

*Conf :
Colloque "Día mundial de la Meteorología"
Instituto Nacional de Meteorología e Hidrología
INAMHI, Quito, 03/1988.*

HOMOGENEISATION DES DONNEES PLUVIOMETRIQUES

par Patrick LE GOULVEN *

mars 1988

Nota Bene

Ce texte fait largement référence aux théories développées par Yves Brunet-Moret (4) et Gérard Hiez (6). Ces 2 hydrologues de l'ORSTOM se sont intéressés tout particulièrement au comportement régional des précipitations annuelles dans le but d'élaborer des vecteurs représentatifs de chaque région.

hydrologue ORSTOM, Mission ORSTOM, CP 17-11-06596, Quito, Equateur.

INTRODUCTION

L'inventaire des ressources hydriques et leur utilisation rationnelle sont des préoccupations d'un grand nombre de pays afin d'augmenter la production agricole par l'irrigation, de pourvoir en eau les agglomérations et industries toujours plus concentrées et de satisfaire la demande d'énergie.

Néanmoins, ces objectifs sont étroitement liés à une bonne connaissance des facteurs du climat parmi lesquels la pluviométrie tient un rôle prépondérant.

Les précipitations sont un phénomène très variable dans le temps et l'espace, surtout dans un pays au relief aussi contrasté que l'Équateur.

Sa connaissance, surtout quantitative, s'obtient par des mesures permanentes sur un réseau d'observations ponctuelles. Malheureusement, toute mesure est entachée d'erreurs diverses qui proviennent tant du système d'acquisition des données que du caractère individuel de la mesure.

Cette constatation démontre la nécessité d'une sévère critique des données obtenues qui peuvent être éliminées quand elles s'écartent trop de la réalité, corrigées avec précaution lorsque la nature systématique des erreurs détectées le permet ou bien simplement signalées quand leur valeur anormale peut être considérée comme localement probable.

L'ensemble de ces opérations constitue l'homogénéisation des données.

Ici on ne traitera que des données pluviométriques et dans la conclusion on donnera des recommandations pour les autres facteurs climatiques.

1. DÉFINITION DU PROBLÈME

1.1. Homogénéité d'une série chronologique de précipitations annuelles

Une série chronologique de valeurs observées est de caractère aléatoire simple si toutes les valeurs sont indépendantes et si elles proviennent d'un tirage au sort parmi une même population statistique. Le caractère aléatoire peut être altéré par :

- *un effet de persistance* : une valeur n'est pas indépendante de la (ou des) précédente (s) (processus de MARKOV ou autre), mais la série est stationnaire et les paramètres de distribution de la population ne varient pas avec le temps ;
- *un effet de tendance* : les valeurs observées proviennent de différentes populations dont les probabilités mathématiques augmentent ou diminuent avec le temps ;
- *des effets cycliques ou pseudo-cycliques* : pour une valeur donnée, la probabilité mathématique dépend de la chronologie, mais la probabilité des moyennes des séries peut être considérée comme stationnaire ;
- *des erreurs systématiques (ou non) d'observation et de mesure* : seules ces erreurs modifient l'homogénéité d'une série chronologique de précipitations annuelles qui peut, en plus, présenter des effets de persistance, de tendance (au sud de la région côtière) ou de pseudo-cycles (phénomène du Niño) qui détermineront son organisation interne.

1.2. Les principales causes d'hétérogénéité

Modification de l'environnement du site de mesure

- Par déplacement de l'appareil

Ceci est un cas fréquent et souvent la station conservera son nom tout en suivant les déménagements de son observateur. Généralement, ces déplacements sont faibles (de l'ordre du kilomètre) mais dans un pays avec tant de relief, ils peuvent provoquer de grandes différences dans la série de mesures si on modifie l'exposition de l'appareil ou si le changement d'altitude est important.

- Par modification de l'environnement lui-même

Cette modification peut être brusque (construction proche) ou progressive (croissance d'arbres à proximité de l'appareil).

- Par changement de la hauteur de l'appareil

On peut trouver des différences de l'ordre de 50 % entre un appareil situé à 1,20 m du sol et un autre placé au sol. De telles différences sont dues aux turbulences qui se forment autour de l'appareil dans les lieux très exposés aux vents [1], [2], [3].

Erreurs dues à l'appareil

- Modification de la surface réceptrice par construction, échange ou déformation

Si les engins « standard » ont une surface réceptrice constante et connue, il n'en est pas de même des pluviomètres totalisateurs de fabrication artisanale.

- Erreurs d'étalonnage

Un vice de construction ou une erreur d'étalonnage peuvent se produire dans le cas des pluviographes.

Erreurs de mesure ou d'enregistrement

- Au niveau de l'éprouvette (ou de la réglette)

Les erreurs peuvent provenir de précisions différentes d'une éprouvette à l'autre ou de lectures incorrectes lorsque l'éprouvette n'est pas verticale ou bien de confusion de chiffres, etc. Le cas le plus typique est celui de confusion d'éprouvette. Dans les pays utilisant le système métrique, les pluviomètres les plus courants ont une superficie de 200, 314 ou 400 cm². Il est alors assez fréquent que les mesures soient faites avec une éprouvette inappropriée.

Cette erreur est décelée facilement quand on connaît les différentes surfaces mais elle peut aussi être confondue avec un changement d'emplacement, surtout dans les régions montagneuses où existent de forts gradients altitudinaux des précipitations.

Un autre cas difficile à détecter se présente lorsque l'observateur dispose, par sécurité, de deux éprouvettes (différentes évidemment) et utilise l'une et l'autre.

- Au niveau de l'enregistrement

- manque d'encre,
- erreurs de dates,

- notation erronée des hauteurs mesurées,
- erreurs de transcription, etc.

Erreurs de transmission

Perte du courrier ou mauvais fonctionnement des systèmes de transmission (électrique, électronique, par satellite, etc.).

Erreurs lors de l'archivage et de la publication

Lors de la collecte et de la transcription des données brutes, peuvent se produire des erreurs de copie ou de saisie et la publication des archives donne lieu à toutes les erreurs d'écriture, (oubli de dates, erreurs de stations, etc.).

