Météorologie Nationale

RETOUR SUR L'HOMOGENEISATION DES PLUIES ANNUELLES
PAR VECTEUR REGIONAL.

par Y. Brunet-Moret
Conseiller scientifique au Service
hydrologique de l'ORSTOM

RESUME

Définition de l'homogénéité des suites chronologiques de précipitations annuelles et vérification de cette homogénéité par une suite chronologique d'indices annuels de précipitation. Méthode de calcul de ces indices.

* * *

RETOUR SUR L'HOMOGENEISATION DES PLUIES ANNUELLES
PAR VECTEUR REGIONAL

par Y. Brunet-Moret
Conseiller scientifique au
Service hydrologique de l'ORSTOM

* * *

INTRODUCTION

Caractère aléatoire simple d'une suite chronologique.

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 de la population-mère des valeurs observées ne varient pas dans le temps : entre autres, l'espérance mathématique
- par un effet de tendance : les valeurs observées sont tirées de populations-mères différentes (par un ou plusieurs paramètres) dont les espérances mathématiques croissent ou décroissent avec le temps.
- par des effets (cycliques ou pseudo-cycliques) : l'espérance mathématique, pour une valeur de la suite, est fonction de la chronologie, mais l'espérance mathématique de la moyenne de séries suffisamment longues peut être considérée comme stationnaire.
- par des erreurs, systématiques ou non, d'observation affectant des termes de la suite chronologique.

Définition de l'homogénéité d'une suite chronologique de précipitations annuelles

Ce sont ces erreurs (paragraphe précédent) qui détruisent l'homogénéité d'une suite chronologique de précipitations annuelles, qui peut présenter des caractères de persistance, de tendance, de pseudo-cycles.

Les "tests d'homogénéité" existants (ou futurs) ne pourront jamais (sauf si on connaît a priori les lois de persistance, de tendance, de pseudo-cycles) que vérifier le caractère aléatoire simple d'une suite chronologique.

Le véritable problème, pour vérifier l'homogénéité de suites chronologiques, n'est pas de détecter les effets de persistance, de tendance, de pseudo-cycles, mais de détecter les erreurs d'observation (y compris les erreurs d'interprétation du document original, les erreurs de recopies ...)

En étude de précipitations annuelles, 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 (s'il n'y a pas trop d'erreurs d'observations).

Etant dans la même zone climatique, les stations sont soumises aux mêmes effet de persistance, de tendance, de pseudo-cycles. Leurs précipitations moyennes inter-annuelles sont voisines et leurs coefficients de variation très voisins. Les coefficients calculés d'asymétrie peuvent différer notablement si les suites ne sont pas très longues étant donné la très grande variance de ces coefficients.

Hypothèse de base de l'homogénéité spatiale des précipitations annuelles.

On admet que les stations pluviométriques situées dans la même zone climatique ont des totaux annuels de précipitations pseudo-proportionnels (quelles que soient le fluctuations climatiques de la zone). Cette hypothèse est également l'hypothèse de ba de la méthode des doubles cumuls. C'est-à-dire que pour une année i , les précipitations annuelles x_i et y_i à deux stations X et Y peuvent s'écrire

$$x_i = \frac{\bar{x}}{\bar{y}} y_i + \varepsilon_i$$

$\bar{x}$ et $\bar{y}$ étant les moyennes interannuelles aux deux stations sur une très longue période. La variance de ε_i (terme aléatoire indépendant tant de la valeur de x_i que de celle de y_i) étant d'autant plus faible que le coefficient de corrélation linéaire entre les stations est plus grand et l'espérance mathématique de ε_i étant nulle.

La méthode classique pour vérifier l'homogénéité temporelle de suite de précipitations annuelles est de comparer par la méthode des doubles cumuls ces suites à la suite des précipitations annuelles d'une station, homogène dans le temps, de la même zone climatique. Comme on ne connaît pas a priori quelle est la station d'observations homogènes et, qu'en fait, il faut les soupçonner toutes de défauts, la méthode est longue et difficile à appliquer.

VECTEUR DES INDICES ANNUELS DE PRECIPITATION

Etant donné les difficultés de la méthode des doubles cumuls, il est tentant de créer une suite chronologique de précipitations annuelles fictives à une station fictive "vecteur des indices annuels de précipitation" : suite comprenant les effets de persistance, de tendance, de pseudo-cycles de la zone climatique, mais homogène dans le temps et d'utiliser, en double cumul, cette suite avec chacune des stations de la zone pour vérifier leur homogénéité temporelle. Ce vecteur des indices annuels ne peut, évidemment, être créé qu'à partir des observations aux stations de la zone climatique homogène dans l'espace.

Hypothèses de base.

