

# CLASIFICACIÓN DE ESTRUCTURAS EN LA CIRCULACIÓN GENERAL ATMOSFÉRICA

P. N. Neytchev<sup>1</sup>

C. I. Christov<sup>1,2</sup>

Kh. P. Khristov<sup>1</sup>

N. M. Neykov<sup>1</sup>

L. Latinov<sup>1</sup>

<sup>1</sup> (Instituto Nacional de Meteorología e Hidrología, Academia de Ciencias de Bulgaria)

<sup>2</sup> (Servicio de Predicción Numérica, -INM-)

## RESUMEN

*El problema de la clasificación automática de procesos sinópticos (PS) en la 'región de Europa-Atlántico Norte es abordado mediante una combinación de análisis discriminante y análisis "cluster". Los resultados de una clasificación empírica con seis clases principales son examinados y se encuentra que sólo alrededor del 50% de los casos son reconocidos utilizando análisis discriminante estándar. Las clases empíricas originales son divididas en subclases mediante "K-mean cluster analysis" aplicado a cada grupo, y como resultado la calidad del esquema de clasificación es mejorado. El nuevo catálogo consiste en diecisiete clases y permite la clasificación correcta de modo automático de alrededor del 78% de los PS.*

### 1. Introducción.

Una parte significativa de la circulación general atmosférica se manifiesta a través de PS suaves y relativamente estables. Éstos tienen una organización temporal y unos movimientos de aire espacialmente localizados que tienen una influencia profunda en el tiempo a escala local. Por esto el problema de la detección y reconocimiento del tipo de un PS ha atraído una atención significativa (ver Court, 1957; Katz, 1960). Actualmente se dispone de clasificaciones empíricas de este tipo en varios países. Como puede comprenderse, ni la presión, ni la temperatura, ni ningún otro campo meteorológico puede describir por sí sólo una

situación sinóptica. No obstante, suele utilizarse el geopotencial de la superficie de 500 hPa ya que las estructuras son más estables. Una de estas clasificaciones es debida a Katz (Katz, 1960) y en ella distingue entre PS meridionales y zonales. En (Lund, 1963) son consideradas las estructuras en el mapa de presión al nivel del mar.

El problema de la objetividad en las distintas clasificaciones empíricas resulta de crucial importancia. Aunque no se dispone todavía de un procedimiento objetivo de clasificación de estructuras totalmente automático, los esquemas empíricos pueden ser mejorados utilizando los métodos estadísticos de análisis discriminante, análisis fac-

torial y análisis "cluster". Recientemente, diversos autores han abordado el problema de la objetivización mediante análisis discriminante y/o análisis "cluster", este último aplicado directamente a los datos originales o a los componentes principales (Kirchhofer, 1973; Kruizinga, 1979; Kozulin y Reitenbakh, 1984; McCutchan y Schroeder, 1973).

Consideraremos aquí una adaptación (ver Andreev y otros, 1978) de la clasificación de Katz (Katz, 1960) para la región donde se originan los sistemas atmosféricos que afectan a Bulgaria. Mostramos que la clasificación empírica puede ser mejorada de forma que el 80% de las estructuras de los PS pueden ser correctamente reconocidas.

## 2. Los datos.

Consideraremos un intervalo de 3.650 días desde el 04/01/1978 al 01/01/1988. Se toman los datos diarios (00 GMT) del archivo de análisis objetivos del CEPPM (Centro Europeo de Predicción a Plazo Medio) para el geopotencial en 500 hPa. El campo de geopotencial es considerado en una malla de 65 puntos en la región [60°W, 60°E] x [30°N, 80°N] (Fig. 1.a). Por conveniencia en los cálculos restamos 500 dmgp al campo de geopotencial.

Un importante rasgo en los datos considerados es la presencia de tendencias asociadas a las variaciones estacional y latitudinal. La superficie isobárica se encuentra más alejada de la superficie terrestre en el ecuador que en la proximidad de los polos, y se eleva como un todo durante el verano. Una importante característica de la circulación general, el frente polar, también cambia de posición e intensidad estacionalmente (Vorobyov, 1967). Numéricamente las tendencias son mucho mayores que la variancia, que define la circulación sinóptica. Por esta razón los datos son promediados zonal y mensualmente y los resultados para la tendencia son presentados en la tabla I.

Se ve que los promedios mensuales están muy cerca de las respectivas isohipsas estacionales (Khrabrov, 1957), lo cual sirve para marcar la zona frontal planetaria. El promedio zonal reduce la información para esta zona y los datos de la tabla I dan sólo una idea vaga de la variancia de la zona frontal planetaria a lo largo del año. La media anual se presenta gráficamente en la Fig. 1.b.

La estructura espacio-temporal que se presenta en la tabla I sirve para centrar los datos en sentido estadístico. En cada uno de los 65 puntos se sustrae el valor del geopotencial medio para su mes y su latitud.

**Tabla I**  
**Promedios mensual y zonal del geopotencial para todos los datos y para los 3.650 días**

MES	LATITUD										TOTAL
	75	70	65	60	55	50	45	40	35	30	
1	13,0	15,2	18,8	24,1	31,9	40,7	50,3	58,9	65,4	70,1	35,3
2	16,3	20,0	24,3	28,7	34,7	40,4	47,5	55,7	62,7	68,9	37,0
3	18,3	20,7	24,3	29,6	37,0	44,4	51,9	59,1	65,1	70,2	39,0
4	25,2	28,4	33,0	39,1	46,3	52,2	57,2	62,2	68,5	73,9	46,1
5	37,4	40,3	44,1	48,4	54,2	59,3	64,3	69,2	75,2	80,0	54,9
6	44,6	48,0	51,8	55,5	60,6	66,3	72,4	77,6	83,0	87,1	62,3
7	52,4	54,3	56,8	60,2	65,4	71,8	78,2	83,5	87,9	90,8	67,8
8	49,3	51,5	54,4	58,7	64,6	71,1	77,8	83,5	88,2	91,1	66,5
9	38,7	41,3	44,7	49,4	56,8	65,7	74,3	80,5	85,3	88,2	59,4
10	28,1	31,0	35,4	41,4	49,8	58,6	67,2	73,8	79,2	82,5	51,5
11	18,6	21,8	26,2	32,6	41,8	51,3	60,0	66,9	72,3	76,4	43,4
12	15,4	17,3	20,6	25,7	33,9	43,0	52,7	61,5	68,4	73,4	37,6
1-12	29,9	32,6	36,3	41,2	48,2	55,5	62,9	69,5	75,2	79,4	50,1