1.3. Les conséquences

Comme il l'a été dit auparavant, toutes ces erreurs altèrent l'homogénéité des séries chronologiques des précipitations annuelles, mais de manière différente selon le problème.

Un changement de l'environnement se détecte si l'on dispose de longues séries homogènes avant et après ce changement. Plus le changement est marqué, moins d'années seront nécessaires pour le détecter. L'expérience montre que dans ce cas, on peut corriger les totaux annuels et mensuels pour homogénéiser la série mais qu'il est toujours impossible de modifier les données quotidiennes.

La confusion d'éprouvette est un cas très fréquent et se décèle avec un faible nombre d'années (à condition que l'observateur n'utilise pas en même temps toute une collection d'éprouvettes et de réglottes). Lorsque l'on connaît la date du changement, on reconstitue sans peine les données annuelles, mensuelles et quotidiennes.

Dans le cas de déplacements fréquents de l'appareil (ou d'autres causes produisant le même effet), les hétérogénéités ne peuvent être discernées et les corrections sont impossibles.

1.5. Nécessité d'effectuer des « tests » d'homogénéité

Les diverses causes d'hétérogénéité et les conséquences de celle-ci montrent la nécessité de contrôler rigoureusement les données pluviométriques, ce qui pourrait se faire, dans la plupart des cas, en consultant l'historique de la station.

L'expérience montre qu'un changement de site coïncide généralement avec un changement d'observateur et que la confusion d'éprouvettes (ou réglottes) se produit après une interruption des mesures. Cela signifie qu'un bon historique où soient signalés les changements de site, d'observateurs ou d'appareils et ceux de l'environnement, et une vérification des dimensions des pluviomètres et des éprouvettes (ou réglottes), permettraient de résoudre de nombreux problèmes.

Malheureusement, ces historiques sont en général inexistantes ou d'accès difficile. L'historique peut être partiellement reconstitué à partir des documents originaux envoyés par les observateurs, mais cela est insuffisant pour l'analyse d'une série chronologique complète.

D'où la nécessité d'effectuer des tests d'homogénéité sur la base des simples données annuelles et, ultérieurement, s'il y a un problème difficile, de faire une vérification sur le terrain.

1.4. Les différents tests d'homogénéité

Dans la bibliographie on trouve divers tests d'homogénéité (test des signes, de MANN-WHITNEY, de FISHER-YATES, etc.) qui pourraient servir à la vérification de l'homogénéité d'une série pluviométrique. Ils vérifient généralement le caractère aléatoire d'une série de variables indépendantes en admettant l'hypothèse de l'absence d'une organisation chronologique et donc de pseudo-cycles de l'échantillon observé, ce qui contredit la réalité (succession de périodes sèches et humides, phénomènes du Niño, etc.).

De par leur conception, ces tests peuvent être utilisés après avoir déterminé le moment de rupture séparant des séries homogènes car alors on compare une série provenant de l'échantillon au reste.

En faisant varier la série de valeurs, on pourrait trouver la date la plus probable de la rupture à condition que celle-ci soit unique, mais dans les cas complexes de plusieurs ruptures et de sens contraire, ces tests ne peuvent rien détecter, car ils n'ont pas été élaborés dans ce but.

Leur efficacité est inférieure ou égale au test de STUDENT de comparaison de moyennes provenant d'échantillons correspondant à des problèmes courants.

En résumé, il est facile de démontrer que ces tests ne peuvent détecter les erreurs systématiques et la présence réelle d'une organisation interne des séries pluviométriques correspondant à des effets pseudo-cycliques ou de persistance.

Malheureusement, la plupart des séries pluviométriques contiennent au moins l'un de ces deux effets, ce qui rend l'utilisation des tests classiques d'homogénéité peu intéressante.

2. LES TESTS DES SIMPLES ET DOUBLES MASSES

2.1. Hypothèse de base

On admet que les stations pluviométriques situées dans une même zone climatique ont des totaux pluviométriques annuels pseudo-proportionnels indépendamment des fluctuations climatiques de la zone, c'est-à-dire que pour l'année i , les précipitations annuelles x_i et y_i de deux stations X et Y de la même zone peuvent s'écrire [4] :

$$(1) \quad x_i = \frac{\bar{x}}{\bar{y}} y_i + e_i$$

où \bar{x} et \bar{y} correspondent aux moyennes inter annuelles des deux stations pendant une période suffisamment longue. La variance de e_i (terme aléatoire indépendant tant de x_i comme de y_i) est d'autant plus petite si le coefficient de corrélation linéaire entre les deux stations est plus fort et son espérance mathématique est égale à zéro.

2.2. Graphiques des simples masses

Suivant l'hypothèse précédente, il est raisonnable de penser que l'information climatique annuelle de la zone considérée puisse être représentée par un vecteur V dont les valeurs annuelles seraient la superposition de deux variables asynchrones :

- l'une représentant l'éventuelle fluctuation de la moyenne au long du temps (sur une longue période) ;
- l'autre rendant compte de la variation annuelle autour de la moyenne (sur une courte période, l'année).

STATION X P311MI CAHUASQUI PROVINCE D'IMBABURA PÉRIODE 1963-1985 COEFFICIENT DE CORRECTION 1 000

	ANNÉES	TA (mm)	P							
			+ 3 I	- 2 I	- 1 I	0,0 I	+ 1 I	+ 2 I	+ 3 I	
1	1984	874,7								
2	1983	755,6								
3	1982	1 051,9								
4	1981	730,4								
5	1980	774,5								
6	1979	677,5								
7	1978	509,4								
8	1977	510,0								
9	1976	760,0								
10	1975	878,5								
11	1974	1 260,1								
12	1973	667,2								
13	1972	923,9								
14	1971	850,8								
15	1970	879,1								
16	1969	876,9								
17	1968	777,2								
18	1967	658,1								
19	1966	727,9								
20	1965	933,7								
21	1964	658,8								

