

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

## **HOMOGENEIZACION DE LOS DATOS PLUVIOMETRICOS**

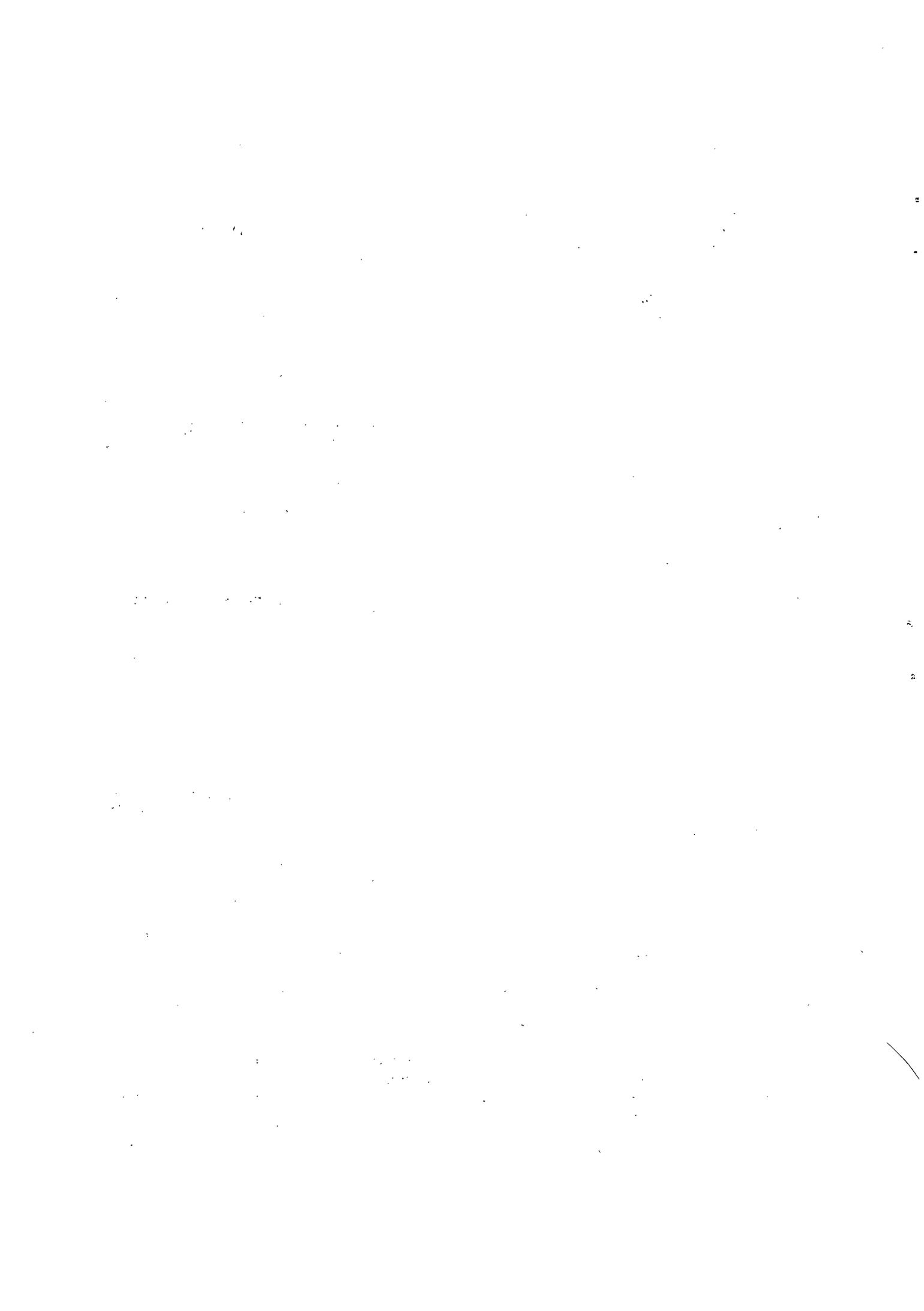
por Patrick Le Goulven \*

marzo de 1988

### **Nota Bene**

Este texto se apoya principalmente en las teorías desarrolladas por Yves Brunet-Moret (4) y Gérard Hiez (6). Estos 2 hidrólogos de ORSTOM se interesaron particularmente en el compartamiento regional de las precipitaciones anuales con el objetivo de elaborar vectores representativos de cada región.

\* hidrólogo ORSTOM, Misión ORSTOM, CP 17-11-06596, Quito, Ecuador.



## INTRODUCCION

El inventario de los recursos hídricos así como su manejo racional son preocupaciones actuales de un sinnúmero de países para intensificar la explotación de los recursos agrícolas mediante riego, abastecer a aglomeraciones urbanas e industriales cada vez más concentradas y satisfacer el consumo energético.

Sin embargo, tales objetivos son estrechamente dependientes de un buen conocimiento de los elementos del clima, entre los cuales la pluviometría juega un papel preponderante.

Las precipitaciones atmosféricas constituyen un fenómeno esencialmente variable en el tiempo y en el espacio, sobre todo en un país de relieve muy irregular como lo es el Ecuador.

Su conocimiento, ante todo cuantitativo, se obtiene con mediciones permanentes a través de una red de observaciones puntuales. Lastimosamente, toda recolección de datos adolece de errores de toda naturaleza provenientes tanto del sistema de obtención como del carácter individual de la medición.

Esa constatación demuestra la necesidad de una severa crítica de los datos reunidos, los mismos que pueden ser eliminados cuando se alejan demasiado de la realidad, corregidos prudentemente cuando la naturaleza sistemática de los errores detectados lo permite o simplemente señalados cuando su valor anormal puede ser considerado como localmente posible.

El conjunto de esas operaciones constituye la homogeneización de los datos.

Consideraremos solamente los valores pluviométricos. En la conclusión se presentan algunas recomendaciones para el tratamiento de los otros parámetros climáticos.