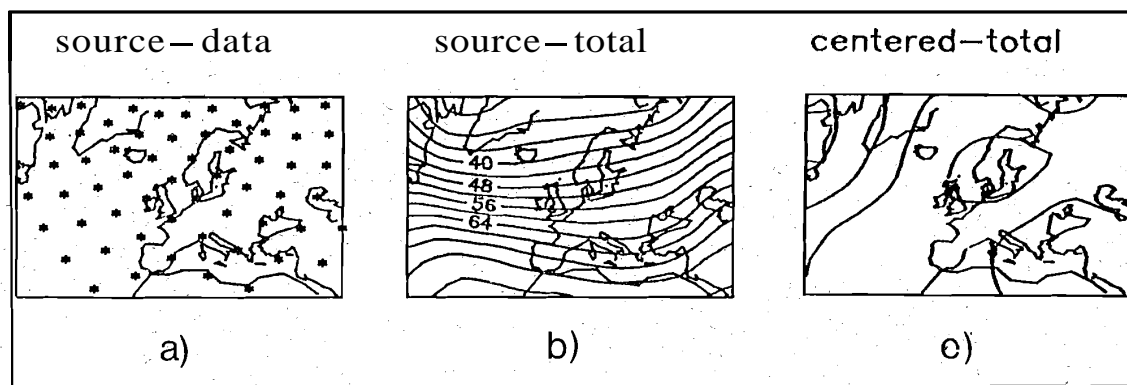


Figura 1

### 3. Los procesos sinópticos.

Consideremos la clasificación sinóptica empírica de (*Andreev y otros, 1978*) que define en el período de los 3.650 días que tenemos en cuenta un total de 778 PS, separados en seis clases principales, digamos, A, B, C, D, E y F. Las cuatro primeras clases son de tipo meridional, F es de tipo zonal y E de tipo mixto. En los tipos meridionales distinguimos dorsales en el geopotencial. Cuando la dorsal es aproximadamente paralela al meridiano  $30^{\circ}\text{W}$  el PS es denominado "dorsal sobre Groenlandia" o de tipo A. Cuando la dorsal está sobre el meridiano de Greenwich, la llamamos "dorsal Occidental" o tipo B. La "dorsal Central" o tipo C está aproximadamente alineada con  $30^{\circ}\text{E}$  y la "dorsal oriental" o tipo D con  $60^{\circ}\text{E}$ . Estos tipos son encontrados de acuerdo con los expertos en 83, 119, 106 y 100 PS, respectivamente. El tipo "Mixto" o tipo E, que tiene lugar cuando dos dorsales (o una "onda" de gran amplitud de la zona frontal planetaria) está presente, es encontrado en 34 de los 778 PS. Finalmente, el resto de los 336 casos son asignados al tipo llamado "Zonal" o F que se caracteriza por una "onda" de amplitud moderada en la zona frontal planetaria. Los nombres reflejan la situación geográfica respecto a Bulgaria.

Haciendo un promedio extendido a todos los días para los que se presenta un tipo particular obtenemos un retrato representativo de ese tipo (sería el "germen" de ese "cluster"). Entonces aplicamos análisis discriminante extendido a todos los días. Resulta que sólo el 50% de los casos son considerados de los tipos prescritos por la clasificación empírica. Este resultado insatisfactorio muestra que en el esquema de clasificación automática deben ser considerados algunos rasgos finos de los PS que el meteorólogo experto

considera intuitivamente. La forma de mejorar nuestra comprensión del esquema de clasificación se presenta en lo que sigue.

Tenemos que hacer notar aquí que con el fin de reducir el tiempo de cálculo, en los cálculos preliminares sólo consideramos 10 puntos de rejilla para construir las funciones discriminantes.

### 4. Variancia temporal y promediado.

La principal característica de los datos es su variancia temporal, en el sentido de que los primeros días de un determinado PS son similares a los últimos del PS que le precede mientras que los últimos días se parecen a los del proceso que le sigue. Considerar la secuencia completa de un PS particular creando un vector compuesto incrementaría la dimensión espacial 65 en un factor 6-8 (el número total de días en un período dado), lo que no es posible con la capacidad de los ordenadores de que disponemos. Por tanto, está claro que estamos limitados a considerar un conjunto de 778 objetos en el cual cada PS está representado por un solo día (vector de dimensión 65). Ahora el problema será cómo seleccionar la estructura representativa de un PS dado.

Discutiremos aquí cuatro tipos de promedio. El primer tipo consiste en la suma extendida a todos los días del PS dividida por el número de días del PS (la media aritmética). El segundo es la media, pero con pesos que son proporcionales al inverso de la norma del cambio total de un día a otro. En el tercer caso el valor del geopotencial en un punto particular toma el valor medio de los valores de los días durante los cuales los cambios son menores. El cuarto tipo consiste en tomar la mediana (se toma el día central

cuando el número de días en ese período es impar y la semisuma de los dos días centrales cuando es par).

El comportamiento de los distintos esquemas de promedio es valorado por medio de análisis discriminante de la misma forma que hemos hecho antes para los datos iniciales. Primero hacemos el promedio de las medianas de los procesos adscritos por los expertos a una de las clases mencionadas A, B, C, D, E y F. De

esta forma obtenemos un retrato representativo del grupo (el cual no tiene que coincidir necesariamente con el retrato obtenido anteriormente para los datos diarios). Las funciones discriminantes son calculadas y el conjunto de los 778 PS son separadas en seis grupos en las proximidades de los retratos representativos. Las estructuras que caen en el grupo prescrito por la clasificación empírica son consideradas como correctamente reconocidas y el porcentaje de estructuras reconocidas correctamente es tomado como una medida

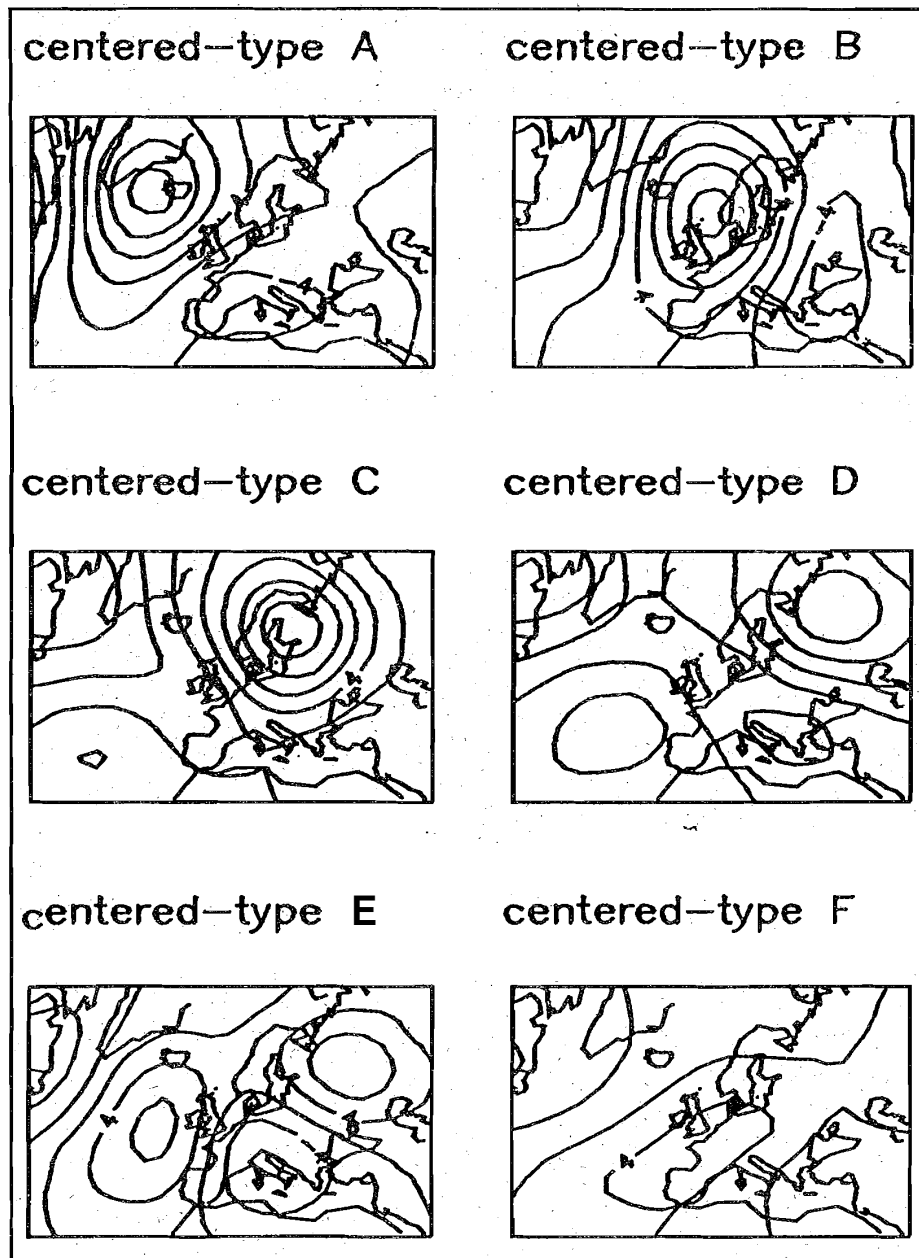


Figura 2

del comportamiento del procedimiento de promedio.

