

Aplicación de la taxonomía numérica en las aguas de la cuenca del río Tajo

RAFAEL BERMÚDEZ VALBUENA (*)

JOSÉ M^o GASCÓ MONTES (*)

RESUMEN La taxonomía numérica se ha aplicado para clasificar las aguas de la cuenca del río Tajo en España antes de la frontera con Portugal. Se han utilizado los datos de 52 estaciones con determinación de 15 variables. Las clases se han obtenido mediante la aplicación de la distancia euclídea entre estaciones y la distancia euclídea entre estaciones y centroide, con variación del número de parámetros y con tipificación o no de los valores de las variables.

Se ha encontrado que el agrupamiento centroeide clasifica las estaciones de una manera más adecuada según los condicionamientos naturales que el agrupamiento por distancias mínimas. La tipificación de las variables no mejora significativamente la clasificación, excepto con el agrupamiento centroeide.

Comparando los dendrogramas y las clases que se originan con los condicionamientos naturales, se ha encontrado que la cuenca del río Tajo se divide en cinco grupos: El grupo 1 incluye los tramos de aguas con mayor influencia **ombrogénica** con las estaciones 18, 24, 25, 26, 27, 28, 29, 33, 34, 35, 36, 41, 46, 47 y 49. El grupo 2 incluye los tramos de aguas **ombrogénicas con influencia de agua soligénica** o de escorrentía con las estaciones 7, 13, 22, 32, 43, 11 y 17. El grupo 3 incluye tramos de aguas con influencia **soligénica** con las estaciones 1, 2, 3, 8, 9, 10, 12, 15, 16, 19, 20, 21, 23, 38, 39, 40, 44, 51 y 52. El grupo 4 incluye tramos de aguas con influencia **litogénica** o de la litología con las estaciones 4, 5, 6, 14, 30, 31, 37, 42, 45 y 48. El grupo 5 incluye tramos con mayor influencia de la escorrentía y de la litología con la estación 50. El grupo 3 varía de una clasificación a otra, resultando clave para la diferenciación entre ellas.

APPLICATION OF THE NUMERICAL TAXONOMY ON THE WATER OF THE TAJO BASIN

ABSTRACT Numerical Taxonomy was used to classify the water of the Tagus basin in Spain before the border with Portugal. We used the information of 52 research stations with determination of 15 variables. The classes were got by means of the application of euclidean distance between research station and centroid, with variation of the number of parameters and with classification or not of the values of the variables.

It was found that the grouping centroid classifies the research in a more appropriate way depending on natural conditioning than the grouping by minimum distance. The classification of the variables does not improve significantly the classification, except with the centroid grouping.

Comparing the Dendrograms and the classes that originate with the natural conditioning, it was found that the Tagus basin is divided in five groups: Group 1 includes the water stretches with the highest **ombrogenic** influence with the research stations 18, 24, 25, 26, 27, 28, 29, 33, 34, 35, 36, 41, 46, 47 and 49. Group 2 includes the **ombrogenic water stretches with influence of soligenic** or runoff water with the research stations 7, 13, 22, 32, 43, 11 and 17. Group 3 includes water stretches with **soligenic** influence with the research stations 1, 2, 3, 8, 9, 10, 12, 15, 16, 19, 20, 21, 23, 38, 39, 40, 44, 51 and 52. Group 4 includes water stretches with **lithogenic** influence or of the lithology with the research stations 4, 5, 6, 14, 30, 31, 37, 42, 45 and 48. Group 5 includes stretches with greater influence of the runoff and of the lithology with the research station 50. Group 3 varies from a classification to another and becomes a clue for the differentiation between them.

Palabras clave: Taxonomía numérica; Dendrograma; Distancia; Clases; Jerarquía indexada; Clasificación jerárquica; Algoritmo de clasificación; Tramos; Ombrogénico; Soligénico y litogénico.

1. INTRODUCCIÓN

El desarrollo de la **taxonomía numérica**, nos ha impulsado a aplicar algunas de sus técnicas a las aguas, definidas por los **análisis hidroquímicos**, con muestras temporales y espaciales de distintas estaciones de una amplia zona, la

cuenca del río Tajo, para ver si su partición en subcuencas y tramos, se corresponde con las obtenidas con otros métodos geoquímicos.

Se trata de agrupar estaciones con características comunes, utilizando métodos de análisis multivariante, que se apoyan en métodos algebraicos como las relaciones de equivalencia y de orden.

Supongamos que para clasificar los ríos de una cuenca utilizamos los caudales como elemento separador o agrupa-

(*) Doctor Ingeniero Agrónomo, Universidad Politécnica de Madrid.

dor. Esta clasificación separa bien los ríos según sus caudales, de manera, que es de utilidad para la planificación hidrológica.

Si además del caudal consideramos su composición química y salinidad, esta clasificación de los ríos es más útil, lo que debiera ser tenido en cuenta en la planificación hidrológica a causa de su vinculación con la calidad de las aguas para la agricultura. Un sistema de análisis, como el aquí realizado contribuye a la mejora de la planificación hidrológica teniendo en cuenta los aspectos de ecología de los cultivos y de biodiversidad, y es instrumento válido para la tramitación del río, basada no sólo en el caudal, sino también en la composición y calidad de las aguas.

El estudio corresponde a la cuenca del río Tajo considerando 52 estaciones en total, correspondiendo a 28 ríos. El número de parámetros o variables analizado es 15 y el número total de datos procesados es de 9.360 para cada promedio.

2. OBJETO DE ESTUDIO

El objeto material de estudio es el agua y el objeto formal o finalidad de este trabajo es el agrupamiento en clases de las aguas de los ríos, según su similitud, en grupos, derivadas de la observación de los **dendrogramas** obtenidos por **taxonomía numérica**. Los ríos y las estaciones utilizadas se exponen en la tabla 1. Las estaciones consideradas tienen un número de la Confederación Hidrográfica del Tajo. Para nuestros estudios y para facilitar la identificación usamos un número de referencia para cada estación.

3. DATOS

Las variables son los datos de partida, los publicados anualmente por la Confederación Hidrográfica del Tajo respecto a: caudal (en m³/seg), pH, conductividad eléctrica y concentraciones en miligramos por litro de: Cl⁻, SO₄²⁻, alcalinidad (CO₃H⁻, CO₃²⁻, OH⁻), Ca²⁺, Mg²⁺, Na⁺ y K⁺. Son las variables. Para todos ellos se calculan: la media aritmética, la mediana, la media aritmética ponderada respecto al caudal, la desviación típica y el coeficiente de variación. Estos datos se transforman de mg/l a miliequivalentes por litro (me/l). La distancia euclídea no se aplica directamente sobre la concentración en miligramos por litro, ya que en este caso los datos se ponderan por el peso molecular.

