

Com :
Vème Congrès d'Hydraulique de l'Equateur
Asociación Ecuatoriana de Hidráulica
AEHID/CICP, Quito, 23-26/11/1988.

HOMOGÉNÉISATION ET RÉGIONALISATION PLUVIOMÉTRIQUE PAR LA MÉTHODE DU VECTEUR RÉGIONAL

par Patrick LE GOULVEN^{*}, Miguel A. ALEMÁN^{}, Iván OSORNO^{**}**

RÉSUMÉ

Après avoir défini le concept d'homogénéité des séries de précipitations annuelles et démontré l'inefficacité des tests statistiques traditionnels pour résoudre ce problème, les auteurs prouvent l'efficacité des tests de doubles masses et surtout de la méthode du vecteur régional.

Ils utilisent en outre cette dernière méthode pour entreprendre une régionalisation pluviométrique en tenant compte de la variation interannuelle des précipitations (régionalisation chronologique) et de leur répartition tout au long de l'année (régionalisation du régime).

Les démonstrations et conclusions sont illustrées à partir de travaux déjà réalisés sur le bassin hydrographique du Mira.

^{*} Ingénieur hydrologue de l'Institut Français de Recherche Scientifique pour le Développement en Coopération (ORSTOM)

^{**} Ingénieurs hydrauliciens de l'Institut Équatorien des Ressources Hydrauliques (INERHI)

INTRODUCTION

L'inventaire des ressources hydriques ainsi que leur gestion rationnelle font partie des soucis actuels de l'Équateur.

Cependant, ces objectifs sont étroitement liés à une excellente connaissance des éléments du climat, parmi lesquels la pluviométrie joue un rôle important. Sa connaissance quantitative est obtenue au moyen de mesures permanentes d'un réseau d'observations ponctuelles.

Malheureusement, toute obtention de données souffre d'erreurs de diverses natures, ce qui démontre la nécessité d'une critique sévère des informations reçues du réseau, celles-ci pouvant être éliminées lorsqu'elles s'éloignent trop de la réalité, corrigées lorsque la nature des erreurs le permet, ou simplement signalées lorsque la valeur anormale peut être considérée comme localement possible.

De plus, le calcul des besoins en eau et des ressources disponibles requiert une information complète et cohérente, même s'il n'existe pas de stations proches.

Pour atteindre cet objectif, il est indispensable de diviser l'espace en zones homogènes, dans lesquelles le régime pluviométrique ainsi que la variation interannuelle de la précipitation soient semblables et bien connus, afin de mettre ensuite en place des modèles hydropluviométriques permettant de gérer la ressource eau selon des critères scientifiques.

L'Institut Équatorien des Ressources Hydrauliques (INERHI), en collaboration avec l'Institut Français de Recherche Scientifique pour le Développement en Coopération (ORSTOM), a entrepris cette tâche dans la Sierra et sur la Côte, afin d'élaborer ultérieurement son Plan Directeur d'Irrigation. Étant donné l'envergure du travail, l'École Polytechnique Nationale (EPN) et l'Institut National de Météorologie et Hydrologie (INAMHI) ont été associés au programme.

Les quatre institutions travaillent dans des régions différentes sur la base d'une méthodologie commune qui est présentée dans cet article.

A - HOMOGÉNÉISATION

1. DÉFINITION DU PROBLÈME

Une série chronologique de valeurs observées a un caractère aléatoire simple si toutes ses valeurs sont indépendantes et proviennent d'un tirage au sort dans une même population statistique. Le caractère aléatoire peut être modifié, selon Y. Brunet-Moret (1979) par :

- **un effet de persistance**
Les valeurs dépendent des valeurs antérieures mais la série est stationnaire et ses paramètres de distribution ne varient pas dans le temps.
- **un effet de tendance**
L'espérance mathématique des valeurs observées varie dans le temps.
- **des effets cycliques et pseudo-cycliques**
L'espérance mathématique d'une valeur observée est fonction de la chronologie, mais la valeur moyenne de séries suffisamment longues (moyennes mobiles) est stationnaire et ses paramètres ne varient pas dans le temps.
- **erreurs d'observation et de mesure**
Elles affectent un ou plusieurs termes consécutifs de la série et peuvent se superposer aux effets précédents.

Une série chronologique de précipitations annuelles peut contenir simultanément tous ces types d'effets. Les trois premiers caractérisent la pluviométrie du site et constituent ainsi l'organisation interne de la série.

Seules les erreurs d'observation et de mesure (ou erreurs systématiques) affectent l'homogénéité de la série.

Le véritable problème de l'homogénéisation consiste alors à détecter et à identifier ces erreurs.

Ceci constitue le premier objectif de cette première partie.

2. LES SOURCES D'HÉTÉROGÉNÉITÉ ET LES CONSÉQUENCES DE CELLE-CI

Les problèmes qui se présentent dans un réseau d'observation pluviométrique sont très variés. Rappelons seulement les plus fréquents :

- **Modification du milieu environnant ou du site de mesure**
 - Par déplacement de l'appareil
C'est un cas fréquent et la station garde généralement le même nom.
 - Par modification du milieu environnant
Elle peut être brusque (construction proche) ou progressive (croissance d'arbres à proximité).
 - Par changement de la hauteur de l'appareil
La quantité d'eau reçue par le pluviomètre se modifie sensiblement surtout lorsque le site est exposé au vent.

- **Erreurs dues à l'appareil**

- modification de la surface de réception par construction ou déformation ;
- défaut d'étalonnage (dans les pluviographes).

- **Erreurs de mesure**

Au niveau de l'éprouvette ou de la réglette les erreurs sont fréquentes. La confusion d'éprouvette est un cas typique. Autrefois, le sud de l'Équateur était équipé de pluviomètres de 100 cm² avec l'éprouvette en verre correspondante. Une fois que celle-ci était cassée, l'observateur pouvait en recevoir une neuve en plastique, mais étalonnée pour un pluviomètre de 200 cm². Les mesures postérieures au changement sous-estimaient de 50 % la hauteur réelle des pluies.

Le tableau suivant présente les valeurs des erreurs commises dans le cas d'une confusion entre éprouvettes et pluviomètres de surfaces les plus fréquentes.

		éprouvette			
p l u v i o m è t r e		100 cm ²	200 cm ²	314 cm ²	400 cm ²
	100 cm ²	1	0,5	0,318	0,25
	200 cm ²	2,0	1	0,637	0,50
	314 cm ²	3,14	1,57	1	0,79
	400 cm ²	4,0	2,0	1,27	1

Tableau n° 1

- **Erreurs de transmission**

Problèmes de courrier, radio, satellite...

- **Erreurs de classement et de publication**

Toutes formes d'erreurs d'écriture qui se produiraient lors de la frappe de l'information brute (ommission de données, erreurs de date, etc.).

Selon leur nature, les erreurs systématiques ont des conséquences différentes tant sur leur détection que sur la correction des valeurs erronées.

Une confusion d'éprouvette nécessite généralement peu d'années pour être détectée s'il n'y pas d'utilisation simultanée de plusieurs éprouvettes ou réglettes. Dans ce cas, il est possible de rétablir les totaux annuels, mensuels et même journaliers si on connaît la date de la confusion.

Une modification du milieu environnant est détectée si l'on dispose d'un bon nombre d'années homogènes avant et après le changement. Plus le changement est fort, moins d'années sont nécessaires. L'expérience montre que, dans ce cas, seuls les totaux annuels et mensuels peuvent être corrigés.

S'il y a des déplacements fréquents de l'appareil (ou d'autres causes produisant le même effet), les hétérogénéités ne sont pas discernables et il est impossible d'effectuer des corrections.

