sum_0^{n-1} z_k u_{k+1}$$

d'où $S = Z - \frac{n-1}{4}$, avec $Z = \sum_{k=0}^{n-1} z_k^2$.

2.3.

La quantité Z est liée à la surface comprise entre l'axe des ℓ et la ligne brisée de la simple masse : plus cette ligne coupe l'axe des ℓ , plus Z est petit.

Z peut s'écrire :

$$- \sum_{j=2}^n \sum_{i=1}^{j-1} (j - i) u_i u_j, \quad j > i$$

et c'est en partant de cette définition qu'il est plus facile de calculer quelques valeurs caractéristiques de la distribution de Z

$$E(Z) = \frac{n^2 - 1}{6}$$

$$\text{Variance}(Z) = \frac{(n - 1)^3 (n - 2) (2n - 1)}{90}$$

$$\text{Coefficient d'asymétrie } \gamma_1(Z) = \frac{2 \sqrt{10}}{7} \frac{(n - 3) (4n + 1)}{(n + 3) \sqrt{(n - 2) (2n - 1)}}$$

La valeur minimale de Z est $\frac{(n - 1)^3}{2(2n - 3)}$.

La valeur maximale de Z croît rapidement avec n, au moins comme la puissance 3 de n. Elle est de 2 pour n = 3, de $\frac{9}{2}$ pour n = 4, de 10 pour n = 5.

2.4.

Il semble préférable de prendre une quantité dérivée de Z comme statistique de test d'homogénéité : prenons, par exemple, une simple masse avec deux cassures : comportant trois suites de moyennes différentes. La valeur de Z dépend trop de l'ordre dans lequel se produisent les cassures comme le montre la figure 2 ci-dessous où dans les trois cas sont utilisées les mêmes trois suites de moyennes différentes, mais dans un ordre différent.

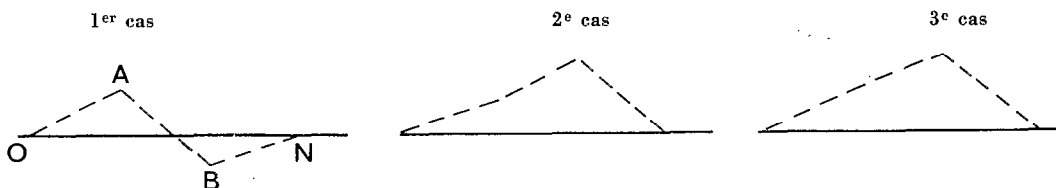


Fig. 2.

2.5.

Pour éliminer cet effet, il suffit de sommer toutes les quantités Z obtenues en faisant passer l'axe des ζ par les différents points $z_0, z_1 \dots z_k \dots z_{n-1}$ (fig. 1) et on peut prendre comme statistique de test d'homogénéité la quantité ZZ .

$$n(n-1)ZZ = \sum_0^{n-1} Z_k$$

$$Z_k = \sum_{j=0}^{n-1} (z_k - z_j)^2$$

3. ÉTUDE DE ZZ

3.1.

La valeur ZZ peut se mettre sous la forme :

$$ZZ = \frac{-2}{n(n-1)} \sum_{j=2}^n \sum_{i=1}^{j-1} (j-i)(n-j+i) u_i u_j, \quad j > i$$

ou sous la forme :

$$ZZ \text{ centré} = ZZ - E(ZZ) = \frac{(n-2)}{12} + \sum_{k=1}^{m-1} \left[2k(n-k) - \frac{n^2}{2} \right] A_k \cdot \frac{1}{n(n-1)}$$

n pair : $2m$, n impair : $2m-1$.

Les A_k sont des polynômes de n termes $u_i, u_j, i \neq j$, de coefficients tous égaux à $+1$, tels que, si $i < j, k = j - i$ ou si $i > j, k = n + j - i$ le polynôme A_k est le produit scalaire $\vec{V}_1 \cdot \vec{V}_k$ du vecteur $\vec{V}_1 (u_1, u_2 \dots u_n)$ et du vecteur $\vec{V}_k (u_{1+k}, \dots u_n, u_1 \dots u_k)$.

3.2.

