Ciencias Marinas (2001), 27(2): 175–208 http://dx.doi.org/10.7773/cm.v27i2.467

A CRITICAL EVALUATION OF THE USE OF CLUSTER ANALYSIS TO IDENTIFY CONTAMINATED SEDIMENTS IN THE RÍA DE VIGO (NW SPAIN)

UNA EVALUACIÓN CRÍTICA DEL ANÁLISIS *CLUSTER* PARA IDENTIFICAR SEDIMENTOS CONTAMINADOS EN LA RÍA DE VIGO (NW ESPAÑA)

B. Rubio
M.A. Nombela
F. Vilas

Departamento de Geociencias Marinas y Ordenación del Territorio
Facultad de Ciencias
Universidad de Vigo
36200 Vigo, España

Recibido en agosto de 2000; aceptado en febrero de 2001

ABSTRACT

The indiscriminate use of cluster analysis to distinguish contaminated and non-contaminated sediments has led us to make a comparative evaluation of different cluster analysis procedures as applied to heavy metal concentrations in subtidal sediments from the Ría de Vigo, NW Spain. The use of different cluster algorithms and other transformations from the same departing set of data lead to the formation of different clusters with a clear inconclusive result about the contamination status of the sediments. The results show that this approach is better suited to identifying groups of samples differing in sedimentological characteristics, such as grain size, rather than in the degree of contamination. Our main aim is to call attention to these aspects in cluster analysis and to suggest that researchers should be rigorous with this kind of analysis. Finally, the use of discriminant analysis allows us to find a discriminant function that separates the samples into two clearly differentiated groups, which should not be treated jointly.

Key words: heavy metals, contamination, cluster analysis, discriminant analysis, subtidal sediments.

RESUMEN

El uso indiscriminado del análisis *cluster* para distinguir sedimentos contaminados y no contaminados nos ha llevado a realizar una evaluación comparativa entre los diferentes procedimientos de estos análisis aplicada a la concentración de metales pesados en sedimentos submareales de la ría de Vigo, NW de España. La utilización de distintos algoritmos de *cluster*, así como otras transformaciones de la misma matriz de datos conduce a la formación de diferentes *clusters* con un resultado inconcluso sobre el estado de contaminación de los sedimentos. Los resultados muestran que esta aproximación se ajusta mejor para identificar grupos de muestras que difieren en características sedimentológicas, tal

como el tamaño de grano, más que en el grado de contaminación. El principal objetivo es llamar la atención sobre estos aspectos del análisis *cluster* y sugerir a los investigadores que sean rigurosos con este tipo de análisis. Finalmente, el uso del análisis discriminante nos permitió encontrar una función discriminante que separa las muestras en dos grupos claramente diferenciados y que no debieran ser tratados conjuntamente.

Palabras clave: metales pesados, contaminación, análisis *cluster*, análisis discriminante, sedimentos submareales.

INTRODUCTION

Confined and semi-confined areas, such as the Galician rias, are prone to heavy metal pollution (Nombela *et al.*, 1994; Rubio *et al.*, 1995; Álvarez-Iglesias *et al.*, 2000; among others). Particularly, in the Ría de Vigo, Rubio *et al.* (2000) recently determined the distribution of heavy metals and assessed their degree of pollution by traditionally calculating geo-accumulation indexes and enrichment factors. However, in other parts of the world, a number of recent studies have made use of multivariate statistical analysis in order to identify locally elevated element concentrations in marine sediments (González and Torres, 1990; Giordano *et al*., 1992; Cortesao and Vale, 1995; Chen *et al*., 1997; Emmerson *et al.*, 1997; DelValls and Chapman, 1998; DelValls *et al.*, 1998; Angelidis and Aloupi, 2000; etc.). These anomalies are usually attributed to anthropogenic contamination. The most common approach involves the use of cluster analysis (CA) to identify groups of samples by their similarity. This is followed by a comparison of the average elemental composition of the different clusters.

Nevertheless, CA can vary substantially, depending on the similarity coefficients and clustering algorithms used (Romesburg, 1990; Swan and Sandilands, 1995), and the use of unstandardized or standardized data.