## 1. DEFINICION DEL PROBLEMA

### 1.1. Homogeneidad de una serie cronológica de precipitaciones anuales

Una serie cronológica de valores observados es de carácter aleatorio sencillo si todos los valores son independientes y si provienen de una misma población estadística por sorteo al azar. El carácter aleatorio puede ser alterado por:

- *un efecto de persistencia*: un valor no es independiente del o de los anteriores (proceso de MARKOV u otro), pero la serie es estacionaria y los parámetros de distribución de la población no varían con el tiempo;
- *un efecto de tendencia*: los valores observados provienen de poblaciones diferentes cuyas probabilidades matemáticas aumentan o disminuyen con el tiempo;
- *efectos cíclicos o pseudo-cíclicos*: la probabilidad matemática para un valor dado depende de la cronología, pero la probabilidad de promedios de series suficientemente largas puede ser considerada como estacionaria;
- *errores sistemáticos (o no) de observación y de medición*: sólo esos errores alteran la homogeneidad de una serie cronológica de precipitaciones anuales, la cual puede además presentar efectos de persistencia, de tendencia (sur de la Costa) o pseudo-cíclicos (el Niño) que determinan su organización interna.

## 1.2. Las principales causas de heterogeneidad

### *Modificación del medio circundante al sitio de medición*

- Por desplazamiento del aparato

Es un caso frecuente y a menudo la estación conserva su nombre mientras el aparato sigue al observador en sus cambios de lugar de vivienda. Generalmente son desplazamientos pequeños (del orden del kilómetro), pero en un país de relieve muy contrastado pueden ocasionar grandes cambios en las mediciones si se modifica la exposición del aparato en relación al relieve o si se cambia considerablemente de altitud.

- Por modificación del medio circundante en sí

Tal modificación puede ser brusca (construcción muy cerca) o progresiva (crecimiento de árboles a proximidad del aparato).

- Por cambio de la altura del aparato

Entre un aparato ubicado a 1,20 m del suelo y otro ubicado a nivel del suelo, los valores encontrados pueden variar en más del 50 %. Tal diferencia se debe a las turbulencias que se forman alrededor del aparato en los sitios muy expuestos al viento [1], [2], [3].

### *Errores debidos al aparato*

- Modificación de la superficie receptora por construcción, cambio o deformación

Si bien los aparatos « estándar » tienen una superficie receptora bien definida y constante, ese no es siempre el caso de los totalizadores, de fabricación artesanal.

- Defectos de calibración

En el caso de los pluviógrafos, tales defectos pueden deberse a una falla de construcción o a una calibración mal realizada.

### *Errores de medición o de grabación*

- A nivel de la probeta (o de la regleta)

Pueden provenir de una precisión diferente de una probeta a la otra, de lecturas mal realizadas cuando la probeta no es vertical, de confusión de cifras, etc. El caso más típico es la confusión de probeta. En los países que utilizan el sistema métrico, los pluviómetros más utilizados tienen una superficie de 200, 314 ó 400 cm<sup>2</sup>. Es entonces muy frecuente que las mediciones efectuadas en un aparato no se hagan con la probeta correspondiente.

Este último error es fácilmente detectable cuando se conocen las diferentes superficies utilizadas pero a veces da lugar a confusiones al haber un cambio de sitio en las regiones montañosas en donde existen fuertes gradientes altitudinales de las precipitaciones.

Otro caso difícil de detectar se presenta cuando el observador, por seguridad, tiene dos probetas (diferentes evidentemente) y utiliza tanto la una como la otra.

- A nivel de la grabación

- falta de tinta,
- error de fechas,
- escritura errónea de la altura medida,
- error de copia, etc.

### *Errores de transmisión*

Pérdida de correo o defectos en los sistemas de transmisión (eléctricos, electrónicos, radio, satélite, etc).

### *Errores de archivo y publicación*

Pueden producirse al archivarse los datos en bruto (errores de fecha, de copia, de digitación, etc.) o al publicarse los archivos en cuyo caso se pueden presentar todo tipo de errores de escritura (omisión de datos, confusión de estaciones, etc.).

## **1.3. Consecuencias**

Como ya lo hemos señalado, todas las fallas mencionadas anteriormente alteran la homogeneidad de las series cronológicas de precipitaciones anuales pero de manera diferente según la naturaleza del problema.

Una modificación del medio circundante se detecta con un buen número de años homogéneos antes y después del cambio. Mientras más marcado es el cambio, menos años se necesitan. La experiencia demuestra que en este caso se pueden corregir los totales anuales y mensuales para homogeneizar la serie, pero en ningún caso se pueden modificar los datos diarios.

La confusión de probetas es un caso muy frecuente y necesita pocos años para ser detectada (cuando no hay utilización concomitante de varias probetas o regletas). Se puede restablecer sin problema la homogeneidad de las series de totales anuales, mensuales y también diarias cuando se conoce la fecha de la confusión.

En casos de desplazamientos frecuentes del aparato (u otras causas que producen el mismo efecto), las heterogeneidades no son discernibles y no hay posibilidad de efectuar corrección alguna.

## **1.4. Necesidad de efectuar « pruebas » de homogeneidad**

Las diversas causas de heterogeneidad y sus consecuencias demuestran la necesidad de controlar de manera rigurosa los datos pluviométricos, control que se podría hacer en la mayoría de casos consultando el historial de la estación.

La experiencia muestra que un cambio de sitio coincide generalmente con un cambio de observador y que la confusión de probetas (o regletas) se produce después de alguna interrupción de las mediciones, lo que significa decir que un buen historial que registre los cambios de sitio, de observadores o de aparatos, los cambios del medio circundante y una\* verificación de las dimensiones de los pluviómetros y probetas (o regletas), permitiría resolver gran número de problemas.

Lastimosamente, por lo general, tales historiales no existen o son difícilmente accesibles. Parte del historial puede a veces reconstruirse en base a los documentos en bruto enviados por los observadores, pero ello es insuficiente como para analizar una serie cronológica completa.