A partir de los datos de cada una de las estaciones, se construye una tabla, la n^o 2, que en cada una de las 52 estaciones, considera la media aritmética, mediana y media aritmética ponderada respecto al caudal de cada variable durante un período de 30 años.

Tipificando los datos anteriores se tienen las tablas de datos que se disponen en matrices, en las que las filas son las quince variables o parámetros. Las columnas son los valores que toman las variables citadas anteriormente en cada una de las estaciones de muestreo de la Confederación Hidrográfica del Tajo consideradas.

Geométricamente, las matrices anteriores se pueden interpretar como un espacio de quince dimensiones, en el que cada eje representa un valor de una variable. En él, podemos situar cada estación, cuyas coordenadas serían los valores que toman cada una de las variables. Según este esquema, todas las variables influyen por igual cuando se trata de determinar la similitud entre dos estaciones.

Considerando que en nuestro estudio las concentraciones de iones deben tener más peso, hemos introducido dos

Nº DE REFERENCIA	Nº DE ESTACIÓN DE C.H.T.	RÍO/SITUACIÓN
1	1	Tajo en Peralejos de las Truchas
2	5	Tajo en Trillo
3	7	Tajo en Bolarque
4	11	Tajo en Aranjuez
5	14	Tajo en Toledo
6	15	Tajo en Talavera
7	19	Tajo en Embalse de Alcantara
8	30	Gallo en Vinosana
9	41	Guadaleja en Alcantud
10	45	Escabas en Priego
11	51	Jarama en Algote
12	52	Jarama en Mejorada
13	54	Guadaleja en Pezadilla
14	60	Henares en Bujalaro
15	61	Henares en Humanes
16	62	Henares en Espinillos
17	67	Sorbe en Beleña
18	69	Manzanares en Embalse de Santillana
19	70	Manzanares en Parque Sindical
20	80	Tajuba en Masegoso
21	82	Tajuba en Orusco
22	100	Guadarrama en Villaiba
23	102	Guadarrama en Burgos
24	113	Alberche en Embalse de Picadas
25	115	Alberche en Embalse de Cazalegas
26	140	Alagón en Coria
27	146	Jerte en El Torno
28	147	Jerte en Galisteo
29	148	Arroyo en Embalse de Barbalón
30	151	Tajo en Embalse de Castrejón
31	152	Tajo en Embalse de Valdecañas
32	153	Jarama en Valdepeñas de la Sierra
33	161	Tietar en Arenas de San Pedro
34	162	Ribera Gato en Mondéjar
35	168	Almorte en Morroy
36	169	Solor en Membria
37	175	Jarama en Puente Largo
38	176	Manzanares en La China
39	177	Manzanares en Vaciamadrid
40	179	Guadarrama en Navalcarnero
41	184	Tietar en La Bozagona
42	186	Tobogue en Priego
43	187	Manzanares en El Pardo
44	193	Torote en Torote
45	230	Algodar en Villamejor
46	231	Alberche en Navalunga
47	238	Arroyo en Huelga
48	239	Tajo en Fuente de la Barca
49	240	Cuerpo de Hombre en Béjar
50	241	Canamilla en Rincónada
51	274	Arroyo Vega en Alcobendas
52	275	Arroyo Valdebebas en Barajas

TABLA 1. Índice de estaciones de muestreo.

