

MINISTERE de l'AGRICULTURE

Direction de l'Equipement
et de l'Hydraulique

Stage d'interprétation hydrologique

STRASBOURG - 2-7 Octobre 1967

CONTROLE et HOMOGENEISATION des DONNEES PLUVIOMETRIQUES

par

Jacques HERBAUD

Maître de Recherches à l'ORSTOM

S O M M A I R E

	Page
1. <u>DEFINITION et NECESSITE de l'HOMOGENEISATION des DONNEES PLUVIOMETRIQUES</u>	1
2. <u>NATURE des HETEROGENEITES AUXQUELLES on PEUT PORTER REMEDE</u>	2
2.1 - <u>Absence de données</u>	2
2.2 - <u>Données défectueuses</u>	2
2.3 - <u>Données correspondant à deux ou plusieurs séries homogènes non défectueuses, groupées sous le nom d'une même station</u>	2
2.3.1 - <u>Déplacements des pluviomètres</u>	2
2.3.2 - <u>Changements d'environnement des pluviomètres</u>	3
3. <u>PREMIERES VERIFICATIONS des DONNEES PLUVIOMETRIQUES</u>	4
3.1 - <u>Anomalies rencontrées dans les "originaux" d'un poste relativement isolé</u>	4
3.2 - <u>Anomalies rencontrées dans les copies des "originaux"</u>	5
4. <u>VERIFICATION de l'HOMOGENEITE des SERIES de TOTAUX ANNUELS. PRINCIPE de la METHODE des COURBES de COMPARAISON des TOTAUX CUMULES</u>	5
4.1 - <u>Caractère de la liaison existant entre les totaux annuels de deux postes pluviométriques</u>	5
4.2 - <u>Représentation graphique de cette liaison</u>	6
4.3 - <u>Interprétation du graphique : hypothèses possibles</u>	6
4.4 - <u>Interprétation du graphique : choix entre les hypothèses possibles</u>	8
4.4.1 - <u>Nouvel examen des originaux</u>	8

4.4.2 - Enquêtes diverses	9
4.4.3 - Considération de l'ensemble des données régionales	9
5. <u>DETERMINATION de SERIES HOMOGENES CORRESPONDANT aux PERIODES d'OBSERVATION :</u> <u>APPLICATION de la METHODE des COURBES de COMPARAISON des TOTAUX CUMULES</u>	9
5.1 - <u>Tracé des droites</u>	10
5.2 - <u>Choix d'une série de base</u>	10
5.2.1 - Choix des séries de base destinées aux premiers essais de vérification d'homogénéité	10
5.2.2 - Choix de la série destinée à l'étude systématique des postes d'une région	11
5.2.3 - Choix de la série de base destinée à corriger une hétérogénéité	12
6. <u>HOMOGENEISATION des PERIODES d'EXTENSION des SERIES PLUVIOMETRIQUES</u>	12
6.1 - <u>Méthode à employer</u>	12
6.2 - <u>Précautions à prendre</u>	13
6.3 - <u>Choix des séries à étendre</u>	13
6.4 - <u>Choix de la période d'homogénéisation, et des séries de base</u>	14
6.4.1 - Choix de la période d'homogénéisation	14
6.4.2 - Choix des séries de base	15
6.4.2.1 - Constitution d'un groupe de base	15
6.4.2.2 - Exemples de schémas d'opérations	15
6.4.2.3 - Critères pour le choix d'un schéma	15
6.5 - <u>Conduite des calculs pour l'extension des séries de totaux pluviométriques annuels</u>	16

7. ESTIMATION des TOTAUX MENSUELS ou SAISONNIERS

21

ANNEXES - EXERCICES d'APPLICATION NUMERIQUE des METHODES de CONTRÔLE et d'HOMOGENEISATION des DONNEES PLUVIOMETRIQUES

1. EXERCICE d'APPLICATION NUMERIQUE de la DETECTION et de la CORRECTION d'une HETEROGENEITE par la METHODE de COMPARAISON GRAPHIQUE des TOTAUX ANNUELS CUMULES

1

2. EXERCICE d'APPLICATION NUMERIQUE de l'EXTENSION d'une SERIE PLUVIOMETRIQUE A PARTIR d'une REGRESSION LINEAIRE entre TOTAUX ANNUELS de DEUX SERIES

3

3. EXERCICE d'APPLICATION NUMERIQUE de l'ESTIMATION des TOTAUX PLUVIOMETRIQUES SAISONNIERS d'une SERIE ETENDUE

8

1. DEFINITION et NECESSITE de l'HOMOGENEISATION des DONNEES PLUVIOMETRIQUES

Toute étude climatique ou hydrologique est basée sur l'exploitation de séries de données recueillies pendant des périodes plus ou moins longues continues ou discontinues.

En particulier, les données pluviométriques sont très généralement des relevés journaliers effectués à un pluviomètre. Cet appareil est relativement facile à mettre en place et à utiliser ; cela explique, sans le justifier, que l'on a souvent changé l'emplacement d'un pluviomètre, ou bien qu'on l'a confié successivement à de nombreux observateurs plus ou moins qualifiés et mal rémunérés. Il en résulte que les séries de données présentent des lacunes particulièrement fréquentes, et que l'on n'est jamais parfaitement sûr qu'elles représentent, comme disent les statisticiens, un échantillon d'une seule population.

A plus forte raison si l'on s'intéresse à la pluie moyenne d'un bassin, fournie par les données d'un certain nombre N de postes, le calcul de cette moyenne peut toujours être effectué, par exemple année par année, en utilisant les données disponibles ; il se peut alors que pendant 5 ans la moyenne soit fournie par les pluviomètres A, B, C, pendant les 10 années suivantes par B, C, D, E, F, ensuite par A, B, E, etc ... De plus, le pluviomètre A a pu fournir par exemple deux séries de données n'ayant pas la même signification, avant et après l'interruption constatée dans les relevés, ou de part et d'autre d'une date quelconque à laquelle on a omis de faire figurer un renseignement utile sur la feuille de relevés ou dans les registres établis par la suite.

Il est évident que dans ces conditions, la série des valeurs de la pluie moyenne calculée n'est pas très satisfaisante. On se propose de réduire les risques d'erreurs systématiques en faisant ce qui s'appelle une homogénéisation, c'est-à-dire la détermination de N séries de données de même extension dans le temps, et autant que possible exemptes de changements de signification parmi les termes d'une même série.

L'amélioration obtenue ne conduira pas forcément à des valeurs très différentes pour les moyennes interannuelles des pluviométries annuelles saisonnières ou mensuelles, ou des caractéristiques climatiques ou hydrologiques qu'on peut en déduire par corrélation, mais elle conduira à des conclusions bien plus exactes quand on cherchera à définir des valeurs extrêmes apparaissant avec une probabilité donnée dans une période donnée. Elle conduira aussi à des valeurs plus exactes pour l'intervalle de confiance dans lequel on a $p\%$ de chances de trouver les valeurs moyennes interannuelles déterminées.

2. NATURE des HETEROGENEITES AUXQUELLES on PEUT PORTER REMEDE

2.1 - Absence de données

C'est l'hétérogénéité la plus apparente, à laquelle on remédie par l'extension des données de la station défaillante B à partir d'une corrélation entre les données de cette station et celles d'une station de référence A.

2.2 - Données défectueuses

Si les relevés n'ont pas été toujours effectués dans de bonnes conditions, on peut être amené à les éliminer (et on est alors remené, pour une ou plusieurs périodes, au cas précédent), ou à faire subir aux données une correction qui les rendra utilisables, au moins dans une certaine mesure.

2.3 - Données correspondant à deux ou plusieurs séries homogènes non défectueuses, groupées sous le nom d'une même station

On se trouve dans ce cas si le pluviomètre a été déplacé ou si l'environnement immédiat a changé.

