

**ALGUNOS METODOS ESTADISTICOS PARA TIPIFICACION DE SITUACIONES SIROPTICAS:
DISCUSION METODOLOGICA**

Rosa Hilda Compagnucci

Instituto Argentino de Nivología y Glaciología

Conicet, Mendoza

Luis Fornero

Universidad Nacional de Cuyo

Walter Mario Vargas

Departamento de Meteorología - F.C.E.y N.- UBA

Conicet, Buenos Aires

RESUMEN

Se realiza una discusión metodológica comparativa de algunos de los métodos estadísticos conocidos, utilizables para la clasificación, tipificación y zonación de variables meteorológicas. Para ello se tomaron como referencia los estudios realizados para una serie de campos de presión de julio (5 años) en la región sud de Sudamérica. Los métodos involucrados son: componentes principales, Lund (método de correlación) y McQuitty (análisis de encadenamiento simple). Los resultados mostraron que componentes principales, a pesar de una mayor complejidad, tiene mejores perspectivas de éxito en análisis (objetivo) de tipos de situaciones sinópticas.

ABSTRACT

A comparative methodological discussion comparing some of the known statistical methods, used for classification, clustering and map typing of meteorological variables. The references taken were the studies made on a series of pressure fields in the southern part of South America in July (5 years). The methods used were principal components, Lund (correlation method) and McQuitty (simple linkage analysis). It can be expected that the principal components method, in spite of its complexity will show better results when objectively analyzing different types of synoptic situations.

1. INTRODUCCION

Un problema involucrado en la climatología sinóptica es la obtención y construcción de prototipos o modelos de la distribución de variables meteorológicas. En Barry y Perry (1973) se encuentra una revisión exhaustiva de metodologías apropiadas para este fin. Los métodos están divididos en tres clases: de correlación, de "linkage" o encadenamiento simple y múltiple, y análisis de factores o componentes principales (CPs).

Este trabajo surgió de la necesidad de determinar el método apto para clasificar los campos béricos de superficie, para luego asociarlos a fenómenos significativos. Cabe señalar resultados de carácter metodológico previamente obtenidos. Vargas y Compagnucci (1982) discutieron la utilización de matrices de entrada del modo-S y modo-T con el fin de tipificar campos béricos, concluyendo que el modo-T es más ventajoso para este fin. Vargas y Compagnucci (1983) analizaron la red de estaciones a usarse en el caso de tipificar campos reales para el cono sur de Sudamérica. Esta resultó ser redundante, pues los tipos obtenidos no variaron al reducir el número de estaciones en aproximadamente el 50%. Asimismo, se infirió que no variarían los resultados al interpolarse los datos no consignados en algunas estaciones, en especial al hacerlo en base al campo bérico. Estas conclusiones fueron tenidas en cuenta en el desarrollo del presente trabajo.

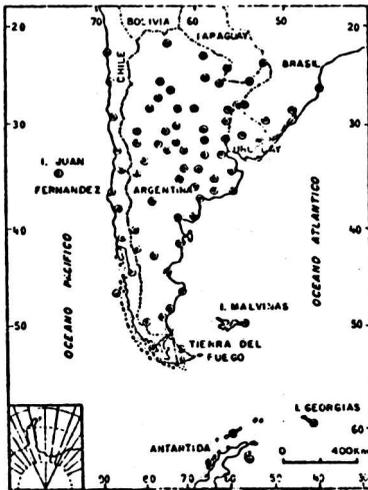
Es imposible presentar aquí todos los resultados que se obtuvieron de la aplicación de los métodos de McQuitty (1957), Lund (1963) y CPs no-rotadas, no obstante se incluyeron algunos de ellos para ejemplificar las conclusiones. Asimismo, se incluyó una breve reseña de los métodos y un análisis de las ventajas y desventajas que los mismos ofrecen. Por otra parte, se describió la metodología empleada por Blasing (1975) y se discutió su aplicabilidad, aunque la misma aún no ha sido utilizada por los autores.

2. DATOS Y CONSIDERACIONES SOBRE LA MATRIZ DE ENTRADA

Se utilizó una muestra de mapas de presión de superficie de la hora 12Z (9 H.O.A.) del Servicio Meteorológico Nacional para los 186 días del mes de julio de 1972 a 1977. Se formó así la base de datos a partir de los valores de presión de 81 estaciones de la red. Esta fue utilizada por Vargas y Compagnucci (1983), con la inclusión en esta circunstancia de las estaciones de Antártida y de algunas islas.