De allí la necesidad de efectuar pruebas de homogeneidad a partir de los simples datos anuales y, posteriormente, en caso de haber un problema bastante complicado, una verificación en el campo.

## **1.5. Las diferentes pruebas de homogeneidad**

Varias pruebas estadísticas existen en la literatura (test de los signos, de MANN-WHITNEY, de FISCHER-YATES, etc.) que podrían ser considerados para verificar la homogeneidad de una serie pluviométrica. Se trata generalmente de pruebas sobre el carácter aleatorio de una serie de

variables independientes, lo cual implica la hipótesis de la inexistencia de una organización cronológica de la muestra observada y por lo tanto la inexistencia de pseudo-ciclos. Esto está en contradicción con la realidad (sucesión de períodos secos y húmedos, Niño, etc.)

Por su concepción, tales pueden utilizarse una vez determinada la época de la ruptura que se para a las series homogéneas y comparando una serie de valores provenientes de la muestra con el resto.

Haciendo variar la serie de valores, se podría detectar la fecha más probable de ruptura a condición de que sea única, pero en casos complejos en donde existen varias rupturas y de sentido contrario, esas pruebas nada pueden detectar pues no han sido elaboradas para ese efecto.

Su eficacia es inferior o igual a la del test de STUDENT de comparación de medias provenientes de muestras que corresponden a problemas comunes.

Para resumir, es fácil demostrar que esas pruebas no pueden distinguir los errores sistemáticos ni la existencia real de una organización interna de la pluviometría correspondiente a efectos pseudo-cíclicos o de persistencia.

Desgraciadamente, la mayoría de series pluviométricas anuales contienen por lo menos uno de esos dos efectos, lo que disminuye aún más el interés de aplicar a la pluviometría las pruebas clásicas de homogeneidad.

## 2. LAS PRUEBAS DE SIMPLES Y DOBLES MASAS

### 2.1. Hipótesis de base

Se admite que las estaciones pluviométricas ubicadas en una misma zona climática tienen totales pluviométricos anuales pseudo-proporcionales, independientemente de las fluctuaciones climáticas de la zona. Esto significa que para un año  $i$ , las precipitaciones anuales  $x_i$  y  $y_i$  en dos estaciones  $X$  y  $Y$  de la misma zona pueden escribirse [4]:

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

en donde  $\bar{x}$  et  $\bar{y}$  corresponden a los promedios interanuales de las dos estaciones en un período suficientemente amplio. La varianza de  $e_i$  (término aleatorio independiente tanto de  $x_i$  como de  $y_i$ ) es más reducida si el coeficiente de correlación lineal entre las dos estaciones es mayor; su esperanza matemática es igual a cero.

### 2.2. Gráfico de simples masas

Según la hipótesis anterior, es razonable concebir que la información climática anual de la zona considerada sea representada por un vector  $V$  cuyos valores anuales sean la superposición de dos variables asincrónicas:

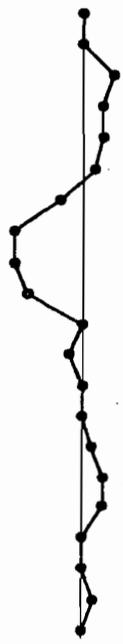
- una representa la fluctuación eventual de la media a lo largo del tiempo (período largo);
- la otra corresponde a la variación anual alrededor de la media (período corto; el año).

Los componentes  $w_i$  de dicho vector representarán para mayor comodidad los valores de la diferencia entre los totales anuales y la media expresada en términos de desviación estándar

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

ESTACION X P311MI CAHUASQUI PROVINCIA DE IMBABURA PERIODO 1963-1985 COEFICIENTE DE CORRECCION 1.000

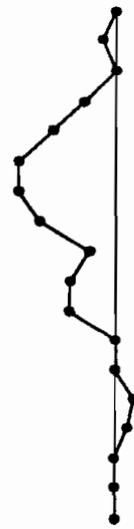
AÑOS		TA (mm)	P + -3 I	-2 I	-1 I	0,0 I	+1 I	+2 I	+3 I
1	1984	874,7							
2	1983	755,6							
3	1982	1.051,9							
4	1981	730,4							
5	1980	774,5							
6	1979	677,5							
7	1978	509,4							
8	1977	510,0							
9	1976	760,0							
10	1975	878,5							
11	1974	1.260,1							
12	1973	667,2							
13	1972	923,9							
14	1971	850,8							
15	1970	879,1							
16	1969	876,9							
17	1968	777,2							
18	1967	658,1							
19	1966	727,9							
20	1965	933,7							
21	1964	658,8							
N	AN	TA	P + -3 I	-2 I	-1 I	0,0 I	+1 I	+2 I	+3 I



PROMEDIO DE LOS 21 AÑOS = 797,0  
 DESVIACION ESTANDAR = 171,9  
 COEFICIENTE DE VARIACION = 0,216

PRUEBA DE SIMPLES MASAS

ESTACION X P302MI		MIRA	PROVINCIA DE CARCHI					PERIODO 1964-1984			COEFICIENTE DE CORRECCION 1.000		
AÑOS		TA (mm)	P+ -3 1	-2 1	-1 1	0,0 1	+1 1	+2 1	+3 1				
1	1983	632,2											
2	1981	524,5											
3	1980	629,3											
4	1979	429,7											
5	1978	436,1											
6	1977	440,9											
7	1976	602,4											
8	1975	754,4											
9	1974	876,8											
10	1973	534,1											
11	1972	597,3											
12	1971	851,0											
13	1970	697,3											
14	1969	703,4											
15	1968	530,2											
16	1967	543,2											
17	1966	583,6											
18	1965	645,5											
N	AN	TA	P+ -3 1	-2 1	-1 1	0,0 1	+1 1	+2 1	+3 1				



PROMEDIO DE LOS 18 AÑOS = 611,7  
 DESVIACION ESTANDAR = 129,4  
 COEFICIENTE DE VARIACION = 0,212

PRUEBAS DE SIMPLES MASAS

En ese caso, los totales anuales de una estación  $X$  que pertenece a dicha zona climática podrán expresarse de la siguiente manera:

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

en donde  $r_x$  representa el coeficiente de correlación lineal entre la estación y el vector, y  $a_i$  las fluctuaciones propias de la estación.