2.3.1 - Déplacements des pluviomètres

Ce n'est pas forcément l'éloignement des deux sites successifs qui est responsable de l'hétérogénéité d'une série de données : c'est plutôt un ensemble de caractéristiques de ces sites (altitude, forme du terrain, orientation du relief, végétation, hydrographie, sols, urbanisation éventuelle). On peut seulement dire que plus la distance de deux sites est grande, plus il y a de chances pour que l'ensemble de ces caractéristiques accuse de nombreuses différences. A partir d'un certain degré de différenciation géographique, on peut parler de changement de régime pluviométrique, mais il est très difficile, et sans grand intérêt pour l'hydrologue dans l'état actuel de sa science, de définir exactement et universellement ce qu'est un changement de régime. Il suffit de savoir que dans la pratique :

- a) certaines caractéristiques de la pluviométrie varient graduellement quand on change de région climatique (par exemple en ALSACE, on passe d'un

climat océanique à l'Ouest à un climat continental à l'Est, et ce sont les coefficients mensuels pluviométriques qui changent) ;

- b) d'autres caractéristiques peuvent présenter des variations plus nombreuses (nombre de jours de pluie, et surtout pluviométrie moyenne annuelle de certains sites où les mouvements ascendants sont plus fréquents, à cause de la topographie, de la direction des vents ou de la présence de fumées, etc ...).

Comme on le comprendra en étudiant la signification statistique des moyennes, on peut très rarement, à partir des données disponibles, prouver avec une faible probabilité d'erreur que les pluviométries moyennes interannuelles de deux sites très voisins sont différentes, et par conséquent que tel ou tel facteur géographique, qui en l'espèce serait le seul à différencier ces sites, a telle ou telle influence sur la pluviométrie.

2.3.2 - Changements d'environnement des pluviomètres

Parmi les caractéristiques d'un site que l'on a citées au paragraphe précédent, celles qui peuvent être modifiées sont les suivantes :

- a) végétation à proximité immédiate (croissance d'un arbre qui devient progressivement un obstacle et peut même rendre les données défectueuses), ou bien encore dans une région plus ou moins étendue autour du poste (plantation ou déplantation de forêts, mise en culture sèche ou irriguée) ;
- b) hydrographie (création de plans d'eau, assèchement de marais ou simplement drainage de zones très humides) ;
- c) urbanisation (surfaces bâties, surfaces aplanies, création de zones industrielles, pollution et réchauffement de l'atmosphère).

Cette liste n'est évidemment pas limitative. Cependant, l'intervention humaine doit assez rarement modifier l'environnement d'un poste pluviométrique au point de créer une hétérogénéité dans la série des données recueillies à ce poste. Par contre, quand cela se produit, c'est généralement assez grave car il s'agit alors d'une action qui rend presque défectueuse les données recueillies par la suite : à la limite, s'il s'agit de la construction d'un immeuble de 30 m de hauteur à une distance très inférieure à 60 m du pluviomètre, les données recueillies ultérieurement peuvent présenter des anomalies plus ou moins faciles à corriger (Cf. Supra, paragraphe 2.2).

3. PREMIERES VERIFICATIONS des DONNEES PLUVIOMETRIQUES

Avant de vérifier l'homogénéité des séries de totaux annuels, il y a lieu de détecter toutes les erreurs grossières dont la correction peut parfois être faite avec un peu de bon sens mais qui pourraient passer inaperçues si l'on se contentait d'examiner des totaux annuels ou même mensuels.

Autant que possible, une étude de la pluviométrie doit donc commencer par l'examen de tous les relevés journaliers, sur les feuilles de relevés des observateurs. En pratique, en FRANCE, le Service Météorologique procède à cet examen critique, corrige éventuellement certaines données, renonce à la publication de certaines autres jugées trop erronées, et classe ensuite tous les "originaux", bons ou mauvais, qui deviennent dès lors assez difficiles à consulter. On peut le regretter dans la mesure où certaines données ont été jugées mauvaises un peu trop sévèrement, mais en compensation, si l'on se contente des données publiées, on peut les utiliser en toute confiance. Le problème de la vérification des "originaux" se pose surtout pour les données ne provenant pas d'un correspondant du Service Météorologique.

Les différentes causes d'erreurs dans la mesure de la pluviométrie ont été signalées dans une précédente conférence du cycle de NIMES. On les reconnaîtra plus ou moins bien en découvrant telle ou telle anomalie quand elle peut être interprétée comme une erreur.

3.1 - Anomalies rencontrées dans les "originaux" d'un poste relativement isolé

Si le poste considéré ne fait pas partie d'un réseau particulièrement dense comme le réseau d'un ensemble de bassins expérimentaux, il est impossible de comparer efficacement dans le détail les totaux journaliers avec ceux qui ont été relevés ailleurs. On doit se contenter de vérifier la vraisemblance des données :

- a) nombre de jours de pluie mois par mois, estimé d'un coup d'oeil sur la feuille et comparé à ce qu'on trouve en général dans la région,
- b) position et durée approximatives des séquences pluvieuses dans le mois (les jours de pluie sont en effet très généralement groupés),
- c) fréquences, estimées également grosso modo, des précipitations de moins

de 1 mm et de plus de 10 mm (ces fréquences doivent avoir des valeurs raisonnables, sinon il est permis de supposer par exemple - soit que les faibles pluies n'ont pas été mesurées et que l'eau recueillie s'est évaporée, si le temps s'y prêtait, avant la précipitation suivante - soit que le pluviomètre avait une fuite, etc ...)

- d) fréquences d'apparition de chacun des chiffres 0 à 9 comme dernier chiffre significatif de la mesure (que celle-ci soit effectuée directement en mm ou bien en cm^3) : on verra si l'observateur n'a pas eu tendance à arrondir les valeurs, d'une façon plus ou moins logique, ce qui dénoterait un manque de soin (en particulier si le chiffre 0 apparaît abusivement) ;
- e) indices divers, prouvant indirectement un manque de soin ou de compétence de l'observateur (30 relevés en Février, 31 en Avril, etc ...).

3.2 - Anomalies rencontrées dans les copies des "originaux"

On peut rencontrer, s'il s'agit d'une copie, des erreurs supplémentaires : des chiffres peu lisibles ont pu être mal interprétés par le copieur, la virgule a pu être omise, l'ordre chronologique des feuilles mensuelles a pu être mal reproduit, etc ...

La répétition des mêmes relevés, jour pour jour, pendant deux mois consécutifs ou non, est le signe indubitable d'une erreur, et on ne peut qu'éliminer une des deux séries en cause.

4. VERIFICATION de l'HOMOGENEITE des SERIES de TOTAUX ANNUELS. PRINCIPE de la METHODE des COURBES de COMPARAISON des TOTAUX CUMULES

La méthode exposée ci-dessous est souvent appelée improprement méthode des "doubles-masses", par suite de la mauvaise traduction du terme anglais "double-mass curve" (to mass : cumuler).

4.1 - Caractère de la liaison existant entre les totaux annuels de deux postes pluviométriques

Si les deux postes considérés sont situés dans la même grande région climatique, et à une distance relativement faible à l'échelle d'un

continent, il existe une liaison positive significative entre les totaux annuels pluviométriques, c'est-à-dire qu'il y a seulement 5 % de chances, par exemple, pour que les séries soient complètement indépendantes, ou présentent des variations de sens contraires.

D'autre part, les séries présentent une distribution statistique à peu près normale, en particulier en FRANCE où la loi de GAUSS est même sans doute la plus adéquate. Dans ces conditions, la régression entre deux séries est linéaire (voir le texte de la conférence "analyse du régime des débits des cours d'eau", paragraphe 2.3.1). Il en sera de même pour toute combinaison linéaire des termes correspondants de ces séries, et en particulier, des totaux cumulés à partir d'une même origine des temps arbitraires.