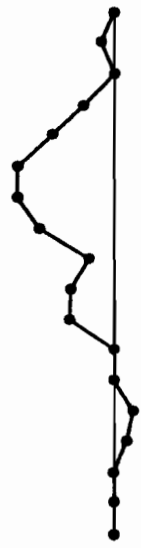


MOYENNE DES 21 ANS = 797,0
 ÉCART-TYPE = 171,9
 COEFFICIENT DE VARIATION = 0,216

TEST DES SIMPLES MASSES

STATION X P302MI MIRA PROVINCE DE CARCHI PÉRIODE 1964-1984 COEFFICIENT DE CORRECTION 1 000

ANNÉES	TA (mm)	P + -3 -2 -1 0,0 +1 +2 +3							
		I	I	I	I	I	I	I	
1	1983	632,2							
2	1981	524,5							
3	1980	629,3							
4	1979	429,7							
5	1978	436,1							
6	1977	440,9							
7	1976	602,4							
8	1975	754,4							
9	1974	876,8							
10	1973	534,1							
11	1972	597,3							
12	1971	851,0							
13	1970	697,3							
14	1969	703,4							
15	1968	530,2							
16	1967	543,2							
17	1966	583,6							
18	1965	645,5							
N	AN	T A	P + -3	-2	-1	0,0	+1	+2	+3
			I	I	I	I	I	I	I



MOYENNE DES 18 ANS = 611,7
 ÉCART-TYPE = 129,4
 COEFFICIENT DE VARIATION = 0,212

TEST DES SIMPLES MASSES

Pour plus de commodité, les composantes w_i de ce vecteur représenteront les valeurs de la différence entre les totaux annuels et la moyenne exprimés en termes d'écart-type

$$w_i = \left(v_i - \frac{v}{ETV} \right)$$

Dans ce cas, les totaux annuels d'une station X appartenant à cette zone climatique peuvent s'écrire de la façon suivante :

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

où r_x représente le coefficient de corrélation linéaire entre la station et le vecteur et a_i les fluctuations propres de la station.

Si on cumule les données jusqu'à l'année n , on obtient :

$$(3) \quad \sum_1^n x_i = n\bar{x} + \sigma_x \left(r_x \sum w_i + \sum a_i \sqrt{1 - r_x^2} \right)$$

et en divisant par la moyenne :

$$(4) \quad X_n = \frac{1}{\bar{x}} \sum x_i = n + C_x \left(r_x \sum w_i + \sum a_i \sqrt{1 - r_x^2} \right)$$

où C_x représente le coefficient de variation de la variable x .

X_n est proportionnel à n , plus un terme aléatoire obéissant au processus de MARKOV.

Le graphique de la relation $X_n = f(n)$ est un graphique de simples masses relatives.

Dans la pratique, on utilise en coordonnées :

- le rang i de l'année dans la série (croissante ou décroissante et en omettant les années sans données) ;
- la somme $\sum x_i$ des précipitations annuelles.

Le tracé oscillera (effet de Slutsky) entre deux droites parallèles qui délimitent une bande sensiblement parallèle à la première bissectrice.

Si la moyenne est affectée par une déformation systématique due aux erreurs signalées en 1.2., le second terme de l'équation précédente aura une valeur significativement différente de zéro pour cette période et le tracé des points se situera dans une bande de pente différente à celle de la première bissectrice.

Ce changement de pente est facilement décelé.

Le graphique des simples masses a donc l'avantage de trouver toutes les périodes pendant lesquelles la moyenne a été perturbée de manière significative, ce que les tests classiques ne peuvent faire.

Malheureusement, dans la pratique, il ne permet pas de distinguer avec certitude d'éventuelles déviations systématiques d'un changement de tendance climatique bien marqué (comme par exemple dans les cas de Cahuasquí, Mira et FF-CC-Carchi dont on verra les doubles masses plus loin).

On ne l'utilisera que pour une première analyse pour détecter des erreurs importantes et le diagnostic final sera réservé aux tests des doubles masses.

2.3. Graphique des doubles masses

On considère deux stations d'une même zone climatique représentée par le vecteur V décrit ci-dessus.

Les précipitations annuelles de chaque station peuvent s'écrire de la façon suivante :

$$(5) \quad x_i = \bar{x} + \sigma_x \left(w_i \cdot r_x + a_i \sqrt{1-r_x^2} \right)$$

$$(6) \quad y_i = \bar{y} + \sigma_y \left(w_i \cdot r_y + b_i \sqrt{1-r_y^2} \right)$$

Si on suppose qu'elles sont liées à la tendance climatique par les coefficients r_x et r_y et qu'elles ont des fluctuations propres et indépendantes a_i et b_i , en cumulant les variables jusqu'à l'année n et en divisant par la moyenne, on obtient :

$$(7) \quad X_n = n + C_x \left(r_x \sum w_i + \sum a_i \sqrt{1-r_x^2} \right)$$

$$(8) \quad Y_n = n + C_y \left(r_y \sum w_i + \sum b_i \sqrt{1-r_y^2} \right)$$

où C représente les coefficients de variation respectifs.

En éliminant n entre les deux équations et en regroupant les différents termes, on obtient :

$$(9) \quad Y = X_n + \sum w_i (r_y \cdot C_y - r_x \cdot C_x) + \left(C_y \sum b_i \sqrt{1-r_y^2} - C_x \sum a_i \sqrt{1-r_x^2} \right)$$

L'équation traduit une relation linéaire altérée par deux termes dont le dernier est une variable aléatoire suivant un processus de MARKOV [6].

Dans la pratique, les points de coordonnées (X_n, Y_n) seront compris dans une bande définie par deux droites parallèles de pente positive et égale à 1. Si les variables sont indépendantes, la largeur de la bande dépendra uniquement des coefficients de corrélation de chaque station.

Dans le cas contraire, elle dépendra également des valeurs des coefficients de corrélation et sera d'autant plus étroite que les coefficients seront plus forts.

TOUTE VARIATION DE L'UNE DES MOYENNES PAR MODIFICATION DES CONDITIONS D'OBSERVATION SE TRADUIRA PAR UNE MODIFICATION CONCOMITANTE DE LA PENTE DE LA BANDE.