Se obtienen resultados similares para los cuatro esquemas de promediado, pero el procedimiento de la mediana tiene la ventaja de que manteniéndose en los límites del 60% de los casos correctamente reconocidos consigue una mejora relativa en el reconocimiento de los casos que en un principio tienen los menores porcentajes de casos reconocidos, es decir E y F. Esto propor-

ciona cierta homogeneidad a los datos y es la razón por la que tomamos las "medianas" como procedimiento de promedio en este trabajo.

En la Figura 2 se presentan los retratos promedio ("germen" del grupo) de las distintas clases. Visualmente son claramente distinguibles. Para poder comparar, las mismas clases pero sumándole las tendencias son presentadas en la Figura 3. Como es natural, las estructuras que incluyen las tendencias son más oscuras.

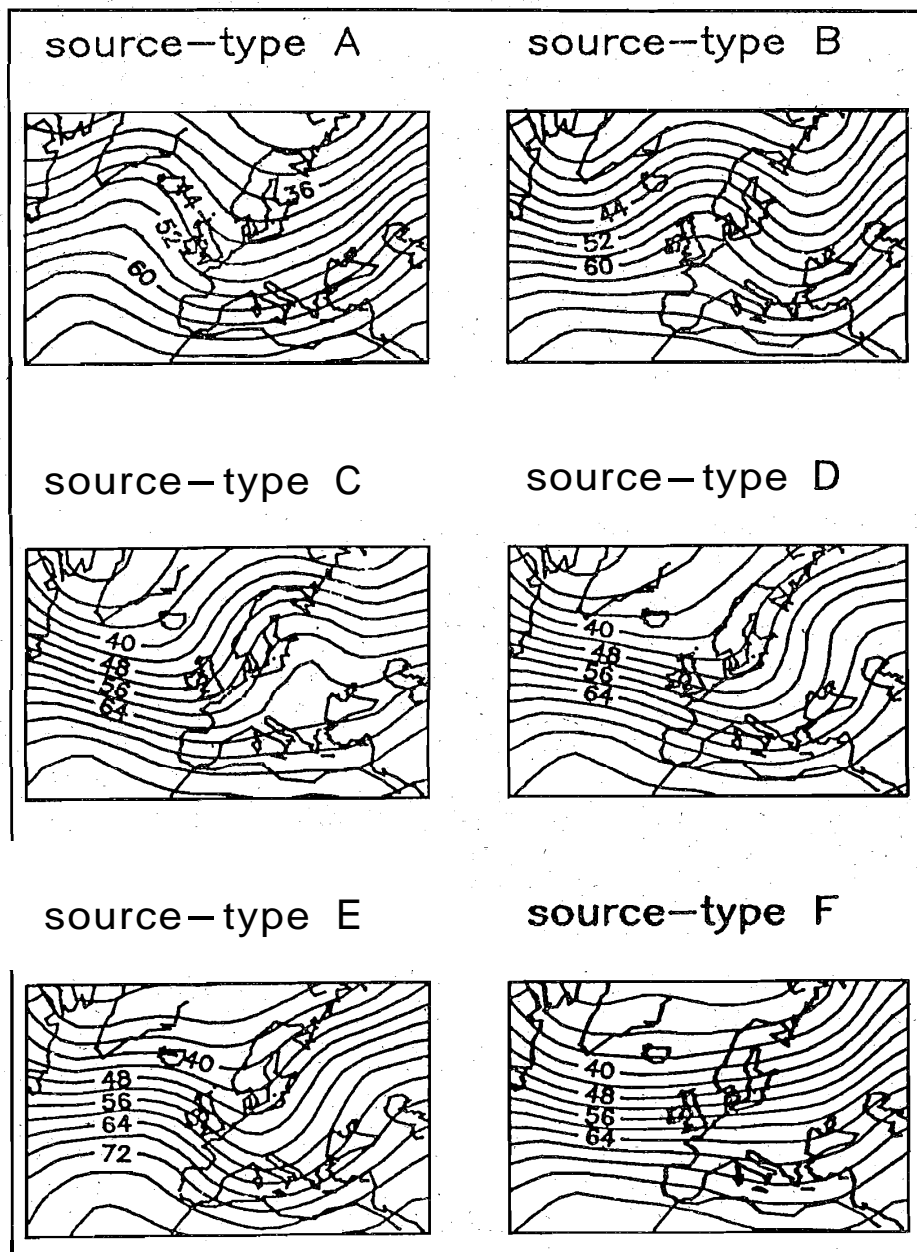


Figura 3

**Tabla II**  
**Clasificación automática de las medianas para 6 tipos en 10 puntos**

tipo	PS número	número de procesos clasificados como tipo:						correctam. clasificad. %
		A	B	C	D	E	F	
A	83	56	5	5	5	7	5	67,5
B	119	7	77	17	1	6	11	64,7
C	106	6	4	67	14	9	6	63,2
D	100	2	5	9	63	16	5	63,0
E	34	5	1	3	5	19	1	55,9
F	336	18	47	39	27	27	178	53,0
total	778	94	139	140	115	84	206	59,1

**5. Reducir las dimensiones de los datos mediante análisis factorial.**

Otra característica importante de los datos es su fuerte correlación espacial. La forma usual de mostrar esta correlación es calcular los auto-vectores (componentes principales) de la matriz de correlación. El análisis factorial reveló que 14 componentes principales (factores) explican el 88% de la dispersión de los datos (ver tabla III).

Es interesante hacer notar aquí resultados análogos a los que llegan otros autores, p. ej., en (Stone, 1989) se encuentra que entre 10 y 13 componentes explican aproximadamente el 85% de la variancia del tiempo en Brisbane, Australia entre 1967 y 1981. En cuanto al primer componente principal, no es sorprendente que explique una parte tan grande de la variancia total del campo. En (Klink y Willmott, 1989) se encuentra que el primer componente principal explica, en

**Tabla III**  
**Representación en factores de los datos iniciales**

Factor núm.	Parte de la dispersión	Parte acumulada de la dispersión	
		en los datos	en los factores
1	10,0501	0,1546	0,1760
2	8,8627	0,2910	0,3312
3	7,0757	0,3998	0,4551
4	5,4365	0,4835	0,5503
5	4,3115	0,5498	0,6258
6	3,7204	0,6070	0,6910
7	3,4240	0,6597	0,7509
8	3,2570	0,7098	0,8080
9	2,5574	0,7492	0,8528
10	2,3489	0,7853	0,8939
11	2,0494	0,8168	0,9298
12	1,5785	0,8411	0,9574
13	1,2876	0,8609	0,9800
14	1,1428	0,8785	1,0000