4.2 - Représentation graphique de cette liaison

Sur le graphique des totaux annuels comparés, il apparaît une certaine dispersion. On pourrait examiner sur ce graphique si les points représentatifs des années successives présentent des écarts simplement aléatoires autour de la droite de régression, ou si jusqu'à telle année les écarts oscillent autour d'une valeur positive et ensuite autour d'une valeur négative ou vice-versa. Le procédé est d'autant plus difficile à appliquer que les séries sont plus longues et leur liaison plus lâche.

Le graphique des totaux cumulés comparés est bien plus lisible car ces totaux sont tous deux des fonctions monotones du temps. C'est ce graphique qui est utilisé.

4.3 - Interprétation du graphique : hypothèses possibles

Supposons que la pluviométrie des stations A et B soit connue à partir du 1er Janvier de l'année X_1 jusqu'au 31 Décembre de l'année X_n . Le graphique des totaux cumulés B en fonction des totaux cumulés A présentera $(n + 1)$ points qui, si les séries sont homogènes, seront tous situés à proximité d'une droite de pente m voisine du rapport des moyennes interannuelles $\frac{P_B \text{ moy.}}{P_A \text{ moy.}}$.

Par contre, si l'homogénéité de la série B est rompue à partir du 1er Janvier de l'année X_i , le graphique présente $i + 1$ points situés

à proximité d'une première droite de pente m , et $(n - i + 1)$ points à proximité d'une seconde droite de pente m' , ces deux ensembles de points ayant évidemment un point commun.

On peut alors faire plusieurs hypothèses :

- a) Un changement instrumental intervenu le 1er Janvier de l'année X_i a induit l'observateur en erreur, et ses relevés sont systématiquement erronés à partir de cette date et peuvent être corrigés jour par jour.

Citons l'exemple bien connu que donne Marcel ROCHE dans son cours professé à l'ENGREF :

"On a eu fabriqué des pluviomètres dont le récepteur avait un diamètre de 200 mm, donc une surface de 314 cm^2 . Des éprouvettes dont la graduation en $1/10^{\text{ème}}$ de mm correspondait à cette surface avaient été distribuées aux observateurs. Un jour, un des observateurs casse son éprouvette ; mais entre-temps, le service dont il dépend a adopté une surface réceptrice de 400 cm^2 et fourni une éprouvette correspondant à cette nouvelle surface. Il est rare que l'observateur s'en aperçoive et, à partir de ce jour, il fournira des hauteurs pluviométriques 1,27 fois trop fortes. Ceci se traduira sur la courbe des totaux cumulés comparés par une cassure et la mise en évidence de deux droites d'ajustement dont le rapport des pentes sera précisément égal ou voisin de 1,27. Dans ce cas, il est possible de corriger toutes les hauteurs journalières observées durant la période fautive."

- b) Le 1er Janvier de l'année X_i , le pluviomètre B a été déplacé. Le déplacement n'a pas été assez important pour qu'on change de région climatique, mais le nouveau site peut être en moyenne plus ou moins arrosé que l'ancien (Cf. Supra, paragraphe 2.3.1 b). La série B peut être rendue homogène en effectuant une correction : tous les totaux annuels des années X_i à X_n seront multipliés par le rapport $\frac{m}{m'}$, ou bien au

contraire ceux des années X_1 à X_{i-1} seront multipliés par $\frac{m'}{m}$. On choisira celle de ces deux solutions qui paraîtra la plus judicieuse ; la première par exemple :

- si le 1er emplacement est situé plus loin que le second des obstacles naturels ou artificiels pouvant perturber localement la répartition de la pluie,
- si au contraire il représente une catégorie de sites dont on veut connaître la pluviométrie,

- ou encore si la pluviométrie moyenne interannuelle adoptée permet de tracer un réseau d'isohyètes plus simple.

L'étude de certaines corrélations hydropluviométriques en ALSACE tend à prouver qu'on a intérêt à utiliser une expression de la pluviométrie qui n'est pas forcément la plus voisine possible de la moyenne spatiale vraie. Si l'étude de la pluviométrie est faite dans le but d'étendre une série de débits par corrélation hydropluviométrique, on peut donc choisir indifféremment l'une des deux corrélations possibles à faire subir aux données d'un poste tel que B.

Si l'étude de la pluviométrie doit servir à celle des besoins en eau d'irrigation, il vaut mieux adopter la correction qui conduira aux conclusions les plus pessimistes.

Dans tous les cas, une correction destinée à homogénéiser des données pluviométriques correspondant à deux emplacements successifs, ne peut porter que sur les totaux annuels et en général sur les totaux mensuels, mais pas sur les totaux journaliers.

- c) Le 1er Janvier de l'année X_i , on a modifié l'environnement immédiat du pluviomètre. Si l'on sait que l'état actuel des lieux est le même que pendant la période des années X_i à X_n , on peut vérifier s'il est satisfaisant et dans l'affirmative corriger les données de la période plus ancienne. Si l'on ignore quel était l'état des lieux pendant chacune des deux périodes, on est ramené au cas (b) ci-dessus.

4.4 - Interprétation du graphique : choix entre les hypothèses possibles

4.4.1 - Nouvel examen des originaux

Si l'on a détecté une hétérogénéité, il est bon de vérifier si elle apparaît, et de quelle façon, dans les relevés journaliers. Dans certains cas, un indice de la mauvaise qualité générale des observations peut être ainsi découverte, après être passé inaperçu au premier examen ; en pareil cas, si la lecture du graphique des totaux cumulés comparés laissait planer un doute sur la correction à effectuer - cas (b) ou (c) - ce doute est levé automatiquement car la cause physique supposée être la cause principale de l'hétérogénéité s'avère alors comme une cause secondaire.

4.4.2 - Enquêtes diverses

On peut chercher à connaître, auprès des personnes qui détiennent les relevés anciens ou celles qui sont chargées actuellement des relevés, les dates de changement possible d'emplacement de l'appareil, de modifications possibles de l'environnement, de changement de l'observateur (qui peut ne pas être toujours la personne chargée officiellement des relevés, mais un membre de sa famille), etc ...

4.4.3 - Considération de l'ensemble des données régionales

Comme on l'a dit au début du paragraphe 4.3, les diverses hypothèses émises au sujet du poste B reposent sur la certitude a priori que c'est la série B et non la série A qui est hétérogène. En pratique, on a rarement cette certitude.

Il faut donc compléter l'information fournie par le graphique "B en fonction de A" par celle des graphiques "C, D, ..., en fonction de A" (C, D, ..., étant d'autres stations de la même région). Si ces derniers graphiques n'accusent pas de "cassure", les séries A, C, D, ... sont très probablement homogènes et B est hétérogène. Si tous les graphiques présentent une cassure à la même date, et si les rapports des pentes $\frac{m}{m'}$ sont les mêmes quelle que soit le poste B, C, D, ..., on peut dire que c'est la série A qui est hétérogène.

Il faut pourtant se méfier des dates où un bouleversement général a pu se produire dans la région : un certain nombre de postes B, C, D, ..., peuvent avoir été perturbés alors que A a été épargné (il est vrai que dans ce cas, les perturbations produisent généralement des cassures caractérisées par des rapports $\frac{m}{m'}$ différents).

5. DETERMINATION de SERIES HOMOGENES CORRESPONDANT aux PERIODES d'OBSERVATION : APPLICATION de la METHODE des COURBES de COMPARAISON des TOTAUX CUMULES

Le procédé de détection des hétérogénéités et de leur correction a déjà été décrit dans ses grandes lignes au chapitre 4.

Il reste à préciser comment on doit ajuster une droite sur un graphique de totaux cumulés comparés, et quelle série de base on doit choisir pour corriger une série hétérogène donnée.

