

**UN METODO GRAFICO DE CLASIFICACION:
"LA MATRIZ ORDENABLE" DE J. BERTIN.**

Severino ESCOLANO UTRILLA

**Departamento de Geografía y Ordenación del
Territorio**

Universidad de Zaragoza

Ciudad Universitaria

E-50009 Zaragoza

"En la base misma de cualquier proceso de medida se encuentra la más humilde y modesta de las operaciones intelectuales y lingüísticas: la de identificar algo como perteneciente a un tipo dado... Este primer paso tan "sencillo" es también, a menudo, el último paso de un proceso de medida, y requiere un sutil y complejo juego combinado de teoría y observación". Marx W. Wartosky (1983). *Introducción a la filosofía de la ciencia*. Alianza Universidad: 206.

**1. EL CONTEXTO METODOLOGICO DE LA
CLASIFICACION EN GEOGRAFIA.**

El desarrollo reciente de la denominada "Geografía cuantitativa" se ha caracterizado por su constante actitud de renovación, epistemológica y conceptual, de métodos e instrumentos, para adaptarse a los rumbos actuales de la geografía. A este fin ha contribuido notablemente, junto a la postura crítica de otros enfoques, la saludable y dinámica actividad de revisión interna practicada, sin acritud pero con vigor, por los miembros adscritos a dicha corriente o, con más propiedad, usuarios de métodos cuantitativos, como para España se desprende de la lectura de las Actas de las anteriores reuniones del Grupo de Métodos Cuantitativos de la A.G.E.

La reflexión y el debate permanente han puesto de manifiesto la carencia sustancial de métodos adecuados a las perentorias necesidades de la teoría y praxis geográficas de hoy; el recurso a instrumentos ya contrastados, pero no totalmente apropiados a los nuevos cometidos geográficos. La dirección de los esfuerzos parece que ha de tener por norte "... la necesidad como respuesta a la preocupación investigadora en la geografía humana actual, de crear nuevos datos y nuevas técnicas más de acuerdo con la ampliación del campo de la geografía humana actual" (Estébanez, 1989: 131).

En lo atinente a la ampliación de métodos y técnicas ha habido, y persiste, una intensa labor de adaptación y creación de herramientas operativas, algunas de singular interés como las reputadas "blandas", frente a las "clásicas" o "duras", más eficientes para evidenciar estructuras de datos, sin violentar su organización original. Están ya disponibles técnicas numéricas y gráficas, intuitivas, flexibles y sencillas para casi todas las necesidades de los análisis geográficos, constituyendo un corpus importante las agrupadas bajo el nombre de "Análisis Estadístico Exploratorio de Datos" (EDA). (Bosque, 1986) que integra variados procedimientos de este tipo para múltiples propósitos.

Sin embargo, para un tema tan trascendental desde el punto de vista epistemológico como recoge el párrafo que encabeza esta comunicación, y tan útil y universal en la práctica como es la clasificación, se sigue apelando al uso de complejos y sofisticados métodos (análisis factorial, taxonomías numéricas, etc.).

Sin duda, una de las causas del favor y proliferación de empleo de estos procedimientos es la facilidad de lograr resultados con programas *ad hoc* que funcionan en microordenadores. Más allá del ejercicio estadístico-matemático que comportan, se persigue

derivar de su aplicación cierto marchamo de "objetividad"⁵¹ (Rainelli,1983: 15); este argumento ha sido discutido largamente, e implica una identificación de lo "real" con lo "aparente", representado por los datos, y la aceptación de la hipótesis, de dudosa validez, de que la realidad es reproducible por las mismas reglas de transformación que los números. Esta pretensión confunde, también, "objetividad" con "exactitud" o "precisión": ¿acaso es más objetivo, entendido como más real, aplicar el criterio de la media aritmética (\bar{x}) y la desviación típica (s) para clasificar un fenómeno (jerarquía de ciudades, valores de consumo, índices de desarrollo...) que otro método menos preciso?. Desde luego, la "exactitud" alcanzada con estos métodos no es razón suficiente para avalarlos como "objetivos" o, incluso, como más "objetivos". Más aún, no siempre son congruentes con los sujetos a que se aplican y, en otras ocasiones, no descubren nada que no sea ya conocido por otros medios habituales (Martínez et al. 1988; 280).