La précipitation annuelle $x_{\alpha i}$ de l'année i à la station α (pour laquelle $\bar{x}_\alpha$ est l'espérance mathématique de la population-mère des précipitations annuelles) est liée à l'indice z_i de l'année i par (quelle que soit la statio parmi les μ existantes)

$$\frac{x_{\alpha i}}{\bar{x}_\alpha} = z_i + e_{\alpha i}$$

$e_{\alpha i}$ étant une variable aléatoire indépendante de z_i et les variables E_α étant indépendantes entre elles : pour une station α la suite chronologique des $e_{\alpha i}$ est de caractère aléatoire simple . Nous exprimons ainsi l'hypothèse de pseudoproportionalité des totaux annuels de précipitations et la condition d'établissement du vecteur des indices annuels en ajoutant :

espérance mathématique de Z égale à 1 dans sa population-mère

espérance mathématique de E_α égale à 0 dans sa population-mère

variances des E_α ... E_μ égales entre elles et à var E

Nous pouvons admettre que pour chaque année i les $e_{\alpha i}$ ... $e_{\mu i}$ ne sont pas indépendants entre eux (proximité géographique des stations) mais nous admettrons que, dans les populations-mères, la somme des covariances des écarts E est nulle (Σ covar $(E_\gamma, E_\delta) = 0$, $\gamma \neq \delta$, γ et δ de α à μ) ce qui veut dire que les coefficients de corrélation linéaire entre stations ne sont pas tous égaux entre eux, mais que leur valeur moyenne (dans les populations-mères) est de $\frac{\text{var } Z}{\text{var } Z + \text{var } E}$

Les hypothèses ci-dessus conduisent à la valeur de $\sqrt{\frac{\text{var } Z}{\text{var } Z + \text{var } E}}$ pour la valeur du coefficient de corrélation linéaire entre l'indice et une station, quelle que soit la station, dans les populations-mères. Cette valeur étant supérieure à la valeur moyenne des coefficients de corrélation entre stations, on comprend que les doubles cumuls entre indice et stations seront plus précis que les doubles cumuls entre stations.

PRINCIPE DU CALCUL DES INDICES ANNUELS DE PRECIPITATION

La matrice des observations $A(n,\mu)$ est composée des précipitations annuelles de

μ stations sur n années : et peut ne pas être complète (observations manquante certaines années à certaines stations).

Dans un espace à μ dimensions, l'équation

$$\frac{X_\alpha}{\bar{x}_\alpha} + \frac{X_\beta}{\bar{x}_\beta} + \dots + \frac{X_\mu}{\bar{x}_\mu} - Z = 0$$

représente un hyperplan perpendiculaire à la droite

$$\frac{X_\alpha}{\bar{x}_\alpha} = \frac{X_\beta}{\bar{x}_\beta} = \dots = \frac{X_\mu}{\bar{x}_\mu}$$

Nous allons chercher à déterminer les valeurs $\bar{x}_\alpha$ ... $\bar{x}_\mu$ pour chaque station et les valeurs z_i pour chaque année pour que les hyperplans représentatifs des observations de chaque année soient parallèles entre eux et perpendiculaires à la droite

$$\frac{X_\alpha}{\bar{x}_\alpha} = \dots = \frac{X_\mu}{\bar{x}_\mu} = 1$$

Compte tenu des hypothèses ~~2.2.~~ appliquées à l'échantillon (matrice des observations) c'est-à-dire

valeur moyenne des z_i égale à 1
valeur moyenne des écarts E nulle
et homoscedasticité : variances des E_α égales à var E quelle que soit la station α
on peut utiliser la méthode des moindres carrés : minimiser

$$\sum_{\alpha}^{\mu} \sum_{1}^{n} \left(\frac{x_{\alpha i}}{\bar{x}_\alpha} - z_i \right)^2$$

pour déterminer $\bar{x}_\alpha \dots \bar{x}_\mu$, $z_1 \dots z_n$

Comme on est, en fait, obligé de dériver par rapport à $\frac{1}{\bar{x}_\alpha}$, l'erreur "moyenne" sur chaque $\frac{1}{\bar{x}_\alpha}$ est la même pour tous ;

l'erreur "moyenne" sur chaque $\bar{x}_\alpha$ est proportionnelle à $\bar{x}_\alpha^2$, l'erreur "relative moyenne" est proportionnelle à $\overline{x_\alpha}$. Ce défaut n'est pas grave : d'une part nous travaillons dans une zone climatique homogène, les $\bar{x}_\alpha$ ne sont pas très différents les uns des autres et, d'autre part, l'erreur moyenne est petite lorsque les observations sont homogènes. On peut admettre que la variance de $\bar{x}_\alpha$ est toujours nettement inférieure à $\frac{1}{n}$ x (variance des observations de la station α) lorsque les observations sont homogènes.

La méthode ci-dessus n'est pas équivalente à celle qui consisterait à déterm ner, par les moindres carrés, un vecteur ligne-stations $C(1,\mu)$ "précipitations annuelles moyennes aux stations" et un vecteur colonne-années $L(n,1)$ "suite chronologique de précipitations fictives" tels que

$$L\,C = A\,(n,\mu) - F\,(n,\mu)$$

F matrice de résidus minimisée au sens des moindres carrés.

Dans ce cas, pour chaque année $z_i - \frac{1}{\mu}\,\Sigma\,\frac{x_{\alpha i}}{\bar{x}_\alpha}$ est nulle, pour chaque station $\Sigma\,(\frac{x_{\alpha i}}{\bar{x}_\alpha} - z_i)$ est nulle et les hyperplans déterminés pour chaque année i par les points d'abscisses sur leurs axes respectifs $x_{\gamma i} - f_{\gamma i}$ (γ de α à μ) ne sont pas parallèles.