5.1 - Tracé des droites

La théorie mathématique de la méthode n'a, semble-t-il, pas encore été faite, mais l'expérience a montré que l'ajustement graphique d'une droite doit être fait de la façon suivante : la droite doit avoir pour pente la valeur modale (c'est-à-dire la plus fréquente) du rapport $(\frac{P_B}{P_A})_i$ des totaux annuels des deux postes. En pratique, les années pour lesquelles ce rapport est très voisin de sa valeur modale se suivent par 3 ou 4 ; les points représentatifs sont alignés, et l'on peut tracer la droite d'ajustement parallèle à ces alignements. On remarquera que la pente de cette droite n'est pas exactement égale, en général, au rapport des totaux cumulés en fin de période $\frac{\sum P_B}{\sum P_A}$, ni même au rapport des moyennes interannuelles vraies dont l'expression ci-dessus n'est qu'une estimation. La théorie mathématique de la méthode montrerait sans doute que ce fait est lié à la dissymétrie de la loi de distribution du rapport $(\frac{P_B}{P_A})_i$.

Le procédé d'ajustement des droites, indiqué ci-dessus, n'est pas aussi facile à appliquer que s'il suffisait de joindre le 1^{er} et le i^o ou le n^o point du graphique, ou s'il fallait tracer la droite de régression de $\sum P_B$ en $\sum P_A$ ou de $\sum P_A$ en $\sum P_B$. Le procédé conseillé a donc l'inconvénient de laisser intervenir un facteur personnel, mais on verra à l'usage qu'il permet de déterminer avec moins d'ambiguïté la date d'une cassure ; de plus, le résultat de l'ajustement est moins influencé par les valeurs exceptionnellement fortes ou faibles du rapport $\frac{P_B}{P_A}$ qui, certaines années, peuvent provenir d'un défaut impossible à déceler dans les observations de l'un des postes.

5.2 - Choix d'une série de base

Le choix d'une série de base telle que la série des totaux cumulés P_A détermine toute la suite des opérations. Nous abordons seulement ici la question de ce choix car on ne peut le justifier sans avoir effectué plusieurs essais d'ajustement par le procédé indiqué au paragraphe précédent.

5.2.1 - Choix des séries de base destinées aux premiers essais de vérification d'homogénéité

Il arrive souvent que l'on ait à effectuer tout un ensemble

d'opérations de vérification, pour les postes d'un réseau présentant des périodes d'observation très diverses. On a intérêt à représenter ces périodes d'observation sous une forme condensée et de lecture facile, en groupant les postes par sous-région (voir modèle ci-joint, graphique 1).

On peut alors choisir un ou plusieurs postes de base dont les périodes d'observation simultanée avec la plupart des autres sont assez longues. Plusieurs postes de base peuvent être nécessaires si l'on veut que les vérifications portent sur l'ensemble des données.

5.2.2 - Choix de la série destinée à l'étude systématique des postes d'une région

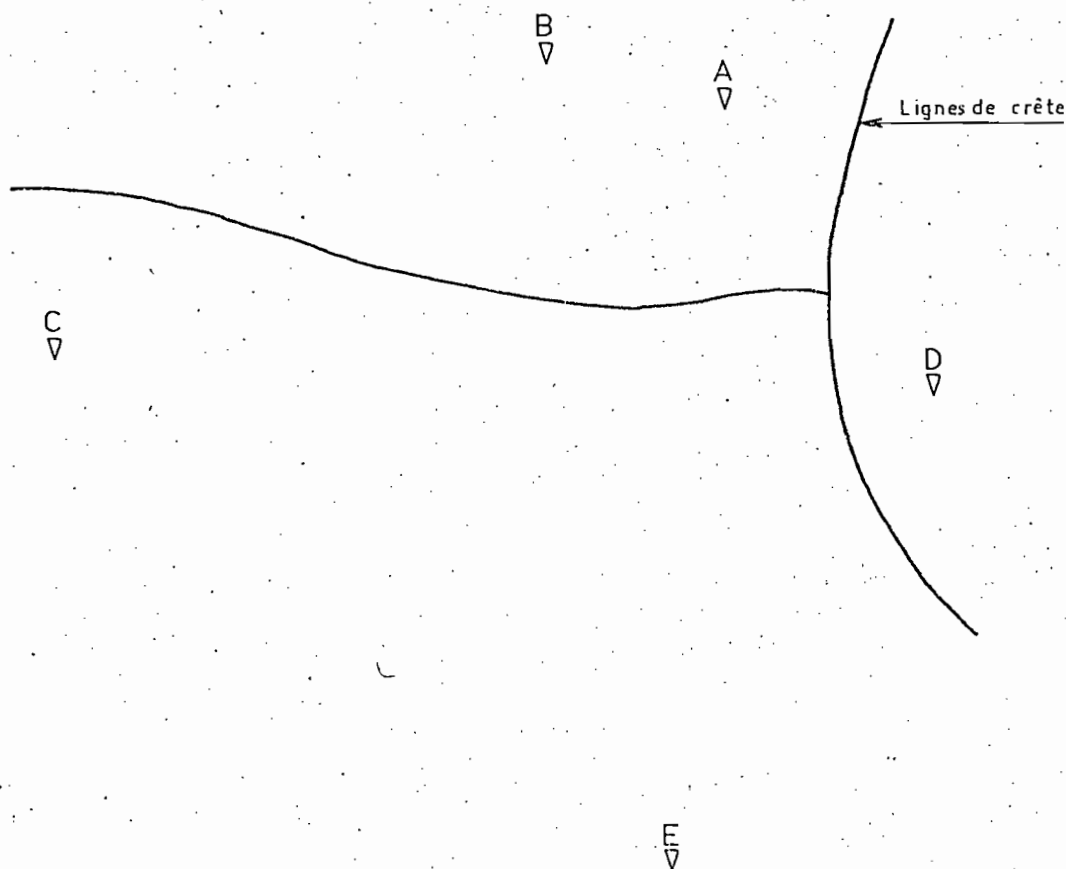
Lorsqu'un graphique accuse un décrochement très important, autrement dit que l'un des rapports $\frac{PB}{PA}$ est très différent des autres, on peut examiner en détail les relevés de l'année en cause et éventuellement éliminer les données de l'une des deux stations si elles sont aberrantes. Cependant, il est bien plus fréquent de ne pas pouvoir prendre cette décision, et d'être obligé de considérer que les données sont exactes mais correspondent à un phénomène rare.

On peut alors se demander si la fréquence d'apparition des décrochements, sur le graphique étudié, ne serait pas diminuée environ de moitié en remplaçant une des deux séries, la série de base A, par celle qui précéderait de la moyenne des données de plusieurs postes. On ne peut donner de règles générale pour l'application de ce procédé, sinon rappeler que la constitution d'un groupe de base doit être précédée de la vérification minutieuse de l'homogénéité des séries composantes, qui doivent être comparées graphiquement deux à deux. On peut être amené à préférer une série donnée par un seul poste de base, si la fréquence des pluviométries exceptionnelles enregistrées à ce poste est plus faible que partout ailleurs.

La série de base étant choisie pour l'étude systématique de tous les postes d'une région, il est possible de grouper les graphiques de comparaisons totaux cumulés en un seul, la série de base étant reportée sur l'axe des abscisses et chaque millésime étant indiqué en face de l'intervalle correspondant. Les graphiques peuvent être décalés par changement d'origine sur l'axe des ordonnées, afin d'éviter que leurs droites d'ajustement se croisent. Si l'on prend soin de choisir judicieusement

Représentation graphique des périodes d'observation de postes pluviométriques

Noms des postes pluviométr.	ANNÉES																															
	196..								195..								194 ..								193 ..							
	6	5	4	3	2	1	0	9	8	7	6	5	4	3	2	1	0	9	8	7	6	5	4	3	2	1	0	9	8	7	6	5
A																																
B																																
C																																
D																																
E																																



l'ordre du tracé des différents graphiques, en groupant les postes par sous-région géographique ou climatique, on verra apparaître des caractères communs aux séries des postes d'une même sous-région : même allure des fluctuations, autour de la droite d'ajustement, de la ligne brisée joignant les points représentatifs.