Aucune valeur u ne pouvant devenir infinie ZZ possède une borne supérieure, probablement de l'ordre de a , a étant une constante. Aucune valeur de Z ne pouvant être nulle ou négative, ZZ possède une borne inférieure, probablement de l'ordre de $\frac{1}{2}$.

L'intervalle de variation de ZZ étant borné inférieurement et supérieurement par des valeurs finies, la connaissance de tous les moments (ou cumulants) de ZZ détermine la fonction de distribution de ZZ (I p. 110).

3.3.

On peut, théoriquement, calculer tous les moments de ZZ car on peut calculer les espérances mathématiques des produits $u_i^\alpha u_j^\beta \dots u_k^\gamma$ quel que soit le nombre de termes. Il ne semble pas possible en fait de le faire, même en utilisant des semaines d'ordinateur puissant.

Nous nous sommes contenté d'aller jusqu'au 4^e ordre.

$$E(ZZ) = \frac{n+1}{6}$$

$$\text{Variance} = \frac{(n-2)(n-3)}{90}$$

$$\text{Cumulant d'ordre 3} = \frac{(n-2)(n-3)(n-5)(2n-1)}{945(n+3)}$$

$$\text{Cumulant d'ordre 4} = \frac{(n-2)(n-3)(3n^4 - 39n^3 + 112n^2 - 9n - 91)}{5 \times 945(n+3)(n+5)}$$

La valeur limite du coefficient d'asymétrie γ_1 est de $\frac{2\sqrt{10}}{7}$ si $n \rightarrow \infty$, et la valeur limite du coefficient d'aplatissement est $\frac{36}{7}$: la distribution de ZZ ne tend pas vers une distribution normale.

3.4. APPROXIMATION DE LA DISTRIBUTION DE ZZ

Étant donné l'allure de la variation de γ_1 et de γ_2 avec n , il semble qu'une distribution β incomplète généralisée puisse, a priori, représenter convenablement la distribution de ZZ : fonction de densité :

$$\frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} u^{\alpha-1} (1-u)^{\beta-1} du \quad \text{avec } u = \frac{ZZ - x_0^{1/\delta}}{s}$$

Nous avons été conduit à choisir :

$$\alpha = 1 \quad \beta = \left(\frac{n-1}{4}\right)^2$$

δ , x_0 et s sont des fonctions, compliquées, de n dont nous n'avons pas cherché d'approximation. Le programme FORTRAN qui calcule une probabilité d'une valeur ZZ calcule en premier lieu les valeurs δ , x_0 , s et β pour que, n étant connu, les trois premiers cumulants de l'approximation soient égaux aux trois premiers cumulants de la distribution de ZZ.

Il s'en suit que l'erreur sur le coefficient d'aplatissement γ_2 , nulle ou quasiment nulle pour $n = 5, 6, 7$ croît jusqu'à $+0,0726$ pour $n = 14$ (γ_2 de la distribution approchée supérieur au γ_2 de la distribution réelle), décroît ensuite pour être quasiment nulle pour $n = 29$ puis augmente en valeur absolue lorsque n croît ($-0,0568$ pour $n = 40$, $-0,1755$ pour $n = 80$, $-0,2764$ pour $n = 200$).

L'erreur absolue relative du coefficient γ_2 de la distribution approchée par rapport au coefficient γ_2 de la distribution de ZZ est inférieure, ou égale, à 5%, de $n = 15$ à $n = 140$.

Il semble que la distribution approchée soit très largement suffisante dans la gamme des valeurs de n qui nous intéressent (vérification par quelques séries de tirages au hasard).

3.5. ESSAIS SUR ZZ

Quelques essais ont été faits, par tirage au hasard, sur le comportement du test ZZ par rapport aux tests de Mann et de Spearman. Nous donnons les résultats d'une série d'essais, sans les valeurs correspondant au test de Spearman qui sont très voisines de celles correspondant au test de Mann (cf. tableau).

Colonne A1 répartition de 200 échantillons chronologiques sans effets de persistance, de tendance, ou pseudo-cycliques, et sans erreurs systématiques. Il est normal de trouver environ 10 échantillons dans chaque classe de probabilité, par le simple fait du hasard, que le classement soit fait d'après le test de Mann ou le test ZZ.