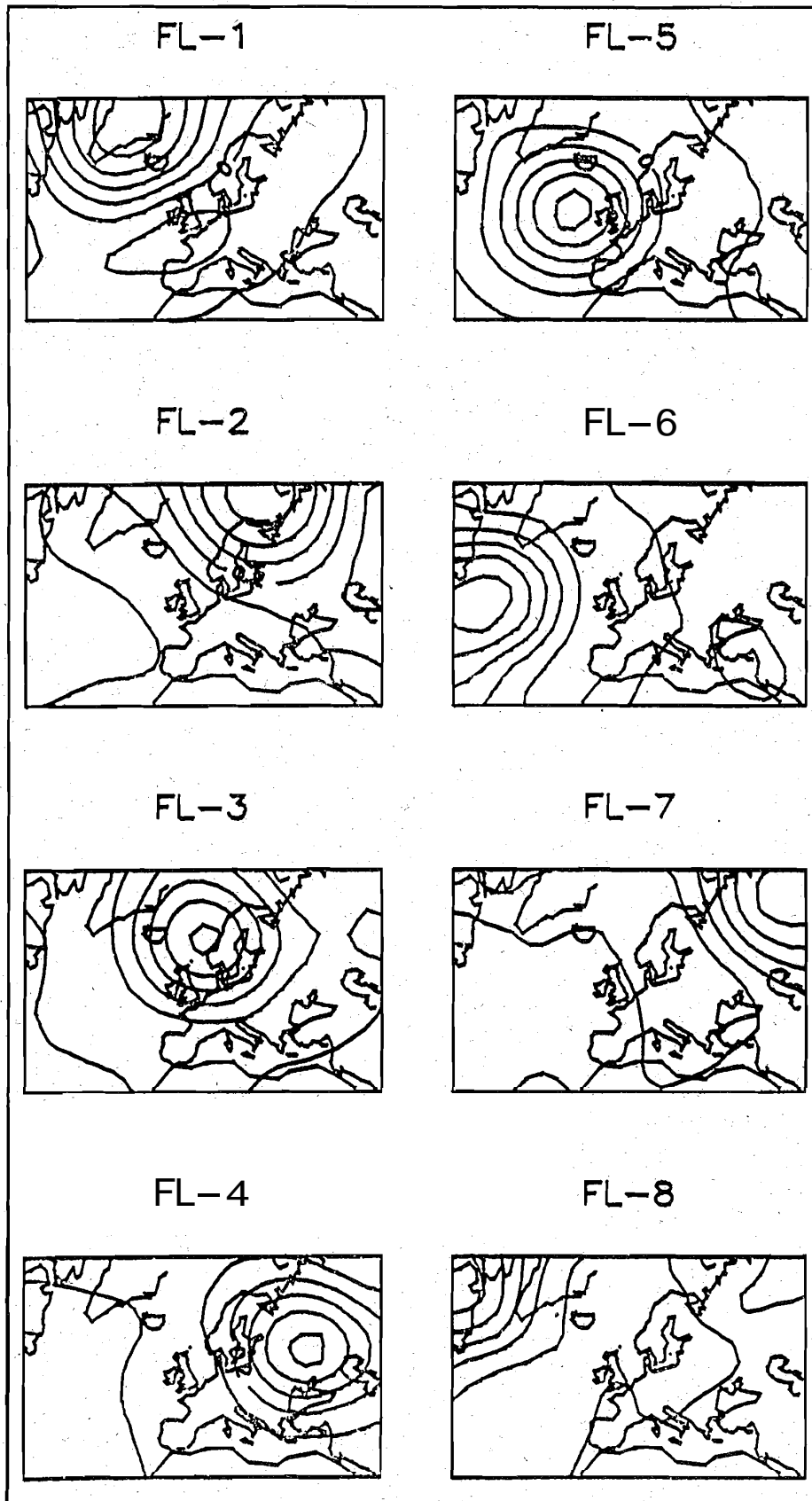


Figura 4

un problema similar, el 22% de la variancia. En nuestra opinión el "número mágico" de 10-15 factores, está conectado con la dimensión del atractor para los procesos macrosinópticos. Para un estudio como el nuestro es esencial poder considerar un número de factores como el mencionado porque esto permite filtrar las componentes del campo de alta frecuencia que no tienen significado para la escala macrosinóptica considerada. La fuerte correlación entre los valores del geopotencial dentro de un factor dado tiene una

interpretación clara (ver Fig. 4) y cada factor está conectado con una región físico geográfica concreta.

Para poder chequear las propiedades de la representación en factores del campo aplicamos análisis discriminante sobre los factores. La calidad de la clasificación es prácticamente la misma (tabla IV) lo que significa que con la representación en términos de componentes principales no se pierde información en cuanto a la clasificación.

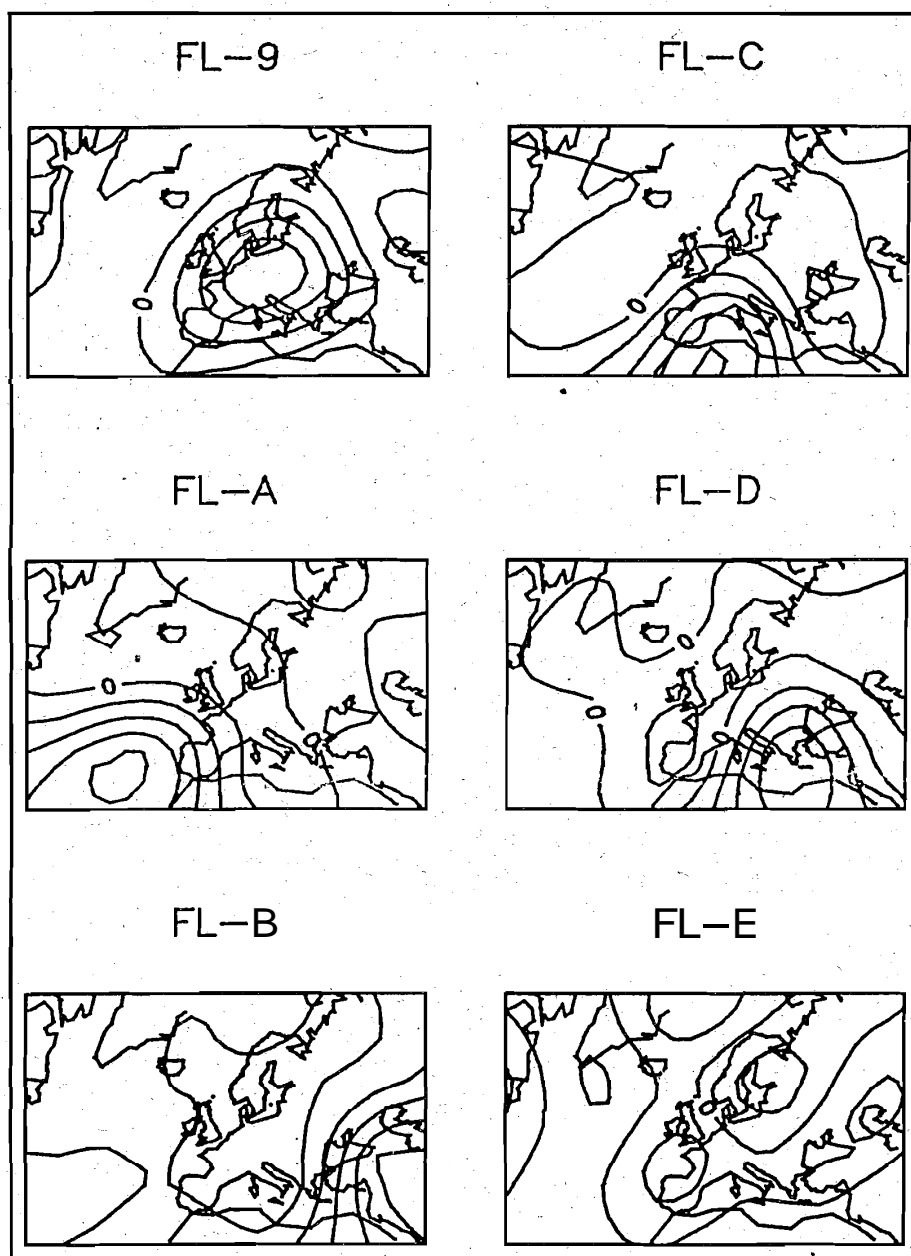


Figura 4 (continuación)