Respecto a la base de datos y a la medida de similitud empleada, cabe señalar lo siguiente:

a) Frente a la posibilidad de una red regular cuyos datos son valores interpolados, se optó por un enrejado formado con estaciones de la red meteorológica existente. Esto significa una densidad de información y medidas de similitud sesgadas debido a distintos "pesos" por áreas. Lo último se aceptó como un error me-



Mapa 1: Enrejado

nor que el que se ocasionaría usando una red regular, pues la medida de similitud se vería afectada por las interpolaciones subjetivas del campo bórico o por las aproximaciones fuertes dadas por un análisis de contorno a través de "puntos fijos" o datos de las estaciones. Por otra parte, los analistas de las oficinas de pronóstico generan los modelos de los campos diarios de presión en base a datos de una red que en su mayor parte es la aquí considerada.

b) Para construir la matriz de entrada procesada por los métodos de clasificación se pueden usar distintas medidas de similitud, Barry y Perry (1973) señalan como factibles a coeficientes de asociación,

de correlación y diversas medidas de distancia. De todos ellos, se decidió utilizar el coeficiente de correlación.

c) En general, para cualquier método de clasificación, es posible tener seis formas de modos distintos como matriz de entrada (Fig. 1, de Green, 1978).

La discusión metodológica encarada en este trabajo está basada en los resultados obtenidos al aplicar los métodos de clasificación a dos de ellas, el modo-T y el modo-S. Sin embargo, las conclusiones obtenidas podrán en su mayoría generalizarse para los otros modos y diferentes tipos de datos.

3. METODOLOGIAS

Se define como grupo al conjunto de casos que se consideran "conceptualmente" similares, y como tipo al modelo teórico o al caso real de la muestra, que "representa" o define al grupo.

3.a Método de correlación - Lund (1963): utilizó el método de correlación para hallar tipos de campos de presión sobre el nivel del mar. Por este método los prototipos obtenidos son casos de la muestra. A partir de su identificación, y por otro proceso matemático, se generan grupos para cada tipo identificado. En síntesis, el desarrollo es el siguiente:

- i. Obtención de la matriz de correlación.
- ii. Determinación del nivel de tolerancia o valor umbral de r (Lund utilizó un coeficiente de correlación $r \geq 0.7$).
- iii. Localización del día para el cual hay mayor cantidad de casos con correlación igual o superior al nivel de tolerancia.
- iv. Ese día es denominado Tipo-A, y se lo extrae de la matriz de correlación al