3. IMPORTANCE DES TESTS D'HOMOGENÉITÉ

Tout ce qui précède fait apparaître la nécessité d'un contrôle rigoureux des données pluviométriques, qui pourrait être effectué en étudiant l'historique des stations.

En général, un changement de site coïncide avec un changement d'observateur et une confusion d'éprouvette se produit après une interruption des mesures. Un bon historique comportant des notes sur les changements d'observateurs ou de site, d'équipements ou de milieu environnant, permettrait alors de résoudre un grand nombre de problèmes.

Malheureusement, si les inspecteurs connaissent ces détails de la vie d'une station, celles-ci ne figurent que rarement par écrit dans un document, et la disparition des « gens qui savent », fait également disparaître l'information.

Une partie de l'historique peut être reconstituée à partir des documents bruts envoyés par les observateurs, mais cela est généralement insuffisant.

Pour cette raison, a surgi l'intérêt général d'élaborer, à partir des seules données, des tests d'homogénéité.

Il existe dans la littérature de nombreux tests statistiques : test des signes, test de corrélation sérielle de Wald, Wolfowitz et Anderson, test des pointes et des phases, de corrélation de rang de Mann et Sperman, de Fisher-Yates, etc.

Ces tests ne peuvent que prouver le caractère aléatoire d'une série de variables indépendantes, ce qui implique l'hypothèse de l'inexistence d'une organisation interne de l'échantillon observé.

Par construction, ils comparent une partie des valeurs de l'échantillon au reste, et il leur faut donc connaître à l'avance le moment qui sépare les séries homogènes.

Pour résumer, ces tests ne peuvent distinguer les erreurs systématiques d'une organisation interne réelle de la série : ils n'ont pas été élaborés pour cela.

Malheureusement, la plupart des séries pluviométriques annuelles a une organisation interne, ce qui diminue encore l'intérêt d'appliquer à la pluviométrie les tests classiques d'homogénéité.

4. LES TESTS DES SIMPLES MASSES

Pour utiliser cette méthode, les coordonnées sont :

- le numéro l de l'année dans la série (croissante ou décroissante), en omettant les années manquantes ;
- la somme $\sum x_l/x_m$ où x_l qui représente les précipitations annuelles et x_m la pluviométrie moyenne de toutes les années utilisées.

Dans le cas d'une série chronologique à caractère aléatoire simple, le graphique des simples masses oscillera autour de la première bissectrice.

Dans la pratique, le travail étant très long, on emploie des programmes qui utilisent une variable transformée de type $x_l/x_m - l$, ce qui permet de convertir la première bissectrice en un axe vertical, facilitant ainsi le tracé des graphiques par ordinateur.

Si la série a été affectée par un ou plusieurs des effets énumérés au paragraphe 1, le graphique présentera un changement de pente significatif. Ce changement est facilement détectable et permet en outre d'identifier les années affectées, ce que les tests traditionnels ne pouvaient pas faire.

Malheureusement, les tests des simples masses ne peuvent distinguer avec certitude les erreurs systématiques d'éventuels effets faisant partie de l'organisation interne de la série observée.

La figure 1, élaborée automatiquement par le programme MASA utilisé dans le projet, illustre bien cette assertion.

Le graphique des simples masses de la station Mira n'oscille pas autour de l'axe vertical et semble avoir trois périodes de pentes différentes.

La période 1965-1968 est quasiment parallèle à l'axe, ce qui signifie qu'elle est composée d'années normales (proches de la moyenne multi-annuelle).

La période 1969-1976 correspond à une série de totaux annuels forts et la dernière période (1977-1983) à une série de précipitations « faibles » (inférieures à la moyenne multi-annuelle).

Le cas de la station Los Cerros (Montecristi), située dans la province du Manabí, est encore plus spectaculaire.

Le graphique est composé de 3 périodes principales (1964-1971, 1972-1976, 1977-1982) et l'on remarque nettement la valeur anormalement forte de l'année 1983 qui altère complètement le graphique et qui représente évidemment le phénomène du Niño.

Tous ces changements de pente correspondent à des effets qui viennent modifier le caractère aléatoire des deux séries, mais le problème consiste à savoir quels sont ces effets.

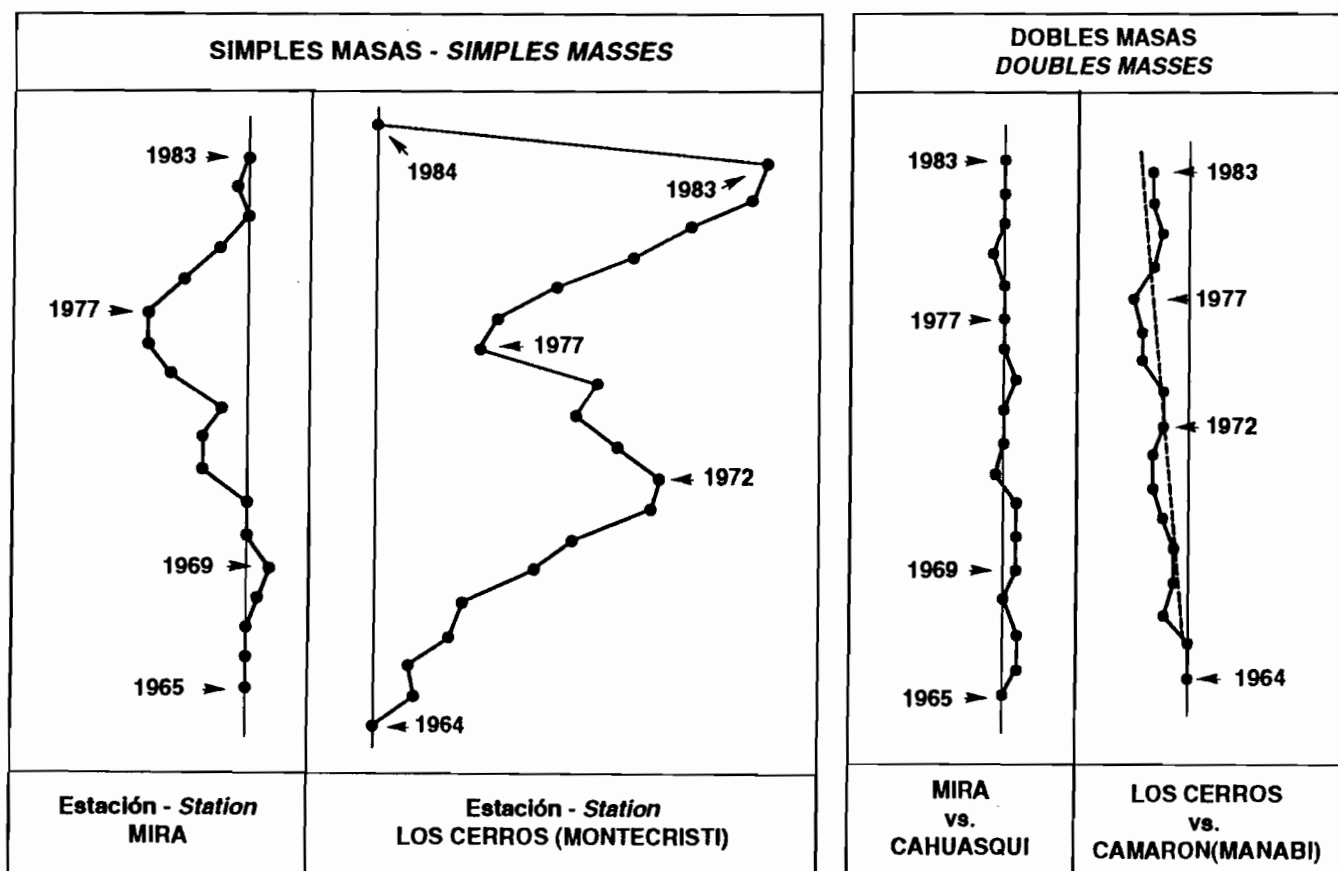


Figure 1

Figure 2

La seule utilisation des simples masses ne permet pas de répondre objectivement et n'autorise qu'une remarque : il existe dans les deux stations un changement similaire de pente entre les années 1976 et 1977. Ce changement doit alors correspondre à un phénomène normal d'organisation interne.