Tabla IV  
Clasificación de las medianas utilizando los factores

PS		número de procesos clasificados como tipo:						reconocidos correctamen.
tipo	número	A	B	C	D	E	F	%
A	83	59	5	3	4	8	4	71,1
B	119	7	81	16	1	3	11	68,1
C	106	1	7	73	10	8	7	68,9
D	100	1	4	8	59	20	8	59,0
E	34	2	2	3	8	19	0	55,9
F	336	20	43	37	36	30	170	50,6
total	778	90	142	140	118	88	200	59,3

#### 6. Catálogo detallado vía análisis discriminante pseudolineal.

De lo visto hasta aquí podemos concluir que con el sistema de seis tipos adoptado no podremos modificar más las funciones discriminantes de forma que la clasificación automática se aproxime a la clasificación empírica. Por tanto el primer objetivo del presente artículo ha sido alcanzado: podemos decir que dada la clasificación (Andreev y otros, 1978), los procedimientos automáticos de clasificación son sólo capaces de reconocer el 59-60% de los casos clasificados por los expertos. Esto puede ser valorado como un resultado muy bueno para una clasificación empírica, que está sujeta a ciertas decisiones subjetivas. Creemos que para el propósito de un esquema de clasificación robusto debemos de encontrar maneras de mejorar la calidad del esquema de clasificación.

La forma de mejorar la clasificación es separar las clases iniciales "groseras" en subclases que puedan ser mejor distinguidas por el análisis discriminante. Sin pretender generalidad, presentamos uno de los procedimientos a seguir.

Analizando los casos de clasificación errónea (ver tabla II) encontramos que la confusión tiene lugar principalmente en los casos tipo F para los cuales el máximo de la "onda" de la Zona Frontal Planetaria (ZFP) está desplazado 30° al Este o al Oeste. Está claro, que dependiendo de la amplitud y la fase de la onda de la ZFP, los respectivos procesos zonales pueden diferir considerablemente del retrato medio del tipo F es-

tando más próximos, en el sentido del análisis discriminante, a uno de los tipos meridionales. Para verificar esta conjetura aplicamos "K-mean cluster analysis" al conjunto de PS de tipo F. Establecemos 5 subclases que se corresponde con el número de tipos meridionales. Como resultado del análisis "cluster" obtenemos las subclases F1, F2, F3, F4 y F5 que comprenden 98, 48, 59, 62 y 69 PS respectivamente. La subdivisión obtenida dentro del tipo F es verificada por medio de análisis discriminante (ver tabla V).

En la tabla VI se muestra cómo afecta esta subdivisión a la calidad del esquema global de clasificación una vez incluidos los nuevos tipos. Podemos ver que se produce un aumento significativo del número de procesos zonales correctamente clasificados y un mejoramiento general del esquema de clasificación. Sin embargo se produce un empeoramiento en la clasificación de los tipos D y E, cuyo reconocimiento tampoco era suficientemente bueno con el catálogo original. Para el caso mixto E, la situación se agrava por la presencia de dos dorsales en el campo de geopotencial, de forma que dependiendo de la posición de la dorsal de mayor amplitud, el análisis discriminante puede confundir a uno de estos PS con uno de los tipos meridionales A, B, C o D. Esto puede comprenderse ya que el caso mixto E puede verse como una superposición de la "dorsal Oriental" (tipo A o B), el último aplicando un factor de escala. Esto significa que cuando la amplitud del último componente en el tipo mixto es suficientemente grande, el análisis discriminante puede confundirlo con uno de los casos de "dorsal Occidental". Lo contrario tam-

**Tabla V**  
**Verificación de la clasificación en subclases del tipo F**

PS		número de procesos clasificados como:					reconocidos correctamen.
tipo	número	F1	F2	F3	F4	F5	%
F1	98	96	0	0	2	0	98,0
F2	48	1	47	0	0	0	97,9
F3	59	1	0	57	0	1	96,6
F4	62	0	0	1	61	0	98,4
F5	69	0	0	2	0	67	97,1
total	336	98	47	60	63	68	97,6

**Tabla VI**  
**Calidad del esquema de clasificación cuando el tipo F es dividido en 5 subtipos**

PS		número de procesos clasificados como tipo:										reconocidos correctamen.
tipo	núm.	A	B	C	D	E	F1	F2	F3	F4	F5	%
A	83	51	1	5	6	6	3	4	2	4	1	61,4
B	119	4	67	13	1	2	5	0	11	8	8	56,3
C	106	2	6	57	7	6	10	1	8	7	2	53,8
D	100	2	3	6	47	17	9	4	1	6	5	47,0
E	34	3	1	2	5	16	2	0	2	3	0	47,1
F1	98	2	3	9	7	1	75	0	0	0	1	76,5
F2	48	2	1	0	2	1	0	42	0	0	0	87,5
F3	59	1	4	4	0	2	1	0	46	0	1	78,0
F4	62	1	4	3	2	4	0	0	1	47	0	75,8
F5	69	1	0	2	2	2	0	0	0	0	62	89,9
total	778	69	90	101	79	57	105	51	71	75	80	65,6

bién es verdad: cuando alguna perturbación está presente en la región este del tipo "dorsal Occidental", éste podría ser confundido con el tipo zonal si la amplitud de la onda de este último es suficientemente grande y viceversa.

Siguiendo en la misma línea de razonamiento, unimos los dos grupos originales D y E en uno sólo que llamamos tipo DE. Encontramos que la división óptima del nuevo tipo DE se encuentra al dividirlo en tres subclases, que

denominamos DEI, DE2, DE3 y que engloban a 57, 29 y 48 casos, respectivamente. Esta subdivisión se traduce en una nueva mejora del número total de casos correctamente reconocidos pero se manifiestan algunas deficiencias en la clasificación del tipo C. Por esto, lo separamos en subclases y el análisis discriminante muestra que, una vez más, el número óptimo de subclases es tres: C1, C2, C3 con 46, 33 y 27 casos, respectivamente. La calidad total del esquema vuelve a mejorar, y ahora las deficiencias se encuentran en

la clasificación del grupo B. De la misma manera descubrimos que la división óptima del grupo B es en B1, B2 y B3 con 31, 44 y 44 casos. El esquema de clasificación vuelve a mejorar, pero ahora es el tipo A el que presenta la necesidad de una nueva división. Una vez más la subdivisión óptima incluye tres subclases: A1, A2 y A3 con 19, 22 y 42 casos, respectivamente.