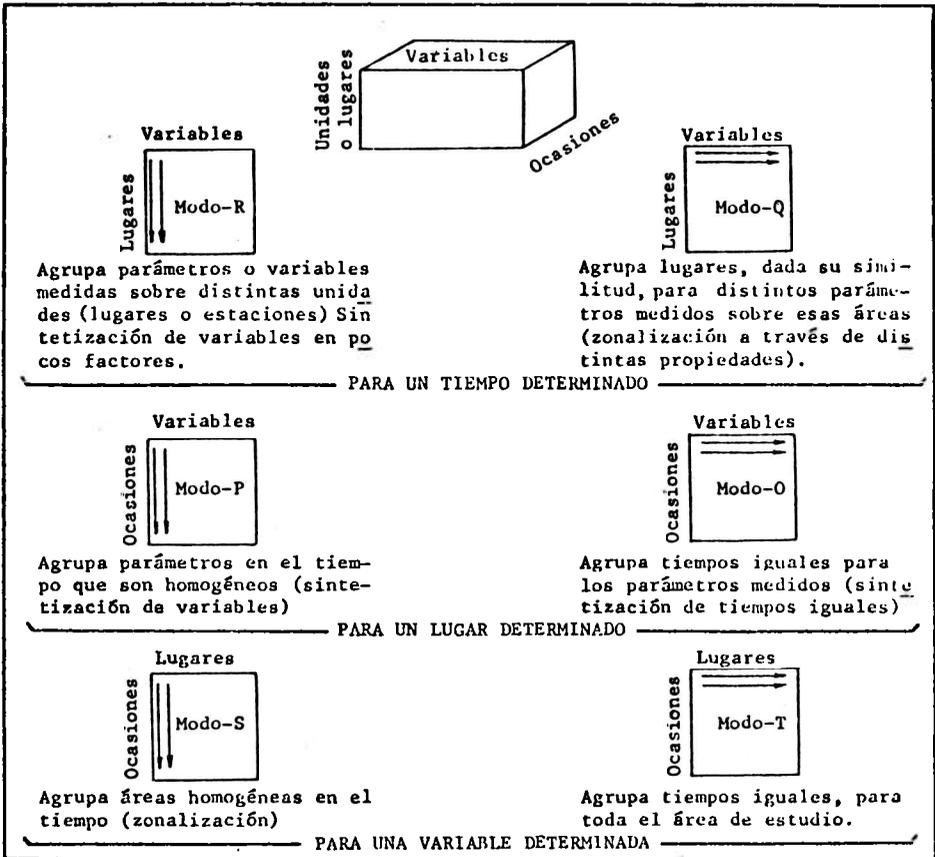


Fig. 1: Se ilustran los seis modos de presentar la matriz de entrada de datos, las flechas indican el sentido en que son tomados los pares para calcular los coeficientes de similitud.

igual que todos aquellos que están correlacionados con él, en mayor o igual valor que el umbral.

- v. Para la matriz restante se localiza el siguiente día con mayor cantidad de cantidad de casos correlacionados con él con un valor de $r \geq r_c$. Se lo denomina Tipo-B, y repitiéndose los pasos en forma sucesiva, hasta no hallar más tipos.
 - vi. Los días que se relacionen con valores menores que r_c con todos los casos, quedan sin agrupar y se los extrae de la matriz formando la clase de los no-agrupados.
 - vii. Se reagrupan los casos restantes de vi, buscando para cada uno de ellos cuál es el tipo con el cual tienen correlación mayor, se los agrupa bajo ese tipo.
- Características inferidas del método de Lund: en la Fig. 2 se muestra la canti-

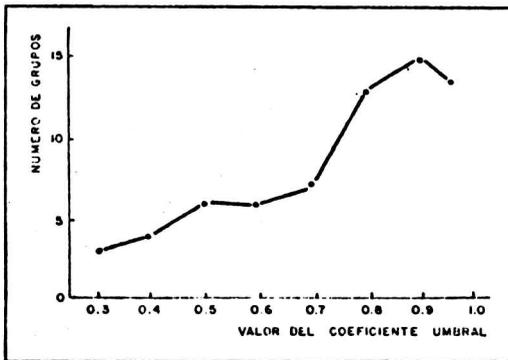


Fig. 2: Número de grupos en función de r_c .

muestra. Como se podría suponer, la cantidad de días no-agrupados, aumenta al aumentar el valor de r_c .

Este método puede ser aplicado a cualquiera de los seis modos de matriz de entrada. Compagnucci y Vargas (1984a) lo utilizaron para regionalizar las variaciones de presión donde la matriz de entrada fue del modo-S.

El tipo es un caso real que depende del valor del nivel de tolerancia y del tamaño de la muestra tomada. Posiblemente se obtendrían distintos tipos si se agregan casos a la muestra y surgirán por ende, distintos grupos. En este sentido es el más inestable de los métodos de clasificación. Para obviar este inconveniente Barry y Perry (1973) sugirieron calcular los tipos a partir de una muestra inicial y luego manteniéndolos como modelo de base, agregar los nuevos casos en los grupos correspondientes de acuerdo a la mayor correlación.

El valor de tolerancia elegido queda a criterio del investigador, con lo cual se introduce un factor de subjetividad. Algunos autores como Hartranfvej y otros (1970) sugirieron un $r_c \geq 0.7$ para campos de presión de superficie y $r_c = 0.9$ para los mapas de altura.

Cada uno de los casos de la muestra es ubicado en forma precisa dentro de alguno de los grupos o no se clasifican, no habiendo casos ambiguos en donde el investigador deba decidir la pertenencia a algún grupo. Sin embargo, es posible encontrar para algunas muestras casos agrupados bajo algunos tipos que tengan en realidad su mayor correlación con casos de otros grupos. Esto significa un factor de ambigüedad en el método. No se puede garantizar la independencia de los grupos pues los tipos o modelos no son necesariamente ortogonales.

Es posible obtener, a partir del número de casos de cada grupo, el porcentaje relativo respecto del total aunque esto no es igual que determinar la varianza que explica cada uno de ellos. El método es de aplicación simple y no requiere gran capacidad de cálculo ni procedimientos complejos de computación.

dad de tipos que se formaron aplicando la metodología antecedente a los 186 días del mes de julio de 1972 a 1977 (con el modo-T) usando diversos niveles de tolerancia. Se puede ver que el número de tipos aumenta de tres (para $r_c = 0.3$) hasta quince ($r_c = 0.9$) disminuyendo luego a trece ($r_c = 0.95$). Podemos inferir que existe un valor máximo en la cantidad de grupos para algún valor de r_c , dependiendo esto de las características de la

Terence Blasing (1975): utilizó un método de correlación basado en el de Lund al cual aplicó modificaciones. Analizó la misma área y grupos de datos para los cuales Kutzbach en 1970 calculó las componentes (CPs) principales. Concluyó que la obtención de tipos mediante el método de correlación parecería tener ventajas sobre el análisis de CPs cuando es utilizado con propósitos descriptivos, en oposición a lo que ocurre cuando se utiliza con propósitos predictivos.