Pour conclure, il est indispensable de consulter les graphiques des doubles masses (figure 2), réalisés sur des stations voisines, où les différentes périodes ont disparu, ce qui prouve que tous les changements de pente antérieurs sont dus à des phénomènes normaux et caractéristiques des régions climatiques où se trouvent les deux stations.

Si les tests des simples masses sont intéressants en comparaison des tests statistiques, ils ne peuvent apporter des critères véritablement objectifs en ce qui concerne l'homogénéisation des données pluviométriques.

Il semble en outre impossible de détecter avec certitude les erreurs systématiques à partir d'une seule série chronologique. Heureusement, dans l'étude des précipitations, on dispose généralement de plusieurs stations situées dans une même région.

5. LES TESTS DES DOUBLES MASSES

5.1 Hypothèse de base

On admet l'existence de zones climatiques où l'on dispose de plusieurs stations dont les séries chronologiques sont liées par des coefficients de corrélation linéaire significativement positifs.

Appartenant à une même zone climatique, les stations sont soumises aux mêmes effets de persistance, de tendance, ou pseudo-cycliques. Leurs coefficients de variation sont donc similaires et leurs précipitations annuelles pseudo-proportionnelles.

C'est-à-dire que, pour une année i , les totaux annuels x_i et y_i de deux stations de la même zone ont entre eux la relation suivante :

$$(1) \quad x_i/x_m = y_i/y_m + e_i$$

où x_m et y_m correspondent aux moyennes multi-annuelles des deux stations, calculées sur une période suffisamment longue. La variance de e_i (terme indépendant tant de x_i que de y_i) est d'autant plus réduite que le coefficient de corrélation entre les deux stations est plus fort. Son espérance mathématique est égale à zéro.

5.2 Théorie des doubles masses (G. Hiez, 1977)

L'hypothèse précédente implique alors qu'une station X est liée à une seule tendance climatique régionale dont les valeurs annuelles t_i peuvent s'exprimer par une variable w_i réduite et normalisée comme suit :

$$(2) \quad w_i = (t_i - t_M)/\sigma_t$$

où t_M correspond à la moyenne de la tendance et σ_t a son écart-type.

En considérant qu'une station a des facteurs de site propres (moyenne x_M , écart-type σ_x), représentatifs du site de mesure et que ses observations ont des fluctuations aléatoires (u_i) dont la grandeur varie selon la relation existant entre la station et sa tendance, les précipitations x_i de la station X pourrait s'écrire selon la formule :

$$(3) \quad x_i = x_M + \sigma_x \cdot (r_x \cdot w_i + u_i \cdot \sqrt{1 - r_x^2})$$

où r_x représente le coefficient de corrélation entre la station et sa tendance.

En cumulant les données jusqu'à l'année n , on a :

$$(4) \quad \sum x_i = n \cdot x_M + \sigma_x \cdot \left(r_x \cdot \sum w_i + \sum u_i \cdot \sqrt{1-r_x^2} \right)$$

En divisant par la moyenne :

$$(5) \quad X_n = \sum x_i / x_M = n + C_x \left(r_x \cdot \sum w_i + \sum u_i \cdot \sqrt{1-r_x^2} \right)$$

où C_x correspond au coefficient de variation de la station X . Le graphique des simples masses illustre cette relation.

Considérons maintenant une autre station Y appartenant à la même zone climatique. En effectuant les calculs précédents, on obtiendra l'équation suivante :

$$(6) \quad Y_n = n + C_y \cdot \left(r_y \cdot \sum w_i + \sum v_i \cdot \sqrt{1-r_y^2} \right)$$

En éliminant n et en regroupant les termes :

$$(7) \quad Y_n = \begin{cases} X_n \\ + \sum w_i \cdot (r_y \cdot C_y - r_x \cdot C_x) \\ + (C_y \cdot \sum v_i \cdot \sqrt{1-r_y^2} - C_x \cdot \sum u_i \cdot \sqrt{1-r_x^2}) \end{cases}$$

L'équation précédente montre l'existence, entre les totaux cumulés X_n et Y_n , d'une relation linéaire altérée par 2 termes complémentaires, **A** et **B**.

Cela signifie que le graphique des doubles masses (ou des doubles cumulés) oscillera autour d'une droite de pente positive égale à 1.

Si les stations sont indépendantes, l'ampleur de l'oscillation du graphique dépendra uniquement de C_x et C_y . Dans le cas contraire, l'oscillation diminuera lorsque les coefficients de corrélation tendront vers 1.

Toute variation d'une des moyennes se traduira par une modification simultanée de la pente du graphique, comme dans le cas des simples masses.

Selon l'équation (7), si les 2 stations sont suffisamment proches pour obéir strictement à une même tendance climatique, C_x et C_y seront sensiblement égaux de même que r_x et r_y . Dans ce cas, le terme **A** de l'équation tendra vers zéro et la relation linéaire existant entre les deux sommes cumulées X_n et Y_n sera indépendante d'éventuelles fluctuations climatiques et uniquement altérée par des facteurs aléatoires lorsque les données sont homogènes.

DANS CE CAS, TOUTE MODIFICATION DE LA PENTE DU GRAPHIQUE CORRESPONDRÀ UNIQUEMENT À DES ERREURS SYSTÉMATIQUES.

C'est le grand avantage des doubles masses par rapport aux simples masses dont l'équation (5) comporte toujours les valeurs w_i relatives à la tendance climatique de la zone considérée et pouvant cumuler d'autres effets tel qu'on l'a vu dans la figure 1.

L'équation (7) permet l'utilisation de la méthodologie pour n'importe quelle station, mais dans la pratique, la détection d'erreurs peu importantes est possible à condition que l'oscillation soit moins ample, c'est-à-dire lorsqu'il existe une corrélation suffisante entre les stations.

raison, l'utilisation de la méthodologie est restreinte à des stations proches, bien que cette notion de proximité soit définie plus par la valeur du coefficient de corrélation que par la distance absolue.

5.3. Analyse des graphiques

Les figures 2 et 3 illustrent bien les propos précédents.

Dans la première, la station Mira est comparée à celle de Cahuasquí et la station Los Cerros (Montecristi) à celle de Camarón (Manabí.) Les changements de pente que présentaient les deux stations en simples masses ont disparu, ce qui démontre qu'ils n'étaient dus qu'à des effets particuliers (w_i) de la tendance climatique de chaque station.

Les courbes oscillent légèrement autour d'une droite. Dans le cas de la station Los Cerros, cette droite diffère de l'axe représenté par la première bissectrice, par l'effet du Niño de 1983, dont le caractère exceptionnel fait que l'hypothèse de pseudo-proportionnalité n'est pas totalement respectée.

La figure 3 montre des graphiques de doubles masses altérés par des erreurs systématiques.