5.2.3 - Choix de la série de base destinée à corriger une hétérogénéité

Le graphique général établi pour la région permet de mettre en évidence des hétérogénéités, en conservant à peu près la même sévérité pour l'ensemble des postes, c'est-à-dire en considérant comme hétérogénéité les seules "cassures" qui affectent l'allure des graphiques de façon nettement plus accusée que les fluctuations à caractère sous-régional évoquées au paragraphe précédent.

On obtiendra cependant une meilleure précision pour déterminer la correction à apporter à une série hétérogène, en la comparant avec une série homogène de la même sous-région, à condition que la période commune d'observation des deux postes soit assez étendue de part et d'autre de la date de la "cassure" du graphique, car c'est bien là la principale difficulté pratique d'application de l'ensemble des procédés indiqués dans cette note : la diversité des périodes d'observations.

6. HOMOGENEISATION des PERIODES d'EXTENSION des SERIES PLUVIOMETRIQUES

6.1 - Méthode à employer

Comme on l'a dit plus haut (Cf. paragraphe 2.1), on peut considérer, pour simplifier l'inventaire des difficultés rencontrées dans l'étude de la pluviométrie, que l'absence de données est une hétérogénéité. Or, on a indiqué également que l'extension des données devait être effectuée à l'aide de corrélations.

On peut se demander pourquoi la méthode graphique décrite aux chapitres 4 et 5 n'est pas conseillée également pour effectuer cette extension, d'autant plus qu'elle conduirait forcément à des résultats un peu différents.

On écarte la méthode graphique pour la raison principale suivante : elle ne permet d'estimer directement que des valeurs annuelles et des moyennes interannuelles ; la variance des termes d'une série ainsi

étendue ne peut pas être estimée directement à partir de celles des données observées. Rien n'empêche d'estimer la variance de la série étendue à partir des termes estimés, et d'en déduire par exemple l'intervalle de confiance de la moyenne interannuelle, mais le "facteur personnel" intervenu dans l'ajustement graphique de la droite aura ici des conséquences plus graves que dans l'estimation des valeurs annuelles et des moyennes.

En attendant qu'une théorie plus complète permette l'exploitation, par le calcul, des totaux cumulés comparés, on utilisera donc des régressions linéaires entre totaux annuels.

Ce procédé n'est satisfaisant que si les séries obéissent à la loi de GAUSS (mais on a vu - Cf. Supra, paragraphe 4.1 - que cette condition est réalisée), et si le phénomène de persistance n'apparaît pas (mais il n'apparaîtrait que si l'on étudiait par exemple des totaux journaliers).

6.2 - Précautions à prendre

Il est essentiel de vérifier tout d'abord l'homogénéité des séries pour la période d'observation simultanée. Si l'on veut étendre la série B observée pendant 30 ans à partir de la série A observée pendant 15 ans de plus, mais si l'étude graphique de la série des 30 premières années révèle une "cassure" en deux séries homogènes de 10 et 20 ans, le calcul de la régression linéaire ne peut porter que sur un des deux échantillons, et on choisira le plus long (ici 20 ans).

D'autre part, si dans l'échantillon retenu il se trouve des valeurs particulièrement douteuses, parmi celles qui provoquent des "décrochements" sur les graphiques de contrôle, on peut soit les éliminer de cet échantillon, soit les conserver, dans l'optique de la recherche d'une sécurité plus grande (on choisit par exemple la solution qui conduit à la plus faible estimation de la moyenne interannuelle).

6.3 - Choix des séries à étendre

Si l'on veut dresser une carte d'isohyètes interannuelles, les données de tous les postes peuvent être utiles, même si pour certains la série observée est très courte. Le calcul de l'intervalle de confiance de la moyenne estimée sur la période étendue, évitera de mal interpréter certains écarts et dans l'ensemble le tracé des isohyètes y gagnera de la précision.

Par contre, si les données pluviométriques sont destinées à établir des régressions hydropluviométriques en vue de préciser les caractéristiques du régime d'une rivière observée pendant k années, il est inutile de chercher à étendre les séries pluviométriques ne présentant pas, au départ, beaucoup plus de k années d'observation.

6.4 -- Choix de la période d'homogénéisation, et des séries de base

Si l'on dispose de moyens de calcul automatique, on peut déterminer l'ensemble des opérations qui conduisent au gain d'information maximal : l'ordinateur peut effectuer tous les calculs à partir de chacun des schémas qu'on lui propose et choisir ainsi le meilleur schéma.

Si les calculs doivent être faits à l'aide de simples machines électromécaniques, on doit choisir a priori le schéma des opérations qui est vraisemblablement le meilleur.

6.4.1 -- Choix de la période d'homogénéisation

On peut commencer par traduire l'inventaire des observations sous la forme d'un tableau. Reprenons l'exemple cité par Marcel ROCHE dans son Cours professé à l'ENGREF (bassin du MONO, au TOGO).

On peut utiliser 33 postes observés de 9 à 44 ans :

pour 4, la période est de 44 ans			pour 1, la période est de 20 ans		
" 2	"	35 "	" 2	"	19 "
" 1	"	33 "	" 1	"	18 "
" 1	"	31 "	" 1	"	17 "
" 1	"	29 "	" 1	"	15 "
" 2	"	28 "	" 2	"	12 "
" 1	"	25 "	" 4	"	11 "
" 2	"	23 "	" 5	"	10 "
			" 2	"	9 "

On a choisi une période de 29 ans couverte par 9 séries.

Il est probable que si l'on essayait d'étendre toutes les séries à une période plus longue, le gain d'information serait négligeable par suite de la proportion plus faible de données observées dans les séries.

6.4.2 - Choix des séries de base

6.4.2.1 - Constitution d'un groupe de base

Parfois, il est possible de trouver quelques postes présentant une série d'observations complète pour la période d'homogénéisation (cas du bassin du MONO évoqué ci-dessus), mais de toute façon on doit résoudre un à un les problèmes posés par chaque poste.

6.4.2.2 - Exemples de schémas d'opérations

Considérons le poste B observé de 1955 à 66 inclus.

On cherche à étendre la série à la période 1935-66, à l'aide des données des postes A, C, D, E (exemple du graphique 1).

Les valeurs de la pluviométrie en B peuvent être estimées :

- de 1951 à 54 à partir de A, C, D ou E
- de 1949 à 50 " C, D ou E
- de 1941 à 48 " C ou E
- de 1935 à 40 " E

Cependant, on ne peut envisager le calcul d'une régression linéaire pour des échantillons inférieurs à une dizaine de couples de valeurs. L'estimation de P_B de 1935 à 40 doit donc être effectuée en deux étapes :

- estimation de P en C ou en D à partir de E
- estimation de P en B à partir de C ou D

6.4.2.3 - Critères pour le choix d'un schéma

Parfois, les séries ont déjà été comparées deux à deux par la méthode graphique exposée plus haut : on verra entre les deux combinaisons, BA et BC par exemple, celle qui correspond au graphique le plus satisfaisant, et si BA est préférable, on estimera P_B de 1951 à 54 à partir de A et seulement de 1941 à 50 à partir de C : schéma (1). Même si les deux graphiques sont aussi satisfaisants, on adoptera le même schéma de calcul (1), car la longueur de l'échantillon est de 12 années pour le couple BA contre 10 pour le couple BC.