Por medio de esta metodología se obtienen los tipos (modelos teóricos) y asignan los casos a cada grupo, mediante los procedimientos siguientes:

- i. Se calcula la matriz de correlación.
- ii. Igual que en el método de Lund, se examina la matriz para determinar qué días están correlacionados con valores iguales o mayores que un r_c .
- iii. Si de esta manera se logra en la matriz más de cinco tipos que a su vez formen grupos de más de cinco casos, ese r_c es usado como valor crítico, y se repite la prueba hasta que el número de grupos y la cantidad de elementos en cada uno de ellos sea aceptable.
- iv. Una vez fijado el r_c se analiza por filas en la matriz de correlación cuáles son los días que tiene un valor de r mayor o igual que el r_c . Para cada una de las filas se promedian estos coeficientes así elegidos obteniéndose, entonces, tantos r promedios como filas tenga la matriz.
- v. Se identifica la fila que tenga el valor de r promedio mayor, el día correspondiente a esa fila y se promedia con los otros cinco días más correlacionados con él, formando así el primer modelo o mapa patrón. La cantidad de días a promediar para formar los tipos, es arbitraria (Blasing utilizó 5) y depende del número de casos que formen la muestra y del grado de resolución que se desee en los tipos.
- vi. Se correlaciona el mapa tipo así obtenido, con todos los casos de la muestra y se incluye dentro de un grupo a todos aquellos que tengan un coeficiente de correlación mayor o igual que el crítico previamente fijado. Efectuado lo anterior se los extrae de la matriz para repetir el proceso.
- vii. Se reitera el paso iii para el mismo valor de correlación crítica. Si no se logra que el número de grupos y de integrantes para cada uno de ellos sea aceptable se baja el valor de r_c y se repite hasta definir un nuevo r_c . A continuación se realiza nuevamente el proceso completo hasta formar todos los modelos y grupos asociados que hay en la muestra.
- viii. Una vez obtenidos los modelos, de acuerdo a lo anterior, se redefinen los grupos asociados, correlacionando nuevamente cada caso con todos los modelos semiteóricos. A posteriori se asignan los casos al grupo representado por el modelo con quien tenga la mayor correlación.

Características indefinidas del método de Blasing: el tipo obtenido mediante esta metodología no es un caso real de la muestra sino más bien un modelo semi

teórico (promedio de casos reales). Al aumentar o disminuir el tamaño de la muestra, se altera el modelo teórico resultante en una medida menor que en el caso de Lund. Los r_c están elegidos arbitrariamente así como también la cantidad de casos a promediar para obtener el tipo. Los valores de r_c que se elijan definirán, como en el caso de Lund, la cantidad de grupos que se formen. A valores más altos de r_c mayor será la cantidad de grupos y la posibilidad de que sus tipos sean similares. Los casos de la muestra son agrupados, de acuerdo al tipo con el cual están más altamente correlacionados, lo que disminuye la ambigüedad en la clasificación. Los tipos no son necesariamente ortogonales, con lo cual los grupos no son independientes. También en este caso es posible obtener el porcentaje de casos por grupo, pero no la varianza explicada por cada uno. El método iterativo aunque no requiere una gran capacidad de memoria es más complicado que el de Lund en cuanto a su programación.

3.b Análisis de encadenamiento simple - McQuitty (1975): describe, discute y utiliza un método de encadenamiento simple o análisis de linkage, para la agrupación de individuos en base a valores obtenidos de la aplicación de distintos "tests". El mismo método es usado por Racine y Raymond (1973) para regionalización urbana.