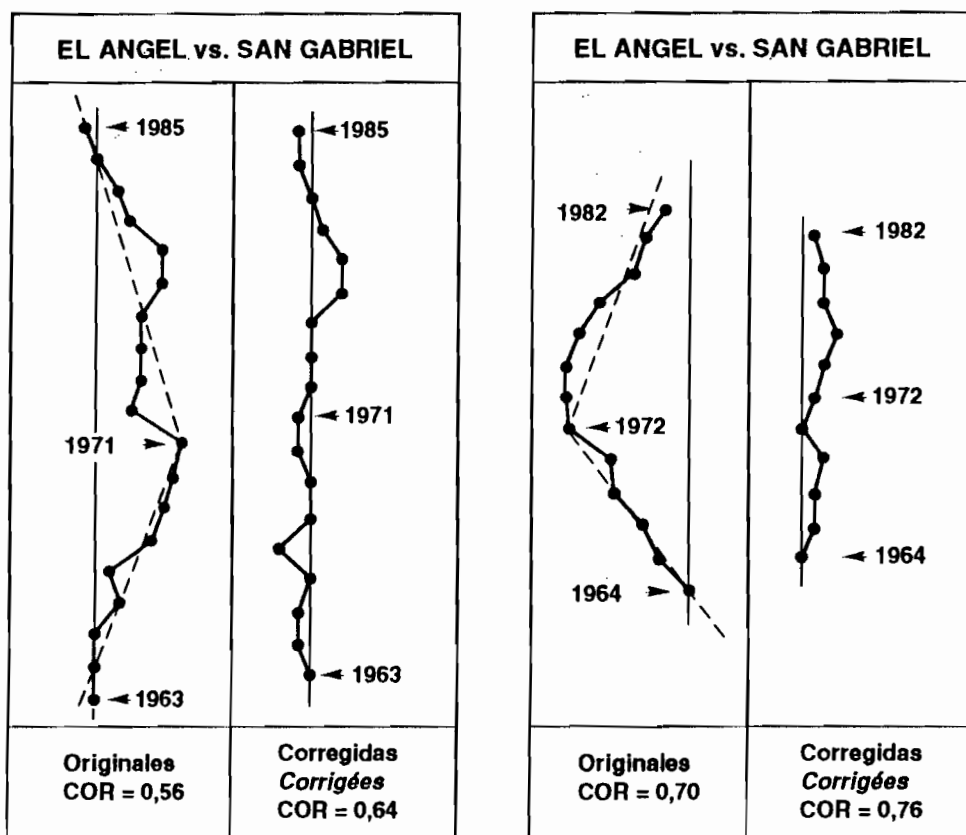


Figure 3

La courbe représentant la comparaison entre les stations El Angel et San Gabriel a 2 périodes différentes (1963-1970 et 1971-1985) dont les pentes ont une relation égale à 1,20. Cette valeur ne correspond à aucune valeur du tableau n° 1 ; on a donc diagnostiqué un changement de site.

En multipliant la première période par la valeur de 1,2 afin d'obtenir une série chronologique homogène et représentative du dernier site de l'appareil, on obtient une courbe beaucoup mieux adaptée.

Cependant, après avoir vu la station et étudié son historique à l'INAMHI (que nous remercions de nous avoir permis la consultation), on a pu confirmer le changement de site de la station climatique, mais à une date antérieure.

Cette erreur dans le diagnostic est due à la valeur particulièrement faible du coefficient de corrélation (0,56) et vient corroborer ce qui est dit dans le paragraphe précédent.

Les doubles masses entre les stations Zuleta et Pimampiro montrent un changement de pente à partir de l'année 1972. La relations entre les pentes des deux périodes est de 1,36, ce qui ne correspond non plus à aucune valeur du tableau n° 1. Un changement de site a également été diagnostiqué.

Une visite de terrain a permis de vérifier la date du changement, mais l'observateur nous a signifié un changement d'éprouvette.

Ces deux stations ont également des coefficients de corrélation peu significatifs (0,70 et 0,76 pour les valeurs corrigées), ce qui peut expliquer l'erreur de diagnostic.

Dans ces deux cas, il a été nécessaire de vérifier les stations par rapport à toutes les autres des environs afin d'émettre un diagnostic correct, ce qui constitue une limite importante de cette méthodologie.

5.4. Conclusions sur les graphiques des doubles masses

Pour utiliser cette méthode, on doit tout d'abord déterminer de manière empirique des groupes de stations appartenant à la même zone climatique, en tenant compte de la distance entre elles, des effets du relief, du type de régime intra-annuel, etc.

On compare ensuite chacune des stations à toutes les autres du groupe, ce qui signifie un nombre de doubles masses égal à $m(m - 1) / 2$ pour un groupe de m stations. Dans ces conditions, le traitement de la pluviométrie d'une région ou d'un pays exige la manipulation d'une quantité considérable d'information, possible seulement par l'automatisation des opérations de traitement.

Le programme MASA utilisé dans le projet, constitue un bon exemple de tracé automatique des graphiques. Son utilisation facilite considérablement les comparaisons entre stations en supprimant la plupart des opérations manuelles et en apportant une précieuse aide au travail d'interprétation.

La seule figure 2 ne permet pas non plus de détecter les sources des erreurs apparaissant dans les graphiques élaborés à partir de données brutes. Les changements de pente peuvent provenir de Mira ou de Cahuasquí, de Los Cerros (Montecristi) ou de Camarón (Manabí). Plusieurs tests seront nécessaires pour découvrir l'origine des problèmes.

La meilleure manière d'utiliser la méthode des doubles masses et de sélectionner, dans la zone climatique considérée, quelques stations ayant des données de bonne qualité et de longue période (stations de base). En les comparant aux stations voisines et en effectuant les corrections au fur et à mesure de l'apparition de changements de pente, il est possible de constituer d'autres stations de base et d'avancer ainsi, lentement, en vérifiant chaque fois les diagnostics émis à l'aide de toutes les stations de base.

En opérant minutieusement et de cette manière, on peut arriver à des diagnostics justes et confirmés. La méthodologie est utilisée dans plusieurs continents où son efficacité a été prouvée.

Cependant, la réalité est plus complexe. Il n'est pas toujours possible de trouver à l'avance des stations de base et dans ce cas, l'interférence entre des erreurs provenant de stations différentes rend très difficile le choix de critères de décision réellement objectifs.

En outre, il est fréquent de disposer de stations de base ayant des périodes incomplètes d'observation justement pour les années intéressantes à comparer.

On constate alors l'intérêt d'élaborer un vecteur de référence :

- homogène,
- couvrant toute la période considérée,
- automatiquement,
- à partir des données existantes.

Disposant d'un vecteur élaboré dans des conditions optimales, sa simple comparaison à une station permettrait de conclure, sans avoir à passer par toute une série de déductions qui mènent rapidement à confusion lorsque le nombre de stations à traiter augmente.

6. ÉLABORATION D'UN VECTEUR DE RÉFÉRENCE

6.1 Intérêt principal

Si Z désigne le vecteur de référence représentatif de la tendance climatique régionale décrite dans le paragraphe 5.2. et dans laquelle se trouvent les stations à vérifier, ce vecteur n'aura pas de fluctuations aléatoires et sa relation avec la tendance pourra être traduite par l'équation suivante (G. Hiez, 1977) :

$$(8) \quad z_i = z_M + \sigma_z \cdot w_i$$

où z_i représente les valeurs annuelles, z_M la moyenne et σ_z l'écart-type.

En cumulant les variables z_i jusqu'à l'année n et en divisant par la moyenne, l'équation précédente s'écrit :

$$(9) \quad Z_n = n + C_z \cdot \sum w_i$$

où C_z correspond au coefficient de variation de z_i .

En comparant ce vecteur à une station X appartenant à la même zone climatique et en éliminant n entre les équations (5) et (9), on obtient :

$$(10) \quad X_n = Z_n + \sum w_i \cdot (r_x \cdot C_x - C_z) + C_x \cdot \sum u_i \cdot \sqrt{1 - r_x^2}$$

ou pour simplifier :

$$(11) \quad X_n = Z_n + A + B$$

Comme on l'a signalé dans le paragraphe 5.2, dans une même zone climatique, le terme A tend vers zéro, alors que le terme B , dont dépend l'ampleur de l'oscillation des doubles masses, a comme variance :

$$(12) \quad \sigma_v^2 = C_x^2 + C_z^2 - 2 \cdot C_x \cdot C_z \cdot r_x$$

relation où C_x^2 et C_z^2 correspondent au carré des coefficients de variation C_x et C_z .