Si les couples BA et BC n'ont pas fait l'objet d'un contrôle graphique d'homogénéité, les séries A, B, C doivent au moins figurer, chacune pour son compte, sur le graphique général de comparaison à une série de base régionale (Cf. Supra, paragraphe 5.2.2), et l'examen de ce document suffit : le schéma (1) indiqué dans l'exemple ci-dessus sera effectivement choisi sauf si l'allure du graphique de contrôle est très différente pour la série A d'une part, et les séries B et C d'autre part.

Notons que PB aurait pu être estimé de 1951 à 54 à partir de D, mais nous avons admis implicitement jusqu'ici que le critère "allure des graphiques de contrôle" n'était pas favorable au choix de ce schéma. Même si ce critère n'est pas déterminant, le même schéma (1) sera toujours adopté car la distance BA est plus courte que BD, et parce qu'une crête isole la station D des autres.

Supposons maintenant, pour simplifier, que le critère "allure des graphiques de contrôle" ne soit déterminant pour le choix d'aucun schéma, et nous voyons que, en appliquant les critères déjà énoncés au passage dans l'étude de notre exemple, P_B devra être estimé à partir de D pour 1949 et 50, à partir de C pour 1935 à 48, P_C ayant été estimé au préalable à partir de E pour 1935 à 40.

En pratique, cependant, pour alléger les calculs sans perdre une quantité d'information appréciable, P_B pourra être estimé à partir de C de 1935 à 50.

En résumé, les critères de choix entre divers schémas sont par ordre d'importance :

- 1° L'intensité de la liaison des séries, combinée à l'étendue des échantillons (si l'on traitait le problème par le calcul, ce critère deviendrait : valeur du coefficient de corrélation ayant telle probabilité donnée, par exemple 95 %, d'être dépassée).
- 2° La distance des postes deux à deux (ou en cas d'égalité, leur différence d'altitude, ou de situation géographique).
- 3° La réduction du volume des calculs, si elle peut être consentie sans s'écarter sensiblement du choix dicté par les autres critères.

6.5 - Conduite des calculs pour l'extension des séries de totaux pluviométriques annuels

Les estimations des valeurs annuelles et des paramètres

statistiques d'une série Y de k années observées au poste B, à partir d'une série X de n valeurs observées en A, sont fournies par les formules suivantes (les k années d'observation de Y étant incluses dans les n de X) :

$$a) \quad y_x = \bar{y}_k + k^{r_{xy}} \frac{k^s_y}{k^s_x} (x - \bar{x}_k)$$

y_x moyenne conditionnelle de y liée à x (par exemple moyenne des valeurs possibles de la pluviométrie au poste B pour l'année 1952 sachant qu'elle était $x = 854$ mm en A).

$\bar{x}_k$ et $\bar{y}_k$ moyennes interannuelles de x et de y calculées respectivement à partir des k valeurs de x et de y observées simultanément.

$k^{r_{xy}}$ coefficient de corrélation entre x et y estimé à partir des k couples de valeurs observées simultanément.

k^s_x et k^s_y écarts-types de x et y estimés à partir des échantillons de k valeurs.

$$b) \quad \hat{\bar{y}} = \bar{y}_k + k^{r_{xy}} \cdot \frac{k^s_y}{k^s_x} (\bar{x}_n - \bar{x}_k)$$

$\hat{\bar{y}}$ estimation de la moyenne de y (dont l'espérance mathématique est $\bar{y}$, toujours inconnue) à partir de la première estimation $\bar{y}_k$, des autres paramètres statistiques des échantillons de valeurs observées et du coefficient de corrélation entre x et y.

$\bar{x}_n$ moyenne interannuelle de x calculée à partir des n valeurs observées.

$$c) \quad \hat{\sigma}_y^2 = k^{s^2}_y + k^{r^2}_{xy} \frac{k^{s^2}_y}{k^{s^2}_x} (n^{s^2}_x - k^{s^2}_x)$$

$\hat{\sigma}_y^2$ estimation de la variance de y (dont l'espérance mathématique est σ_y^2), à partir de la première estimation $k^{s^2}_y$, et des estimations de la variance de x et du coefficient de corrélation entre x et y.

$n^{s^2}_x$ variance de x estimée à partir de l'échantillon de n valeurs.

$$d) \hat{\rho} = k_{xy}^r \frac{k_y^s}{k_x^s} \frac{n_x^s}{\hat{\sigma}_y}$$

estimation du coefficient de corrélation entre x et y (dont l'espérance mathématique est ρ), à partir de la première estimation k_{xy}^r , des écarts-types des échantillons de valeurs observées et de l'estimation $\hat{\sigma}_y$ définie ci-dessus.

$$e) E = 1 + \left(1 - \frac{k}{n}\right) \frac{1 - (k-2) \hat{\rho}^2}{k-3} \quad \text{expression comprise entre } \frac{k}{n} \text{ et } 1$$

E efficacité relative de $\bar{y}_k$ et de $\bar{y}$, définie par le rapport de la variance de $\bar{y}$ à celle de $\bar{y}_k$ (plus E est grand, moins l'extension apporte d'information supplémentaire, autrement dit la définition de E n'est pas très judicieuse).

$$f) \text{ Nombre d'années "efficaces" } n' = \frac{k}{E}$$

Ce nombre n' (toujours compris entre k et n) doit être utilisé pour le calcul de la variance de $\bar{y}$ à partir de σ_y^2 par l'expression $\frac{\sigma_y^2}{n'}$ (qui est à rapprocher de la formule générale $\sigma_u^2 = \frac{\sigma_u^2}{N}$ si l'on veut justifier l'appellation de n'). En effet, la variance de $\bar{y}_k$ est $\frac{\sigma_y^2}{k}$ et celle de $\bar{y}$ est donc :

$$\text{-- d'après la définition de } E \quad \frac{E \sigma_y^2}{k}$$

$$\text{-- ou encore, d'après la définition de } n' \quad \frac{\sigma_y^2}{n'}$$

g) Si l'on veut connaître l'intervalle de confiance à 95 % de $\bar{y}$, on supposant que $n' \geq 30$ pour pouvoir admettre que $\bar{y}$ est distribué normalement :

-- on consulte les tables de GAUSS si l'on a oublié la valeur bien connue de la variable réduite qui correspond à :

$$F(t) = \frac{1-0,95}{2} = 0,475 \quad \text{ou à } \Pi(t) = \frac{1+0,95}{2} = 0,975$$

$$\text{soit } t = 1,96$$

on en déduit qu'il y a 95 % de chances pour que $\bar{y}$ soit compris entre $\bar{y} \pm 1,96$ fois l'écart-type $\frac{\hat{\sigma}_y}{\sqrt{n}}$.

Remarque 1

On désire toujours donner des résultats avec un certain coefficient de sécurité, chaque fois que les calculs reposent sur des hypothèses simplificatrices, ce qui est très souvent le cas en calcul statistique. En particulier, on doit s'inquiéter si l'estimation $\hat{\sigma}_y^2 < k s_y^2$, car le résultat donné par la formule (c) est celui que l'on trouverait en appliquant la formule de définition de la variance

$$s_y^2 = \frac{\sum (y_i - \bar{y}_n)^2}{n - 1}$$

à la série composite des k valeurs y_i observées et des $(n - k)$ valeurs estimées par la formule (a) ; les $(n - k)$ valeurs estimées pourraient être lues sur le graphique de régression grâce à la droite de régression de y en x d'équation (a), et ceci montre bien que l'estimation de la variance $\hat{\sigma}_y^2$ est sûrement trop faible, la dispersion des valeurs estimées étant plus faible que celle des valeurs observées.