Colonne A2 répartition des 200 échantillons chronologiques affectés d'une erreur systématique : dans chaque échantillon de A1, chacune des 20 dernières valeurs chronologiques a été multipliée par 1,075. Le test de Mann interprète l'erreur systématique comme une tendance et est un peu supérieur au test ZZ.

Colonne A3 répartition des 200 échantillons chronologiques affectés d'une erreur systématique : dans chaque échantillon des A1, chacune des 20 valeurs au milieu (de x_{16} à x_{35}) a été multipliée par 1,075. Le test ZZ a sensiblement la même puissance au seuil de 0,05 qu'en A2. Le test de Mann ne détecte rien.

Colonne A4 répartition des 200 échantillons chronologiques affectés d'une erreur systématique : dans chaque échantillon de A1, chacune des 10 premières et des 10 dernières valeurs a été multipliée par 1,075. Le test ZZ a sensiblement la même puissance qu'en A2 et A3. Le test de Mann ne détecte rien.

Colonne A5 répartition des 200 échantillons chronologiques affectés d'une erreur systématique : dans chaque échantillon de A1, chacune des 20 valeurs du milieu a été augmentée de 100 (= écart-type). Le test de Mann ne détecte rien.

Colonne A6 répartition des 200 échantillons chronologiques affectés d'une erreur systématique : dans chaque échantillon de A1, chacune des 20 valeurs du milieu a été augmentée de 40 ($= \frac{2}{5}$ écart type). Le test de Mann ne détecte rien.

TABLEAU

SÉRIE D'ESSAIS SUR 200 ÉCHANTILLONS CHRONOLOGIQUES DE TAILLE $n = 50$.
Chaque échantillon de variables aléatoires, indépendantes,
provenant d'une population mère distribuée normalement, de moyenne 1 000 et d'écart type 100

Limites de classes de probabilité au dépassement 0,00	Nombre d'échantillons trouvés dans chaque classe de probabilité, par essai													
	A1		A2		A3		A4		A5		A6		A9	
	ZZ	Mann	ZZ	Mann	ZZ	Mann	ZZ	Mann	ZZ	Mann	ZZ	Mann	ZZ	Mann
,05	11	13	94	109	87	10	95	10	146	10	30	14	64	106
,10	8	12	34	22	34	9	29	12	17	7	16	14	16	22
,15	11	11	8	13	20	14	25	14	10	10	18	8	14	12
,20	12	12	10	6	7	7	14	10	9	14	10	7	16	8
,25	5	11	8	9	11	9	7	10	3	9	15	12	15	9
,30	15	14	8	6	5	15	3	12	4	11	14	19	9	7
,35	10	9	8	9	6	16	5	9	1	17	13	8	4	5
,40	13	10	3	7	9	10	4	12	0	11	16	10	7	9
,45	12	9	3	4	4	9	0	7	1	6	6	7	7	3
,50	12	13	5	3	1	13	3	9	1	7	6	13	2	3
,55	8	4	6	0	1	5	2	14	1	9	8	4	2	2
,60	5	7	3	2	5	10	1	12	2	12	8	6	6	2
,65	16	14	2	0	1	7	1	7	1	6	7	15	12	3
,70	11	12	0	0	1	5	3	12	0	11	8	6	5	0
,75	14	9	2	3	3	15	2	5	0	9	7	10	3	1
,80	7	5	4	1	1	8	1	5	0	6	2	10	5	1
,85	7	11	0	2	1	8	2	17	0	12	6	14	5	1
,90	6	6	1	3	1	7	0	4	0	15	4	5	2	3
,95	4	10	0	0	1	13	0	9	0	9	2	10	3	3
1,00	13	8	0	1	1	10	2	10	1	9	4	8	3	0

Colonne A9 répartition des 200 échantillons chronologiques affectés d'une tendance faible : chaque valeur de rang chronologique i a été augmentée de $2i$, c'est-à-dire que la 50^e et dernière a été augmentée de 100. Le test de Mann est nettement supérieur au test ZZ.