There are three main categories of similarity measurement: (1) distance coefficients, (2) correlation coefficients and (3) association

INTRODUCCIÓN

Las áreas confinadas y semiconfinadas, tal como las rías gallegas, son propensas a contaminación por metales pesados (Nombela *et al.*, 1994; Rubio *et al*., 1995; Álvarez-Iglesias *et al*., 2000; entre otros). En particular, en la ría de Vigo, recientemente Rubio *et al.* (2000) han determinado la distribución de metales pesados y valorado su grado de contaminación de modo tradicional, calculando índices de geo-acumulación y factores de enriquecimiento. Sin embargo, en otras partes del mundo, gran número de estudios recientes han usado el análisis estadístico multivariante para identificar concentraciones anómalas de metales en sedimentos marinos (González y Torres, 1990; Giordano *et al*., 1992; Cortesao y Vale, 1995; Chen *et al*., 1997; Emmerson *et al.*, 1997; DelValls y Chapman, 1998; DelValls *et al.*, 1998; Angelidis y Aloupi, 2000; etc.). Estas anomalías generalmente se atribuyen a contaminación antropogénica. La estrategia más común implica el uso del análisis *cluster* (CA) para identificar grupos de muestras por su similitud y comparar posteriormente los valores medios de los diferentes grupos o *clusters*.

Sin embargo, un CA puede variar sustancialmente dependiendo de los coeficientes de similitud y los algoritmos de agrupamiento usados (Romesburg, 1990; Swan y Sandilands, 1995), así como del uso de datos estandarizados o no estandarizados.

Existen tres categorías principales de medidas de similitud: (1) coeficientes de

coefficients. The latter are restricted to binary data; in fact, CA was primarily developed only for these types of data. However, the simplicity in the calculation with the increased number of statistical packages has led to the indiscriminate application of CA for any type of data. Correlation coefficients are usually preferred for *R*-mode analysis (Simeonov and Andreev, 1989; Badarudeen *et al*., 1996; Karbassi and Nadjafpour, 1996). Most researchers use Euclidean distances as the similarity coefficient, mainly in *Q*-mode analysis (Barreiro-Lozano *et al*., 1988; Albani *et al.*, 1989; Macías-Zamora *et al.*, 1999; Angelidis and Aloupi, 2000; among others), although numerous researchers do not indicate the similarity coefficient used in their studies (González and Torres, 1990; Giordano *et al.*, 1992; Huang *et al*., 1994; Carral *et al.*, 1995; Segovia-Zavala *et al*., 1998; Ruíz *et al.*, 1998).

Regarding the clustering algorithms, the methods differ in the way that the similarity between such new clusters and all the other objects/clusters should be calculated. Unweighted pair-group method arithmetic averages (UPGMA) is the method most researchers favour in their applications (Karbassi and Nadjafpour, 1996; Bebianno and Machado, 1997; Chen *et al.*, 1997; Angelidis and Aloupi, 2000). However, other clustering methods are also useful, although some researchers do not indicate the algorithm used in their studies (Carral *et al*., 1995; Ruíz *et al.*, 1998; among others). The main alternatives to UPGMA are the single linkage clustering method (SLINK), also known as nearest neighbour linkage and recently used by Giordano *et al*. (1992) and Segovia-Zavala *et al*. (1998), and the complete linkage clustering method (CLINK), also known as furthest neighbour linkage and recently used by Macías-Zamora *et al*. (1999) and Soares *et al.* (1999).

Finally, some authors (e.g., Chen *et al.*, 1997; Emmerson *et al.*, 1997) consider that

distancia, (2) coeficientes de correlación y (3) coeficientes de asociación. Estos últimos se restringen sólo a datos binarios; de hecho, el CA fue inicialmente desarrollado sólo para este tipo de datos. Sin embargo, la simplicidad en el cálculo de estos análisis, debido a un gran incremento en el número de paquetes estadísticos, ha conducido a una aplicación indiscriminada de CA para cualquier tipo de datos. Los coeficientes de correlación se usan generalmente para análisis en modo *R* (Simeonov y Andreev, 1989; Badarudeen *et al*., 1996; Karbassi y Nadjafpour, 1996). Para los análisis en modo *Q*, la mayoría de los investigadores usan principalmente distancias euclidianas como coeficiente de similitud (Barreiro-Lozano *et al*., 1988; Albani *et al.*, 1989; Macías-Zamora *et al*., 1999; Angelidis y Aloupi, 2000; entre otros), si bien numerosos investigadores no indican el coeficiente de similitud usado en sus estudios (González y Torres, 1990; Giordano *et al.*, 1992; Huang *et al*., 1994; Carral *et al.*, 1995; Segovia-Zavala *et al*., 1998; Ruíz *et al.*, 1998).