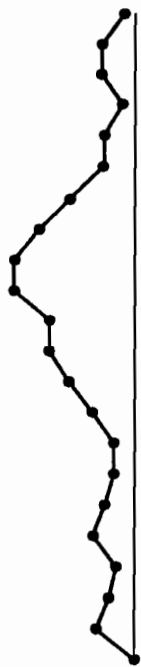
Le changement de pente correspondra à la variation de la moyenne altérée.

Si les deux stations appartiennent réellement à la même zone climatique, les coefficients de variation ainsi que les coefficients de corrélation auront des valeurs semblables, ce qui implique que le second terme sera proche de zéro. Cette constatation démontre l'avantage des doubles masses : dans une même zone climatique, la relation entre deux variables pluviométriques cumulées est indépendante d'éventuelles fluctuations climatiques et n'est affectée que par les fluctuations aléatoires si les données sont homogènes.

Cela signifie que les changements de pente correspondent seulement aux erreurs systématiques décrites en 1.2., ainsi que le montrent les graphiques de FF-CC Carchi/Mira et Mira/ Ca-huasquf.

STATION X P301MI FF-CC-CARCHI PROVINCE DE CARCHI PÉRIODE 1956-1985 COEFFICIENT DE CORRECTION 1 000

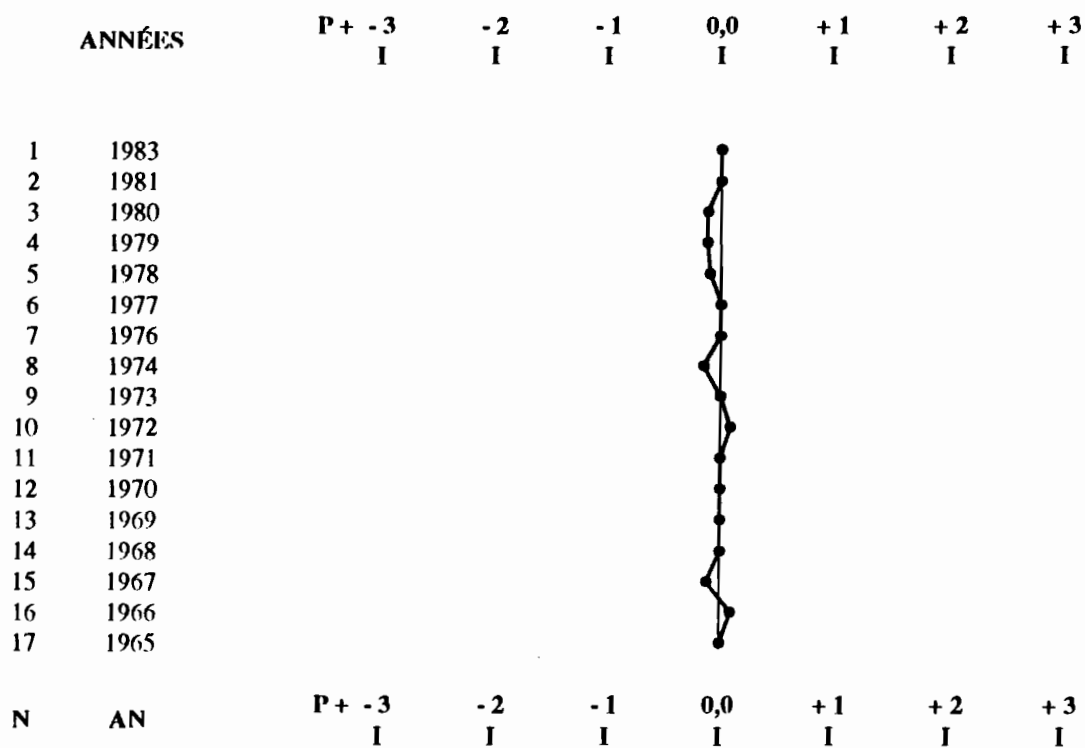
	AÑOS	TA (mm)	COEFFICIENT DE CORRECTION 1 000						
			P + -3 I	-2 I	-1 I	0,0 I	+1 I	+2 I	+3 I
1	1985	495,8							
2	1984	425,7							
3	1983	556,1							
4	1982	663,6							
5	1981	449,2							
6	1980	533,8							
7	1979	394,0							
8	1978	352,2							
9	1977	486,4							
10	1976	516,2							
11	1974	747,5							
12	1973	542,4							
13	1972	671,2							
14	1971	636,2							
15	1970	662,2							
16	1969	575,1							
17	1968	492,5							
18	1967	460,4							
19	1966	657,8							
20	1965	516,5							
21	1962	517,0							
22	1956	783,5							
N	AN	TA	P + -3 I	-2 I	-1 I	0,0 I	+1 I	+2 I	+3 I



MOYENNE DES 21 ANS = 551,6
 ÉCART-TYPE = 112,2
 COEFFICIENT DE VARIATION = 0,203

TEST DES SIMPLES MASSES

STATION X P301MI	FF-CC-CARCHI	PROVINCE DE CARCHI	PÉRIODE 1964-1984	COEFFICIENT DE CORRECTION 1 000
STATION X P302MI	MIRA	PROVINCE DE CARCHI	PÉRIODE 1964-1984	COEFFICIENT DE CORRECTION 1 000