3.6. CONCLUSION SUR LA VALEUR DU TEST ZZ

Compte tenu de quelques autres séries d'essais faits avec des échantillons de taille $n = 20$ et $n = 50$, le test ZZ semble intéressant. Il ne détecte pas de changements dans la variance (1.6), et nous pensons que pour les séries chronologiques relativement courtes qui nous intéressent, il est illusoire de chercher ces changements.

Sa puissance est facile à étudier par tirages au hasard dans le cas où un seul segment de la suite chronologique est affecté d'une erreur systématique de moyenne. Elle dépend :

- de la longueur de la suite chronologique ;
- de la longueur du segment affecté d'erreur systématique du rapport de l'erreur systématique de moyenne à l'écart type de la suite.

4. UTILISATION PRATIQUE DU TEST ZZ AVEC LES PRÉCIPITATIONS ANNUELLES

4.1. CHOIX DE LA VARIABLE

4.1.1.

Aucun essai sur précipitations annuelles de stations n'a encore été effectué. On peut espérer que la suite chronologique des variables

$$v_i = x_i - \frac{y_i}{\bar{y}} \bar{x} \quad (\text{cf. 1.4})$$

sera constituée de variables aléatoires et indépendantes dans le cas où il n'y a pas d'erreurs systématiques d'observation, et que ces variables pourront être considérées comme tirées d'une population normale de moyenne zéro. La variable u_i , réduite, à utiliser pour le calcul de ZZ est

$$u_i = v_i / \sqrt{\text{variance de } \bar{V}}.$$

Dans la pratique, l'existence d'erreurs systématiques de mesures conduit à une valeur du coefficient de variation surestimée par rapport à sa valeur réelle (qu'il aurait sans la présence d'erreurs systématiques) et à une valeur du coefficient de corrélation sous estimée par rapport à sa valeur réelle, la variance de \bar{V} est surestimée, la valeur du test ZZ sous estimée. En supposant que X est une station homogène, il est probable qu'on aura avantage à estimer la variance de \bar{V}

$$2(1 - \rho) \text{ variance de } x$$

en calculant ρ avec $\sqrt{\text{variance de } \bar{y}} = \bar{y}$ coefficient de variation de x .

4.1.2.

Quelques essais ont été effectués entre précipitations annuelles de stations et le vecteur Hiez calculé d'après ces stations. c_i étant la composante de l'année i du vecteur Hiez, la variable C peut être considérée comme homogène, mais son coefficient de variation est, nécessairement inférieur aux coefficients de variation des stations. La variable V ci-dessus n'est plus utilisable.

Nous avons essayé entre autres la variable

$$a_i = \frac{x_i}{c_i}$$

qui semble répartie à peu près normalement et aussi la variable

$$b_i = \frac{x_i}{\sqrt{\text{var } x}} - \frac{c_i}{\sqrt{\text{var } C}}.$$

Ces deux variables semblent conduire aux mêmes résultats pour la valeur de ZZ.

La variance de B est $2(1 - \rho)$, ρ coefficient de corrélation entre la station X et le vecteur Hiez. Si la station X n'est pas homogène, la valeur de ρ sera inférieure à la valeur du coefficient de corrélation entre le vecteur Hiez et une station X qui serait homogène, et la variance de B est surestimée. Il semble bien que, lorsque toutes les stations sont homogènes, les valeurs des coefficients de corrélation entre les stations et le vecteur Hiez soient d'une part élevées, d'autre part du même ordre de grandeur. Il semble donc possible de prendre pour valeur, dans variance de B, du coefficient de corrélation entre le vecteur Hiez et une station X (pour laquelle le ρ calculé est faible et qui n'est sûrement pas homogène) la valeur moyenne des quelques coefficients de corrélation les plus forts.

4.2.

L'important dans la recherche de la signification du test ZZ est de ne pas se laisser obnubiler par un seuil « de signification ».

4.2.1.

Si on utilise le test ZZ entre une station parfaitement homogène et 200 autres stations parfaitement homogènes (on est dans le cas de la colonne A1 du tableau) il sera normal, et heureux, de trouver une dizaine de fois une valeur de ZZ supérieure à la valeur correspondant à la probabilité de 0,05 au dépassement.

4.2.2.