Dans la comparaison station par station en doubles masses, la variance du terme **B** de l'équation (7) est :

$$(13) \quad \sigma_{xy}^2 = C_x^2 + C_y^2 - 2 \cdot C_x \cdot C_y \cdot r_x \cdot r_y$$

Avec des coefficients de variation similaires, σ_{xy}^2 est supérieur à σ_v^2 , ce qui signifie que les doubles masses entre stations et vecteur auront une oscillation plus réduite que les doubles masses entre stations. En outre, comme le vecteur n'a pas de fluctuations aléatoires, son coefficient de variation sera plus faible que celui de la plupart des stations et cette particularité sera encore plus marquée.

Ceci signifie que les doubles masses entre stations et vecteur de référence seront plus précises que celles entre stations, ainsi qu'on pourra le constater dans des exemples postérieurs.

6.2 Conditions de l'élaboration

L'élaboration du vecteur devra s'ajuster à la condition de pseudo-proportionnalité des totaux annuels et devra également tenir compte de certaines considérations complémentaires :

- Cette élaboration ne doit pas exiger d'hypothèses préalables sur la distribution des totaux pluviométriques annuels.
- Pour une année donnée, le total annuel d'une seule station ne peut, *a priori*, servir de valeur de référence.
- Il est supposé que dans l'ensemble de l'information globale, il existe une valeur estimative de la tendance climatique régionale, plus représentative que l'information partielle de chacune des stations.
- Le processus d'élaboration doit être conçu de façon à ce que toute l'information disponible soit utilisée sans que les données erronées aient une influence sensible sur le résultat.

Cependant, dans la pratique, si les données sont réellement de mauvaise qualité, il sera nécessaire de corriger les plus grosses erreurs.

Mais, *a priori*, on n'éliminera pas des stations de périodes courtes (5 ans par exemple), dont les valeurs peuvent aider à la construction d'une composante annuelle du vecteur.

6.3 Analyse de la matrice des données

L'ensemble des totaux annuels constitue une matrice **A** dont les lignes sont les stations (θ) et les colonnes les années (i).

$$(14) \quad \mathbf{A} = [x_{\theta i}] = \begin{matrix} \theta = \alpha, \dots, \mu \\ i = 1, \dots, n \end{matrix}$$

Les μ données obtenues au cours d'une année i définissent les coordonnées d'un vecteur-pluie dans l'espace des stations.

Si la pseudo-proportionnalité était stricte, l'équation pourrait s'écrire :

$$(15) \quad x_{\alpha i} / x_{\alpha i-1} = \dots = x_{\mu i} / x_{\mu i-1}$$

L'ensemble des n vecteurs constituerait alors un ensemble de vecteurs colinéaires et la matrice **A** pourrait être représentée par une matrice **B** dégénérée.

En réalité, la matrice **A** s'éloigne de ce modèle théorique car il existe, dans chaque station, des fluctuations aléatoires qui doivent être prises en compte. On a ainsi :

$$(16) \quad \mathbf{A} = \mathbf{B} + \mathbf{E}$$

B représente l'information théorique souhaitée.

6.4 Analyse de quelques possibilités

Une possibilité consiste à calculer, pour chaque année, la valeur du vecteur en utilisant la moyenne arithmétique de l'ensemble des totaux observés au cours de l'année concernée.

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

Malheureusement, les variables s'éloignent souvent d'une distribution normale par :

- le nombre réduit de données, ce qui donne lieu à des distributions asymétriques ;
- l'existence de séries hétérogènes régies par des distributions pluri-modales ;
- la présence de valeurs anormales qui altèrent le calcul de la moyenne (« contamination de la moyenne par les valeurs extrêmes »).

Une estimation précieuse peut être obtenue à partir des médianes, mais les calculs deviennent problématiques lorsque les séries ont plusieurs années sans observation.

À partir des considérations précédentes, G. HIEZ [6] a entrepris l'élaboration d'un vecteur régional en tenant compte des valeurs modales. Ce vecteur est utilisé au Brésil où il donne de bons résultats. Malheureusement, l'algorithme est assez long en temps de calcul et son application réservée, pour le moment, à des ordinateurs puissants.

En explorant dans une direction différente, un autre chercheur (Y. BRUNET-MORET, [2], [3], [4]) a conçu un « vecteur des indices annuels » (ou vecteur YBM.), qui estime, pour chaque année, la moyenne la plus probable.

Le projet Plan National d'Irrigation utilise ce dernier vecteur dont la théorie et les résultats sont présentés ci-après.

Il faut signaler que les deux vecteurs ont été essayés dans plusieurs pays avec des résultats similaires, mais sans jamais être appliqués dans un pays à un relief aussi contrasté que l'Équateur.

7. PRÉSENTATION DU VECTEUR YBM

7.1 Hypothèse de base

Si **Z** est le vecteur de la zone climatique étudiée, selon les équations (1) et (14), pour l'année **i**, la précipitation annuelle $x_{\theta i}$ de la station θ (des μ existantes) est liée à l'indice z_i du vecteur par la relation :

$$(17) \quad x_{\theta i} / x_{M\theta} = z_i + e_{\theta i}$$

$x_{M\theta}$ = espérance mathématique de X_{θ} dans sa population-mère
 $e_{\theta i}$ = variable aléatoire indépendante de z_i

Les variables E_{θ} (de α à μ) sont indépendantes entre elles.

Comme il a été dit au début, pour une station θ , la série chronologique de $x_{\theta i}$ a un caractère aléatoire simple, altéré par les effets de tendance, persistance et pseudo-cycliques. Ces effets sont déjà représentés dans le vecteur ; la série chronologique des $e_{\theta i}$ a donc un caractère aléatoire simple.

Il faut rappeler que, dans la zone climatique considérée, on admet l'hypothèse que les totaux pluviométriques annuels sont pseudo-proportionnels.

Tout ce qui précède (hypothèse et relation) implique que :

- l'espérance mathématique de $E_{\theta} = 0$ dans son univers ;
- l'espérance mathématique de $z = 1$ dans son univers ;
- les variances de $E_{\alpha}, \dots, E_{\mu}$ sont égales entre elles et à $\text{Var } E$.

Pour chaque année i , les $e_{\alpha i}, \dots, e_{\mu i}$ ne sont pas indépendants entre eux (proximité géographique), mais il est admis que, dans leurs univers, la somme des covariances est égale à zéro :

$$\sum \text{covar}(E_{\theta}, E_{\delta}) = 0, \theta \neq \delta$$

Ceci implique que les coefficients de corrélation entre stations ne sont pas égaux mais ont une valeur moyenne de :

$$(18) \quad \text{CORXYM} = \text{Var } Z / (\text{Var } Z + \text{Var } E)$$

Avec ces hypothèses, le coefficient de corrélation entre le vecteur et une station quelconque aura dans les univers une valeur égale à :

$$(19) \quad \text{CORXVM} = \sqrt{\text{CORXYM}}$$

Cette valeur étant supérieure à CORXYM , il est compréhensible que les doubles masses entre vecteur et station soient généralement plus précises que les doubles masses entre stations.

Ceci implique aussi que la valeur du coefficient de variation de chaque station dans son univers soit égale à :

$$(20) \quad C_x = \sqrt{\text{Var } Z + \text{Var } E}$$

7.2 Algorithme de calcul

Le programme examine les totaux annuels des stations choisies à partir d'une année de début (ID) et jusqu'à une année finale (IF), et retient uniquement les stations possédant au moins **MINIAÑO** années sur la période considérée, ainsi que les années de la période où il existe au moins **MINIEST** totaux pluviométriques.