En este método se definen los grupos dentro de los cuales se podría considerar como tipos al par de casos recíprocos que son el núcleo del grupo. La aplicación se puede describir mediante los pasos siguientes:

- i. Obtención de la matriz de correlación o asociación.
- ii. Determinación del mayor valor de correlación para cada columna (variable) de la matriz de entrada.
- iii. Definición del primer par recíproco. Será aquel para el cual ambas variables tengan el mayor valor de correlación, una en la otra y viceversa, que coincide con el máximo valor en la matriz.
- iv. Se genera el primer grupo a partir de este par recíproco, buscando por fi las cuáles son las variables que tienen su correlación mayor con las del par y la unión con éste, forma el primer nivel de encadenamiento.
- v. Se buscan ahora las variables que tienen su mayor correlación con algunas de las pertenecientes al primer nivel de encadenamiento y se las une a éstas, formando el segundo nivel de encadenamiento. Se continúa hasta no hallar variables que permitan formar un nuevo nivel.
- vi. Se eliminan de la matriz las variables que constituyen este primer grupo de encadenamientos y con la matriz restante se repite el proceso descrito en iii, iv, v y vi, hasta hallar todos los grupos dentro de la muestra.

Características indefinidas del método de McQuitty: el método en sí no determina tipos sino grupos dentro de los cuales hay una jerarquización de relaciones. Es posible, sin embargo, considerar como tipos a uno de los dos integrantes

del par recíproco o a un promedio de ambos. Se podrían obtener agrupaciones distintas si se varía el tamaño de la muestra, obligando a una continua re-clasificación.

Se objeta del método que en algunos casos, el mayor coeficiente de correlación de alguna variable puede ser inferior a un valor significativo, incluyéndose igualmente en alguno de los grupos cuando, en realidad, es independiente de todos. Para evitar esto los autores tomaron un coeficiente de correlación como límite y si los valores para alguna variable estaban debajo de este mínimo, se la consideraba no agrupada e independiente.

Los casos son agrupados en forma precisa, una variable sólo puede pertenecer a un grupo. Los mismos tienen una gran homogeneidad interna ya que los elementos que lo integran poseen dentro de él su máxima correlación y una relación menor con cualquier elemento de otro grupo.

Dentro de cada grupo hay una jerarquización y el número de ellos en una muestra es único. En este trabajo para el modo-T, 186 días como variables, quedaron definidos 31 grupos de dos o más integrantes y 6 casos quedaron sin clasificar, para valores de correlación mayores que 0.7. Según McQuitty (1957) los grupos así constituidos tienden a ser independientes, pero se pudo ver en este caso que los resultados no apoyaron esa inferencia. La independencia de los grupos dependerá de las características de la muestra. En el presente caso se encontró que dos o más grupos podrían a su vez reagruparse. Los máximos valores de correlación que determinan que las variables pertenezcan a un grupo, no impiden que esas mismas variables se correlacionen bien con los integrantes de otros grupos. Se puede concluir que, en general, este procedimiento genera mayor cantidad de "familias" que las existentes en la realidad.

Se puede obtener, una vez formados los grupos, el porcentaje relativo de casos en cada uno de ellos, pero no la varianza explicada. El método es de aplicación simple y rápida y una vez realizada la matriz de asociación, la elaboración no requiere de cálculos complicados.

3.c Análisis factorial: el análisis por componentes principales, autovectores y autovalores, o de funciones ortogonales empíricas, fue ampliamente utilizado, en especial en los últimos años, para tipificar campos de variables meteorológicas continuas.

El método genera modelos teóricos y señala para cada paso de la muestra, cuanto explica el modelo correspondiente. En apretada síntesis, los pasos que se siguen son:

- i. Formación de la matriz de correlación (o de algún otro coeficiente de asociación).
- ii. Transformación de esta matriz en una matriz de factores a partir de la resolución de la siguiente ecuación:

$$\det | X - \lambda I | = 0$$

X es una matriz de $N \times N$

N = número de variables

con $i = 1, 2, \dots, N$

λ_i son los autovalores de la matriz

I es la matriz identidad

Los autovalores o factor loadings son las raíces de la ecuación polinomial correspondiente.

iii. De los autovalores se obtienen los autovectores por medio de:

$$X(\vec{F}_i) = \lambda_i(\vec{F}_i) \quad \vec{F}_i \text{ son los } i \text{ factores}$$

Obteniendo tantos factores como autovalores (o variables) se tenga.

iv. Resolviendo el sistema de ecuaciones dados por el paso iii se encuentran los multiplicadores o factor score, que expresan una representación de las variables originales en las nuevas componentes.

$$X = F \cdot M \quad M \text{ matriz de multiplicadores}$$

v. Según el modo usado, los factores brindan los prototipos o modelos.