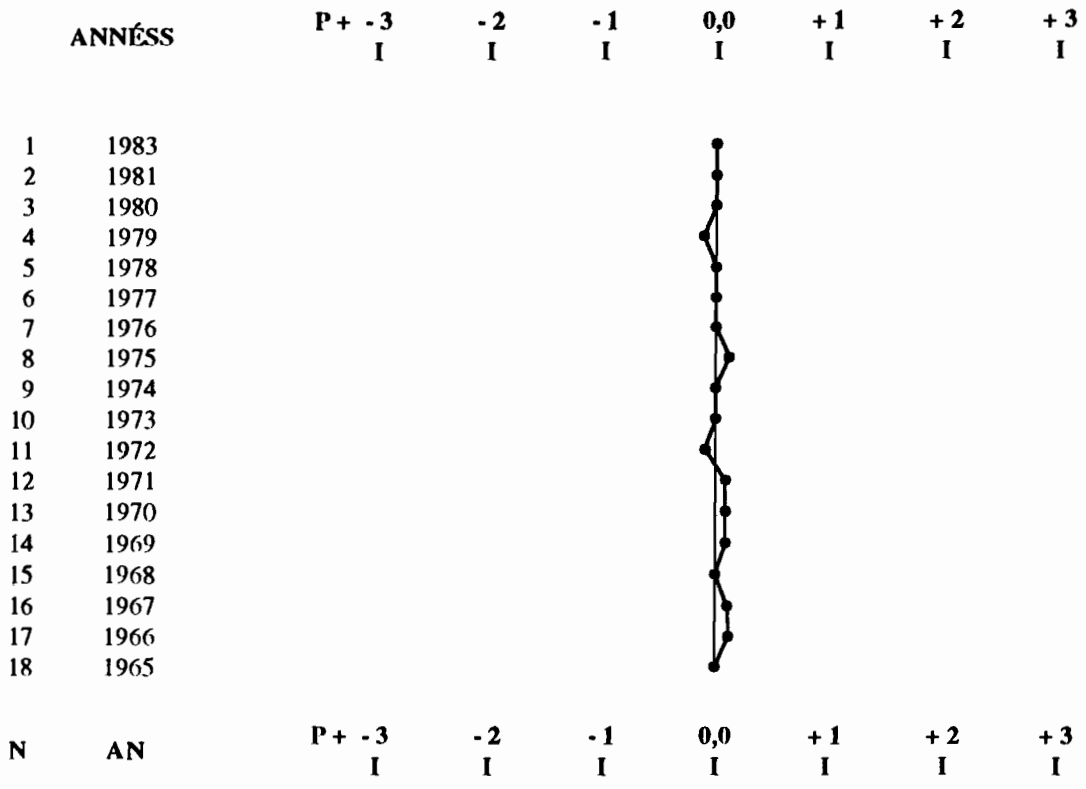


XM = 544,1 XM/YM = 902
 YM = 603,3 YM/XM = 1 109
 COEFFICIENT DE CORRÉLATION = 0,800

TEST DES DOUBLES MASSES

STATION X P302MI MIRA PROVINCE DE CARCHI PÉRIODE 1964-1984 COEFFICIENT DE CORRECTION 1 000

STATION X P311MI CAHUASQUI PROVINCE DE IMBABURA PÉRIODE 1964-1984 COEFFICIENT DE CORRECTION 1 000



XM = 611,7 XM/YM = 778
 YM = 786,2 YM/XM = 1 285
 COEFFICIENT DE CORRÉLATION = 0,833

TEST DES DOUBLES MASSES

2.4. Considérations sur les graphiques des simples et doubles masses

La première opération consiste à déterminer, de façon empirique et en se basant sur les distances entre elles, les effets du relief, le type de régime intra-annuel des pluies, des groupes de stations qui appartiennent à une même zone climatique.

Ensuite, à l'intérieur des groupes, on compare chacune des stations à toutes les autres, ce qui signifie que pour un groupe de m stations, le nombre des doubles masses atteint $m(m-1)/2$.

On remarque que le traitement de la pluviométrie d'une région ou d'un pays demande la manipulation d'une grande masse d'informations qui n'est possible que grâce à l'automatisation du traitement des données.

Le programme MASA utilisé actuellement par l'INERHI est un bon exemple de traceur automatique des graphiques des simples et doubles masses.

Son utilisation rend plus facile la comparaison entre les stations en supprimant la plupart des opérations manuelles et en rendant possible une facile interprétation visuelle.

La meilleure méthode pour utiliser les doubles masses est de rechercher, dans une zone climatique déterminée, quelques stations ayant de longs registres dont on est assuré de l'homogénéité (stations de base).

Ensuite, on les compare aux stations voisines que l'on corrige à chaque changement de pente. À chaque correction, celle-ci est vérifiée par comparaison avec les stations de base.

On peut ainsi créer d'autres stations de base et progresser lentement en vérifiant chaque fois les résultats en les comparant à ceux des stations les plus fiables.

On est assuré, opérant minutieusement de cette manière, de pouvoir détecter les changements de pente et d'en faire un diagnostic généralement vérifié soit sur le terrain soit dans l'historique de la station. Cette méthodologie est utilisée dans de nombreux pays et plusieurs générations d'hydrologues ont prouvé son efficacité.

Néanmoins, la réalité est plus complexe. Il n'est pas toujours possible de trouver dès le début des stations fiables et, dans ce cas, la diversité des anomalies d'une ou de l'autre station rend difficile de trouver des critères réellement objectifs.

Il est donc intéressant de construire, à partir des stations situées dans une même zone climatique et sans intervention manuelle, une variable de référence sans années incomplètes.

3. CONSTRUCTION D'UN VECTEUR DE RÉFÉRENCE

3.1. Principales caractéristiques

La comparaison des stations qui appartiennent à une même zone climatique et à un vecteur de référence représentatif de la tendance climatique de cette zone permet de réduire le nombre d'analyses. Dans une zone qui englobe m stations on ne calculera que m doubles masses (au lieu de $m(m-1)/2$ comme auparavant), ce qui diminue significativement le volume des calculs.

Si z représente ce vecteur de référence, on aura :

$$(10) \quad z_i = \bar{z} + \sigma_z w_i$$

En cumulant les variables jusqu'à l'année n et en divisant par la moyenne \bar{z} , on obtient :

$$(11) \quad z_n = \frac{1}{\bar{z}} \sum z_i = n + C_z \sum w_i$$

En comparant ce vecteur avec une station X et en éliminant n dans les deux équations comme il a été fait auparavant pour les équations (7) et (8), on obtient :

$$(12) \quad X_n = Z_n + \sum w_i (r_x \cdot C_x - C_z) + C_x \sqrt{1 - r_x^2} \sum a_i$$

Comme il l'a été dit précédemment, le deuxième terme du second membre de l'équation tend vers 0 et le troisième terme aura pour variance :

$$(13) \quad V_1 = C_x^2 - 2C_x \cdot C_z \cdot r_x + C_z^2$$

Lors de la comparaison station par station, la variance du troisième terme de l'équation (9) peut s'écrire :

$$(14) \quad V_2 = C_x^2 - 2C_x \cdot C_y \cdot r_x \cdot r_y + C_y^2$$

Avec des coefficients de variation similaires, on vérifiera que $V_1 < V_2$, ce qui signifie que la largeur de la bande sera plus réduite et que, par conséquent, les doubles masses entre station et vecteur seront plus précises que celles entre stations.