Les variables ID, IF, **MINIAÑO** et **MINIEST** sont des variables originelles définies par l'utilisateur.

Les valeurs sélectionnées constituent la matrice d'observation **A** (n, μ) des précipitations annuelles de μ stations sur n années.

Cette matrice est généralement incomplète (observations manquantes) et constitue la base de calcul du vecteur.

Dans un espace à μ dimensions, l'équation :

$$(21) \quad X_{\alpha} / x_{M\alpha} + X_{\beta} / x_{M\beta} + \dots + X_{\mu} / x_{M\mu} - \mu \cdot z_1 = 0$$

représente un hyperplan perpendiculaire à la droite :

$$(22) \quad X_{\alpha}/x_{M\alpha} = X_{\beta}/x_{M\beta} = \dots = X_{\mu}/x_{M\mu}$$

Le calcul consiste à déterminer les valeurs $x_{M\theta}$ pour chaque station et les valeurs de z_i pour chaque année, afin que les hyperplans représentatifs des observations de chaque année soient parallèles entre eux et perpendiculaires à la droite (22).

Selon les hypothèses précédentes appliquées à l'échantillon (moyenne de $z = 1$, moyenne de $E = 0$, $\text{Var } \theta = \text{Var } E$), on peut utiliser les moindres carrés pour minimiser l'expression :

$$(23) \quad \sum \sum (x_{\theta i}/x_{M\theta} - z_i)^2$$

et déterminer $x_{M\theta}, \dots, x_{M\mu}, z_1, \dots, z_n$.

En dérivant cette expression selon z_i , s'il existe μ_1 stations observées dans l'année i , on obtient n équations :

$$(24) \quad z_i = 1/\mu_1 \sum (x_{\theta i}/x_{M\theta})$$

En dérivant l'expression (23) selon $1/x_{M\theta}$, s'il existe n_1 années observées dans cette station θ , on obtient μ équations :

$$(25) \quad \sum x_{\theta i} (x_{\theta i}/x_{M\theta} - z_i) = 0$$

En réalité, compte tenu du manque de données, z_i doit être remplacé dans l'expression (23) par sa valeur trouvée dans l'équation (24) avant de dériver selon $1/x_{M\theta}$, afin d'obtenir un système de μ équations à μ inconnues ($1/x_{M\theta}$) :

$$(26) \quad f(\alpha, \theta)(1/x_{M\alpha}) + \dots + f(\mu, \theta)(1/x_{M\mu}) = 0$$

Ce système linéaire est homogène et a une solution simple :

$$1/x_{M\alpha} = \dots = 1/x_{M\mu} = 0$$

Pour sa solution, il doit être réduit à $(\mu - 1)$ équations linéaires à $(\mu - 1)$ inconnues et dont le second membre est différent de zéro.

En fixant $1/x_{M\mu} = 1$, on calcule alors, par les moindres carrés (régression multiple), les $(m-1)$ valeurs de $1/x_{M\theta}$ qui s'ajustent le mieux aux μ relations entre les quantités f de l'équation (26).

On connaît ainsi les valeurs $C/x_{M\alpha}, \dots, C/x_{M\mu}$, C étant un coefficient multiplicateur inconnu, mais égal pour toutes les variables.

Les valeurs de z_i sont calculés avec les équations (24). En exprimant l'hypothèse que la moyenne des z_i est égale à 1, on calcule le coefficient C et on peut ainsi déterminer exactement les valeurs $x_{M\alpha}, \dots, x_{M\mu}$.

Le calcul précédent montre que :

$$\bullet \text{ pour chaque année } i, \quad z_i - 1/\mu_1 \sum x_{\theta i}/x_{M\theta} = 0 \quad (27)$$

$$\bullet \text{ pour chaque station } \theta, \quad \sum (x_{\theta i}/x_{M\theta} - z_i) \neq 0 \quad (28)$$

$$\bullet \sum \sum (x_{0i}/x_{M0} - z_i) = 0 \quad (29)$$

APPLICATION À DES CAS CONCRETS

L'algorithme précédent donne le même poids à chacune des valeurs observées et aussi aux valeurs altérées qu'on souhaite détecter et qui peuvent avoir une influence sur les valeurs de x_{M0} et z_i .

Il convient alors d'énoncer une dernière hypothèse en admettant que pour chaque année et dans chaque station, les valeurs erronées constituent une minorité.

Le calcul s'effectue selon le processus suivant :

- * Une première itération est effectuée avec les valeurs réelles de la matrice **A** et on calcule une première série de valeurs pour les x_{M0} et z_i .
Pour chaque année et pour chaque station, x_{0i}/x_{M0} est comparé à z_i . Si x_{0i}/x_{M0} n'est pas compris dans un intervalle donné, x_{0i} sera remplacé par $x_{M0} \cdot z_i$.
- * Les nouvelles valeurs de la matrice **A** sont utilisées pour une deuxième itération qui calcule de nouvelles valeurs de x_{M0} et z_i qui sont comparées à nouveau à l'intervalle et remplacées, si nécessaire, selon la procédure précédente.
- * Une série d'itérations est effectuée jusqu'à ce que toutes les valeurs de $x_{0i}/x_{M0}/z_i$ soient à l'intérieur de l'intervalle, ce qui se produit rapidement.

Il est logique que les valeurs délimitant l'intervalle dépendent de la variance de **E** dont on a une estimation (par excès si l'homogénéité spatiale n'est pas réalisée). Dans la réalité, il est difficile de tenir compte de cette observation lorsqu'il existe des séries hétérogènes.

Après plusieurs essais, tant dans le couloir interandin que sur la Côte, il a été décidé de prendre 0,7 et 1,5 comme des valeurs délimitant l'intervalle de comparaison.

Ces quantités sont similaires à $2/\pi$ et à $\pi/2$, valeurs dues aux erreurs les plus fréquentes entre des éprouvettes et des pluviomètres (tableau n° 1).

7.3 Tests de sensibilité

On considère un groupe de 12 stations fictives observées durant 20 ans (1966-1985 par exemple) et dont les totaux annuels varient entre 999 et 1 001 mm (très souvent égaux à 1 000 mm).

Les données de 3 stations sur une période de 10 ans (1966-1975) sont multipliées par 5.

Un vecteur élaboré à partir de la moyenne arithmétique serait altéré présentant 2 périodes dont les pentes différentes auraient une relation égale à 2.

Avec le vecteur, la relation est de 1,25, ce qui démontre le peu de sensibilité du vecteur à des confusions d'appareils ou à des changements de site.

Sur ces mêmes stations, la précipitation de 1985 sur une station est multipliée par 13.

Pour cette année, la valeur de la moyenne atteint 2 000 mm, alors que celle du vecteur est de 1 025 mm. Ceci montre que le vecteur est pratiquement insensible à des valeurs extrêmes comme des erreurs de saisie, etc.

Dans les deux cas précédents (un peu extrêmes), la comparaison des stations à la moyenne n'aurait pas permis de détecter les erreurs, elle-même étant altérée.

8. RÉSULTATS OBTENUS

Depuis plusieurs mois, l'équipe de travail effectue, avec la méthodologie décrite ci-dessus, le traitement de l'information de plus de 700 stations de la *Sierra* et de la Côte.

Le bassin du Mira est terminé et les diagnostics émis ont été confirmés par des vérifications de terrain et une révision des historiques à l'INAMHI.

Dans la partie sud du pays et sur les bassins du Guayas et du Pastaza, le travail de traitement est fini et il reste seulement à vérifier les diagnostics.

L'homogénéisation est bien avancée dans la province du Manabí et dans le bassin de l'Esmeraldas.

Il manque uniquement l'analyse des stations situées dans les bassins du Cañar et du Zapotal.