Si se efectúa el cúmulo de los datos hasta el año  $n$  tenemos:

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

y dividiendo por la media,

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

en donde  $C_x$  representa el coeficiente de variación de la variable  $x$ .

$X_n$  es proporcional a  $n$ , más un término aleatorio que obedece a un proceso de MARKOV.

El gráfico que traduce la relación  $X_n = f(n)$  es un gráfico de simples masas relativas.

En la práctica utiliza en coordenadas:

- el número  $i$  del año en la serie (creciente o decreciente y omitiendo los años faltantes);
- la suma  $\sum x_i$  de las precipitaciones anuales.

(El trazado oscilará (efecto SLUTZKY) entre dos rectas paralelas que van a determinar un corredor) <sup>??</sup>  
sensiblemente paralelo a la primera bisectriz.

Si la media es afectada por una deformación sistemática debida a los errores señalados en 1.2., el segundo término de la ecuación anterior tendrá un valor bien diferente a cero durante ese período y el trazado de los puntos se ubicará en un corredor de pendiente diferente a la primera bisectriz.

Ese cambio de pendiente es fácilmente detectable.

El gráfico de simples masas tiene entonces la ventaja de encontrar todos los períodos en los cuales la media ha sido afectada de manera significativa lo que no podían hacer las pruebas clásicas.

Desafortunadamente, en la práctica, no permite distinguir con certeza eventuales desviaciones sistemáticas de un cambio de la tendencia climática a veces bien marcada (como en el caso de las estaciones Cahuasquí, Mira y FF-CC-Carchi cuyas pruebas de dobles masas veremos más adelante).

Se lo utiliza sólo un el primer análisis para detectar los errores graves, reservando el diagnóstico final para las pruebas de dobles masas.

### 2.3. Gráfico de dobles masas

Consideramos dos estaciones que pertenecen a una misma zona climática representada por el vector  $V$  descrito anteriormente.

Las precipitaciones anuales de cada estación pueden escribirse de la siguiente manera:

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

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

Si se supone que se relacionan con la tendencia climática por los coeficientes de correlación  $r_x$  y  $r_y$  y tienen fluctuaciones propias independientes  $a_i$  y  $b_i$ , acumulando las variables hasta el año  $n$  y dividiendo por la media, tenemos:

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

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

en donde  $C$  representa los respectivos coeficientes de variación.

Eliminando  $n$  entre las dos ecuaciones y agrupando los diferentes términos, obtenemos:

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

La ecuación traduce una relación lineal alterada por 2 términos, el segundo de los cuales constituye una variable aleatoria que obedece a un proceso de MARKOV [6].

2.2: ( En la práctica, los puntos de coordenadas  $(X_n, Y_n)$  estarán al interior de un corredor definido por dos rectas paralelas de pendiente positiva e igual a 1. Si las variables son independientes, el ancho del corredor dependerá únicamente de los coeficientes de correlación de cada estación. En caso contrario, dependerá también del valor de los coeficientes de correlación y será más estrecho si los coeficientes son más fuertes. )

TODA VARIACION DE UNA DE LAS MEDIAS POR MODIFICACION DE LAS CONDICIONES DE OBSERVACION SE TRADUCIRA EN UNA MODIFICACION SIMULTANEA DE LA PENDIENTE DEL CORREDOR.

La variación de pendiente corresponderá a la variación de la media afectada.

Si las dos estaciones pertenecen estrictamente a la misma zona climática, los coeficientes de variación así como los coeficientes de correlación tendrán un valor similar, lo que implica que el segundo término será cercano a cero. Esta constatación demuestra la ventaja del método de las dobles masas: en una misma zona climática, la relación existente entre los acumulados de dos variables pluviométricas es independiente de eventuales fluctuaciones climáticas y es modificada únicamente por fluctuaciones aleatorias si los datos son homogéneos.

Esto significa que los cambios de pendiente corresponden únicamente a los errores sistemáticos descritos en 1.2., como se puede comprobar en los gráficos de FF-CC-Carchi/Mira y Mira/Cahuasquí.

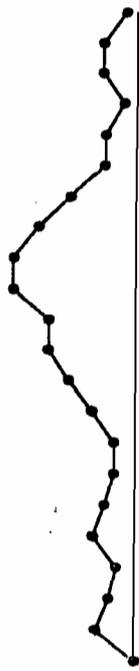
#### 2.4. Consideraciones sobre los gráficos de simples y dobles masas

La primera operación consiste en determinar de manera empírica grupos de estaciones que pertenecen a la misma zona climática teniendo en cuenta la distancia entre ellas, los efectos de relieve, el tipo de régimen intra-anual, etc.

Posteriormente, se compara cada una de las estaciones con todas las demás del grupo, lo que significa que para un grupo de  $m$  estaciones, el número de dobles masas alcanza  $m(m-1)/2$ .