Nº	Estación	Caudal	pH	Cond	Cl ⁻	SO ₄ ²⁻	Alc	S. An	S. Cr	Ca ²⁺	Mg ²⁺	Na ⁺	K ⁺	SAR	AdSAR	ISP Ad
1	Tajo en Perales	5,68	8,0	314	1,44	1,02	3,75	6,20	6,32	3,65	1,37	1,28	0,02	0,8	1,7	2,7
2	Tajo en Trillo	12,91	7,9	536	0,97	2,00	3,87	6,84	6,75	4,18	1,72	0,82	0,03	0,5	1,1	1,7
3	Tajo en Salasque	17,14	7,8	730	0,68	5,91	2,92	9,51	9,41	6,41	2,44	0,52	0,04	0,3	0,6	0,9
4	Tajo en Aranjuez	16,85	7,8	1335	2,63	12,91	2,96	18,51	18,06	10,42	3,70	3,81	0,14	1,4	3,4	5,0
5	Tajo en Toledo	47,35	7,3	1465	3,26	12,57	3,84	19,67	19,24	9,20	4,60	5,23	0,20	2,0	4,9	7,1
6	Tajo en Talavera	47,53	7,5	1187	2,76	9,06	3,59	15,41	15,13	7,31	3,49	4,06	0,26	1,7	4,1	6,0
7	Tajo en Emb. Alcantara	68,82	7,7	361	0,90	2,24	1,26	4,40	4,38	2,14	1,01	1,15	0,08	0,9	1,3	2,1
8	Gallo en Ventosa	1,47	7,8	698	2,10	2,16	4,68	8,95	9,32	5,34	1,87	2,03	0,07	1,1	2,5	3,9
9	Guadaleja en Alcantara	10,86	7,7	657	0,57	3,45	4,43	8,45	8,43	5,72	2,17	0,51	0,03	0,3	0,6	0,9
10	Escobas en Priego	5,68	7,8	913	0,27	7,43	3,98	11,68	12,21	9,79	2,26	0,14	0,03	0,1	0,1	0,2
11	Jarama en Algarriz	1,92	7,2	481	0,47	2,99	2,64	6,10	6,14	3,99	1,41	0,68	0,06	0,4	0,8	1,3
12	Jarama en Mejorada	14,47	7,4	895	2,38	4,39	4,12	10,89	10,45	4,54	2,43	3,28	0,21	1,7	3,9	5,8
13	Guadaleja en Pesadilla	0,33	7,5	300	0,48	0,56	2,64	3,68	3,92	2,30	0,86	0,51	0,05	0,4	0,7	1,2
14	Herales en Bujalaro	4,84	7,8	1350	6,15	5,61	5,01	16,77	17,14	7,73	3,56	5,74	0,11	2,4	6,0	8,7
15	Herales en Huanca	19,66	7,6	714	2,38	9,30	2,63	8,31	8,60	4,77	1,58	2,21	0,05	1,2	2,5	3,8
16	Herales en Espinillos	5,27	7,3	968	3,10	5,15	4,35	12,60	12,06	4,92	2,82	4,11	0,21	2,1	4,9	7,1
17	Corbe en Beleso	6,58	7,5	435	0,42	3,83	1,29	5,54	5,64	4,74	0,46	0,40	0,02	0,2	0,4	0,6
18	Manzaneros en E. Sanlúcar	1,00	6,9	74	0,25	0,22	0,36	0,84	0,81	0,44	0,15	0,20	0,02	0,4	0,1	0,2
19	Manzaneros en P. Sindical	0,41	7,1	491	1,66	1,39	2,38	5,43	5,37	1,98	0,89	2,32	0,18	1,9	3,4	5,0
20	Tajuña en Masegosa	2,70	7,9	507	0,87	1,21	4,48	6,55	6,72	4,12	1,85	0,70	0,05	0,4	0,9	1,5
21	Tajuña en Orisca	4,46	7,8	767	0,58	5,04	4,43	10,04	10,31	7,48	2,34	0,42	0,06	0,2	0,5	0,7
22	Guadama en Villalba	0,83	7,2	317	0,95	0,91	1,66	3,52	3,38	1,48	0,48	1,25	0,17	1,2	1,8	2,8
23	Guadama en Bargas	3,07	7,4	574	1,43	1,28	3,31	6,02	6,11	2,47	0,98	2,45	0,22	1,9	3,6	5,4
24	Alberche en Emb. Pinedas	1,00	6,7	78	0,23	0,18	0,39	0,80	0,76	0,37	0,11	0,24	0,04	0,5	0,1	0,2
25	Alberche en Emb. Cazaleja	1,00	6,3	123	0,34	0,35	0,69	1,39	1,35	0,56	0,23	0,50	0,05	0,8	0,6	0,9
26	Alagón en Corta	91,23	7,0	90	0,36	0,15	0,52	1,02	1,00	0,40	0,20	0,38	0,04	0,7	0,3	0,5
27	Jerte en El Torno	1,70	6,9	44	0,21	0,07	0,18	0,46	0,44	0,17	0,03	0,23	0,02	0,7	0,0	0,0
28	Jerte en Galisteo	11,27	6,8	79	0,32	0,09	0,40	0,82	0,84	0,34	0,13	0,32	0,05	0,7	0,2	0,7
29	Anago en Emb. Barbalón	1,00	7,0	40	0,24	0,09	0,19	0,51	0,50	0,14	0,07	0,23	0,06	0,7	0,0	0,0
30	Tajo en Emb. Castrejón	0,97	7,6	1409	3,23	11,32	3,99	18,53	18,13	8,83	4,05	5,05	0,19	2,0	4,8	7,0
31	Tajo en Emb. Valdecañas	1,00	7,9	1066	2,56	8,58	2,85	13,99	14,06	6,61	3,34	3,96	0,17	1,6	4,0	5,9
32	Jarama en Valdepeñas Sierra	2,03	7,7	321	0,23	2,10	1,83	4,16	4,24	3,01	1,07	0,14	0,02	0,1	0,2	0,3
33	Tietar en Arenas de S. Pedro	0,90	6,9	77	0,27	0,14	0,42	0,83	0,83	0,35	0,12	0,35	0,03	0,7	0,2	0,3
34	Ribera Gata en Moraleja	2,14	6,9	104	0,41	0,14	0,61	1,16	1,17	0,35	0,33	0,42	0,07	0,7	0,4	0,7
35	Almaraz Morroy	1,43	7,1	91	0,34	0,11	0,60	1,06	1,02	0,34	0,31	0,34	0,04	0,6	0,4	0,6
36	Salor en Membrio	2,10	7,5	208	0,84	0,28	1,33	2,45	2,45	0,71	0,64	1,01	0,09	1,2	1,4	2,2
37	Jarama en Puente Largo	30,74	7,1	1300	3,17	8,32	5,05	16,54	16,85	7,22	3,17	5,01	0,44	2,1	5,3	7,7
38	Manzaneros en La China	5,03	7,3	652	1,71	2,38	2,75	6,85	6,44	2,06	1,14	2,85	0,39	2,2	4,1	6,1
39	Manzaneros en Yacamadrid	6,22	7,2	925	2,42	3,24	4,12	9,79	9,31	3,14	2,06	3,61	0,48	2,3	4,8	7,0
40	Guadama en Navacerrada	5,11	7,2	560	1,46	1,56	2,62	5,64	5,60	2,07	0,69	2,56	0,27	2,2	3,9	5,7
41	Tietar en La Bazagosa	9,17	6,8	70	0,28	0,12	0,35	0,73	0,74	0,31	0,12	0,27	0,03	0,6	0,1	0,2
42	Trabaje en Priego	0,67	7,7	1446	0,29	15,80	4,17	20,26	20,31	17,08	2,98	0,19	0,06	0,1	0,2	0,3
43	Manzaneros en El Pardo	0,13	7,2	348	1,25	0,84	1,73	3,82	3,82	1,62	0,56	1,54	0,10	1,5	2,2	3,3
44	Torote en Torote	0,20	7,3	577	1,43	1,02	4,20	6,65	6,65	2,45	1,88	2,21	0,12	1,5	3,2	4,8
45	Algodar en Villanueva	0,81	7,9	1531	3,94	13,34	3,75	21,03	21,13	10,35	5,11	5,57	0,10	2,0	4,9	7,2
46	Alberche en Navacerrada	4,27	6,5	40	0,18	0,06	0,18	0,42	0,42	0,17	0,09	0,14	0,02	0,4	0,0	0,0
47	Anago en Huelga	4,48	7,1	137	0,57	0,23	0,66	1,45	1,42	0,48	0,40	0,48	0,06	0,7	0,5	0,9
48	Tajo en Puente de la Barca	69,32	7,3	1308	2,89	10,03	3,82	16,74	16,73	8,75	3,44	4,37	0,18	1,7	4,2	6,2
49	Cuervo de Hombre en Bajar	4,30	6,9	86	0,29	0,25	0,36	0,79	0,83	0,24	0,05	0,36	0,19	0,9	0,1	0,1
50	Comanilla en Rincónada	0,04	7,3	2192	17,54	4,49	6,48	28,51	28,24	4,43	3,67	19,75	0,39	10,4	23,1	25,4
51	Arroyo Yego en Alcobendas	0,32	7,4	823	2,21	2,53	2,70	7,46	7,56	2,53	0,90	4,24	0,47	3,5	6,2	8,9
52	Arroyo Valdebebas en Barajas	0,23	7,4	636	1,65	1,62	3,27	6,55	6,63	2,87	0,74	2,72	0,30	2,1	4,0	5,9

TABLA 2. Medias aritméticas. Datos en meq/l.



FOTO 1. Sorbe, antes de la confluencia con Henares, Diciembre 1994. Clasificación: Ombracológica. Nieva.

variables que son la suma de aniones y la suma de cationes, que además nos indicará el equilibrio iónico o neutralidad eléctrica. Como en el estudio de la calidad del agua de riego tiene importancia la permeabilidad del suelo y la toxicidad iónica, se han introducido las variables SAR, SAR ajustado y ESP (porcentaje de sodio intercambiable) ajustado.

