

## Comunicación B-2

# CLASIFICACIÓN AUTOMÁTICA DE SONDEOS. SITUACIONES DE MAL TIEMPO EN EL MEDITERRÁNEO OCCIDENTAL

**Elisa Tudurí**

GPV de Baleares del INM

**Clemente Ramis**

Grupo de Meteorología. Departamento de Física.  
Universitat de les Illes Balears

### RESUMEN

*En este trabajo se analizan 313 sondeos de Palma simultáneos a cinco tipos de fenómenos convectivos: granizo, tormenta con y sin lluvia, lluvia únicamente y tornados. Se realiza un análisis de «clusters» a los sondeos definidos por 34 variables (estructuras térmica y de humedad, CAPE, helicidad y agua precipitable) y aparecen cuatro tipos de ambientes. Un análisis de componentes principales sobre la misma base de datos muestra que un nuevo «clustering» sobre seis variables (76% de la varianza) da resultados semejantes. El objetivo, la clasificación de sondeos, buscará la mayor correlación entre éstos y el representante de cada «cluster».*

### 1. Introducción

El Mediterráneo occidental, así como sus costas, es una zona donde no son infrecuentes los fenómenos convectivos, como las lluvias con grandes cantidades de precipitación (en ocasiones superiores a 200 mm/día), las tormentas y en menor cuantía los granizos y tornados. Se han realizado muchos estudios sinópticos (como García-Dana y otros, 1982) y más recientemente de diagnóstico (Ramis y otros, 1994) o estudios de características ambientales (Gayá y Soliño, 1993) para este tipo de situaciones. No obstante, desde el punto de vista operativo, la diagnosis de estas situaciones incluye tanto el marco sinóptico y el mesoescalar como el estudio de los datos de radiosondas en diagramas termodinámicos. En este sentido se han desarrollado gran cantidad de índices y otros parámetros (CAPE (Weisman y Klemp, 1982) y Helicidad (Lilly, 1986)) para ayudar a determinar el comportamiento convectivo del ambiente y la potencia del fenómeno posible en él. La mayor parte de estudios realizados han utilizado índices pero las estructuras verticales de viento y variables energéticas han sido menos utilizadas.

Aunque este trabajo ha considerado los valores de los índices, la CAPE y helicidad para varios casos en que ocurrieron fenómenos convectivos, el objetivo se centra en el desarrollo de técnicas para la identifica-

ción de ambientes asociados a ellos. Para su realización se ha hecho uso de técnicas de *clustering* en dos sentidos: directamente de las variables que definen un ambiente y de las que resultan de un análisis de componentes principales.

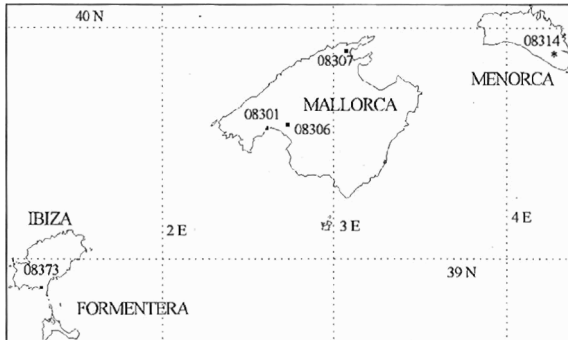


Fig. 1. Los observatorios utilizados en el estudio con su indicativo sinóptico

## 2. Base de datos

Entre los archivos climatológicos de los principales observatorios que elaboran partes cifrados en las Islas Baleares se han elegido series de diez años (1984-1993) entre las que se han seleccionado los días en que se observó algún fenómeno convectivo significativo según determinados criterios. Cada uno de los días elegidos ha sido caracterizado por el sondeo de Palma de fecha y hora más próximos al fenómeno observado. El resultado son 313 casos a los que añadimos otros 23 sin fenómeno significativo.

Fenómeno	Criterio	Símbolo	Núm. de casos
Granizo	Observación	Ha	77
Tormenta + Lluvia fuerte	Tormenta observada y prec. > 30 mm/h durante 10 min	R1	114
Lluvia fuerte	Prec. > 30 mm/h durante 10 min	R2	52
Tormenta seca	Tormenta observada y prec. < 30 mm/h y rachas > 14 m/s	T	65
Tornado	Observación y registro de fuertes rachas	W	5
Buen tiempo	Sin fenómenos significativos	F	23

## 3. Estudio de inestabilidad

Para realizar un primer análisis de los datos e intentar una primera clasificación de los ambientes, se utilizaron los sondeos seleccionados para calcular algunos índices de inestabilidad: *Lifted Index* (Galway, 1956), *Lifted Parcel* (Sanders, 1986), *Showalter* (1953), *K-Index* (1960), Total de Totales (1972) y *Windex* (McCann, 1994). Buscando posibles relaciones entre fenómenos y valores de índices se observa que, en general, los valores medios de los índices no presentan diferencias muy grandes entre diferentes fenómenos y únicamente se pueden distinguir los casos de buen tiempo de todos los demás.