Otra razón de su abundante uso es la carencia de suficientes soluciones alternativas encuadradas en los nuevos enfoques "blandos"; aunque existen desde, hace tiempo, propuestas útiles como la elaborada por la "semiología gráfica", bien conocida por los geógrafos a partir de la obra de J. Bertin (1967), más citada que leída, y desde luego, seguida.

2. EL PROCEDIMIENTO DE LA "MATRIZ ORDENABLE"

Para la confección de clasificaciones la sistemática de J. Bertin tiene varias opciones en función de la

⁵¹ Tan es así que incluso el concepto "objetivo" alcanza al título de algunos trabajos, por ejemplo: LLORCA, A. y RUIZ, M. (1987). "Apuntes para una metodología objetiva de sectorización territorial". *Estudios Territoriales*, 25: 107-119. La conclusión indica la contribución de estos métodos numéricos para reducir la subjetividad, pero su incapacidad para eliminarla. No obstante, cabría preguntarse si, en ciencias sociales, es bueno, deseable o de interés la consecución de la "objetividad".

"ordenabilidad" o no de individuos y variables⁵² (Bertin, 1967: 220-221) siendo la más ventajosa para fines geográficos la recogida aquí. Su objetivo es el establecer tipologías, o sea, agrupar a los individuos en clases lo más homogéneas posible en cuanto a la semejanza que presentan en los valores de sus atributos; el procedimiento es enteramente gráfico e intuitivo, sin andamiaje matemático que transforme los datos originarios.

2.1. Preparación de los datos.

Un requisito imprescindible para trabajar con este método es disponer de datos en escala ordinal. Si la información ya está expresada en esa escala se puede proceder de inmediato; pero, si por el contrario, los datos vienen dados en escala superior, de intervalos o ratios, es necesario reducirla, no sin antes evaluar si la pérdida de información es o no significativa.

La categorización de las variables ha de ser idéntica para que los resultados sean comparables; no importa el número intervalos en que se dividan, pero conviene recordar que cuantos más haya, mayor será la acuracidad y más grande la dificultad si la búsqueda de la solución se lleva a cabo manualmente.

La introducción de numerosas variables y categorías, como suele ser habitual aunque no estrictamente necesario, entraña un proceso muy laborioso de comparación de los "perfiles" de los individuos para observar sus trazos comunes, debido a la enorme cantidad de variantes posibles, no siempre cubiertas con casos reales. Así, con 2 variables y 3 categorías cada una, las "siluetas" individuales distintas pueden ser hasta 7, como se muestra en el gráfico.

⁵² La combinatoria de los elementos considerados y su ordenación da lugar a tres soluciones: a) individuos y variables son considerados diferentes y, por tanto, ordenables según sus frecuencias; b) sólo las variables se estiman diferentes y ordenables, y c) individuos y variables se consideran ordenados.

Representación en "perfiles"			Representación matricial		
Individuos	Variables	Intens.	Indiv.	Variables	Intensidad
A	1 2	Alta	A	1 2	Alta
B			B		
C		Media	C		Media
D			D		
E			E		
F		Baja	F		Baja
G			G		

Representación en "perfiles" y matricial de todos los individuos posibles que se forman con dos variables y tres categorías.

La ordenación matricial de individuos en las filas y variables en las columnas - o al revés- es la más ágil para trabajar; en las celdas se consignan las intensidades correspondientes a través de una gama de gris.

2.2. Diagonalización.

La mecánica consiste en comparar los "perfiles" individuales y, según su grado de parecido, redistribuir la matriz permutando hileras y columnas, hasta obtener áreas de intensidades homólogas, que tengan sentido geográfico; a esta manipulación se denomina "diagonalización". Aquí no tiene el significado matemático del álgebra matricial, sino el de dividir la matriz aproximadamente en dos mitades, con valores altos a un lado y bajos a otro de dicha diagonal. La tarea de "diagonalizar" es muy minuciosa si el número de individuos y variables es elevado.

2.3. Clasificación.

Como en toda clasificación la meta perseguida es hallar grupos de individuos similares en esos atributos; la labor, si principal, no es fácil cuando las matrices son

grandes, puesto que, al igual que en la realidad, es corriente que se conformen transiciones, medios tonos insensibles a la percepción visual, entre grupos extremos más netos. Eso sí, la solución escogida, siempre provisional, está abierta al investigador para realizar el número de grupos que desee, y, por tanto, variar el nivel de semejanza o diferencia. Asimismo, se pueden "diagonalizar" aisladamente subconjuntos de atributos para examinar su comportamiento dentro del total de variables.