El riesgo de salinización o medida de la salinidad, se estima con la conductividad eléctrica del agua, CE, expresada en milimhos por centímetro a 25° C (mmhos/cm 25° C) o en decisiemens por metro a 25° C (dS/m 25° C). Las dos medidas son equivalentes. El SAR es la relación de absorción de sodio, que es un índice del riesgo de sodificación del suelo. El ESP es el porcentaje de sodio intercambiable del suelo en equilibrio con la solución derivada del agua de riego utilizada a largo plazo.

Las quince variables estudiadas: 1: Caudal, 2: pH, 3: Conductividad eléctrica, 4: Cl^- , 5: SO_4^{2-} , 6: Alcalinidad, 7: S.An, 8: S.Ct, 9: Ca^{2+} , 10: Mg^{2+} , 11: Na^+ , 12: K^+ , 13: SAR, 14: Adj.SAR y 15: ESP.Adj. y se han considerado tanto para las medias aritméticas como para las medianas y las medias aritméticas ponderadas con el caudal, los siguientes casos:

1. Tipificadas las 15 variables.
2. Se han tipificado nueve variables: 4: Cl^- , 5: SO_4^{2-} , 6: Alcalinidad, 7: S.An, 8: S.Ct, 9: Ca^{2+} , 10: Mg^{2+} , 11: Na^+ , 12: K^+ , es decir, las que ocupan los lugares 4 a 12.
3. Igual que el caso 1.1 pero sin tipificar las 15 variables.
4. Igual que el caso 1.2 pero sin tipificar las 9 variables.

4. MÉTODOS MATEMÁTICOS

La noción de distancia estadística junto con sus propiedades constituyen una importante herramienta, tanto en la estadística matemática como en el análisis de datos. En el primer caso porque mediante una distancia se pueden construir contrastes de hipótesis, estudiar propiedades asintóticas de estimadores, comparar parámetros, etc. En

el segundo caso, porque la distancia es un concepto muy intuitivo, que permite obtener representaciones geométricas, fáciles de entender, ofreciendo al investigador una importante ayuda para interpretar la estructura de los datos.

Mediante la distancia estadística entre individuos y poblaciones, los n individuos de una población (quedan descritos por una matriz de datos $X(n \times p)$, donde p es el número de variables estadísticas: cuantitativas, cualitativas, binarias o categóricas. El número n suele ser el tamaño de una muestra de una población, pero puede darse el caso de que Ω sea una población finita de n elementos. Una distancia $\delta_{ij} = \delta(i, j)$ entre dos individuos o elementos i, j de Ω es una medida simétrica no negativa que cuantifica la diferencia entre ambos en relación con las variables. δ queda determinada a través de la matriz de distancias

$$\Delta = \begin{pmatrix} \delta_{11} & \delta_{12} & \dots & \delta_{1n} \\ \delta_{21} & \delta_{22} & \dots & \delta_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{n1} & \delta_{n2} & \dots & \delta_{nn} \end{pmatrix} \text{ siendo } \delta_{ii} = 0, \delta_{ij} = \delta_{ji}$$

En muchas aplicaciones interesa representar el conjunto Ω con la distancia δ , es decir, (Ω, δ) , mediante el espacio geométrico modelo (V, d) , donde V es un conjunto geométrico (espacio euclídeo, variedad de Riemann, grafo, curva, etc.) y d es una distancia sobre V . Según la técnica de representación utilizada (análisis de componentes principales, análisis de coordenadas principales, análisis de proximidades, análisis de correspondencias, análisis cluster, etc.), la distancia d puede ser euclídea, ultramétrica, aditiva, no euclídea, riemanniana, etc.

En nuestro caso se ha utilizado la distancia euclídea.

Una distancia euclídea δ sobre un conjunto Ω es una aplicación de $\Omega \times \Omega$ en $\mathbb{R}$, tal que a cada par de (i, j) hace corresponder un número real $\delta(i, j) = \delta_{ij}$, cumpliendo

1. $\delta_{ij} \geq 0$
2. $\delta_{ii} = 0$
3. $\delta_{ij} = \delta_{ji}$
4. $\delta_{ij} \leq \delta_{ik} + \delta_{kj}$



FOTO 2. Heceres en Humones, junto a La Muela, Diciembre 1994. Clasificación: Soligénica I, No: medio; SO₄: bajo. Vegetación higrófila: *Phragmites communis* "cañizo". *Juncus* sp. "Junco".

5. δ_0 es euclídea, es decir, que existen dos puntos $x_i = (x_{i1}, \dots, x_{ip})^T$, $x_j = (x_{j1}, \dots, x_{jp})^T$ de R^p tales que

$$\delta_{ij}^2 = (x_i - x_j)^T (x_i - x_j)$$

es decir, δ_0 es la distancia euclídea entre los puntos x_i, x_j . Entonces (Ω, δ) puede representarse mediante el espacio euclídeo (R^n, d) .

Una distancia sobre datos cuantitativos requiere que los valores observados para p variables sobre n individuos, sean cuantitativos, formando la matriz de datos

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}$$

Entonces cada individuo i puede representarse como un punto $x_i \in R^p$. La distancia más familiar entre dos individuos i, j es la distancia euclídea

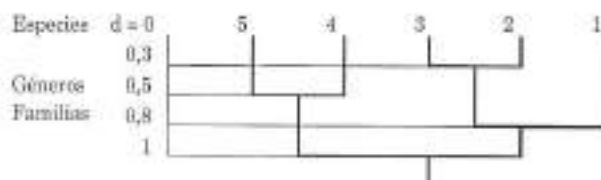
$$d_2(i, j) = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2}$$

En líneas generales, la taxonomía numérica intenta construir clasificaciones "naturales", basadas en semejanza fenotípica de los individuos (o de las clases), que se valora partiendo de una adecuada elección de un coeficiente de similitud.

El esquema de una clasificación jerárquica es el siguiente:



Una jerarquía indexada, resultado de una clasificación, está representada gráficamente por un dendrograma



La figura es una ilustración de una clasificación jerárquica de cinco especies hipotéticas.

El algoritmo fundamental conduce a una clasificación jerárquica si la disimilitud d entre los objetos es ultramétrica. Desgraciadamente en un problema real la disimilitud no es, en general, ultramétrica.

Un algoritmo de clasificación consiste en transformar razonablemente la disimilitud inicial para convertirla en ultramétrica y seguidamente poder construir la jerarquía indexada.



Los algoritmos de clasificación usados son el método del **centroide de Sokal y Michener**, cuya expresión es:

$$\bar{d}(h_k, h_i \cup h_j) = \frac{n_i}{n_i + n_j} d(h_i, h_j) + \frac{n_j}{n_i + n_j} d(h_i, h_k) - \frac{n_i n_j}{n_i + n_j} 2d(h_i, h_j)$$

donde intervienen el número de individuos n_i de h_i y n_j de h_j , y el método del **mínimo de Johnson** de expresión:

$$\bar{d}(h_k, h_i \cup h_j) = \min \{d(h_i, h_k), d(h_j, h_k)\}$$

6. RESULTADOS

Tomando como valores de las variables los distintos promedios: medias aritméticas, medianas y medias aritméticas respecto al caudal, tipificadas o no, número de variables consideradas y método de agrupamiento consideramos 24 casos. Después de estudiados y analizados los resultados obtenidos se ha llegado a la conclusión que las medias aritméticas tipificadas con las variables 4 a 12 y el agrupamiento centroide, es el que mejor representa las características ambientales sobre las aguas.