El promedio de conjunto dentro de cada uno de los 17 nuevos grupos define sus retratos representativos, los cuales se muestran en la

Figura 5. Comparando la Fig. 5 con la Fig. 3 se puede apreciar el significado sinóptico de la clasificación detallada. De hecho, ésta incluye los rasgos principales del esquema original pero proporciona una división más precisa dentro de un tipo dado. Es interesante destacar que las subclases obtenidas para los tipos meridionales A, B y C, coinciden bastante bien con las variantes del PS principal distinguidas por los expertos (ver *Andreev y otros, 1978*). Las subclases están conectadas con las diferentes orientaciones de las dorsales (ángulo formado con el meridiano). De

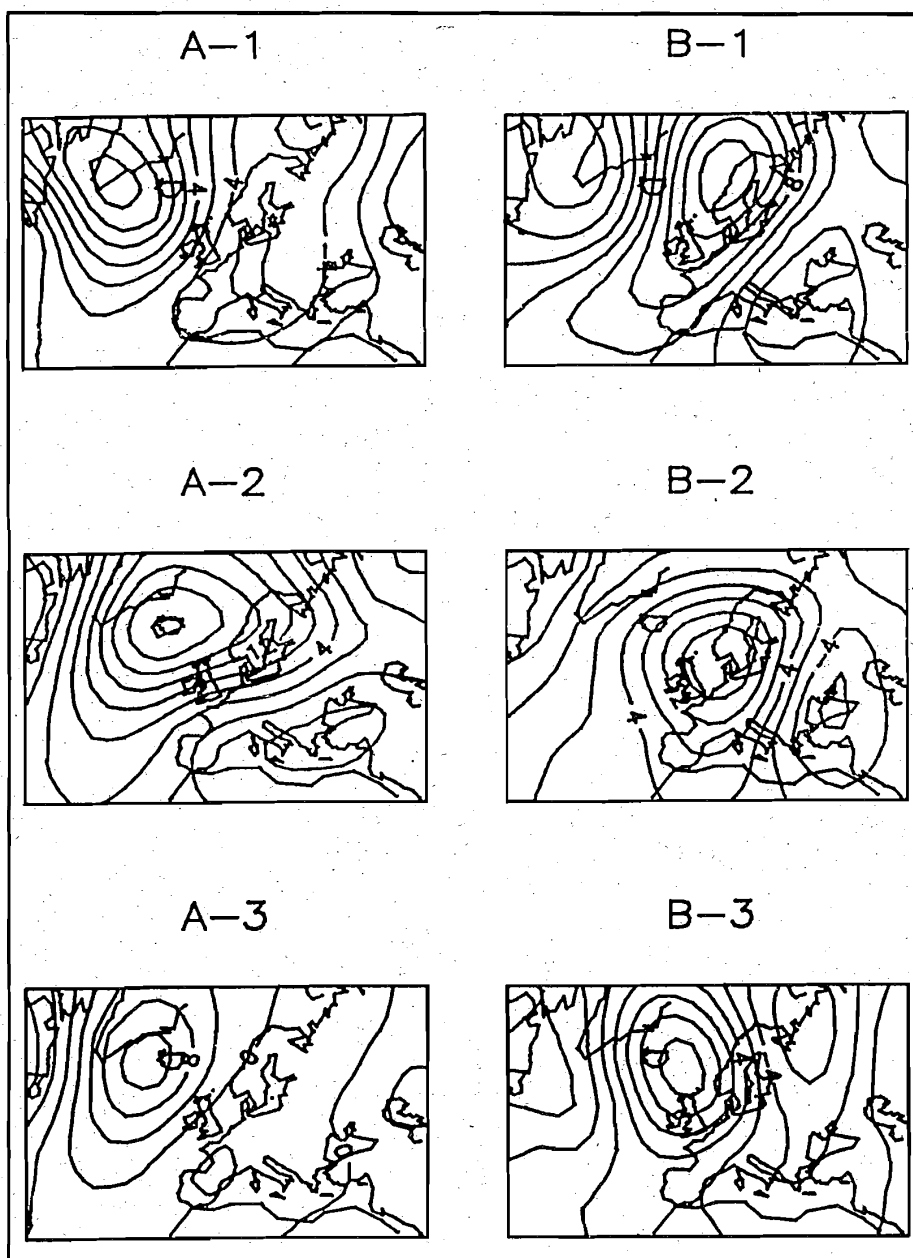


Figura 5

acuerdo con la terminología de (*Andreev y otros, 1978*) las clases pueden ser denominadas subtipos polar, simétrico y ultrapolar, respectivamente.

Para el tipo zonal F, nuestro catálogo enriquece significativamente el esquema de clasificación, y se aparta bastante de la clasificación empírica.

La tabla VII muestra, utilizando los valores que toman los factores, que el nuevo catálogo permite la clasificación automática de más del

69% de los casos. La mejora de la clasificación original es debida a que al introducir las nuevas subclases, estamos introduciendo cierta superposición lineal de las funciones discriminantes que aproxima a la función desconocida y generalmente no lineal.

La mejora del esquema en función de los componentes principales puede ser considerada cuantitativamente muy buena. Tenemos que recordar que hemos utilizado sólomente 14 factores lo cual se traduce en un suavizado de la estructura

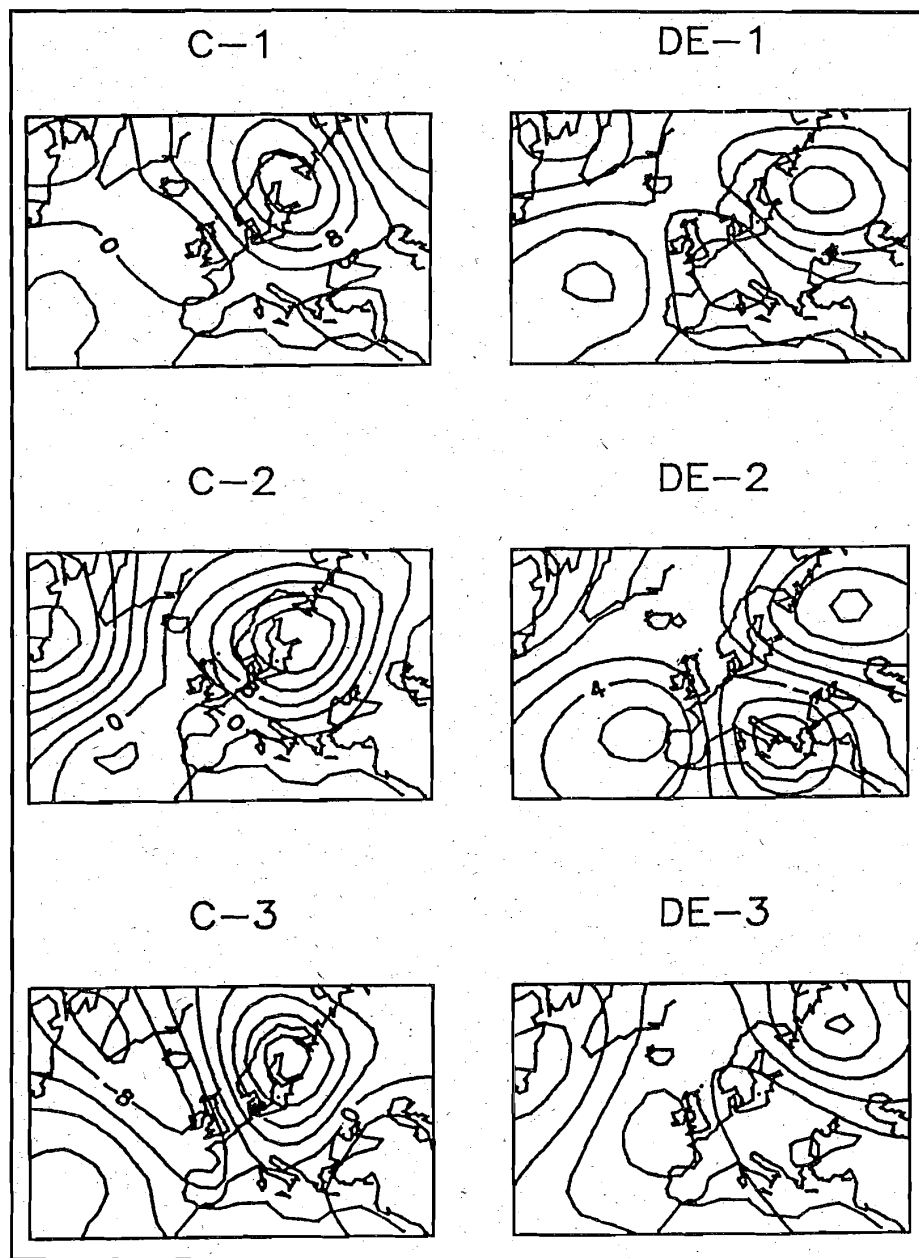


Figura 5 (continuación)

fin del campo que puede resultar importante para distinguir entre casos con diferencias más sutiles. Para poder valorar la influencia de esta estructura fina en la calidad del esquema de clasificación, volvemos a considerar los datos originales (de dimensión 65). Además, éste es el método a seguir cuando el esquema de reconocimiento de tipos es incorporado como parte de un sistema más general de clasificación automática (ver "A pattern recognition...", 1989). Por tanto, aplicamos análisis discriminante a los datos originales, pero con el nuevo catálogo de 17 tipos. Los resultados se

presentan en la tabla VIII. Se ve que la calidad del esquema aumenta hasta el 77,8%, y si no tenemos en cuenta como erróneos los casos en que el proceso adscribe un subtipo a otro de la misma clase, entonces la calidad del sistema supera el 80%.