ESTACION X P30IMI    FF-CC-CARCHI    PROVINCIA DE CARCHI    PERIODO 1956-1985    COEFICIENTE DE CORRECCION 1.000

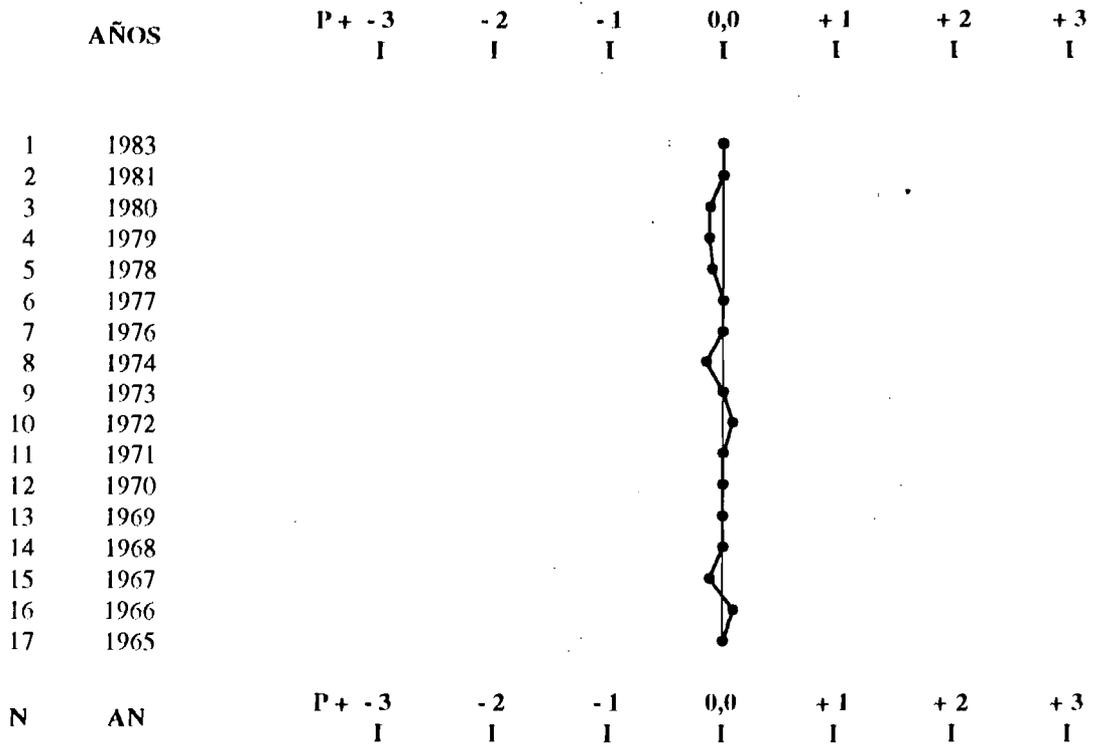
AÑOS	TA (mm)	COEFICIENTE DE CORRECCION 1.000							
		P+ -3 I	-2 I	-1 I	0,0 I	+1 I	+2 I	+3 I	
1	1985	495,8							
2	1984	425,7							
3	1983	556,1							
4	1982	663,6							
5	1981	449,2							
6	1980	533,8							
7	1979	394,0							
8	1978	352,2							
9	1977	486,4							
10	1976	516,2							
11	1974	747,5							
12	1973	542,4							
13	1972	671,2							
14	1971	636,2							
15	1970	662,2							
16	1969	575,1							
17	1968	492,5							
18	1967	460,4							
19	1966	657,8							
20	1965	516,5							
21	1962	517,0							
22	1956	783,5							
N	AN	TA	P+ -3 I	-2 I	-1 I	0,0 I	+1 I	+2 I	+3 I



PROMEDIO DE LOS 21 AÑOS = 551,6  
 DESVIACION ESTANDAR = 112,2  
 COEFICIENTE DE VARIACION = 0,203

PRUEBAS DE SIMPLES MASAS

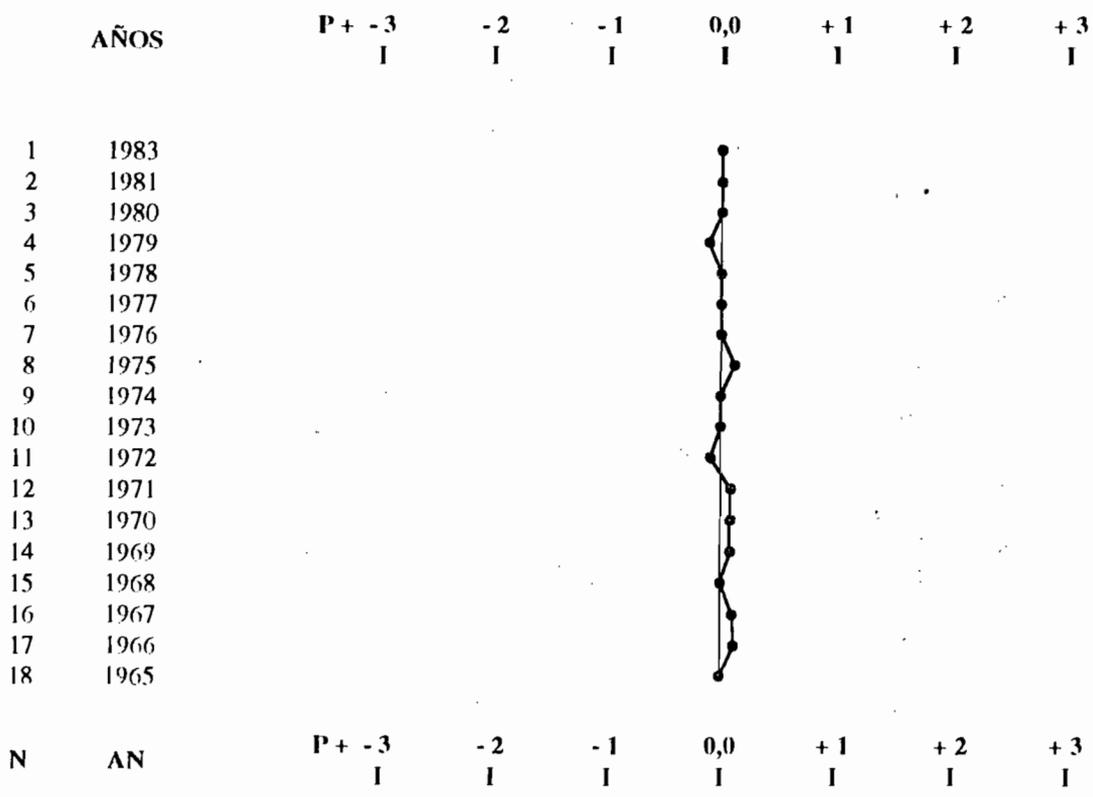
ESTACION X P301MI	FF-CC-CARCHI	PROVINCIA DE CARCHI	PERIODO 1964-1984	COEFICIENTE DE CORRECCION 1.000
ESTACION X P302MI	MIRA	PROVINCIA DE CARCHI	PERIODO 1964-1984	COEFICIENTE DE CORRECCION 1.000