El dendrograma correspondiente está en la figura nº 3.

Observando en el dendrograma la secuencia del agrupamiento, se distinguen las siguientes clases:

Clase 1. Formada por las subclases siguientes:

1.1. Las estaciones 18, 33, 41, 24, 28, 26, 35, 25, 46, 27, 29, 47, 34 y 49, siendo la distancia ultramétrica más alta entre sus elementos 1,199, entre las estaciones 49 y 18. (**AMARILLO**).

1.2. Las estaciones 17, 11, 32, 13, 22, 43, 36 y 7, siendo la distancia ultramétrica más alta entre sus elementos 1,197, entre las estaciones 17 y 7. (**SEPIA**).

La distancia entre ambas subclases es 1,409, entre las estaciones 18 y 7.

Clase 2. Formada por las subclases siguientes:

2.1. Las estaciones 1, 20, 2, 44, 8 y 15, siendo la distancia ultramétrica más alta entre sus elementos 1,305, entre las estaciones 1 y 15. (**AZUL**).

2.2. Las estaciones 10, 9, 21 y 3, siendo la distancia ultramétrica más alta entre sus elementos 1,249, entre las estaciones 10 y 3. (**ROSA**).

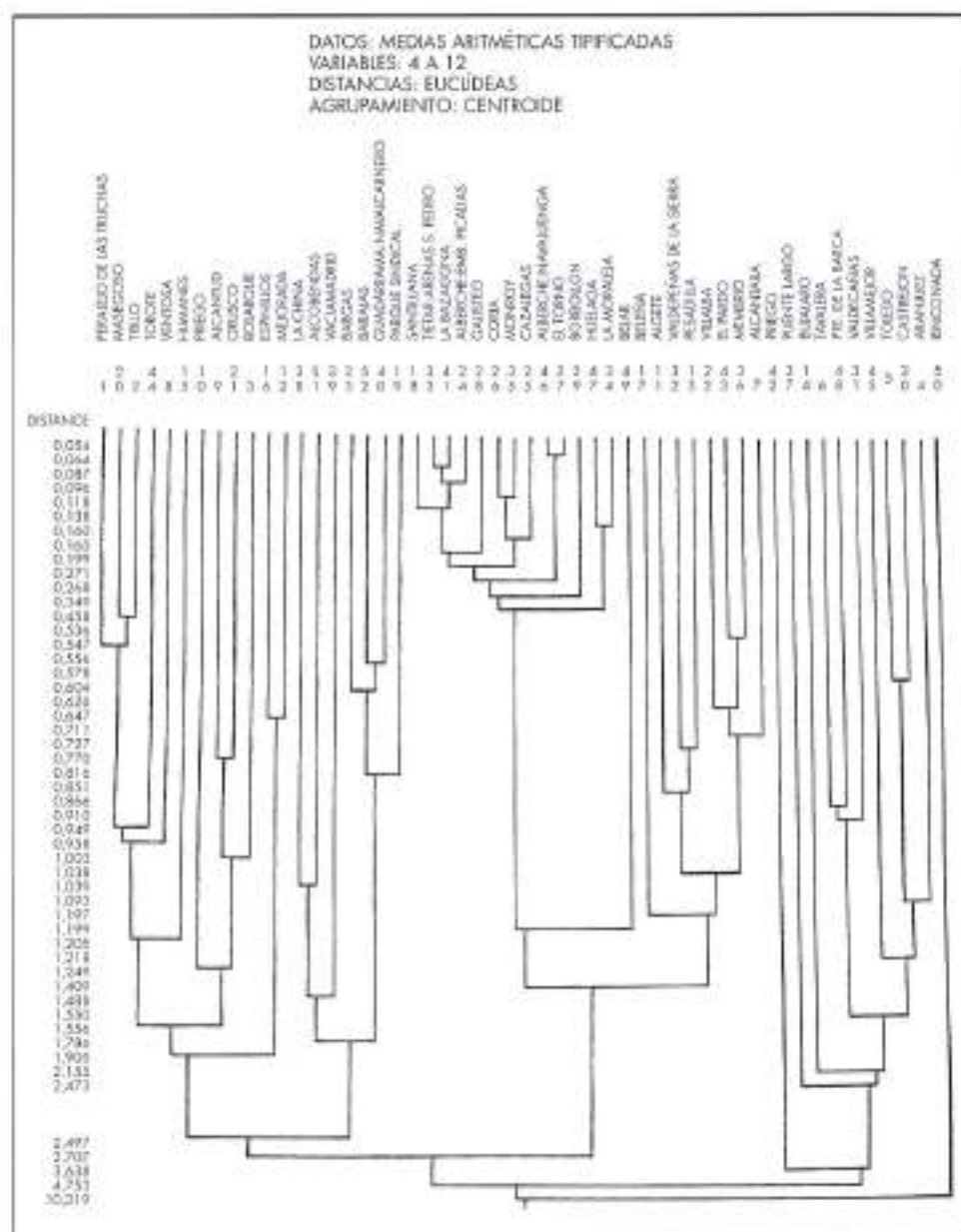


FIGURA 3.



FOTO 3. Jarama en Puente largo, cerca de Aranjuez, Diciembre 1994. Clasificación: Litomórfica I, Na^+ , Mg^{2+} , SO_4^{2-} alto. Vegetación higrófila: *Phragmites communis* "cañizo".

La distancia entre ambas subclases es 1,566, entre las estaciones 1 y 3.

2.3. Las estaciones 16 y 12, siendo la distancia ultramétrica 0,647. (VERDE-ROSA).

Siendo la distancia entre las subclases 1,905, entre las estaciones 1 y 12.

2.4. Las estaciones 38, 51, 39, 23, 52, 40 y 19, siendo la distancia ultramétrica más alta entre sus elementos 1,786, entre las estaciones 38 y 19. (VERDE).

La distancia entre las subclases es 2,497, entre las estaciones 1 y 19.

Clase 3. Formada por las estaciones 42, 37, 14, 6, 48, 31, 45, 5, 30 y 4, siendo la distancia ultramétrica más

alta entre sus elementos 3,638, entre las estaciones 42 y 4. (ROJO).

Clase 4. La estación 50. (AÑIL).

La distancia a las otras clases es 10, 319.