3.2. Conditions de la construction

Comme il l'a été signalé auparavant, l'utilisation des doubles masses requiert l'hypothèse de pseudo-proportionnalité dans un espace donné et le vecteur de référence devra s'y ajuster mais devra également prendre en compte quelques considérations supplémentaires :

- sa construction ne doit pas exiger d'hypothèses préliminaires sur la distribution des totaux pluviométriques annuels ;
- le total pluviométrique d'une station pour une année donnée ne peut servir, *a priori*, de valeur de référence ;
- on suppose que dans l'ensemble des données existe une valeur estimative de la tendance climatique régionale qui est plus représentative que l'information partielle de chacune des stations ; ceci n'exclut pas l'existence d'anomalies purement locales (phénomène du Niño, par exemple), mais ces anomalies ne doivent pas être utilisées pour la construction du vecteur ;
- la construction du vecteur sera conçue de façon à ce que toute l'information disponible soit utilisée sans que les données fausses aient un effet sensible sur le résultat, mais, dans la pratique on est amené à éliminer les données les plus fausses lorsque le résultat n'est pas satisfaisant ; par contre, on n'éliminera pas systématiquement les stations avec des registres très courts (3 à 5 années) car leurs données peuvent contribuer à la construction d'une composante annuelle du vecteur.

3.3. Analyse de la matrice des données

On peut considérer l'ensemble des données pluviométriques comme une matrice A dont les colonnes représentent les stations et les lignes les années.

$$A = [X_{ij}] = \begin{matrix} & \left| \begin{array}{c} i = 1, \dots, n \\ \dots \\ j = 1, \dots, m \end{array} \right. \end{matrix}$$

Les m données obtenues durant une année i représentent les coordonnées d'un vecteur (pluie) annuel dans l'espace des stations.

Si la pseudo-proportionnalité était stricte, l'ensemble des n vecteurs constitueraient un ensemble de vecteurs colinéaires. Alors la matrice A pourrait être remplacée par une matrice B produit d'un vecteur ligne L et d'un vecteur colonne C .

$$(16) \quad B = L \times C$$

Dans la pratique, les données réelles ne suivent pas ce modèle théorique et existent des fluctuations aléatoires dans chaque station (1) dont il faut tenir compte et on obtient alors :

$$(17) \quad A = B + E$$

B représentant l'information théorique recherchée.

3.4. Construction d'un vecteur par la moyenne arithmétique

Une des possibilités est de calculer pour chaque année un élément du vecteur en utilisant la moyenne arithmétique de tous les éléments de cette année.

En cas d'homogénéité préalable de toutes les données et de distribution normale, cette moyenne correspond effectivement à un paramètre à tendance centrale.

Malheureusement, les cas de distribution normale sont rares soit à cause du nombre réduit des données, soit à cause de distributions asymétriques ou plurimodales (dans le cas de séries chronologiques hétérogènes).

La présence de valeurs très fortes peuvent également altérer la moyenne comme on le verra dans quelques exemples ultérieurs (contamination de la moyenne par les valeurs extrêmes).

Les mêmes observations s'appliquent à l'écart-type qui, dans le cas de distributions non normales peut donner une idée de la dispersion des données mais non de leur groupement autour de valeurs centrales.

Il est possible d'obtenir de bons résultats par estimation des valeurs moyennes, mais les calculs sont problématiques quand les séries sont trop incomplètes.

Tout ceci amène à éliminer la distance « euclidienne » comme critère de minimisation de $E = A - B$ et a conduit deux chercheurs, G. HIEZ [6] et Y. BRUNET-MORET [4] à calculer les éléments d'un vecteur régional de façons différentes bien que les résultats soient relativement semblables.

3.5. Le vecteur HIEZ [6]

3.5.1. Algorithme de construction

Une bonne estimation de la valeur centrale est la valeur modale, mais il n'existe pas de moyen courant pour son calcul précis.

Dans le cas de nombreuses données, on pourrait les regrouper par classes et prendre comme valeur modale le point central de la classe de plus grande fréquence.

Si ceci n'est pas le cas, il sera difficile de trouver les valeurs modales pour la ligne (L) et du vecteur colonne (C) sans connaître leurs fonctions de densité qui ne sont pas déterminées *a priori*.

Par ailleurs, il existe d'autres formules que nous allons passer en revue rapidement :

- une formule empirique

$$(18) \quad \hat{M} = 3\bar{M} - 2M$$

où \hat{M} est la valeur modale, \bar{M} la médiane et M la moyenne arithmétique ; cette formule est valable pour les distributions unimodales et symétriques ;

- en considérant les moments de second et troisième ordre, on a :

$$(19) \quad \hat{M} + x_0 = M + \bar{M}$$

où x_0 est la valeur qui annule le moment de troisième ordre ; cette formule tient compte de la dissymétrie de la distribution, mais il faut connaître la médiane dont l'estimation n'est généralement pas précise ;

- le développement d'une fonction de distribution quelconque $f(\mu)$ en série limitée permet d'utiliser une autre formule :

$$(20) \quad \hat{M} = \bar{x} + \sigma \cdot \mu_0$$

où μ_0 est la valeur pour laquelle $f(\mu)$ est maximum ; cette formule ne convient non plus aux distributions plurimodales et est sensible aux valeurs extrêmes car elle fait intervenir un paramètre de dispersion.

Finalement l'auteur utilise une méthode originale de traitement lignes-colonnes qui lui permet, par itérations successives et en utilisant une des formules précédentes, de s'affranchir des restrictions mentionnées. La rapidité de convergence de l'algorithme dépend évidemment de la qualité des données.