Si on utilise le test ZZ entre une station d'homogénéité inconnue, à vérifier, et 200 autres stations parfaitement homogènes, et si on trouve 144 valeurs de ZZ supérieures à la valeur correspondant à la probabilité de 0,50 (cas de la colonne A6), le résultat est hautement significatif d'un défaut d'homogénéité.

En effet, on peut considérer qu'on a fait $m = 200$ tirages au hasard dans une urne composée, en proportions,

de $q = \frac{1}{2}$ boules blanches (probabilités $> 0,5$) et $p = 1 - q = \frac{1}{2}$ boules noires (probabilités $< 0,5$). L'application de l'approximation normale à la loi binomiale conduit à la valeur, pour le cas ci-dessus, de la variable réduite de gauss, de $u = \frac{144 - mp}{\sqrt{m pq}} = 6,2$ de probabilité (au dépassement) de l'ordre de 10^{-10} .

4.3.

Dans la pratique, l'application des remarques ci-dessus est plus difficile car toutes les m stations sont soupçonnées de défaut.

Il semble possible d'utiliser la méthode suivante (essais à faire) :

Partir d'une suite chronologique homogène x (soit suite chronologique d'une station sûre, bien vérifiée par doubles masses avec d'autres stations sûres, soit le vecteur régional Hiez obtenu après les principales corrections).

Calculer le coefficient de corrélation linéaire ρ entre la suite x et la suite y douteuse de la même zone climatique. Tirer au hasard m (mettons 100) suites W_ℓ

$$W_{\ell i} = \rho x_i + \sqrt{(1 - \rho^2) \text{var } x} \varepsilon_i$$

ε_i étant une variable normale de moyenne 0, variance 1.

Les suites W_ℓ (ℓ de 1 à m) ont des moyennes, des variances voisines de celles de la suite x à laquelle elles sont liées par des coefficients de corrélation voisins de ρ , et elles sont homogènes (au hasard près) : elles sont soumises aux mêmes effets de persistance, de tendance, ou pseudo-cycliques que la suite x .

Utiliser le test ZZ entre la suite y et les m suites W_ℓ (cf. 4.2) sans oublier que par le simple fait du hasard, on pourra quelquefois trouver des résultats significatifs de défaut (inexistant en fait) d'homogénéité.

4.4.

On peut utiliser ZZ pour chercher les points probables de cassures : en prenant des sous-suites dans la suite chronologique, en faisant varier la première et la dernière année de sous-suite, on peut détecter des points A et B (fig. 2) en cherchant la valeur minimale de la probabilité au dépassement de ZZ. Mais puisque c'est une recherche de valeur maximale du test, la répartition de cette valeur maximale n'est pas celle de ZZ, et il ne semble pas facile de déterminer sa signification.

NOTES

Test de Mann

Si la taille de l'échantillon est supérieure à $n = 10$, la statistique LQ du test peut être considérée comme distribuée normalement, de moyenne : $\frac{n(n-1)}{4}$

$$\text{de variance } \frac{n(n-1)(2n+5)}{72}$$

La valeur minimale du test est 0 et sa valeur maximale

$$\frac{n(n-1)}{2}$$

En FORTRAN, le calcul est simple :

```

LQ = 0
DØ 16 i = 2,N
XT = X (i - 1)
DØ 16 J = i,N
iF (X (J) · LT · XT) LQ = LQ + 1
16 CØNTINUE
LQ = MINO (LQ, (N * (N - 1))/2 - LQ)

```

Test de Spearman

Si la taille de l'échantillon est supérieure à $n = 10$, la statistique RS du test peut être considérée comme distribuée suivant une transformation d'une loi de Student à $(n - 2)$ degrés de liberté en posant t (Student) =

$$RS \sqrt{\frac{n - 2}{1 - RS^2}}$$

En FORTRAN, le calcul est simple :

```

RS = 1.
DØ 18 i = 1,N
R = 1.
DØ 17 J = 1,N
iF (X (J) · LT · X (i)) R = R + 1.
17 CØNTINUE
18 RS = RS - (R - i) * (R - i)/N/(N * N - 1.) * 6.
RS = ABS (RS)

```