3. UN EJEMPLO: LA ESTRUCTURA DEL COMERCIO EN LAS PROVINCIAS ESPAÑOLAS.

Para vislumbrar la utilidad y potencia del método lo hemos aplicado al análisis de la estructura comercial de las provincias españolas, para averiguar de forma sintética su similitud en potencial, perspectivas, etc. comerciales, toda vez que se ha realizado una comparación con los resultados del método factorial. Los comentarios sobre comercio son breves, sólo los necesarios para desbrozar algunos senderos transitables.

3.1. La elección de las variables.

Si bien la "provincia" no es el individuo geográfico más adecuado para valorar el estado y funcionamiento del comercio -quizá es más apropiado la área mercantil-, lo cierto es que se dispone de series estadísticas provinciales homogéneas y comparables que permiten la correcta aplicación del método.

Se han seleccionado 19 variables, capacidad máxima del programa empleado⁵³, medidas en intervalos y reducidas a escala ordinal con cinco categorías en cuanto a su intensidad: muy baja, baja, media, alta y muy alta. Dadas las imprecisiones de algunos datos de comercio, esta modificación no supone

⁵³ Para la diagonalización y representación gráfica se ha empleado el programa Cartographie-2D v. 1.2a, y se ha corregido manualmente algún caso. La conversión de escala se ha realizado para cada

merma grave de información.

El contenido comercial de cada provincia se ha resumido por:

-variables de la actividad comercial: las licencias comerciales (1989, Banesto) clasificadas en nueve grupos: alimentación (1); textil (2); madera (3); química (4); construcción (5); metales (6); transporte (7); comercio ambulante (8); y otro tipo de comercio (9). Obviamente, se han consignado en tantos por ciento de cada rama respecto del total provincial. Además, se ha incorporado una variable de densidad: licencias/1000 habitantes (10), y otra que pretende medir el grado de modernización del comercio: m^2 en autoservicio/1000 habitantes (11).

-variables económicas: son fundamentales para comprender la composición del comercio, habida cuenta de la influencia de algunas de ellas sobre el equipado comercial; se han considerado la renta (12), determinante importante de la capacidad y estructura de compra y consumo; el índice tercero de capacidad de compra (1989, AME, Banesto)(13), que expresa la potencialidad de compra para artículos especializados; el índice de vivacidad (1989, AME, Banesto) (14), que compendia las perspectivas globales coyunturales, y, finalmente, el gasto real en alimentación (15), importante para entender el peso comercial de este sector.

-variables demográficas: imprescindibles por su relación directa con el comercio; se han reunido: la concentración de la población, a través del porcentaje de habitantes en la capital de provincia (16), para estimar la importancia de las aglomeraciones de población en el comercio; la población mayor de 65 años (17), reflejo de la estructura de edades, determinante de tipos y

modalidades de consumo; el crecimiento vegetativo (18), para incluir la dinámica de la población y, por último, la densidad de población (19), indicador por excelencia de la intensidad de la población sobre el territorio.

3.2. Valoración de los resultados. Comparación con el método factorial.

El diseño, a modo de "crucigrama" o "ajedrezado", producto de la "diagonalización" para las 19 variables aparece en la "matriz ordenable" al final de esta comunicación. Del examen de la misma se derivan múltiples posibilidades de clasificación de acuerdo de los objetivos buscados, sin que sea fácil asegurar la validez de una sobre otra.

El programa sugiere hasta un máximo de cinco grupos, que nosotros hemos establecido, tras "correcciones" personales, como figuran en el mapa 1 y la matriz final; desde luego, los límites son discutibles pero también lo son los del mapa 2.

El primer grupo de provincias se caracteriza por tener una proporción elevada de comercio y gasto alimentarios, un acusado envejecimiento de la población, baja densidad y capacidad de compra, así como escasa o mediana presencia de comercio especializado. Coincide este grupo con las provincias interiores -salvo las tres gallegas- y las más deprimidas económicamente; en todas domina el comercio tradicional, con gran importancia del comercio rural y preponderancia del alimentario, y con baja penetración de las "nuevas formas comerciales". Al contrario, el índice de vivacidad tiene aquí las cotas más altas.