$XM = 544,1$        $XM/YM = 902$   
 $YM = 603,3$        $YM/XM = 1.109$   
 COEFICIENTE DE CORRELACION = 0,800

PRUEBA DE DOBLES MASAS

ESTACION X P302MI	MIRA	PROVINCIA DE CARCHI	PERIODO 1964-1984	COEFICIENTE DE CORRECCION 1.000
ESTACION X P311MI	CAHUASQUI	PROVINCIA DE IMBABURA	PERIODO 1964-1984	COEFICIENTE DE CORRECCION 1.000



XM = 611,7      XM/YM = 778  
 YM = 786,2      YM/XM = 1.285  
 COEFICIENTE DE CORRELACION = 0,833

PRUEBA DE DOBLES MASAS

Por lo tanto, el tratamiento de la pluviometría de una región o de un país exige la manipulación de una cantidad considerable de información, solamente posible mediante automatización de las operaciones de tratamiento.

X (El programa MASA utilizado actualmente en el INERHI constituye un buen ejemplo de trazado automático de los gráficos de simples y dobles masas.

Su utilización facilita grandemente las comparaciones entre estaciones suprimiendo la mayor parte de las operaciones manuales y aportando una valiosa ayuda en el trabajo de interpretación visual.

La mejor manera de utilizar el método de dobles masas es buscar, en la zona climática delimitada, algunas estaciones de período largo en las cuales el hidrólogo esté seguro de la homogeneidad de los datos (estaciones de base); compararlas con estaciones vecinas y corregir cada vez que aparezcan rupturas de pendientes. En cada corrección es necesaria una prueba con todas las estaciones de base a fin de comprobar las correcciones efectuadas.

Se pueden así constituir otras estaciones de base y avanzar lentamente verificando cada vez los resultados con las estaciones más seguras.

Ciertamente, operando de esta manera y además minuciosamente se llegan a detectar de exactamente los cambios de pendiente que llevan al diagnóstico, comprobados generalmente con historiales o verificación de campo. La metodología es aplicada en muchos países y varias generaciones de hidrólogos han comprobado su efectividad.

Sin embargo, la realidad es más compleja. No siempre es posible encontrar estaciones seguras de antemano, y en esos casos, la diversidad de anomalías de una u otra estación, hace difícil encontrar criterios de decisión realmente objetivos.

Además, es frecuente que al disponer de estaciones de base, estas tengan períodos incompletos, justamente durante los años interesantes de comparar.

Se percibe entonces, el interés de elaborar una variable de referencia sin años incompletos a partir de un conjunto de datos de precipitaciones que corresponden a estaciones ubicadas en una misma zona climática y ello sin intervención manual.

### 3. ELABORACION DE UN VECTOR DE REFERENCIA

#### 3.1. Principales particularidades

La comparación de las estaciones que pertenecen a una misma zona climática con un vector de referencia representante de la tendencia climática de dicha zona permite reducir el volumen de análisis. En una zona que contiene  $m$  estaciones, sólo se realizarán  $m$  dobles masas — y no  $m(m-1)/2$  como anteriormente) disminuyendo así de considerablemente el volumen de análisis.

Si  $z$  representa este vector de referencia, tendremos:

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

Acumulando las variables hasta el año  $n$ , y dividiendo por la media  $\bar{z}$ ,

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

Comparando este vector con una estación  $X$  y eliminando  $n$  entre las dos ecuaciones como se hizo anteriormente para las ecuaciones (7) y (8), obtenemos:

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

Como se dijo antes, el segundo término del segundo miembro de la ecuación tiende hacia cero y el tercer término tendrá como varianza:

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

En la comparación estación por estación, la varianza del tercer término de la ecuación (9) podría escribirse así:

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

Con coeficientes de variación similares, se puede verificar que  $V_1 < V_2$ , lo que significa que el ancho del corredor será más reducido y que consecuentemente las dobles masas entre estación y vector serán más precisas que las dobles masas entre estaciones.

### 3.2. Condiciones de elaboración

Como lo hemos manifestado anteriormente, la práctica de dobles masas requiere la hipótesis de pseudo-proporcionalidad en un espacio dado y el vector de referencia tendrá que adaptarse a ella, pero se deberá también tener en cuenta algunas consideraciones complementarias:

- su elaboración no debe exigir hipótesis previas sobre la distribución de los totales pluviométricos anuales;
- el total pluviométrico de una estación para un año dado no puede a priori servir de valor de referencia;
- se supone que en el conjunto de la información global de las estaciones, existe un valor estimativo de la tendencia climática regional, más representativa que la información parcial de cada una de ellas; esto no excluye la existencia de anomalías puramente locales (Niño por ejemplo) pero ellas no deben considerarse para la elaboración del vector;
- la elaboración del vector debe ser concebida de tal manera que toda la información disponible sea utilizada sin que la información errónea tenga una influencia sensible en el resultado; sin embargo, en la práctica, cuando los datos son realmente desastrosos, será necesario eliminar los errores mayores, pero no se eliminarán, a priori, estaciones de período muy corto (3 a 5 años) cuyos datos pueden ayudar a la construcción de un componente anual del vector.

### 3.3. Análisis de la matriz de datos

Se puede considerar el conjunto de los datos pluviométricos como una matriz  $A$  en la que las columnas representan las estaciones y las líneas, los años.

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

Los  $m$  datos obtenidos durante un año  $i$  representan las coordenadas de un vector (lluvia) anual en el espacio de las estaciones.