Le concept de pseudo-proportionnalité a été confirmé partout, bien que les zones climatiques rencontrées dans la *Sierra* soient plus réduites et leurs limites plus difficiles à tracer.

Il n'existe donc pas autant de micro-climats que l'on pouvait espérer et les stations obéissent à des effets climatiques similaires dont l'impact sur les précipitations varie selon la situation de la station par rapport au relief (effets d'abri, etc.).

Il vaut donc mieux parler de topo-climats.

Le vecteur détecte également les erreurs avec une bonne fiabilité comme on peut l'observer dans la figure 4 qui représente les stations El Angel et Zuleta déjà présentées pour le test des doubles masses (figure 3).

Les coefficients de corrélation ont significativement augmenté, ce qui permet d'émettre un diagnostic correct par visualisation d'un seul graphique.

Le changement de site de la station El Angel apparaît maintenant entre 1969 et 1970 (conformément à l'historique), et la relation des pentes pour Zuleta est de 1,57, confirmant ainsi le changement d'éprouvette réalisé par l'observateur en 1972.

Les vérifications ont permis de confirmer tous les diagnostics du bassin du Mira pour lequel on a obtenu les résultats suivants :

- 58 stations analysées
- 14 présentant des changements de site ou inadéquation des appareils
- 17 ayant des années anormales

Cela signifie près de 50 % de stations homogènes, ce qui correspond aux résultats obtenus dans des pays voisins.

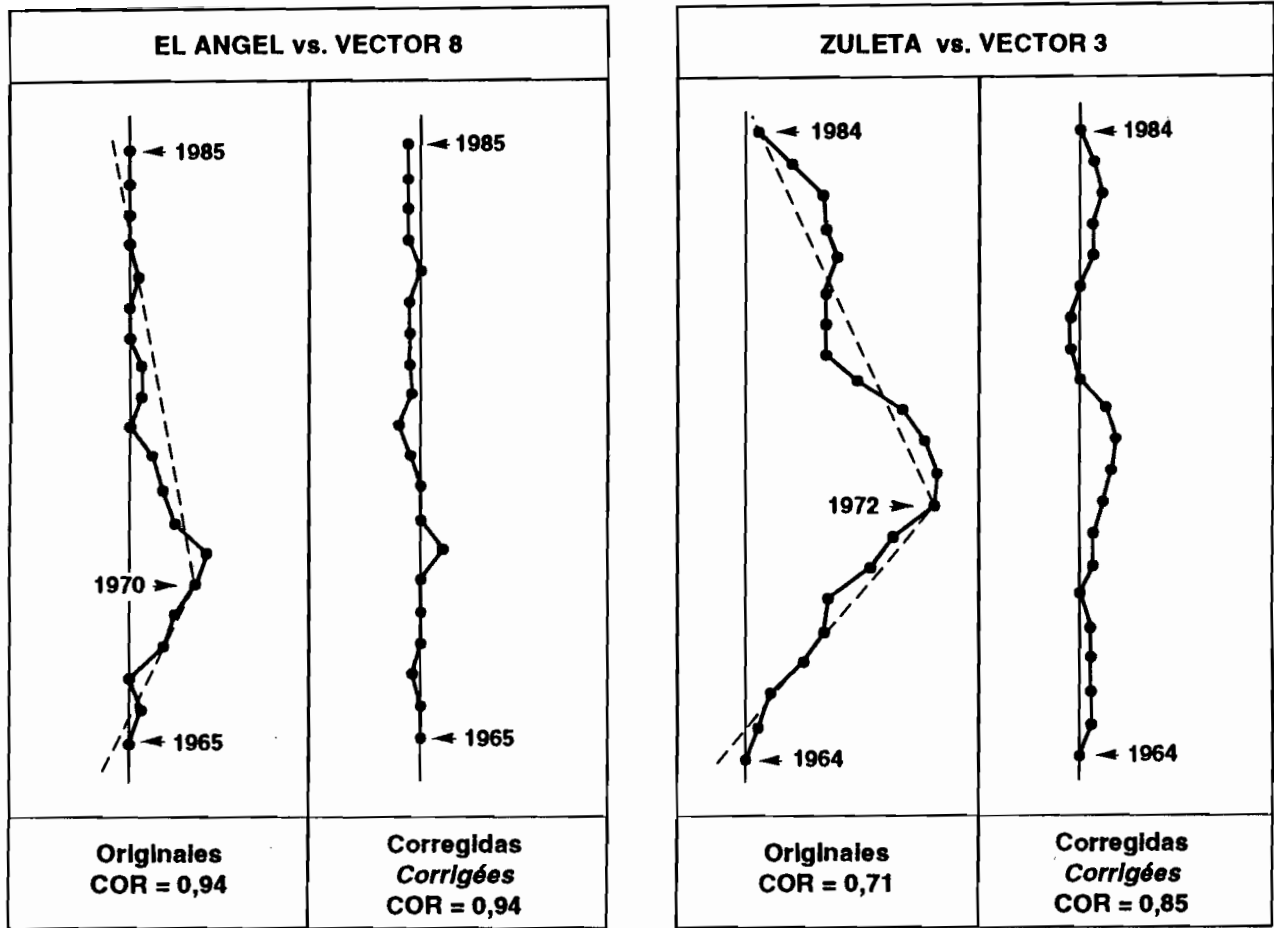


Figure 4

B - RÉGIONALISATION

1. DÉFINITION DU PROBLÈME

La régionalisation pluviométrique, sa conception et la précision de son calcul, dépendent énormément des objectifs recherchés.

Le Plan National d'Irrigation a besoin d'établir, dans n'importe quelle zone du pays, des bilans entre les besoins en eau et les ressources hydriques, qui devront être calculés mois par mois et année par année, afin de prendre en compte les risques (du point de vue statistique) que devraient assumer les périmètres irrigables quant à la disponibilité en eau.

- Dans le cas du calcul des besoins en eau, il faut donc disposer d'une série chronologique mensuelle et annuelle homogène, observée sur une période suffisamment longue et dont la précipitation annuelle moyenne soit égale à celle du périmètre considéré.

L'utilisation d'un vecteur de référence ne permettra pas d'obtenir des données exactes (absence de fluctuations aléatoires au site considéré), mais rendra possible l'estimation des paramètres statistiques corrects.

- Par ailleurs, le calcul de la pluie moyenne d'un bassin versant intègre les précipitations ponctuelles de chacune de ses parties éliminant ou réduisant l'influence des fluctuations aléatoires dues à chaque mesure. Dans ce cas, l'estimation des ressources hydriques par la relation pluie-débit s'approchera plus de la réalité si elle utilise également un vecteur de référence ne contenant pas ces effets.

Ceci montre que les bilans hydriques en un point donné peuvent être établis de manière plus aisée à partir d'un vecteur de référence, représentatif de la région du site considéré (ou de son bassin versant), tant du point de vue tant chronologique (régionalisation chronologique) qu'en ce qui concerne sa répartition intra-annuelle (régionalisation de régime), à condition que la précipitation annuelle moyenne du point considéré ou de son bassin soit connue.

Voyons comment le vecteur YBM peut résoudre ces problèmes.

2. TRACÉ DES ISOHYÈTES

Pour être correctes, les isohyètes doivent être tracées à partir de stations dont les totaux annuels ont été calculés sur une période commune suffisamment longue pour être représentative.

En théorie, le vecteur YBM est élaboré selon l'espérance mathématique des précipitations annuelles de chaque station.

Dans la pratique, le programme de calcul estime chaque espérance mathématique sur la base de l'organisation interne de la tendance climatique régionale.