Para los modos Q, O y T (Fig. 1) los factor scores son los que permiten la construcción de los tipos y los factor loadings analizar cómo esos tipos se manifiestan en el tiempo (modos T y O) o en el espacio (modo-Q). Para los modos R, P y S (Fig. 1), los factor loadings (o autovectores) permiten obtener los prototipos y los factor scores (o multiplicadores) ver cómo se manifiestan en el tiempo (modos P y S) o cómo se distribuyen los grupos de variables en el espacio (modo-R).

vi. Los autovalores dan una medida de la varianza que explica cada modelo.

Características del análisis por componentes principales: en este caso los tipos que están dados por los factores, son modelos teóricos. Los mismos no varían apreciablemente al agregar o quitar algunos casos de la muestra si la misma es representativa. Los modelos podrían no coincidir totalmente con una situación observada, aunque se puede ver que para el modo-T los tipos obtenidos por Compagnucci y Vargas (1984b) para 186 días de julio, se parecen en gran medida a casos reales. Los prototipos resumen la estructura común de la mayor cantidad de situaciones de la muestra, son ortogonales e independientes.

La cantidad de tipos teóricos significativos, que no representan ruido, se puede determinar a partir del valor de los autovalores mediante la aplicación de diversos tests, como lo son el diagrama de LEV descripto y analizado por Cattell (1966) y de Kaiser (1958).

Una vez determinados los modelos teóricos es posible generar los grupos como describieron Compagnucci y Vargas (1984b). La interpretación de la pertenencia o no de un caso a un grupo determinado, cuando no se tiene como resultado una estructura simple, es difícil. En otras palabras esto ocurre cuando algunos casos se pueden representar como una combinación de varios modelos. Es posible, sin em

bargo, obtener a veces una estructura simple mediante la aplicación posterior de rotación ortogonal u oblicua aunque se pierde la ortogonalidad de los modelos en esta última.

Con esta metodología es posible calcular el porcentaje de casos contenidos en cada grupo, además de la varianza explicada por cada modelo. El método es totalmente objetivo en cuanto a la construcción de los modelos, aunque tiene algún grado de subjetividad la asignación de algunos casos a los grupos (Compagnucci y Vargas, 1984b).

El caso de la muestra de campos béricos de julio de 1972 a 1977, para el modo -T, dio como resultado 11 modelos teóricos significativos según el test de Kaiser, 7 o 9 modelos según el diagrama de LEV y los "factors loading" mostraron que sólo 8 modelos tenían peso significativo en más de un día. En Compagnucci y Vargas (1984b) se pueden ver los modelos obtenidos y los grupos representados por ellos, los cuales dan un total de 16.

Esta metodología requiere un considerable tiempo de computación y la cantidad de memoria necesaria aumenta en forma cuadrática con el número de variables a considerar.

4. DISCUSION COMPARATIVA Y CONCLUSIONES

Los tres métodos empleados, Lund, McQuitty y componentes principales, dieron distintos resultados, tanto para el modo-T en los tipos de campos béricos, como para el modo-S en la regionalización del cono sur de Sudamérica para las variaciones de presión de superficie en el mes de julio, este último en Compagnucci y Vargas (1984a).

En la Fig. 3 se muestra un esquema de los tipos asignados a cada uno de los días según el método de Lund y componentes principales. Se puede ver que el tipo uno de componentes principales corresponde aproximadamente bien al tipo A de Lund. El tipo 2 de CPs se asemeja al tipo B de Lund y el tipo 3 ya difiere del tipo C, siendo los demás grupos distintos en un caso y otro. No se presenta el resultado de McQuitty, sólo se muestra el agrupamiento que mayor número de casos tiene. Se puede ver aquí la gran diferencia con el tipo 1 CPs y el A de Lund. Esta discrepancia aumentó en los restantes agrupamientos.

En la Fig. 4 sólo se presenta la zonalización producida por la aplicación de McQuitty y Lund, ya que la de CPs, es igual a la obtenida mediante Lund, Compagnucci y Vargas (1984a). Se puede constatar que las divisiones producidas por estos dos métodos son diferentes. En general el método de Lund para valores de tolerancia de 0.7 o menores, "subclasifica" o da el mismo número de clases que CPs y para mayor valor de tolerancia, deja muchos días sin clasificar. Por su parte, McQuitty "sobre-clasifica" en todos los casos.