Ahora podemos discutir los posibles errores en la clasificación original de los expertos, que no es muy eficiente incluso después de las mejoras introducidas en este trabajo. Como indicador de un error persistente puede servir el hecho de que

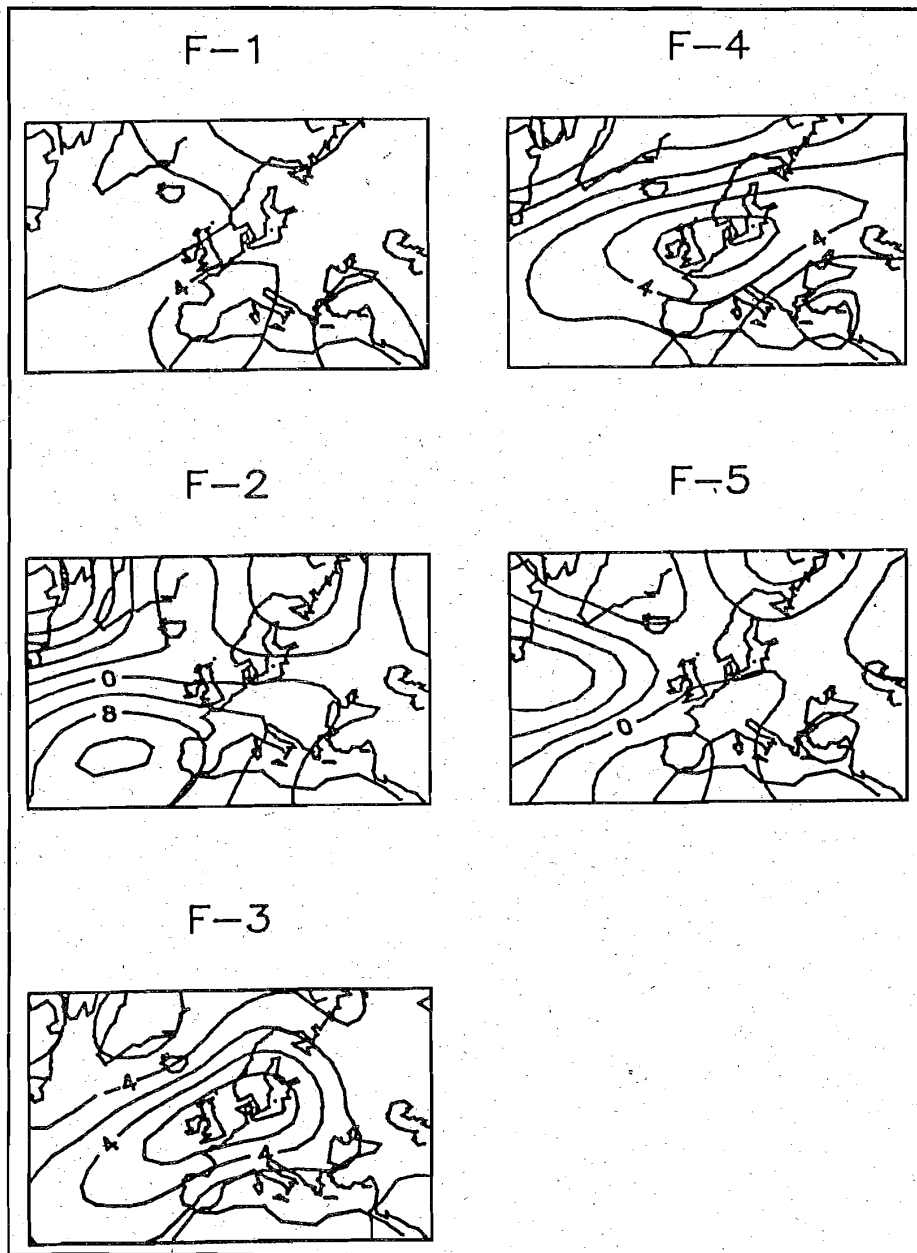


Figura 5 (continuación)

III SIMPOSIO NACIONAL DE PREDICCIÓN

**Tabla VII**  
**Catálogo detallado. Clasificación utilizando los factores**

tipo	número de PS clasificados como tipo:													correctamente reconocidos:					%
	A1	A2	A3	B1	B2	B3	C1	C2	C3	DE1	DE2	DE3	F1	F2	F3	F4	F5		
A1	15	0	0	0	0	2	0	0	0	1	0	0	0	0	1	0	0	78	
A2	0	17	0	1	0	2	1	0	0	0	0	0	0	1	0	0	0	77	
A3	1	0	27	0	0	2	1	1	0	3	0	1	2	2	1	1	0	64	
B1	0	1	0	23	0	0	2	2	2	0	0	0	1	0	0	0	0	74	
B2	0	1	0	1	32	0	0	2	0	0	1	0	0	0	1	4	2	72	
B3	0	1	2	0	0	30	0	0	1	1	0	1	1	0	2	2	3	68	
C1	0	1	1	4	1	0	32	0	0	3	1	1	1	0	1	0	0	69	
C2	0	0	2	2	0	1	1	24	0	0	0	0	0	0	0	2	1	72	
C3	0	0	0	0	2	0	0	0	18	4	1	0	0	1	0	1	0	66	
DE1	2	3	1	0	1	1	4	3	2	30	0	0	0	3	1	4	2	52	
DE2	0	0	0	1	0	0	0	0	1	0	24	0	0	0	0	3	0	82	
DE3	0	1	0	0	3	0	0	2	1	0	0	33	6	0	2	0	0	68	
F1	1	0	2	1	1	3	14	0	4	6	1	10	54	0	0	0	1	55	
F2	0	0	1	0	0	3	0	0	1	1	0	1	0	41	0	0	0	85	
F3	0	0	1	2	1	1	1	4	0	2	2	1	1	0	43	0	0	72	
F4	0	0	1	0	2	2	0	3	0	2	2	1	0	0	2	46	1	74	
F5	4	0	0	0	0	0	0	2	4	1	1	3	1	0	0	1	52	75	
Total																	69,6		