Il n'est donc nécessaire ni de remplir les années manquantes de chaque station sur une période commune ni de calculer la représentativité de cette période. De plus, les relations entre pluie et altitude sont habituellement plus précises à l'intérieur de chaque zone climatique rencontrée.

groupe n°	1	2	3	4	5	6	7	8
coefficient de correction moyen	0,91	0,89	0,91	0,95	0,91	0,98	0,96	0,96

Tableau n° 2

Les limites des zones 4 à 8 ne sont pas totalement définies uniquement par manque de stations proches.

Des résultats similaires ont été obtenus dans les autres régions, et il est possible également de détecter les parties où le réseau pluviométrique est insuffisant.

4. RÉGIONALISATION PAR RÉGIME

Pour regrouper des stations selon leur régime, il est possible de raisonner comme d'habitude, en moyennes mensuelles.

Malheureusement, le régime d'une station varie selon l'année et il existe dans la *Sierra* des stations de régime moyen interandin alors qu'elles ont quelques années de régime pseudo-amazonien, comme le cas de l'année 1975 sur la station Ibarra-Aeropuerto (figure 6). Alors, si la comparaison en régime moyen est valable, elle n'est pas suffisante et il est nécessaire, afin de ne pas perdre d'informations, de considérer les données mensuelles comme une seule série chronologique.

De ce point de vue, les hypothèses et les équations régissant l'élaboration du vecteur n'empêchent pas de considérer une série mensuelle.

Le concept de pseudo-proportionnalité sera plus difficile à trouver, les fluctuations aléatoires seront plus fortes et les ajustements à des valeurs annuelles peut-être complexes ; cependant, l'équipe de travail effectue actuellement des essais dans ce sens, et on espère que des résultats futurs confirmeront ce raisonnement.

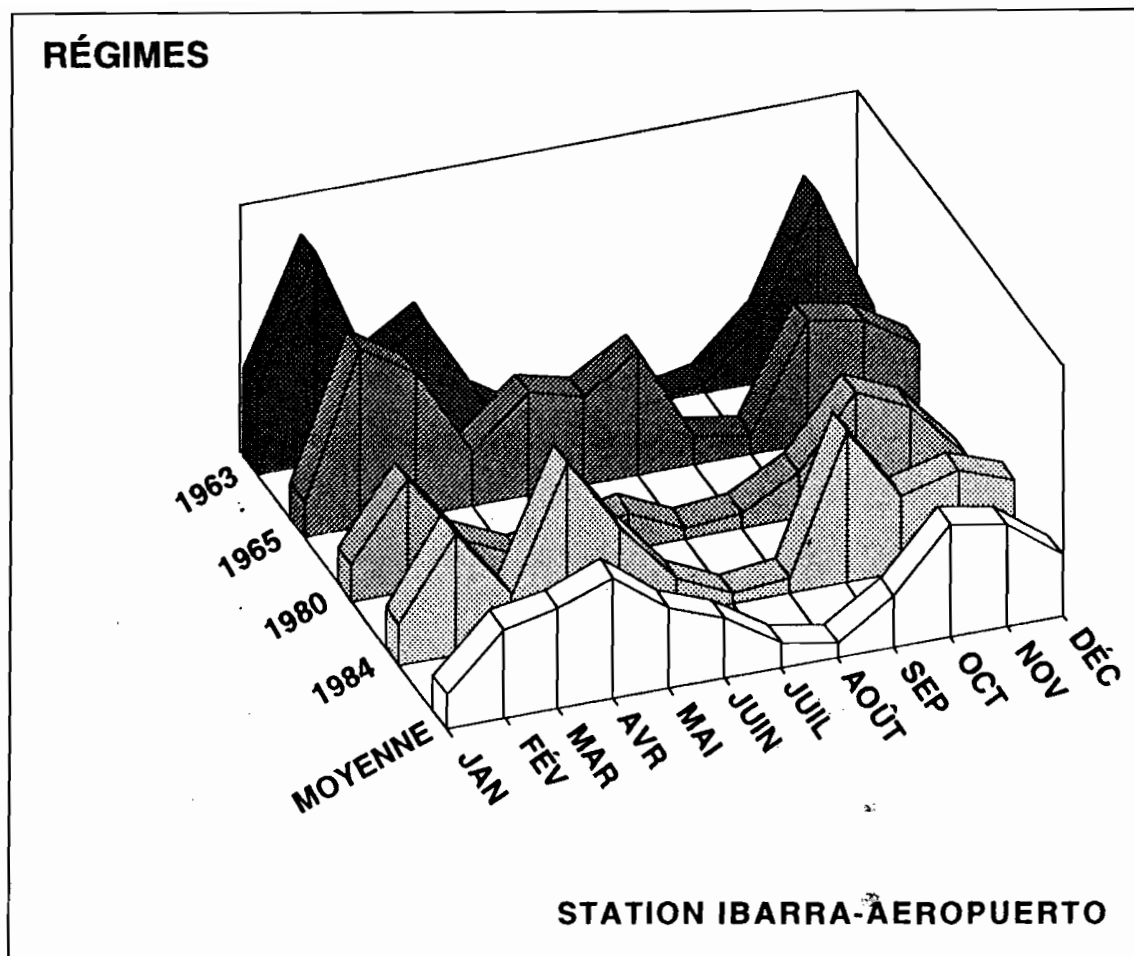


Figure 6

CONCLUSION

La méthode du vecteur régional a démontré sa bonne sensibilité pour détecter les erreurs systématiques qui altèrent les séries chronologiques de précipitations annuelles.

Son utilisation sur ordinateur est relativement simple et rapide par rapport à celle des doubles masses.

Le vecteur YBM trace également des isohyètes et permet de délimiter des zones climatiques homogènes dont il représente les caractéristiques avec précision.

Une synthèse pluviométrique à l'échelle nationale pourra se faire seulement à partir des vecteurs de chaque zone, réduisant ainsi le volume d'analyses et supprimant les effets aléatoires dus à toute mesure ponctuelle.

Les premiers résultats ratifient les hypothèses formulées et confirment, pour la première fois, la validité de la méthodologie en milieu montagneux.

BIBLIOGRAPHIE

- [1] BRUNEL, J.-P., 1986. Comparaison instrumentale des précipitations en climat méditerranéen semi-aride, in *Hydrologie Continentale*, vol. 1, n° 2, p. 87-93.
- [2] BRUNET-MORET, Y., 1971. Étude de l'homogénéité de séries chronologiques annuelles par la méthode des doubles masses, in *Cahiers ORSTOM*, série hydrologie, vol. III, n° 4, p. 3-31.
- [3] BRUNET-MORET, Y., 1977. Test d'homogénéité, in *Cahiers ORSTOM*, série Hydrologie, vol. XIV, n° 2, p. 119-128.
- [4] BRUNET-MORET, Y., 1979: Homogénéisation des précipitations, in *Cahiers ORSTOM*, série Hydrologie, vol. XVI, n° 3-4, p. 147-170.
- [5] CHEVALLIER, P., 1986. Note sur les écarts de mesure observés entre pluviomètres standards et au sol, in *Hydrologie Continentale*, vol. 1, n° 22, p. 111-119.
- [6] HIEZ, G., 1977. L'homogénéité des données pluviométriques, in *Cahiers ORSTOM*, série Hydrologie, vol. XIV, n° 2, p. 129-172.
- [7] L'HÔTE, Y., 1985. *Contrôle de la qualité et de l'homogénéité des données pluviométriques par la méthode du vecteur des indices annuels de précipitation*, Zürich, ETH/IAHS/WMO, Workshop on the correction of precipitation measurements.
- [8] RODDA, J.-C., 1967. The rainfall measurement problem, in *AISH*, publ. n° 78, p. 215-281.
- [9] SEARCY, J.K., HARDISON, C.H., 1983. *Curva de dobles masas*, HIMAT et COLCIENCIAS, Bogota.

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