La clase I está formada por dos subclases:

La subclase 1.1 de origen **ombrogénico**, influenciada principalmente por la aportación de las lluvias, poco mineralizada, formada por estaciones muy homogéneas.

La subclase 1.2, más mineralizada, de origen ombrogénico con influencia del agua de escorrentía, se diferencia de la anterior en tener más elevadas las concentraciones de Ca^{2+} , Alcalinidad y SO_4^{2-} . En ésta se observan dos grupos de



FOTO 4. Tago, junto al Palacio de Aranjuez. Toma de antiguas molinos (salicó) diciembre 1994. Clasificación: Litomórfica I, Na^+ , Mg^{2+} , SO_4^{2-} alto. Vegetación: eutrofizada.



FOTO 5. Tajo en Toledo. Puente de S. Martín, diciembre 1994. Clasificación: litosoligénico II. Na^+ : Mg^{2+} , SO_4^{2-} : alto. Continuidad: espuma.

estaciones separados por las distintas concentraciones de Na^+ y Cl^- , dado por el dendrograma. Son aguas de origen **ombrosoligénico**.

La clase 2 está formada por cuatro subclases de mineralización intermedia, mesotrófica, son de origen **soligénico**, influenciada principalmente por la aportación del agua de escorrentía, con concentraciones variables para cada subclase.

La subclase 2.1 tiene, respecto de la anterior, alta concentración en Alcalinidad, Ca^{2+} y Mg^{2+} . Aguas de origen **soligénico 1**.

La subclase 2.2 tiene la más baja en Na^+ y las más altas en Ca^{2+} y SO_4^{2-} . Son aguas de origen **soligénico 2**.

La subclase 2.3 la concentración es más alta en Cl^- . Aguas de origen **soligénico 3**.

La subclase 2.4 tiene alta concentración en Na^+ y la más baja en Ca^{2+} . Son aguas de origen **litosoligénico 1**, principalmente aguas de infiltración influenciadas por las de escorrentía.

Resumiendo las cuatro subclases de la clase 2 se pueden agrupar por:

- Menos concentración en Na^+ (Subclases 2.1 y 2.2).
- = Alto contenido en Ca^{2+} y SO_4^{2-} y bajo en Cl^- (Subclase 2.2).
- Más concentración de Na^+ (Subclases 2.3 y 2.4).
- = Alto contenido en Cl^- , SO_4^{2-} , CO_3H^- y Ca^{2+} (Subclase 2.3).

Las relaciones $\text{Na}^+/\text{S.Ct}$ son homogéneas en cada subclase.

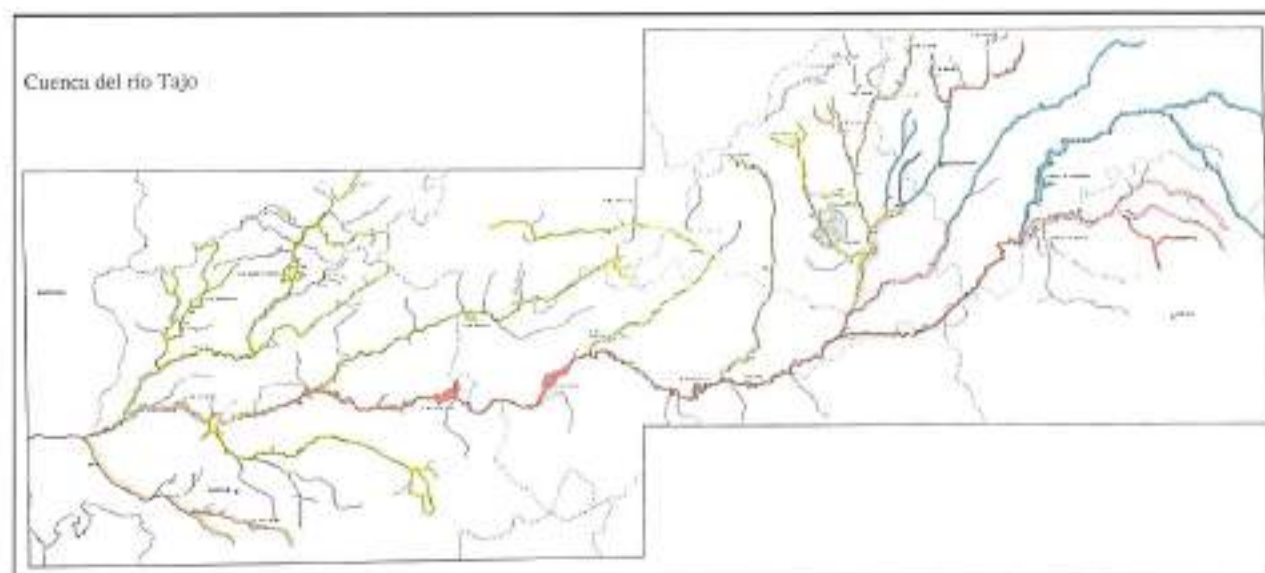


FIGURA 4. Cuenca del río Tago.