Fenómeno índice	Ha Media	R1 Media	R2 Media	T Media	W Media	F Media
LP	-0,1	-1,7	0,8	0,6	-6,0	3,9
LI	2,1	0,5	2,4	2,1	-2,0	6,5
SHI	3,6	2,8	3,9	4,0	1,7	7,5
KI	19,0	23,9	21,7	19,8	25,8	2,8
TT	49,5	47,3	46,8	48,2	48,0	38,2
WI	24,8	26,6	20,1	23,0	39,8	17,2

Este resultado nos sugiere que la mayoría de los índices de inestabilidad no son exactamente aplicables a nuestra zona de estudio por lo que parece necesaria una nueva estrategia.

#### 4. Clasificación de ambientes

Nuestro objetivo es buscar descripciones que relacionen los ambientes con los fenómenos que hemos considerado y hemos optado por un análisis de *cluster.*, utilizando el método *K-means* (Anderberg, 1973).

En primer lugar caracterizamos cada sondeo por 34 variables:

- Temperaturas a intervalos de 50 hPa desde 1 000 a 100 hPa. Éstas no son las temperaturas del ambiente sino desviaciones con respecto a un sondeo medio representativo del día y hora correspondiente a cada uno de ellos (Ramis, 1976). Representan la estructura térmica vertical del ambiente.
- Depresión del punto de rocío a intervalos de 50 hPa desde 1 000 a 500 hPa que, como diferencia entre temperatura y punto de rocío, determina de forma indirecta la distribución vertical de humedad.
- Energía potencial convectiva disponible (CAPE) que se define como la energía que puede desarrollar una partícula que asciende dentro de un ambiente sin perturbarlo. Mide la potencia de la convección, siempre que exista un mecanismo capaz de dispararla. Generalmente se calcula para la partícula de superficie, pero en nuestro caso (y cuando su valor es cero) se aplica el cálculo a niveles superiores donde es posible que su valor sea positivo.
- Nivel de presión en que puede dispararse la convección (PCL), es el nivel más bajo a partir del cual la CAPE tiene un valor positivo.
- Helicidad (H), que se define por el valor integrado en un estrato del ambiente del módulo del producto vectorial del viento horizontal y su cizalladura vertical (Lilly, 1986). Es una medida de la advección térmica y, gráficamente, representa menos dos veces el área encerrada por la hodógrafa entre los límites del estrato, que tomamos como los 3 km más bajos del ambiente a partir del PCL.
- Agua precipitable total entre superficie y 500 hPa (PW), es la altura de agua líquida que se obtendría si todo el vapor, contenido en un determinado estrato con una base de 1 m<sup>2</sup>, condensase.

Una vez definidas estas variables son estandarizadas y se les aplica el método de *clustering* antes mencionado. Este método agrupa en subgrupos (el número de ellos se define *a priori*) los elementos de la base de datos de forma que cada uno de ellos contiene sondeos que presentan características semejantes y que les diferencian del resto de los sondeos. En nuestro trabajo elegimos cuatro grupos y el resultado es un sondeo medio representativo de cada *cluster.*

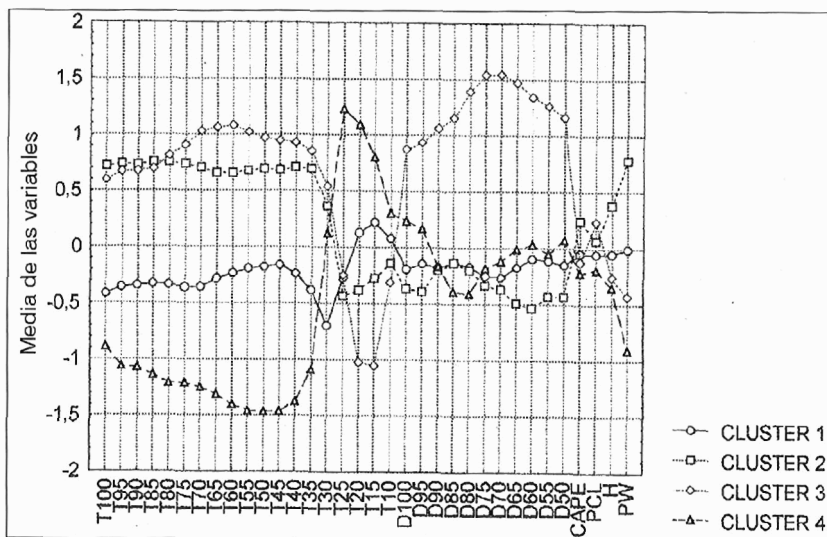


Fig. 2. Valores medios de cada variable estandarizada para cada uno de los cuatro «clusters»