Donc, si le $\hat{\sigma}_y^2$ trouvé est plus faible que $k s_y^2$, ce qui ne se produit tout de même pas toujours, on conserve (dans la suite des calculs qui conduisent à l'intervalle de confiance de y) la valeur $k s_y^2$ à la place de $\hat{\sigma}_y^2$, sauf si le coefficient de corrélation r_{xy} est supérieur à 0,95, car il faut bien admettre une restriction de ce genre (en effet pour $r = 1$, $\hat{\sigma}_y^2$ est forcément une meilleure estimation que $k s_y^2$ puisqu'elle à la valeur que donnerait un échantillon de n valeurs observées).

Remarque 2

On peut calculer, parallèlement à ce qu'on a fait pour l'estimation $\hat{y}$ de la moyenne, la variance de l'estimation $\hat{\sigma}_y^2$ de la variance. Ce calcul permettrait, en principe, de déterminer l'intervalle de confiance de la variance de y , comme on a déterminé celui de la moyenne de y au paragraphe g) ; en réalité, la formule donnant ($\text{var. } \hat{\sigma}_y^2$) est très complexe et de plus, $\hat{\sigma}_y^2$ qui on l'a vu est une estimation biaisée

par défaut, est évidemment distribuée autour d'une moyenne constituant une autre estimation biaisée par défaut, moyenne fournie par une autre formule complexe. Il est donc inutile de calculer un intervalle de confiance centré sur cette moyenne, sachant qu'il y aura toujours une distorsion dans le résultat.

En conclusion, on adopte l'estimation de σ_y^2 la plus forte, entre $\hat{\sigma}_y^2$ et $k s_y^2$ (sauf si $r_{xy} = 0,95$, comme on a dit ci-dessus -remarque 1), et si l'on veut avoir une idée de l'intervalle de confiance de σ_y^2 , on peut dans l'un et l'autre cas utiliser la formule donnant l'écart-type de ce paramètre estimé d'après un échantillon de k valeurs (ou de n valeurs si l'on ne recherche pas une trop grande sécurité) :

$$\text{écart-type } (k s_y^2) = k s_y^2 \sqrt{\frac{2}{k}}$$

Le calcul de l'intervalle de confiance de σ_y peut évidemment être effectué en prenant les racines carrées des bornes de l'intervalle de confiance de σ_y^2 , ou bien en utilisant l'autre formule classique équivalente pour les grands échantillons :

$$\text{écart-type } (k s_y) = k s_y \frac{1}{\sqrt{2k}}$$

Remarque 3

Le calcul des intervalles de confiance de $\bar{y}$ et de σ_y a été exposé jusqu'ici dans l'hypothèse où l'échantillon est assez grand (k ou n supérieur à 30, en pratique).

Pour un petit échantillon :

- l'estimation $\bar{y}_k$ suit une loi de STUDENT à $k - 1$ degrés de liberté
- .. " $\hat{y}$ une même loi à $n - 1$ degrés de liberté.
- .. " $k s_y^2$ suit une loi de χ^2 à $k - 1$ degrés de liberté.

L'application pratique de ces propriétés au calcul des intervalles de confiance sera traité à l'occasion de l'étude des débits (Cf. paragraphe 4.2.3.1 de la note correspondante).

7. ESTIMATION des TOTAUX MENSUELS ou SAISONNIERS

L'estimation des totaux mensuels ou saisonniers ne mérite pas, en général, d'être effectuée avec autant de précision que celle des totaux annuels, d'autant plus que les valeurs estimées doivent être corrigées pour éviter une incohérence désagréable de l'ensemble des résultats (on s'arrange pour que la somme des estimations des totaux pluviométriques, pour les différentes parties d'une année, soit égale à l'estimation du total annuel).

D'ailleurs, il est rare que la distribution statistique des pluies d'un mois donné puisse être considérée comme normale : la régression entre les séries de deux postes n'est pas alors linéaire et la méthode d'estimation ne peut être calquée sur celle qui a été exposée au chapitre 6.

En pratique :

- on commence par effectuer l'extension des séries de totaux annuels par la méthode du chapitre 6 ;
- on établit graphiquement des régressions (une pour chaque mois ou saison) entre les séries considérées, sur la même période d'observations simultanées qui a fourni les séries de totaux annuels (k années) ;
- on estime les $(n - k)$ totaux mensuels ou saisonniers que fournissent chacune des régressions obtenues, à partir des données du poste de base ;
- on totalise les estimations précédentes pour en déduire une nouvelle série de $(n - k)$ totaux annuels estimés ;
- si pour une année j de la période $(n - k)$, ce dernier total est P^j_j , alors que la méthode du chapitre 6 conduit à P_j , il faut multiplier chacune des estimations mensuelles ou saisonnières de l'année j par le rapport $\frac{P_j}{P^j_j}$.

Remarque :

Si l'on effectue une étude qui comporte la recherche de corrélations hydropluviométriques à l'échelle annuelle, on peut avoir intérêt à choisir a priori la date du début de l'année hydrologique. En effet, si ce choix est à peu près imposé par la considération du régime (étage très nettement marqué, faible importance relative des variations du volume

d'eau stocké dans les couches souterraines en relation avec la rivière), il est plus rapide et plus exact d'opérer avec les totaux pluviométriques observés par année hydrologique (desquels on déduit les totaux non observés par extension) que d'opérer avec les totaux calendaires, car l'estimation des totaux non observés, par année hydrologique, reposerait alors en partie sur la méthode graphique que l'on vient d'indiquer dans ce chapitre 7.

Cependant, si la date du début de l'année hydrologique ne peut être choisie qu'à la suite de divers essais de corrélations hydropluviométriques, parmi lesquelles on adoptera celle qui a la meilleure efficacité, le plus simple est alors d'opérer sur les totaux pluviométriques calendaires pour effectuer l'extension à l'échelle annuelle ; ceci est d'autant plus logique qu'il y a de fortes chances pour que la corrélation hydropluviométrique ne soit pas simple, mais fasse appel à un facteur principal et à un ou plusieurs autres facteurs de l'écoulement, facteurs liés principalement à la répartition de la pluviométrie dans l'année hydrologique et à l'état des réserves souterraines lié lui-même aux précipitations antérieures.

A N N E X E S

EXERCICES d'APPLICATION NUMERIQUE des METHODES de CONTROLE
et d'HOMOGENEISATION des DONNEES PLUVIOMETRIQUES

1. EXERCICE d'APPLICATION NUMERIQUE de la DETECTION et de la CORRECTION d'une HETEROGENEITE par la METHODE de COMPARAISON GRAPHIQUE des TOTAUX ANNUELS CUMULES

Les deux séries à comparer sont fournies par des observations effectuées en ALSACE : la série de base correspond au poste de St GILLES (série homogène) et la série à tester correspond au poste de TRUTTENHAUSEN.

On sait que les observations à ce dernier poste ont été confiées successivement à 3 personnes différentes :

- de 1946 à 1949 inclus
- de 1950 à 1952 inclus
- de 1953 à 1960 inclus

Présentation des données

Année	St-GILLES	TRUTTENHAUSEN	Année	St-GILLES	TRUTTENHAUSEN
1960	800	869	1952	820	953
59	549	596	51	841	882
58	858	994	50	732	945
57	540	643	49	459	694
56	657	736	48	522	875
55	677	734	47	540	849
54	702	699	46	511	791
53	393	546			

Opérations à effectuer

- a) Calculer les totaux cumulés des deux séries en remontant le temps (1960, 60 + 59, 60 + 59 + 58, etc ...).
- b) Mettre en place les points représentatifs sur un graphique.
- c) Ajuster une ou plusieurs droites au graphique.

- d) Corriger éventuellement les données d'une ou plusieurs périodes, sachant que l'étude de la pluviométrie du poste de TRUTTENHAUSEN a pour objectif la détermination de besoins en eau d'irrigation.