Si la pseudo-proporcionalidad fuese estricta, el conjunto de los  $n$  vectores constituiría un conjunto de vectores colineales. La matriz  $A$  podría representarse entonces como una matriz  $B$  producto de un vector línea  $L$  y un vector columna  $C$ .

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

En la práctica, los datos reales no obedecen a este modelo teórico y existen fluctuaciones aleatorias en cada estación (1) que deben tenerse en cuenta. Así, tenemos:

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

en donde  $B$  representa la información teórica deseada.

### 3.4. Elaboración de un vector por la media aritmética

Una de las posibilidades es calcular para cada año un elemento del vector utilizando la media aritmética de todos los elementos de ese año.

En caso de homogeneidad previa de todos los datos y de distribuciones normales, la media aritmética corresponde efectivamente a un parámetro de tendencia central.

Lastimosamente, los casos de distribuciones no normales son frecuentes y pueden deberse al escaso número de datos, a distribuciones asimétricas o a distribuciones pluri modales (en los casos de series cronológicas heterogéneas).

La existencia de valores muy elevados vienen también a alterar el cálculo de la media aritmética (contaminación de la media por los valores extremos), como lo veremos en algunos ejemplos más adelante.

Las mismas consideraciones se pueden aplicar a la desviación estándar, la cual en caso de distribuciones no normales puede dar una visión de la dispersión de los datos pero no de su concentración alrededor de valores centrales.

Es posible lograr resultados valiosos por estimación de los valores promedio, pero los cálculos se hacen problemáticos cuando las series tienen muchos datos faltantes.

Todo lo anterior conduce a eliminar la distancia « euclidiana » como criterio de minimización de  $E = A - B$ , y llevó a dos investigadores, G. HIEZ [6] y Y. BRUNET-MORET [4] a calcular los elementos de un vector regional a partir de direcciones diferentes aunque los resultados sean relativamente similares.

### 3.5. El vector HIEZ [6]

#### 3.5.1. Algoritmo de construcción

Consiste en una buena estimación de valor central en el valor modal más probable, pero no existe un método usual para su cálculo preciso.

En caso de numerosos datos, se podría agruparlos por clases y tomar como valor modal el punto mediano de la clase correspondiente a la frecuencia observada más elevada.

En caso contrario, es difícil encontrar los valores modales del valor línea ( $L$ ) y del vector columna ( $C$ ) sin conocer las funciones de densidad de cada uno, lo que no se sabe a priori.

Existen también otras fórmulas que vamos a revisar rápidamente:

- una fórmula empírica

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

en donde  $\hat{M}$  es el valor modal,  $\bar{M}$  la mediana y  $M$  la media aritmética; esta fórmula es válida para distribuciones unimodales y simétricas;

- considerando los momentos de segundo y tercer orden tenemos:

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

en donde  $x_0$  es el valor que anula al momento de tercer orden; esta fórmula tiene en cuenta la disimetría de la distribución pero necesita conocer la mediana cuya estimación carece generalmente de precisión;

- la descomposición de una función de distribución cualquiera  $f(\mu)$  en series limitadas permite llegar a otra fórmula:

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

en donde  $\mu_0$  es el valor para el cual  $f(\mu)$  es máximo; esta fórmula no es tampoco conveniente para las distribuciones plurimodales; es sensible a los valores extremos porque hace intervenir un parámetro de dispersión.

Finalmente el autor utiliza un método original de tratamiento líneas-columnas que permite, por iteraciones sucesivas y utilizando una de las fórmulas anteriores, evadir las restricciones mencionadas. La rapidez de convergencia del algoritmo de cálculo depende evidentemente de la calidad de los datos.

La presencia de valores faltantes no influye en el tratamiento a condición de tener por lo menos tres valores por línea y columna, pero la estimación del vector es mayor\* si los datos son numerosos.

### 3.5.2. Conclusiones sobre este vector

La sensibilidad y confiabilidad del algoritmo ha sido comprobada mediante varias pruebas; la totalidad de la pluviometría del Brasil ha sido comprobada con ese método.

Una vez elaborado el vector, se lo compara con las estaciones por dobles masas pero no se comparan los datos con su valor promedio sino con su valor más probable.

La aplicación de tal metodología requiere respetar algunas reglas:

- utilización del año hidrológico (y no calendario);
- un número de estaciones necesariamente superior a 3, sabiendo que el mínimo aconsejable es 7;
- una zonificación previa de las estaciones, la cual se podrá precisar después de un primer tratamiento.

Se revela evidente que dicho vector es un instrumento poderoso. Su aplicación en todo el Brasil lo comprueba, pero falta analizarlo más precisamente en un medio montañoso como el Ecuador.

### 3.6. El vector de los índices anuales de precipitación [4]

#### 3.6.1. Construcción del vector

El autor considera que una zona climática puede representarse por un vector

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

de modo que su media  $\mathbf{Z}$  sea igual a 1.

La ecuación (17) puede escribirse de la siguiente forma:

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

para la precipitación del año  $i$  de la estación  $p$ , sabiendo que la media de  $E_p$  es igual a cero como lo hemos dicho anteriormente.

La matriz  $\mathbf{A}$  ( $n, m$ ) se compone de las precipitaciones de  $m$  estaciones en  $n$  años y será en la mayoría de casos incompleta.

El algoritmo de cálculo consiste en minimizar la ecuación:

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

para determinar cada  $\bar{x}_j$  y cada  $z_i$ .

Derivando según  $z_i$ , si existen  $p$  estaciones observadas el año  $i$ , obtenemos  $n$  ecuaciones

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

Reemplazando en la fórmula (22) los valores de  $z_i$  (23) y derivando según  $1 / \bar{x}_j$ , obtenemos un sistema de  $m$  ecuaciones de  $m$  variables desconocidas en  $1 / \bar{x}_j$ .

La resolución del sistema permite encontrar  $\bar{x}_j$ , y mediante la ecuación (23), los diferentes componentes  $z_i$ .

El método rápidamente descrito permite considerar los valores observados con el mismo peso.