Del gráfico se deduce: el *cluster* 1 representa masas de aire frías y húmedas, con poca energía y helicidad; el 2, masas de aire templadas y húmedas con mucha energía y helicidad positiva; el 3 reúne masas cálidas y muy secas con poca energía y advección fría (helicidad negativa) y el 4 engloba las masas muy frías pero húmedas en niveles medios, estables y con advecciones frías.

Los casos y porcentajes de cada *cluster* según fenómenos, se representa en el siguiente cuadro:

Cluster	Casos	Ha	%	R1	%	R2	%	T	%	W	%	F	%
1	107	22	28	43	38	17	32	21	32	2	40	2	9
2	109	12	16	53	46	23	44	17	26	3	60	1	4
3	49	10	13	7	6	6	12	4	10	0	0	20	87
4	71	33	43	11	10	6	12	21	32	0	0	0	0

donde vemos que predominan: en el *cluster* 1 tormentas con y sin lluvia, en el 2 tornados y lluvias con y sin tormentas, en el 3 buen tiempo y en el 4 granizo y tormentas secas.

### 5. Análisis de componentes principales

El anterior apartado muestra una clasificación objetiva de sondeos pero maneja una gran cantidad de variables, no todas independientes sino que algunas están correlacionadas entre sí. Esto sugiere que es posible reducir el número de variables que definen un sondeo (Jolliffe, 1990) que matemáticamente se consigue mediante el análisis de componentes principales.

Este análisis comprende el cálculo de valores propios de la matriz de correlación y la elección del número de valores propios significativos. Tras un estudio e interpretación de las posibles componentes significativas hemos seleccionado seis de ellas que explican el 76% de la varianza. Aplicamos una rotación VARIMAX de ejes para la obtención de estructuras simples e interpretamos las seis componentes principales. Esto se consigue analizando los coeficientes de las combinaciones lineales que definen las nuevas componentes principales e identificando las variables primitivas a que modifican.

Nueva variable	Antigua variable
1	Datos de temperatura 1 000 - 350 hPa
2	Punto de rocío 700 - 500 hPa y helicidad (-)
3	Datos de punto de rocío de 1 000 a 750 hPa
4	Temperatura de 300 a 250 hPa
5	Temperatura de 200 a 100 hPa
6	CAPE y agua precipitable

Aplicando el método de *clustering K-means* a los sondeos definidos por las seis componentes principales obtenemos otros valores medios representativos de grupos de sondeos (elegimos también cuatro *clusters*).

Así predominan: en el *cluster* 1 las lluvias, tormentas (con o sin lluvia) y tornados, en el 2 las lluvias y tormentas con lluvia, en el 3 buen tiempo y en el 4 granizo.

#### 6. Asignación de un sondeo a un *cluster*

El objetivo del trabajo es crear una base operativa que dado un sondeo no incluido en la base de datos, lo compare con los valores medios de los grupos y determine a cuál de ellos pertenece y, apriori, obtener información sobre qué fenómenos podrían producirse.

Para esto se obtienen del sondeo las 34 variables ya definidas, se estandarizan utilizando la media y la desviación típica de la base de datos y se hallan los coeficientes de correlación entre las variables de este sondeo y los valores medios de las variables de cada uno de los *clusters*. El sondeo pertenecerá al *cluster* con el que presente mayor correlación y éste indicará qué tipo de fenómeno es más probable con ese ambiente.

Este método se aplicó a 49 sondeos pertenecientes al período 1994-95 entre los que se encontraban todos los fenómenos anteriormente clasificados (salvo los tornados), y los resultados de asignación por este método se distribuyeron de la siguiente forma:

Fenómeno	Casos	Cluster 1	Cluster 2	Cluster 3	Cluster 4
H	10	2	1	2	5
R1	13	3	8	0	2
R2	7	2	3	2	0
T	8	0	4	2	2
F	11	0	0	11	0

A modo de ejemplo presentamos gráficamente dos casos: caso 1 de lluvias y tormentas y caso 2 de granizo.