Los grupos 2,3, y 4 son transiciones graduales y complejas, en las que se alternan valores altos y bajos de las diferentes variables en los distintos individuos, destacándose bien, entre otras, la influencia del turismo. El tipo 5 es más nítido, y comprende las provincias más

avanzadas comercialmente, es decir, las de desarrollo más fuerte de las "nuevas formas de comercio", y más notable presencia de establecimientos especializados, no alimentarios, toda vez que poseen las rentas y capacidad de compra más elevadas.

Con intenciones de comparación, se ha practicado un análisis factorial (por el método de componentes principales) sobre los mismos datos, extrayéndose 5 factores. Los pesos factoriales estandarizados se han sumado para cada provincia y, puesto que la distribución resultante es casi normal -debido a la acumulación de errores redondeos en el proceso de cálculo- con media 0 y desviación típica 1,84 (debería ser 1), se han discretizado 5 grupos, como se recoge en el mapa 2. Las concomitancias con el mapa 1 son grandes, no obstante los movimientos de individuos habidos entre grupos contiguos, perfectamente posibles. Acaso son más llamativos los cambios acaecidos en el grupo 5, ocupado ahora por Lérida y Navarra (que no figuraban en este grupo en el procedimiento anterior, aunque Lérida podría haberse sumado), y la salida de Madrid, Barcelona, Mallorca, Alava, Las Palmas y Zaragoza al grupo 4.

Aunque ésta es sólo una forma de cotejo, y tal vez sencilla, muestra que los resultados son parecidos, igualmente plausibles, aunque con la "matriz ordenable" se tiene una información más comprensible, rica y flexible que permite otros ensayos.

Así por ejemplo se puede ir reduciendo el número de grupos y el grado de similitud (mapa 4), o realizar conglomerados "anidados" (mapas 7 al 11) para encontrar las grandes líneas territoriales de la composición del comercio. Esta última forma "imbricada" es bastante idónea porque permite recoger e interpretar, con algún éxito, las transiciones.

En los mapas 9 al 11, se aprecian tres grandes

áreas con estructura comercial diferente, cada una con dos subconjuntos: el grupo 1 reúne con más "pureza" las propiedades del área, mientras que el 2 constituye la banda que da paso al tipo 1 de la siguiente zona. Así, el área A, compone primer subconjunto con las provincias hasta Lugo, y el segundo hasta Teruel; la zona B forma el primer grupo hasta Guipúzcoa, y el segundo hasta Sta. Cruz de Tenerife, y la C, tiene el límite del primer subconjunto en Lérida y finaliza con Madrid (ver matriz al final). Los rasgos comerciales son fácilmente interpretables, y se reconocen, "grosso modo", por un comercio tradicional, minifundista y rural (A), en otro moderno, con presencia importante de "nuevas formas" y abundancia de grandes superficies -factor turismo- (C), y con características mixtas, el tercer gran conjunto (B).

Todavía se pueden seguir explorando nuevas vías "diagonalizando" grupos temáticos de atributos, para comprobar su incidencia aislada; así se ha hecho con las variables de actividad comercial (1 a 11), con las económicas (12 a 14) y con las demográficas (15 a 19); con las matrices extraídas sólo hemos realizado una división en cinco grupos -para comparar con el mapa 1- que aparecen cartografiados en los mapas 4 a 6. En ellas es igualmente posible un análisis más profundo, reconsiderando otros grupos, o bien por el procedimiento "anidado".

4. A MODO DE APRESURADA CONCLUSION.

La técnica presentada aunque "clásica" puede alinearse en las nuevas orientaciones metodológicas, heurísticas e intuitivas, en la misión establecer clasificaciones. Permite la aprehensión rápida y sintética de tipologías, al menos en el caso tratado⁵⁴, con la

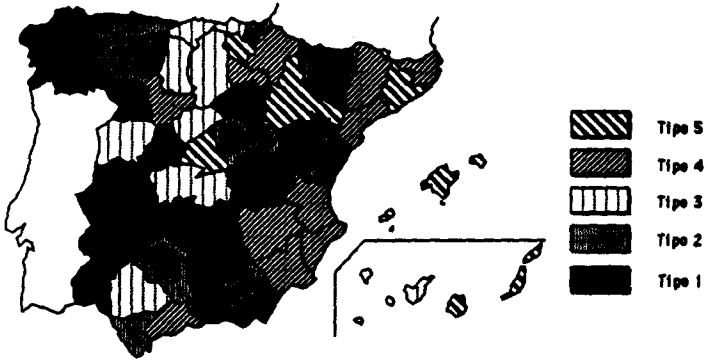
⁵⁴ Una extensa explotación del método aplicada al comercio francés puede encontrarse en COQUERY, M. (1977) *Mutations et structures du commerce de détail en France. Etude géographique*, Cergy, Signe.

misma fiabilidad que otros métodos numéricos complejos.