Se supone que para cada año y en cada estación los valores erróneos son la minoría.

Una primera iteración con los elementos originales de  $\mathbf{A}$  calculan una primera serie de valores. Si un elemento  $x_{ij} / \bar{x}_j - z_i$  no está en un intervalo, es reemplazado por  $\bar{x}_j \cdot z_i$ .

Una serie de iteraciones sucesivas permite eliminar esos problemas rápidamente. Los límites del intervalo dependen de la varianza de  $\mathbf{E}$ .

#### 3.6.2. Utilización del método

Después de haber verificado el historial de las estaciones, suprimido los valores aberrantes (errores de teclado, etc.) y completado los valores que se pueden completar, se efectúa un

primer paso en simples masas para confirmar los grupos representativos de una misma zona climática.

Una vez efectuada esa preparación, se puede hacer un primer paso del programa que sirve para eliminar los valores claramente erróneos.

En función de los coeficientes de correlación se pueden también eliminar algunas estaciones para restringir la zona y obtener así mejores correlaciones.

Los siguientes pasos sirven para detectar los errores menos marcados (cambio de ubicación o modificación del medio circundante, etc.) y llegar así en dos o tres pasos a la construcción de un vector representativo de la región considerada.

#### 4. COMPARACION DE LOS DOS VECTORES

El vector regional HIEZ tiene la ventaja de fundarse desde el inicio en una estimación precisa del valor modal, por lo que es independiente de los errores sistemáticos siempre y cuando estos sean una minoría. Se puede entonces en un solo paso detectar las anomalías y calcular los coeficientes de pendiente correspondientes a una confusión de probeta, un cambio de sitio, etc.

Ha sido en todo el territorio brasileño, lo cual representa un buen número de pruebas que han llevado a la elaboración de normas de utilización.

Lastimosamente, su algoritmo de cálculo es algo difícil y largo y hasta el momento, sólo funciona en computadores grandes.

Recientemente, está siendo adaptado para micro-computación y será interesante ver cómo funciona en el medio montañoso.

El vector de índice anual de precipitación busca el valor más probable para un algoritmo más sencillo, no totalmente insensible a la existencia de errores sistemáticos, de manera que se requieren 2 ó 3 pasos antes de llegar a una estimación correcta.

En el período de elaboración, las estaciones en donde aparecen rupturas de pendiente evidentes, se consideran como estaciones secundarias.

El análisis de la relación entre dos pendientes se efectúa en el paso final, una vez elaborado el vector.

Ha funcionado ya en el Africa y el Caribe, y ha sido ensayado en el sur oriente del Ecuador. Se está probando en la cuenca del Mira y luego será aplicado en la Costa en el marco de la elaboración del Plan Nacional de Riego con el INERHI.

Está disponible en compatibles IBM pero los listados de resultados son tan complicados que será necesario representarlos en gráficos para tener un criterio visual ayudado por criterios numéricos.

Para compararlos con un vector elaborado a partir de la media aritmética, tomaremos el siguiente ejemplo:

Doce estaciones observadas durante veinte años: 1966-1985. Cada año en cada estación los valores varían entre 999,5 y 1.001,5 mm, es decir que todos son muy cercanos a 1.000.

Multiplicamos los datos de tres estaciones por 5 en un período de 10 años (1966-1975).

Un vector elaborado mediante la media aritmética tendría las siguientes características:

1966 - 1975	$x_1 = 2.000$
1976 - 1985	$x_1 = 1.000$

es decir una relación del orden de 2.

Con el vector de índices anuales de precipitación, la relación es de 1,25 lo que indica claramente la poca sensibilidad a las confusiones de probetas o a los desplazamientos.

En esas mismas 12 estaciones, se introduce un año con una precipitación de 13.000 mm. La media aritmética nos indica 2.000 mm, el vector índice 1.015, lo que demuestra la casi insensibilidad a valores extremos como errores de teclado, etc.

En los dos casos anteriores (un tanto extremos), la comparación de las estaciones mediante un vector elaborado a partir de la media aritmética, no hubiera permitido un diagnóstico, porque el vector en sí está muy alterado.

El vector HIEZ es aún más insensible a errores de ese tipo como lo demuestran ejemplos que se presentaron en el Brasil.

Para finalizar, cabe recordar que en el numeral 3.1 se señala que dobles masas entre estación y vector son más precisas que dobles masas entre estaciones. Esa característica puede verificarse en el siguiente ejemplo.

Se calcularon los coeficientes de correlación entre 12 estaciones del norte del Ecuador, comparadas una con otra. Se anotó para cada estación:

- el coeficiente correlación máximo CCMAX;
- el coeficiente de correlación mínimo CCMIN;
- los coeficientes de correlación promedio CCMOY;
- el coeficiente de correlación con el vector índice CCVEC, elaborado a partir de las 12 estaciones sin hacer ninguna corrección y sabiendo que ellas pertenecen a regímenes pluviométricos diferentes.

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

Se puede observar que las correlaciones con el vector son siempre superiores a las correlaciones promedio entre estaciones, casi siempre iguales o superiores a las correlaciones máximas.

Esa característica será de gran ayuda para completar los valores faltantes.

## 5. CONCLUSION

Como lo hemos visto al comienzo, sólo el método de dobles masas acoplado a las simples masas permite detectar y corregir los datos pluviométricos erróneos.

Desafortunadamente, dicho método exige un trabajo amplio y minucioso así como la manipulación de una gran cantidad de información.

El concepto de vector pluviométrico es por eso un gran avance metodológico. Su elaboración no puede contentarse de promedios aritméticos que son alterados por valores extremos.

Los dos vectores presentados representan valores probables por algoritmos diferentes, o sea que son casi insensibles a las alteraciones. Puede parecer un poco exagerado el cuidado con el cual se elaboran, pero no sirven únicamente para homogeneizar los datos sino también para representar lo más exactamente posible todas las características de una zona climática, lo que significa que son de gran ayuda para la regionalización climática.

## BIBLIOGRAFIA

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