Los métodos de Lund, Blasing y McQuitty no necesariamente generan grupos que sean independientes entre sí. Mientras que los tipos obtenidos por CPs no-rotación

DIAS	1972	1973	1974	1975	1976	1977	DIAS	1972	1973	1974	1975	1976	1977
1	1-15	1	1	10-14	1	1	1	B	A	A	C	A	A
2	2	1	1	4	1	1	2	P	A	A	C	A	A
3	3	1	1	4	1	1	3	A	A	A	C	A	A
4	4	3	1	3	1	1	4	C	B	A	A	A	A
5	4	3	1	1	1	1	5	C	B	A	C	A	A
6	5	1	1	1	1-4	1	6	A	A	A	A	A	A
7	7	1	1	1	1	1	7	F	A	A	C	A	A
8	11	1	1	1	1	1	8	F	C	A	C	B	A
9	1-7	1	1	1	1	1	9	A	C	A	C	B	A
10	1	1	1	1	1	8-13	10	A	C	A	A	A	A
11	1	1	1	3	1	8	11	C	A	A	A	A	A
12	2	5	1	1	1	8	12	A	A	A	A	A	A
13	1	1	1	4	1	10	13	A	B	A	A	A	A
14	1	3-16	1	5	1	10-13	14	C	A	A	C	A	A
15	2	2	1	5-16	1	3-5	15	F	D	A	E	A	A
16	1	3	1	3	1	2	16	A	A	A	B	A	E
17	1	3	1	1	1	11	17	A	D	A	A	C	F
18	1	3-6	1	1	1	6-13	18	A	B	A	A	A	B
19	9	6	1	1	1	6	19	A	A	A	A	A	A
20	9	3	3	1	1	8-10	20	A	D	E	A	A	A
21	1	10	3	1	1	8-13	21	A	D	B	A	A	C
22	1	7	3	1	1	4	22	A	A	B	A	A	G
23	1	1	3	1	1	5	23	A	A	B	A	A	G
24	1	1	1	3-5	1	4	24	A	A	B	E	A	A
25	1	1	1-10	3-7	1	4	25	A	A	A	B	A	C
26	1	1	1	6	1	4-12	26	C	A	A	A	A	C
27	4	1	1	1	1	12	27	C	A	A	A	A	C
28	14	1	1	1	1	4-8-16	28	C	A	A	C	A	C
29	14	1	1	1	3	5	29	A	A	A	C	C	C
30	8	9	1	1	3	10	30	A	A	A	A	C	A
31	8	1	1	1-10	3	1	31	A	A	A	A	C	A

Clasificación por componentes principales Clasificación según método de Lund



Fig. 3: Tipos obtenidos para los campos b̄aricos de superficie de 12 Z mes de julio, periodo 1972-1977.

das o rotadas ortogonalmente, son independientes. La falta de ortogonalidad entre los modelos puede ser una desventaja si posteriormente se desea realizar sobre ellos otros análisis estadísticos. Las técnicas que dan modelos ortogonales, permiten en cambio, que los resultados se puedan usar en otros análisis como por ejemplo, en la construcción de ecuaciones de predicción estadística. A su vez, el obligar a los modelos de respuesta a ser ortogonales puede ser una desventaja, si la naturaleza de los datos no se adaptan en realidad a ellos. Esta situación es la que hace necesaria en algunos casos una rotación oblicua para permitir que los resultados se correspondan con modelos físicos reconocibles. Esto no fue necesario en la muestra de 186 días de julio dado que los modelos resultantes se asimilaron satisfactoriamente a casos reales (Compagnucci y Vargas, 1984b). A veces CPs da como resultado, aun rotando oblicuamente, una estructura no-simple donde algunos casos son combinaciones de varios modelos. Lund, McQuitty y Blasing no presentan este inconveniente, ya que generan grupos donde no existen situaciones ambiguas. Aquí los casos se pueden representar por un único tipo o modelo.

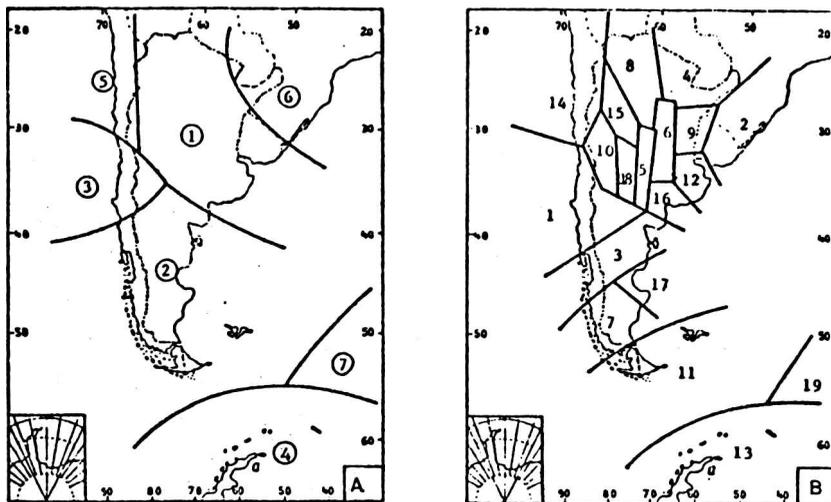


Fig. 4: Zonalización del campo de presión de superficie, mes de julio para el período 1972-1977, método de Lund (A) y método de McQuitty (B).