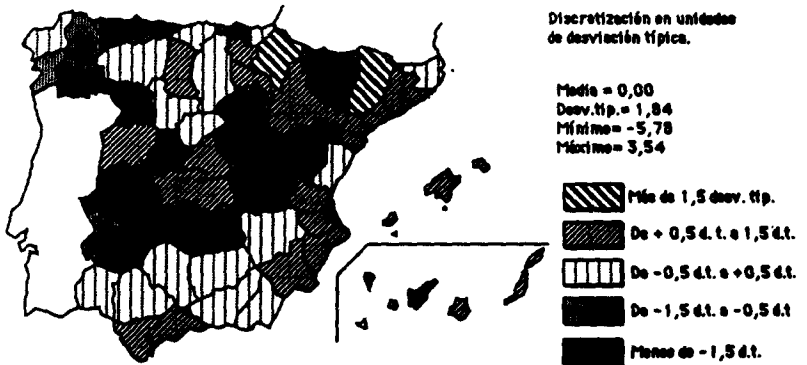
Además aporta sobre ellos, la riqueza de matices, la sencillez y la flexibilidad de intentar nuevas aproximaciones bajo el control directo del investigador hasta encontrar la combinación más conveniente, sin distorsión alguna de los datos. Tal vez el problema de fondo, aparte de la elección del método que es importante, sea definir, meridianamente, el fenómeno o concepto, los individuos y las variables que lo representan; sino se resuelve esta dificultad previa, los métodos de clasificación se agotan en si mismos y sólo proporcionan soluciones muy parciales.

ESTRUCTURA COMERCIAL DE ESPAÑA. Resultados del método de J. Bertin y factorial.

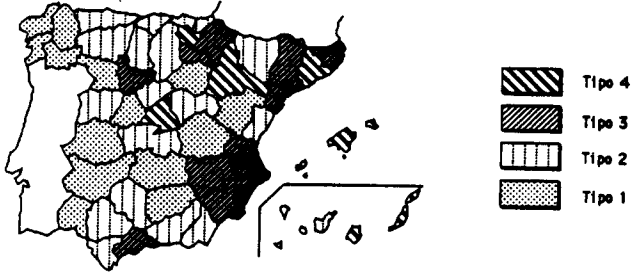
Mapa 1. Tipología según el método de J. Bertin:



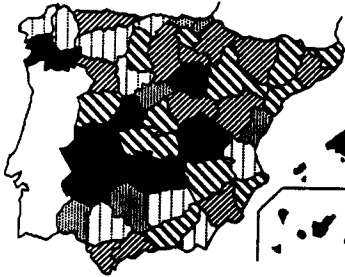
Mapa 2. Tipología según el método factorial.



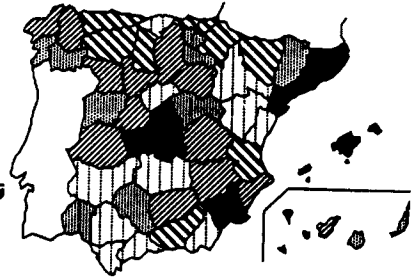
Mapa 3. Tipología según el método de J. Bertin



CLASIFICACIONES PARA SUBGRUPOS DE VARIABLES



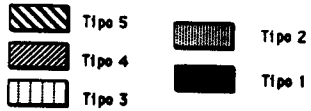
Mapa 4. Variables comerciales.



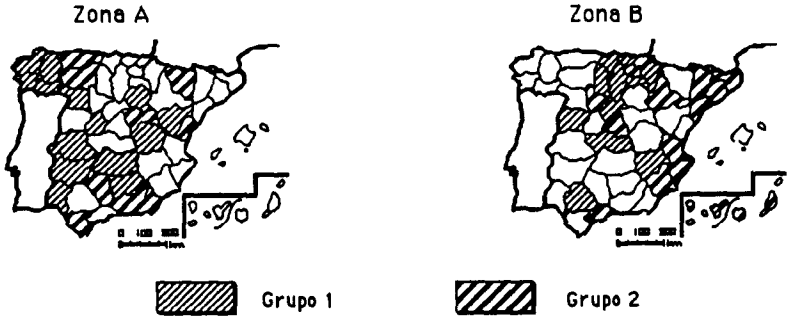
Mapa 5. Variables económicas.