L'absence de valeurs n'influe pas sur le traitement à condition d'avoir pour le moins trois valeurs par ligne et par colonne, mais l'estimation du vecteur sera meilleure si les données sont nombreuses.

3.5.2. Conclusions sur ce vecteur

La sensibilité et la fiabilité de l'algorithme ont été testées sur plusieurs essais et la totalité de la pluviométrie du Brésil a été vérifiée par cette méthode.

Une fois le vecteur construit, on le compare aux stations par doubles masses, mais la comparaison ne porte pas sur les valeurs moyennes mais sur les valeurs les plus probables.

L'application de cette méthode demande de respecter les quelques règles suivantes :

- utiliser l'année hydrologique (et non l'année calendaire) ;
- un nombre de stations strictement supérieur à 3, le minimum conseillé étant de 7 ;
- un zonage préalable des stations lequel pourra être rectifié après un premier traitement.

Il est évident que ce vecteur est un instrument puissant ; son application sur tout le Brésil le démontre, mais il demanderait à être analysé de façon plus précise dans un milieu montagneux comme celui de l'Équateur.

3.6. Le vecteur des indices annuels de précipitation [4]

3.6.1. Construction du vecteur

L'auteur considère qu'une zone climatique peut être représentée par un vecteur

$$\mathbf{Z} = (z_1, \dots, z_1, \dots, z_n)$$

de telle manière que son espérance mathématique soit égale à 1.

L'équation (17) peut s'écrire de la façon suivante :

$$(21) \quad x_{ip} / \bar{x}_p = z_i + e_{ip}$$

pour la précipitation de l'année i dans la station p , sachant que l'espérance mathématique de E_p est égale à 0 comme dit précédemment.

La matrice \mathbf{A} (n, m) se compose des précipitations de m stations durant n années et elle sera, dans la plupart des cas, incomplète.

L'algorithme de calcul consiste à minimiser l'équation :

$$(22) \quad \sum_{i=1}^m \sum_{j=1}^n (x_{ij} / \bar{x}_j - z_i)^2$$

afin de déterminer chaque \bar{x}_j et z_i .

S'il existe p stations observées dans l'année i , en dérivant par z_i on obtient n équations :

$$(23) \quad z_i = \frac{1}{p} \sum_1^p x_{ij} / \bar{x}_j$$

En remplaçant dans la formule (22) les valeurs de z_i (23) et en dérivant par $1/\bar{x}_j$, on obtient un système de m équations à m inconnues en $1/\bar{x}_j$.

La solution du système permet de trouver \bar{x}_j et par l'équation (23) les différents composants de z_i .

Cette méthode rapidement décrite permet de donner le même poids aux différentes valeurs observées.

On suppose que pour chaque année et dans chaque station les valeurs fausses constituent une minorité.

Une première itération avec les éléments originaux de \mathbf{A} donne une première série de valeurs. Si un élément $x_{ij} / \bar{x}_j \cdot z_i$ est en dehors d'un intervalle donné, il est remplacé par $\bar{x}_j \cdot z_i$.

Des itérations successives permettent d'éliminer rapidement ces problèmes. Les limites de l'intervalle dépendent de la variance de \mathbf{E} .

3.6.2. Utilisation de la méthode

Après avoir vérifié l'historique de la station, supprimé les valeurs aberrantes (erreurs de saisie, etc.) et complété les données qui s'y prêtent, on fait un premier traitement par simples masses pour confirmer la représentativité des groupes d'une même zone climatique.

Une fois cette préparation terminée, on effectue un premier passage par le programme qui sert à éliminer les valeurs indubitablement fausses.

En fonction des coefficients de corrélation, on peut également écarter quelques stations pour mieux cerner la zone et améliorer les corrélations.

Les étapes suivantes servent à détecter des erreurs moins frappantes (changement de localisation ou modification de l'environnement, etc.) et en deux ou trois passages on arrive à dégager le vecteur représentatif de la zone considérée.

4. COMPARAISON DES DEUX VECTEURS

Le vecteur régional de HIEZ a l'avantage de se baser sur une estimation précise de la valeur modale, ce qui le rend indépendant des erreurs systématiques si celles-ci ne représentent qu'une minorité. On peut donc en un seul passage déceler les anomalies et calculer les coefficients de pente correspondant à une confusion d'éprouvette, un changement de localisation, etc.

Il a été testé sur tout le territoire du Brésil, ce qui représente un grand nombre d'essais qui ont conduit au choix de normes d'utilisation.

Malheureusement, par son algorithme de calcul pour l'instant quelque peu difficile et long, il requiert d'ordinateurs puissants.

Récemment, il fut adapté pour des micro-ordinateurs et il serait intéressant de le voir fonctionner en région montagnaise.

Le vecteur de l'indice annuel de précipitation recherche la valeur la plus probable par un algorithme plus simple ; il n'est pas entièrement insensible à la présence d'erreurs systématiques ce qui implique 2 ou 3 passages avant d'arriver à une estimation correcte.

Pendant la construction, les stations avec des ruptures de pente sont considérées comme stations secondaires.

L'analyse de la relation entre deux pentes se fait au dernier passage, une fois le vecteur construit.

Il est utilisé en Afrique et aux Caraïbes. En Équateur, il a été essayé dans la région sud orientale et dans le bassin du Mira et ensuite sera testé dans la région côtière dans le cadre du Plan National d'Irrigation de l'INERHI.

Il est disponible pour des ordinateurs compatibles IBM mais les sorties sont tellement compliquées qu'il convient de les présenter sous une forme graphique afin de pouvoir utiliser un critère visuel appuyé par des critères numériques.

Pour les comparer avec un vecteur construit à partir de la moyenne arithmétique, on prendra l'exemple suivant :

Douze stations observées pendant 20 ans (1966-1985). Chaque année dans chaque station, les valeurs varient entre 999,5 et 1 001,5 mm, c'est-à-dire qu'elles sont toutes proches de 1 000.

Multiplions les données de trois stations par 5 pour une période de 10 ans (1965-1975).