**Tabla VIII**  
**Clasificación de los datos originales en los 65 puntos y para el catálogo detallado**

tipo	número de PS clasificados como tipo:													correctamente reconocidos:					%
	A1	A2	A3	B1	B2	B3	C1	C2	C3	DE1	DE2	DE3	F1	M	F3	F4	F5		
A1	17	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	89	
A2	0	18	0	1	0	1	1	0	0	0	0	0	0	1	0	0	0	81	
A3	0	0	31	0	0	0	1	0	0	0	0	3	2	2	0	2	1	73	
B1	0	0	0	28	0	1	0	0	0	0	0	0	1	0	0	1	0	90	
B2	0	0	0	0	38	0	0	1	1	0	0	0	0	0	0	2	2	86	
B3	0	1	0	0	0	32	0	0	0	3	0	1	1	1	1	2	2	72	
C1	0	1	0	2	0	0	36	0	0	2	0	0	3	0	2	0	0	78	
C2	0	0	1	1	0	2	0	26	1	0	0	0	0	0	0	1	1	78	
C3	0	0	1	0	0	0	0	0	21	3	0	0	1	1	0	0	0	77	
DE1	0	2	0	0	1	2	2	0	1	39	1	1	0	2	3	2	1	68	
DE2	0	0	0	0	0	0	0	1	0	0	25	0	0	0	1	2	0	86	
DE3	1	0	0	0	2	2	0	2	0	1	0	36	4	0	0	0	0	75	
F1	0	0	3	3	0	2	8	0	1	3	0	7	65	0	1	1	4	66	
M	0	0	1	0	0	1	0	0	0	0	0	0	0	46	0	0	0	95	
F3	0	0	0	2	2	1	1	6	1	2	1	0	0	0	40	3	0	67	
F4	0	0	3	0	2	2	0	3	0	1	1	1	0	0	0	47	2	75	
F5	2	0	0	0	1	1	1	0	1	0	0	2	0	0	0	1	60	86	
Total																	77,8		

hay grupos de procesos de una subclase que son reconocidos como pertenecientes a una subclase de otro de los tipos principales. Por ejemplo, hay 8 casos F1 reconocidos como pertenecientes a la subclase C1, 7 PS F1 reconocidos como DE3 y 6 procesos de la subclase F3 que son considerados como C3. Un camino posible para mejorar la calidad del esquema sería el realizar una iteración, p. ej., considerar el nuevo catálogo con 17 tipos como una cierta clasificación experta y repetir todos los pasos descritos en este artículo. Creemos que después de 2-3 iteraciones el proceso convergerá y un cierto esquema objetivo podrá ser conseguido.

Para terminar es instructivo dar la distribución mensual de los diferentes tipos en el nuevo catálogo. En función de la estación en la que muestran preferencia en aparecer, pueden ser divididos, a groso modo, en tres grupos (ver tabla IX). Al primer grupo pertenecerían aquellos procesos que aparecen sobre todo en el período invierno-primavera, aparecen menos en otoño y casi nunca en verano. Este grupo comprende los tipos A1, A2, B1, C2, DE2 y F2. El segundo grupo comprende los tipos C1, F1, F3 y F4 que muestran preferencia a aparecer en verano y otoño

y evitan la mitad fría del año. El resto de los tipos forman el tercer grupo y se presentan en cualquier época del año.

**Conclusiones.**

Los métodos de clasificación automática son capaces de reconocer correctamente alrededor del 60% de los casos empíricos. Mediante una combinación del análisis "cluster" y análisis discriminante la clasificación es mejorada drásticamente separando los procesos dentro de un cierto tipo en subclases. Como resultado de esto la calidad del esquema de reconocimiento alcanza casi el 80%. En el futuro el esquema puede ser "iterado" hasta conseguir un nuevo catálogo de estructuras sinópticas que puedan ser mejor distinguidas por los procedimientos del análisis discriminante:

**Agradecimiento.**

El segundo autor quiere agradecer una estancia sabática financiada por el Ministerio de Educación y Ciencia español.

**Tabla IX**  
Frecuencias de los tipos del nuevo catálogo en función del mes del año (en %)

mes	núm.	TIPO:																	
		A1	A2	A3	B1	B2	B3	C1	C2	C3	DE1	DE2	DE3	F1	F2	F3	F4	F5	
1	67	3	8	0	6	6	9	0	4	2	6	10	3	3	19	6	8	8	
2	58	3	10	9	9	10	9	2	7	2	7	0	2	0	7	5	7	12	
3	67	4	3	4	4	4	0	0	12	3	13	8	3	12	10	8	4	6	
4	63	10	3	16	5	8	8	6	0	2	10	5	2	11	3	6	6	0	
5	67	0	0	6	0	2	10	19	6	8	16	5	3	8	3	0	8	8	
6	64	0	0	8	0	9	9	6	5	3	6	2	23	9	0	5	11	3	
7	66	0	0	4	0	11	3	8	0	2	3	0	12	30	0	11	11	6	
8	67	0	0	3	3	3	3	9	2	2	8	2	9	27	2	10	9	10	
9	65	2	0	6	3	2	6	11	2	6	6	2	3	20	2	14	11	6	
10	67	2	2	2	3	4	4	6	3	6	8	2	4	18	2	13	9	13	
11	66	2	8	4	9	6	2	3	6	3	4	4	2	11	6	8	9	14	
12	61	5	2	3	7	3	5	0	5	5	0	7	8	0	22	5	3	20	
1-12	778	%	2	3	5	4	6	6	6	4	4	7	4	6	13	6	8	8	9

### III SIMPOSIO NACIONAL DE PREDICCIÓN

#### Referencias.

- Court, A. (1957). *Climatology: complex, dynamic, and synoptic*. Ann. Assoc. Amer. Geographers, v. 47, pp. 125-936.
- Katz, A.L. (1960). *Variación estacional de la circulación general atmosférica y predicción del tiempo a largo plazo*. Gidrometeoizdat. Leningrad. (En ruso).
- Lund, I.A. (1963). *Map-pattern classification by statistical methods*. J. Appl. Meteor., v. 2, pp. 56-65.
- Kirchthofer, W. (1973). *Clasificación de estructuras en 500 mb*. Arbeitsbericht der Schweizerischen Meteorologischen Zentralanstalt, 45. pp. 16.
- Kruizinga, S. (1979). *Objective classification of daily 500 mb patterns*. Sixth Conference on probability and statistics in atmospheric sciences. American Meteorological Society.
- Kozulin, K.N.; Reitenbakh, R.G. (1984). *Objective classification of the wind fields near the Earth in the Northern Hemisphere via algorithmic analysis of Vangenheim's minimization*. Trudy VNIIGMI-MCD, no. 109. (Statistical analysis of meteorological information). pp. 30-50. (En ruso).
- McCutchan, M.H.; Schroeder, M.J. *Classification of meteorological patterns in southern California by discriminant analysis*. J. Appl. Meteor., vol. 12. pp. 571-577.
- Andreev, V.; Prazdnikov, M.; Pashaliev, V. (1978). *An attempt for medium-range weather forecast of certain meteorological phenomena over Bulgaria*. Hidrology and Meteorology, vol. 2. (En búlgaro).
- Vorobyov, V.I. (1967). *Planetary high frontal zones and jet flows in the Northern Hemisphere*. Gidrometeoizdat. Leningrad. (En ruso).
- Khrabrov, Yu.B. (1957). *The essential traits of evolution of synoptic processes*. Trudy CIP, no. 63. pp.3-19. (En ruso).
- Internal Report of National Institute of Meteorology and Hidrology of Bulgarian Academy of Sciences: Problem 420 for 01.06.1987 to 31.05.1988.
- A pattern recognition scheme for the macro-synoptic processes. (An automated system for finding analogues). Internal Report of the National Institute of Meteorology & Hydrology. Bulgarian Academic of Sciences. (1989).
- Klink, K.; Willmott, C.J. (1989). *Principal components of the surface wind field in the United States: a comparison of analyses based upon wind velocity, direction and speed*. Int. Journal of Climatology, vol. 9. pp. 293-308.
- Stone, R.C. (1989). *Weather types at Brisbane, Queensland: an example of the use of principal components and cluster analysis*. Int. Journal of Climatology, vol. 9. pp. 3-32.