2. EXERCICE d'APPLICATION NUMERIQUE de l'EXTENSION d'une SERIE PLUVIOMETRIQUE à PARTIR d'une REGRESSION LINEAIRE entre TOTAUX ANNUELS de DEUX SERIES

On utilisera ici aussi des données recueillies en ALSACE. Le poste de base est St-GILLES, observé de 1891 à 1965. La série à étendre est celle du poste de GUEBWILLER, observé de 1946 à 1965.

On admettra que les deux séries sont homogènes et que les distributions sont normales, ce qui autorise à calculer une équation de régression linéaire.

Présentation des données

Voir les tableaux 1 et 2 ci-joints.

La moyenne interannuelle de 1891 à 1965 à St-GILLES est de 667 mm, et la variance de la série est 18 541.

Conseils pratiques et résultats intermédiaires fournis

On utilisera les formules suivantes qui dérivent des formules de définition de la variance et du coefficient de corrélation, et son particulièrement recommandées pour le calcul électromécanique (les signes Σ doivent être lus Σ_1^k) :

$$k^s_x^2 = \frac{k \Sigma x^2 - (\Sigma x)^2}{k(k-1)}$$

$$k^s_y^2 = \text{expression analogue en } y$$

On conservera les numérateurs de ces expressions, pour les calculs suivants. En effet l'équation (a) de la droite de régression, donnée dans l'exposé de la méthode, peut être mise sous la forme $y = ax + b$. On a alors :

$$a = \frac{k \Sigma xy - \Sigma x \Sigma y}{k \Sigma x^2 - (\Sigma x)^2}$$

$$b = \frac{\Sigma y \Sigma x^2 - \Sigma x \Sigma xy}{k \Sigma x^2 - (\Sigma x)^2}$$

TABLEAU 1

SAINT-GILLES

Totaux pluviométriques annuels et saisonniers observés

Années	Janv Fév.	Mars	Avril	Mai	Jn-Jt Août	Sept.	Oct	Nov Déc	Totaux annuels
1936	156	21	39	21	353	115	36	84	825
37	152	68	54	34	168	115	27	83	701
38	131	12	10	78	398	65	72	78	844
39	71	63	87	132	274	80	155	101	963
1940	52	40	76	77	197	156	63	160	815
41	185	60	78	85	223	40	71	55	797
42	60	32	42	65	157	98	20	65	539
43	82	16	37	59	137	127	5	98	563
44	75	16	34	16	181	108	63	187	680
1945	80	29	23	46	311	38	23	73	623
46	105	12	10	44	243	42	18	37	511
47	48	77	8	45	151	11	59	141	540
48	98	7	31	45	227	59	38	17	522
49	52	35	38	93	117	46	14	64	459
1950	91	3	62	74	196	65	14	227	732
51	132	40	37	60	269	45	148	110	841
52	116	97	60	40	74	96	123	214	820
53	38	2	31	35	223	38	21	5	393
54	82	17	53	31	239	118	26	136	702
1955	147	33	0	47	316	39	32	62	677
56	45	23	69	89	251	62	65	53	657
57	89	20	12	38	250	47	11	73	540
58	215	25	29	92	225	100	39	133	858
59	122	42	72	40	141	11	34	87	549
1960	125	40	25	57	309	53	86	105	800
61	122	14	78	67	163	24	63	106	625
62	150	75	57	44	71	65	10	96	568
63	54	54	29	34	313	34	21	120	659
64	29	73	49	61	164	51	57	64	548
1965	103	75	44	123	165	79	11	217	817

TABIEAU 2

GULBELLER

Totaux pluviométriques annuels et saisonniers observés

: Années :	Jan Fév	Mars	Avril	Mai	Jn-Jt Août	Sept	Oct	Nov Déc	Totaux annuels
: 1946 :	230	16	11	81	273	71	20	108	810
: 47 :	90	167	15	20	151	17	40	293	793
: 48 :	201	16	67	61	226	41	62	63	737
: 49 :	70	40	76	112	130	53	30	128	639
: 1950 :	201	5	95	98	263	89	25	393	1169
: 51 :	222	120	71	67	255	44	23	200	1002
: 52 :	225	110	56	47	106	126	186	392	1248
: 53 :	78	3	36	37	267	59	24	18	522
: 54 :	138	53	69	32	326	124	50	252	1044
: 1955 :	390	68	1	81	280	101	47	172	1140
: 56 :	100	38	108	92	316	55	91	86	886
: 57 :	253	23	14	23	273	88	9	93	776
: 58 :	466	49	54	119	241	85	51	233	1288
: 59 :	267	75	109	67	161	5	57	207	948
: 1960 :	223	55	20	59	309	45	132	216	1059
: 61 :	212	18	68	90	213	23	81	167	872
: 62 :	252	158	104	86	108	83	13	166	970
: 63 :	93	118	37	36	405	44	46	280	1059
: 64 :	52	87	63	74	159	72	144	111	762
: 1965 :	167	162	61	159	236	145	20	549	1499

$$k^r_{xy} = a \frac{k \sum xy - \sum x \sum y}{k \sum y^2 - (\sum y)^2}$$

Notons aussi que :

$$k^r_{xy} \frac{k^s_y}{k^s_x} = a^2$$

(expression qui entre dans le calcul de $\hat{\sigma}_y^2$ et dans celui de $\hat{\beta}^2$).

Pour faciliter les calculs, nous donnons les valeurs numériques suivantes :

$$\begin{array}{ll} \sum x^2 = 8\,573\,110 & \sum x = 12\,816 \\ \sum y^2 = 19\,264\,082 & \sum y = 19\,082 \\ \sum xy = 12\,751\,417 & \end{array}$$

(sommes étendues aux échantillons de 20 termes).

Opérations à effectuer

- Calculer les variances de x et de y à partir des échantillons de 20 valeurs fournies, l'équation de la droite de régression de y en x et le coefficient de corrélation k^r_{xy} .
- Tracer la droite et placer les points expérimentaux à titre de vérification.
- Estimer la pluviométrie à GUEBIVILLER, année par année, de 1936 à 1945.
- Estimer la pluviométrie moyenne interannuelle à GUEBIVILLER, à partir de celle de St-GILLES estimée par la moyenne des 75 valeurs observées de 1891 à 1965.
- Estimer la variance de la série pluviométrique de GUEBIVILLER en tenant compte de l'extension (calcul de $\hat{\sigma}_y^2$). Comparer $\hat{\sigma}_y^2$ à k^s_y .
- Estimer le coefficient de corrélation entre les 2 séries pluviométriques, en tenant compte de l'extension (calcul de $\hat{\beta}$).

- Calculer l'efficacité relative, le nombre d'années efficaces et l'écart-type de l'estimation de la moyenne interannuelle de GUEBWILLER.
- Donner l'intervalle de confiance à 95 % de cette moyenne.

3. EXERCICE d'APPLICATION NUMERIQUE de l'ESTIMATION des TOTAUX PLUVIOMETRIQUES SAISONNIERS d'une SERIE ETENDUE

Les séries sur lesquelles on opérera sont les mêmes que pour l'exercice précédent.

Les totaux pluviométriques de certains mois ont été groupés afin d'étudier globalement leur importance dans le mécanisme de l'écoulement, et de tester le choix, pour début de l'année hydrologique, du 1er Janvier, du 1er Mars, etc ...

Présentation des données

Voir les tableaux 1 et 2.

Opérations à effectuer

- a) Ajuster graphiquement des droites aux 8 graphiques à établir (un par mois ou par groupe de mois).
- b) Estimer les valeurs non observées de la pluviométrie à GUEBWILLER (de 1936 à 1945).
- c) Corriger les estimations précédentes afin que leur somme, année par année, soit égale au total estimé au cours de l'exercice précédent.