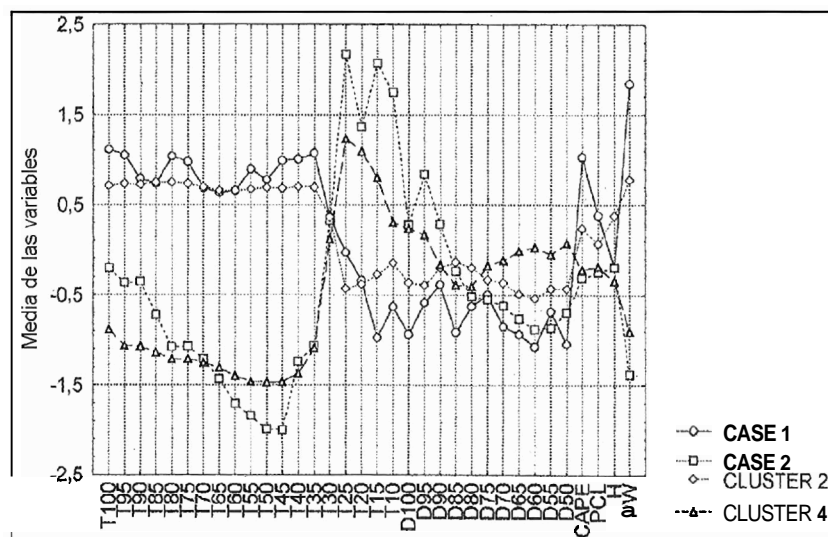


Fig. 4. Comparación de dos sondeos con los grupos clasificados utilizando las 34 variables

Los coeficientes de correlación para estos casos (caso 1 de lluvias con tormentas y caso 2 de granizo) son:

Caso 1	Cluster- 2	0,92	Caso 1	Cluster 4	-0,72
Caso 2	Cluster 2	-0,57	Caso 2	Cluster 4	0,85

## 7. Conclusiones

El método de *clustering* permite una clasificación objetiva de sondeos pues, estudiadas 336 situaciones diferentes, distingue 4 grupos de ambientes diferenciados. En cada grupo predomina algún fenómeno significativo, por tanto, es posible asignar a cada ambiente un fenómeno más probable. El método de componentes principales permite reducir el número de variables a seis, que determinan la estructura del sondeo, explicando el 76% de la varianza. El método de *clustering* aplicado a los sondeos definidos en el espacio de componentes principales, permite obtener también 4 grupos de sondeos, a los que es posible asignar como dominante algún fenómeno significativo. La estructura de cada cluster por los dos métodos es similar. Finalmente es posible tomar un sondeo que no pertenezca a la base de datos y compararlo con los valores medios de los *clusters* para determinar a qué grupo pertenece y, por tanto, cuáles son los fenómenos que, *a priori*, podrían producirse.

## Referencias

- Andenberg, M. R., 1973: *Cluster Analysis for Applications*. Academic Press, 359 pp.
- Galway, J. G., 1956: *The Lifted Index as a predictor of latent instability*. *Bull. Amer. Soc.*, 37, pp. 528-529.
- García-Dana, F.; R. Font y A. Rivera, 1982: *Situación meteorológica durante las lluvias fuertes del levante español en Octubre de 1982*. INM.
- Gayá, M. y A. Soliño, 1993: *Caps de fibló i esclafits a Menorca*. *Revista de Menorca*, 1, pp. 5-18.
- Joliffe, I. T., 1990: *Principal Components Analysis: A beginners guide. I. Introduction and application*. *Weather*, 45, pp. 375-382.
- Lilly, D. K., 1986: *I. The structure, energetics and propagation of rotating convective storms. II. Helicity and storm stabilization*. *J. Atmos. Sci.*, 43, pp. 126-140.
- McCann, D. W., 1994: *WINDEX - A new index for forecasting microburst potential*. *Wea. Forecasting*, 9, pp. 532-541.
- Ramis, C., 1976: *Climatología de la atmósfera sobre Mallorca*. Univ. Barcelona, 60 pp.
- Ramis, C.; M. C. Llasat; A. Genovés y A. Jansá, 1994: *Las riadas de octubre de 1987 en Cataluña: Mecanismo sinóptico y inesoalcar*. *Meteorological Applications*, 1, pp. 337-350.
- Sanders, F., 1986: *Temperatures of air parcels lifted from the surface: Background, application and monograms*. *Wea. Forecasting.*, 5, pp. 570-575.
- Weisman, M. L. y J. B. Kemp, 1982: *The dependence of numerically simulated convective storms on vertical wind shear and buoyancy*. *Mon. Wea. Rev.*, 112, pp. 790-808.

## Agradecimientos

Los autores quieren agradecer al Dr. Charles Doswell III, sus comentarios y ayuda en la versión final de este trabajo y al Prof. Sergio Alonso por sus sugerencias durante el desarrollo del mismo.