Un vecteur construit avec la moyenne arithmétique aura les caractéristiques suivantes :

1966 - 1975 $x_1 = 2\ 000$

1976 - 1985 $x_1 = 1\ 000$

soit une relation de l'ordre de 2.

Avec le vecteur des indices annuels de précipitation, la relation sera de 1,25 ce qui indique la faible sensibilité aux confusions d'éprouvette ou aux déplacements du pluviomètre.

Dans cette même série de 12 stations, on introduit une année avec une précipitation de 13 000 mm. La moyenne arithmétique donnera 2 000 mm et le vecteur indice 1 015, ce qui démontre la faible sensibilité aux valeurs extrêmes souvent produites par des erreurs de saisie.

Dans les deux cas précédents (quelque peu extrêmes), la comparaison des stations avec le vecteur construit par la moyenne arithmétique n'aurait pas permis d'émettre un diagnostic car le vecteur lui-même est trop altéré.

Le vecteur HIF⁷ est encore plus insensible à ce genre d'erreurs comme le démontrent des exemples du Bréasil.

Pour terminer, il faut se rappeler qu'au paragraphe 3.1. on a signalé que les doubles masses entre vecteur et station sont plus précises que les doubles masses entre stations. Cette caractéristique peut être vérifiée à l'aide de l'exemple suivant.

Pour 12 stations du nord de l'Équateur, on a calculé les coefficients de corrélation, chaque station par rapport à toutes les autres. Pour chacune station on a noté :

- le coefficient de corrélation maximum CCMAX
- le coefficient de corrélation minimum CCMIN
- le coefficient de corrélation moyen CCMOY
- le coefficient de corrélation avec le vecteur indice CCVEC, élaboré à partir des 12 stations sans avoir effectué aucune correction et sachant qu'elles appartiennent à des régimes pluviométriques différents.

Station	CCMAX	CCMIN	CCMOY	CCVEC
M054	0,64	0,17	0,41	0,66
M083	0,90	0,18	0,67	0,90
M084	0,91	0,41	0,68	0,86
M101	0,84	0,20	0,44	0,50
M102	0,86	0,20	0,60	0,86
M103	0,83	0,14	0,46	0,65
M303	0,91	0,39	0,66	0,88
M305	0,85	0,17	0,55	0,85
M306	0,90	0,37	0,72	0,93
M307	0,76	0,25	0,46	0,72
M324	0,80	0,30	0,57	0,83
M487	0,62	0,14	0,39	0,60

On observe que les corrélations avec le vecteur sont toujours supérieures aux corrélations moyennes entre stations et presque toujours égales ou supérieures aux corrélations maximum.

Cette caractéristique sera d'une grande aide pour le remplissage des données manquantes.

5. CONCLUSION

Comme on l'a vu au début, seule la méthode des doubles masses couplée à celle des simples masses permet de détecter et corriger les données pluviométriques erronées.

Malheureusement, cette méthode demande un travail long et minutieux ainsi que la manipulation d'une grande quantité de chiffres.

Pour ces raisons, le concept de vecteur régional est une grande avancé méthodologique. Sa construction ne peut se satisfaire que des moyennes arithmétiques qui sont trop sensibles aux valeurs extrêmes.

Les deux vecteurs présentés, calculés à l'aide d'algorithmes différents, donnent des valeurs probables et sont donc pratiquement insensibles aux altérations. Le soin avec lequel ils sont élaborés peut sembler excessif, mais ils servent non seulement à homogénéiser les données mais aussi à représenter, de la manière la plus fidèle possible, les caractéristiques d'une zone climatique. Ceci signifie qu'ils sont une grande aide pour la régionalisation climatique.

BIBLIOGRAPHIE

- [1] RODDA, J.C., 1967. The rainfall measurement problem, in *AISH*, publ. n° 78, p. 215-281.
- [2] BRUNEL, J.P., 1986. Comparaison instrumentale des précipitations en climat méditerranéen semi- aride, in *Hydrologie Continentale*, Vol.1, n° 2, p. 87-93.
- [3] CHEVALLIER, P., 1986. Note sur les écarts de mesure observés entre les pluviomètres standards et les pluviomètres au sol, in *Hydrologie Continentale*, Vol. 1, n° 2, p. 111-119.
- [4] BRUNET-MORET, Y., 1979. Homogénéisation des précipitations, in *Cahiers Hydrologiques*, ORSTOM, Vol. XVI, n° 3 y 4, p. 147-170.
- [5] BRUNET-MORET, Y., 1971. Études de l'homogénéité de séries chronologiques annuelles par la méthode des doubles masses, in *Cahiers Hydrologiques*, ORSTOM, Vol. III, n° 4, p. 3-31.
- [6] HIEZ, G., 1977. L'homogénéité des séries pluviométriques, in *Cahiers ORSTOM*, Vol. XIV, n° 2, p. 129-163.
- [7] SEARCY, J.K., HADIRSON, C.H., 1983. *Curva de dobles masas*, HIMAT y COLCIENCIAS, Bogota, 36 p.

REPUBLICA DEL ECUADOR

MINISTERIO DE AGRICULTURA Y GANADERIA
CONADE - INERHI - ORSTOM

FUNCIONAMIENTO DEL RIEGO PARTICULAR
EN LOS ANDES ECUATORIANOS
Recomendaciones para el Plan Nacional de Riego

FONCTIONNEMENT DE L'IRRIGATION TRADITIONNELLE
DANS LES ANDES EQUATORIENNES
Recommandations pour le Plan National d'Irrigation

COMMUNICATIONS PRESENTEES
SUR L'IRRIGATION TRADITIONNELLE
DANS LES ANDES EQUATORIENNES

Période 1987 - 1992



Document élaboré par :

INERHI

Hugo RIBADENEIRA
Miguel ALEMAN
Ivan OSORNO

Patrick LE GOULVEN

Thierry RUF

ORSTOM

COMMUNICATIONS PRESENTEES
SUR L'IRRIGATION TRADITIONNELLE
DANS LES ANDES EQUATORIENNES

Période 1987 - 1992

Quito, juin 1993