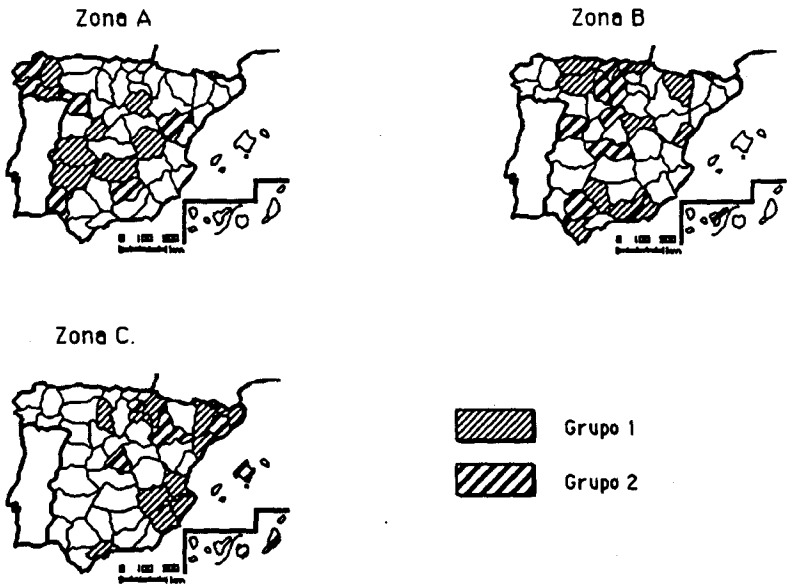
Mapa 6. Variables demográficas



Mapas 7 y 8. Clasificación en dos grandes zonas y dos grupos.



Mapas 9, 10 y 11. Clasificación en tres zonas y dos grupos.



Matriz ordenable por provincias

	19	8	13	10	16	4	11	9	14	12	3	2	6	5	7	18	1	17	15
Caceres																			
Avila																			
Badajoz																			
Drense																			
Soria																			
Cuenca																			
Ciudad Real																			
Lugo																			
Zamora																			
Pontevedra																			
Jaen																			
La Coruña																			
Huelva																			
Teruel																			
León																			
Huesca																			
Cast. de la Plana																			
Granada																			
Cordoba																			
Asturias																			
Almeria																			
Guadalajara																			
Cádiz																			
Gulpúzcoa																			
Segovia																			
Burgos																			
Palencia																			
Salamanca																			
Toledo																			
Sevilla																			
Uizcaya																			
Cantabria																			
Sta. Cruz de Tenerife																			
Albacete																			
Navarra																			
La Rioja																			
Valladolid																			
Valencia																			
Murcia																			
Alicante																			
Málaga																			
Tarragona																			
Gerona																			
Lerida																			
Zaragoza																			
Las Palmas																			
Aleua																			
Mallorca																			
Barcelona																			
Madrid																			

Intensidad de la variable: Muy Baja Baja Media Alta Muy Alta

BIBLIOGRAFIA CITADA.

-BERTIN, J. (1967). *Semiologie graphique. Les diagrammes- Les reseaux- Les cartes*, Gauthier-Villars, Paris.

-BOSQUE, J. (1986). "El Análisis Estadístico Exploratorio y la enseñanza de técnicas cuantitativas en Geografía". *Metodos cuantitativos en geografía: enseñanza, investigación y planeamiento*. Grupo de Métodos Cuantitativos (A.G.E.)

-ESTEBANEZ, J. (1989). "Nuevas tendencias en Geografía Humana y sus repercusiones en Geografía Cuantitativa". *X Congreso Nacional de Geografía. Ponencias* (Zaragoza, 28 sept. - 3 oct. 1987). A.G.E. - Univers. de Zaragoza.

-MARTINEZ, F. J., PABLO, F. y SANCHEZ, J. A. (1988). "Hidrogéochimie des aquifères carbonatés du massif de Guara (Huesca, Espagne)". *Hydrogéologie*, 4: 273-280.

-RAINELLI, P. (1983). "Apports et limites de l'analyse des données". *Economie rurale*, 157: 15-33.