APLICACIÓN DE LA TAXONOMÍA NUMÉRICA EN LAS AGUAS DE LA CUENCA DEL RÍO TAJÓ

Order	W	Location	Gravel	pH	Cond	DT	SDI	Alc	SAc	SD	Ca ²⁺	Mg ²⁺	Na ⁺	K ⁺	SAR	Aq SAR	ESP Adj	Na/SO ₄	Na/Cl ⁻	Cl ⁻ /SO ₄ ²⁻
21	1	Tajo en Paríplejo	3.98	8.8	334	1.44	1.02	3.75	6.28	9.30	3.65	1.37	1.28	0.02	0.4	1.7	2.7	0.20	0.35	1.41
	28	Tajula en Mosagosa	2.70	7.9	507	0.87	1.21	4.46	6.55	6.70	4.12	1.85	0.70	0.65	0.4	0.9	1.3	0.10	0.17	0.72
	3	Tajo en Viljo	12.91	7.9	436	0.67	2.00	3.87	6.88	6.75	4.18	1.72	0.82	0.83	0.5	1.1	1.7	0.12	0.20	0.49
	44	Tajula en Torm	0.26	7.3	577	1.43	1.02	4.20	6.65	6.65	2.48	1.89	2.21	0.12	1.3	2.2	4.8	0.22	0.90	1.40
	8	Gallo en Ventosa	1.47	7.8	496	2.10	2.16	6.68	8.95	9.32	5.24	1.87	2.00	0.87	1.1	2.5	3.8	0.22	0.38	0.97
	15	Hercules en Hercules	19.66	7.6	714	2.38	3.30	2.63	8.21	8.80	4.77	1.58	2.21	0.95	1.2	2.5	3.8	0.26	0.46	0.72
22	10	Escobas en Fraga	3.65	7.8	812	0.27	7.43	2.98	11.69	12.21	9.79	2.26	0.14	0.03	0.1	0.1	0.2	0.01	0.01	0.04
	9	Guadalupe en Alcantad	18.86	7.7	657	0.57	3.45	4.43	8.45	8.43	5.72	2.17	0.51	0.03	0.3	0.6	0.9	0.66	0.09	0.15
	21	Tajula en Chasco	4.46	7.8	767	0.58	9.04	4.43	10.04	10.31	7.48	2.24	0.42	0.08	0.2	0.5	0.7	0.84	0.06	0.12
	3	Tajo en Boleque	17.14	7.8	738	0.68	5.91	2.92	9.51	9.41	6.41	2.44	0.52	0.04	0.3	0.6	0.9	0.86	0.05	0.12
23	18	Hercules en Espinillas	5.27	7.3	968	3.19	3.15	4.35	12.4	12.86	4.92	2.62	4.11	0.21	2.1	4.9	7.1	0.24	0.84	0.60
	12	Jarama en Mejorada	14.67	7.4	885	2.38	4.39	4.72	10.88	10.45	4.54	2.43	2.28	0.21	1.7	3.9	5.8	0.21	0.73	0.54
24	38	Marzoneros en La Chira	5.03	7.3	652	1.71	2.36	2.75	6.85	9.44	3.86	1.14	2.85	0.39	2.2	4.1	6.1	0.46	1.38	0.72
	51	Arroyo Vega en Alcobendas	0.52	7.4	620	2.21	2.55	2.70	7.46	7.16	2.95	0.39	4.24	0.47	3.5	6.2	8.9	0.56	1.66	0.87
	39	Marzoneros en Nájera	6.22	7.2	925	2.42	3.24	4.12	9.79	9.31	3.14	2.88	3.61	0.48	2.3	4.8	7.0	0.39	1.15	0.75
	23	Guadalupe en Bargas	3.07	7.4	574	1.43	1.38	3.31	6.02	6.11	2.47	0.88	2.45	0.22	1.8	3.6	5.4	0.40	0.99	1.32
	52	Arroyo Valdelebas en Bargas	8.23	7.4	456	1.63	1.62	3.27	6.95	6.63	2.87	0.74	2.72	0.20	2.1	4.0	5.9	0.41	0.95	1.32
	40	Guadalupe en Navalcarnero	9.11	7.2	560	1.46	1.58	2.62	5.84	5.60	2.07	0.49	2.56	0.27	2.2	3.8	5.7	0.46	1.24	0.94
	19	Marzoneros en P. Sindical	8.41	7.1	491	1.66	1.39	2.38	5.40	5.37	1.99	0.89	2.32	0.18	1.8	3.4	5.1	0.43	1.17	1.19</

TABLA 4. Medias aritméticas tipificadas. Variables: 4 a 12. Distancias: Euclídeas. Agrupamiento: Centroides. Datos en mg/L

Dentro de estas subclases hay grupos de diferentes concentraciones dados por los dendrogramas.

La clase 3 tiene las más altas concentraciones en todos los iones. Aguas de origen **litosoligénico 2**.

La clase 4 es la última en agruparse, tiene características peculiares. Aguas de origen **solilitigénico**, principalmente aguas de escorrentía influenciada por la litología.

Las clases y sus características están en la tabla nº 4.

El proceso de concentración de sales en clima mediterráneo, como el de la cuenca del río Tajo, provoca la separación en clases siendo correlativo el aumento de cada ión con el de la suma de todos los iones y que se refleja en la conductividad eléctrica de la solución.

De esta manera la clase 1 es oligotrófica, con baja conductividad, menor que 500 $\mu\text{S}/\text{cm}$ y la clase 3 distrófica por su elevada salinidad, mayor que 1.000.

La clase 2 es intermedia entre las clases 1 y 3 respecto al contenido total de sales, se trata de una clase mesotrófica muy influenciada por el suelo y la litología, que se manifiesta en la composición iónica: Na^+ , SO_4^{2-} , Ca^{2+} , Cl^- , resultando la clase que más varía de un método de agrupamiento a otro.

En esta clasificación se observan en el dendrograma cuatro clases, respecto a la calidad de aguas:

La clase 1, con CE inferior a 490. (**AMARILLO + SEPIA**).

La clase 2, con CE entre 500 y 970. (**AZUL + ROSA + VERDE**).

La clase 3, con CE entre 1.065 y 1.530. (**ROJO**).

La clase 4, con CE superior a 2.000. (**ANIL**).

7. CONCLUSIONES

- 1º. El agrupamiento centroides clasifica las estaciones de una manera más adecuada a las condiciones naturales que el agrupamiento por distancias mínimas.
- 2º. Comparando los dendrogramas y las clases que se originan con los condicionamientos naturales, se ha encontrado que la cuenca del río Tajo se divide en cinco grupos: El grupo 1 incluye los tramos de aguas con mayor influencia **ombrogénica** con las estaciones 18, 24, 25, 26, 27, 28, 29, 33, 34, 35, 36, 41, 46, 47 y 49. El grupo 2 incluye los tramos de aguas **ombrogénicas con influencia de agua soligénica** o de escorrentía con las estaciones 7, 13, 22, 32, 43, 11 y 17. El grupo 3 incluye tramos de aguas con influencia **soligénica** con las estaciones 1, 2, 3, 8, 9, 10, 12, 15, 16, 19, 20, 21, 23, 38, 39, 40, 44, 51 y 52. El grupo 4 incluye tramos de aguas con influencia **litigénica** o de la litología con las estaciones 4, 5, 6, 14, 30, 31, 37, 42, 45 y 48. El grupo 5 incluye tramos con mayor influencia de la escorrentía y de la litología con la estación 50. El grupo 3 varía de una clasificación a otra, resultando clave para la diferenciación entre ellas. La figura 4 incluye esta tramificación.
- 3º. Las distintas clasificaciones se diferencian principalmente según las estaciones de la subcuenca 3, que se caracterizan por tener un contenido en sales intermedio y con diferentes composición química, influyendo principalmente el SO_4^{2-} y el Na^+ .
- 4º. Los resultados revelan que el mejor agrupamiento desde la perspectiva de su relación con las condiciones ambientales de la cuenca del río Tajo se obtienen mediante las medias aritméticas tipificadas, con las variables 4 a 12 y el agrupamiento centroides.

5º. Este análisis parece válido para efectos de tramificación de los ríos de acuerdo con sus características naturales.