Las técnicas de Lund y Blasing contienen cierto grado de subjetividad en cuanto a la decisión del nivel de tolerancia y en Blasing, con respecto al número de casos a tomar para hacer el prototipo. McQuitty y CPs son, en cambio, técnicas objetivas.

Los modelos en el caso de Lund y McQuitty son casos reales de la muestra y varían en función del tamaño muestral. Blasing brinda un modelo cuasi-teórico que al ser un promedio de casos reales es bastante más "estable". Por su parte, CPs define modelos teóricos abstractos y "estables" frente a la variación del tamaño muestral. Estas asimismo permiten, al contener las características más sobresalientes de la muestra, obtener prototipos sinópticos útiles aún para comparar con casos ajenos a la muestra inicial. Se debe tener en cuenta, que en el caso de realizar estudios climáticos, es necesario tomar una muestra suficientemente extensa, de al menos 10 años.

A partir de las conclusiones anteriores se ve como más apto para la tipificación de situaciones sinópticas, el método de CPs, aún a pesar de la complejidad de su cálculo. En el caso de desear realizar una regionalización se piensa que sería necesario completar el análisis usando también Lund. Este último permite trazar límites más precisos y CPs explican mejor la varianza en cada área y la relación entre ésta y las otras de la zona de estudio.

Agradecimientos: los autores agradecen al Dr. M. Richman por sus valiosas sugerencias y comentarios, y al personal del Centro de Cómputos del CRICYT por su colaboración.

BIBLIOGRAFIA

- Barry, R.G. y A.H. Perry, 1973: Synoptic climatology Methods and applications; Methuen & Coltd., p.555.
- Blasing, T.J., 1975: A comparison of map-pattern correlations and principal components eigenvector methods for analyzing climatic anomaly patterns; Am. Met.Soc., Preprints "5th Conference Probability and Statistics", pp.96-101.
- Cattell, R.B., 1966: The screen test for the number of factors. Multivariate Behavioral Res., 1, 245-276.
- Compagnucci R.H. y W.M. Vargas, 1984a: Regionalización del campo de presión del mes de julio para la parte austral de Sudamérica; presentado en la 13° Reun. Cient. Arg. de Geofísica y Geodesia., Bariloche, Argentina (a publicar).
- _____, 1984b: Tipificación de los campos béricos de superficie para julio 1972 a 1977. Análisis por componentes principales no-rotadas; presentado en la 13° Reun. Cient. Arg. de Geof. y Geod. , Bariloche, Argentina (a publicar).
- Green, P.E., 1978: Analyzing multivariate data ; The Dryden Press, p.519.
- Hartranft, F.R., J.S. Restivo y R.C. Sabin, 1970: Computerized map typing procedures and their applications in the development of forecasting aids; 4th Weath. Wing, Aerospace Sci. Div., Tech. Paper 70-2, Ent. Air Force Base, Colo., 57pp.
- Kaiser, 1959: The application of electronic computers to factor analysis; Symp. on the application of computer to psychological problem. Am. Psy. Ass.
- Lund, 1963: Map-pattern classification methods; Journ.App.Met. 2 (1), 56-65.
- McQuitty, 1957: Elementary linkage analysis for orthogonal and oblique types and typal relevancies; Ed. Psy. Measurement, 17, p.207-229.
- Racine, J.B. y H. Reymond, 1973: L' analyse quantitative en géographie; Presses Univ. de France, pp.145-162.
- Vargas, W.M. y R.H. Compagnucci, 1982: Factibilidad de análisis de campos de presión en América del Sur mediante el uso de autovectores y autovalores; IV Congremet, Buenos Aires, Argentina (en prensa en la revista Meteorológica)
- _____, 1983: Tipificación de campos de presión usando autovectores y autovalores; Preprints 1st Int. Conf. on Southern Hemisphere Meteorology, Am. Met. Soc., pp.206-209.