8. BIBLIOGRAFÍA

- ANDERBERG, M. (1973). Cluster analysis for applications. Tesis leída en la Universidad de Texas publicada por Academic Press, Londres.
- BASTIN, CH. y BENZÉCRI, J. B. (1981). Pratique de l'analyse des données, Dunod, París.
- BEALE, E. M. L. (1969). Euclidean cluster analysis. Bull. I.S.I. 43, (92-94).
- BENZÉCRI, J. B. (1973). L'Analyse des Données: 1. La Taxonomie. Dunod, París.
- BENZÉCRI, J. B. (1973). L'Analyse des Données: 2. L'Analyse des Correspondances. Dunod, París.
- BENZÉCRI, J. B. (1982). Construction d'une classification ascendente hiérarchique par la recherche en chaîne des voisins réciproques. Cahiers de l'Analyse des Données, 7 (209-218).
- BENZÉCRI, J. B. (1982). Histoire et préhistoire de l'Analyse des données. Dunod, París.
- BENZÉCRI, J. B. y BENZÉCRI, F. (1980). Pratique de l'analyse des données. Analyse des correspondances, exposé élémentaire. Dunod, París.
- BENZÉCRI, J. B. y BENZÉCRI, F. (1985). Introduction à la classification ascendente hiérarchique d'après un exemple de données économiques. Journal Soc. Stat. de París 1 (24-28).
- BERMÚDEZ, R. (1995). Aplicación de la taxonomía numérica en las aguas de la cuenca del río Tajo. Tesis doctoral. Universidad Politécnica de Madrid. E.T.S.I.A.
- BERTIER, P. y BOURUCHE, J. M. (1975). Analyse des données multidimensionnelles. P.U.F., París.
- BMDP Programs (1983). Biomedical Computer Programs, P-series, Dixon Ed., Univ. of California Press.
- BORRELL, S. (1943). Análisis de las aguas del río Tajo. Tesis doctoral. Universidad Central. Facultad de Farmacia.
- CATALÁN, J. (1965). Estudio químico-sedimentológico de algunos afluentes del Tajo. Tesis doctoral. Instituto de Edafología del Consejo de Investigaciones Científicas C.S.I.C. Madrid.
- CATALÁN, J. (1981). Química del Agua. Ed. Blume, Madrid.
- CATALÁN, J.; CABO, J. y MORA, J. (1975). Estudio de la calidad química y biológica de los ríos Manzanares, Jarama y Tajo. Reunión de estudio del medio ambiente, Madrid.
- COMISION DE LAS COMUNIDADES EUROPEAS (17.9.1975). Criterios de calidad de las aguas.
- CORMAK, R. M. (1971). A review of classification. J. Roy. Statist. Soc. A. 134 (321-367).
- CRAMER, H. (1968). Métodos matemáticos de Estadística. Aguilar, Madrid.
- CUADRAS, C. M. (1981). Métodos de Análisis Multivariante. Eunibar, Barcelona.
- CUADRAS, C. M. (1988). Distancias estadísticas. Estadística Española 30 (295 - 378).
- D.G.O.H., Resúmenes de aforos. Dirección General de Obras Hidráulicas del MOPTMA, Madrid.
- DURÁN, B. y ODELL, P. L. (1974). Cluster analysis: A survey. Springer Verlag, Berlín.
- EVERIT, B. (1980). Cluster analysis. Halsted Press, New York.
- FAO (1987). La calidad del agua en la agricultura. Roma.
- FENELON, J. P.; LEBART, L. y MORINEAU, A. (1985). Tratamiento estadístico de datos. Marcombo, Barcelona.

GASCÓ, J. M.; IBÁÑEZ, V.; GUERRERO, F.; DÍAZ ALVAREZ, M. C. (1985). Mapa de calidad de los ríos españoles en relación con su utilización en el riego. Dirección General de Obras Hidráulicas. MOPT.

GARCÍA MOUTON, E. (1988). Nuevas evaluaciones multidimensionales para el tratamiento de las encuestas en el sector agrario. Tesis leída en la E.T.S.I. Agrónomos, Madrid.

GOLUB, G. H. y VAN LOAN, C. F. (1982) Matrix Computations. The Johns Hopkins University Press, Baltimore.

IBÁÑEZ ORT, V.; NAMESNY VALLESPÍR, A. (1993) Calidad para riego de las aguas superficiales de la cuenca del río Tajo, Junta de Castilla-La Mancha, Colegio de Ing. Agrónomos Centro y Canarias.

JAMBU, M. y LEBEAUX, M. O. (1978). Classification automatique pour l'Analyse des données. Tome I. Methodes et algorithmes. Dunod, Paris.

JAMBU, M. y LEBEAUX, M. O. (1980). Classification automatique pour l'Analyse des données. Tome II. Logiciels, Dunod, Paris.

JAMBU, M. y LEBEAUX, M. O. (1983). Cluster Analysis and Data Analysis. North-Holland.

ROUANET, H. y LE ROUX, B. (1993). Analyse des Données Multidimensionnelles. Dunod, Paris.

JUDEZ, L. (1990). Técnicas de análisis de datos multidimensionales. Madrid.

LING, R. F. (1971). Cluster analysis. Tesis leída en la Universidad de Michigan.

MARDIA, K. V.; KENT, J. T. y BIBBY, J. M. (1979) Multivariate Analysis. Academic Press, Londres.

MIGUEL ARENAL, J. L. DE. (1981). Matrices y aplicaciones. S.P.E.T.S.I. Agrónomos, Madrid.

MONTES, M. (1975). Estudio de Métodos de Clasificación Cuantitativa dinámica, basados en el análisis en componentes principales. Tesis leída en la E.T.S. Ingenieros Agrónomos de la Universidad de Córdoba.

NAREDO, J. M. y GASCÓ, J. M. (1994). Las cuencas del agua en España. Dirección General de la Calidad de las aguas. MOPTMA. Madrid.

ROQUERO, C. y GASCÓ, J. M. (1976). La Calidad de las aguas en la cuenca media del Tajo. Provincia, Toledo.

SOKAL, R. R. y SNEATH, P. H. A. (1963). Numerical Taxonomy. W.A. Freeman and Company, San Francisco.

Pº DE LA CASTELLANA, 96, 7º 28046 MADRID TEL. 562 85 07 - FAX 562 82 27

Fábrica de Sabón
Polígono Industrial de Sabón
15142 Arteixo - La Coruña
Tlf. (981) 60 06 75
Fax (981) 60 13 06

AENOR

ER

Empresa
Registrada

ER-238/2/94

HUMO DE SILICE

MICROSILICE

DURABILIDAD

✓COMPACIDAD

✓IMPERMEABILIDAD

✓RESISTENCIAS

**HORMIGONES DE ALTAS
PRESTACIONES**

La Calidad es la Base de Nuestro Prestigio



ENet

*Autovia del Cantábrico. Tramo:
Castro Urdiales-Colindres (Santander)*



*Escollera en el Aeropuerto
de La Palma (Tenerife)*



Presa de La Llosa del Cavall (Lérida)



Centro Universitario CEES (Madrid)



**GINES NAVARRO
CONSTRUCCIONES, S.A.**
Cuestión de Eficacia