En relación con los algoritmos de agrupamiento, los métodos difieren en el modo en que se calcula la semejanza entre los grupos nuevos y los restantes objetos/grupos. Aunque algunos investigadores no indican el algoritmo matemático usado en sus estudios (Carral *et al*., 1995; Ruíz *et al*., 1998; entre otros), el método de las medias aritméticas no ponderadas (UPGMA, del inglés) es el que más suelen usar los investigadores en sus aplicaciones (Karbassi y Nadjafpour, 1996; Bebianno y Machado, 1997; Chen *et al*., 1997; Angelidis y Aloupi, 2000). Sin embargo, se emplean otros métodos de agrupamiento, como el método de las distancias mínimas o del vecino más próximo (SLINK, del inglés), que ha sido utilizado por Giordano *et al.* (1992) y Segovia-Zavala *et al.* (1998), y el método de las distancias máximas o del vecino más lejano (CLINK, del inglés), recientemente usado por

the raw data must be standardized before execution of clustering. Although there is no agreement between authors about the best way of carrying out the standardization, logarithmic (Giordano *et al*., 1992) and additive logarithmic transformations (Emmerson *et al.*, 1997) and *z*-transformation (Simeonov and Andreev, 1989) are the most widely used methods.

The main aim of this study is to determine the validity and variability of the information obtained with the different ways of running a CA. We will establish a comparison with the pre-existent contamination data for the study area by Rubio *et al.* (2000). Also, this contribution tries to show researchers the different conclusions that could be obtained depending on the CA chosen. Finally, due to the recommendation of other researchers to combine multivariate techniques (Ratha and Sahu, 1993; Emmerson *et al.*, 1997), the information obtained with CA is compared to a discriminant analysis (DA) and a principal components analysis (PCA).

MATERIALS AND METHODS

Study area

The Rías Baixas of Galicia are fault-bounded depressions drowned by the sea during the Flandrian transgression (18,000 years ago), forming elongated coastal embayments. The Ría de Vigo is one of the classical rias originally described by Richthofen (1886) when he introduced the term in the geological literature. It is approximately 33 km long and gradually narrows landwards, from 10 km near the mouth to 0.6 km near the Rande Strait. Its outer parts are sheltered from the open sea by the Cíes Isles (fig. 1). The area has a warm, humid climate, with average temperatures of 18.5ºC in summer and 9.5°C in winter. Rainfall is approximately 1000 mm yr^{-1} and is heaviest in winter. The ria is mesotidal, with an average

Macías-Zamora *et al.* (1999) y Soares *et al.* (1999).

Por último, algunos autores (e.g., Chen *et al*., 1997; Emmerson *et al*., 1997) consideran que los datos de partida deben ser estandarizados antes de ejecutar el CA. Aunque tampoco existe acuerdo entre autores sobre el mejor modo de llevar a cabo la estandarización, los métodos más extendidos son transformaciones logarítmicas (Giordano *et al*., 1992), transformaciones logarítmicas aditivas (Emmerson *et al*., 1997) o transformación normal tipificada *z* (Simeonov y Andreev, 1989).

El objetivo principal de este trabajo es determinar la validez y la variabilidad de la información obtenida mediante los distintos modos de realizar un CA. Se comparará con la información existente sobre las zonas contaminadas en el área de estudio (Rubio *et al*., 2000). Además, este trabajo pretende hacer reflexionar a los investigadores sobre las diferentes conclusiones que podrían ser obtenidas dependiendo del tipo de CA realizado. Finalmente, se revisa y compara también esta información con la que nos aportan otras técnicas, como el análisis discriminante (DA) y el análisis de componentes principales (PCA), ya que numerosos autores recomiendan la utilización del CA en combinación con otras técnicas multivariantes (Ratha y Sahu, 1993; Emmerson *et al*., 1997).

MATERIAL Y MÉTODOS

Área de estudio

Las rías Baixas gallegas son depresiones limitadas por fallas e invadidas por el mar durante la transgresión Flandriense (hace 18,000 años), conformando bahías costeras de morfología alargada. La ría de Vigo es una de las clásicas rías descrita originalmente por Richthofen (1886) cuando introdujo este término en la literatura geológica. Tiene

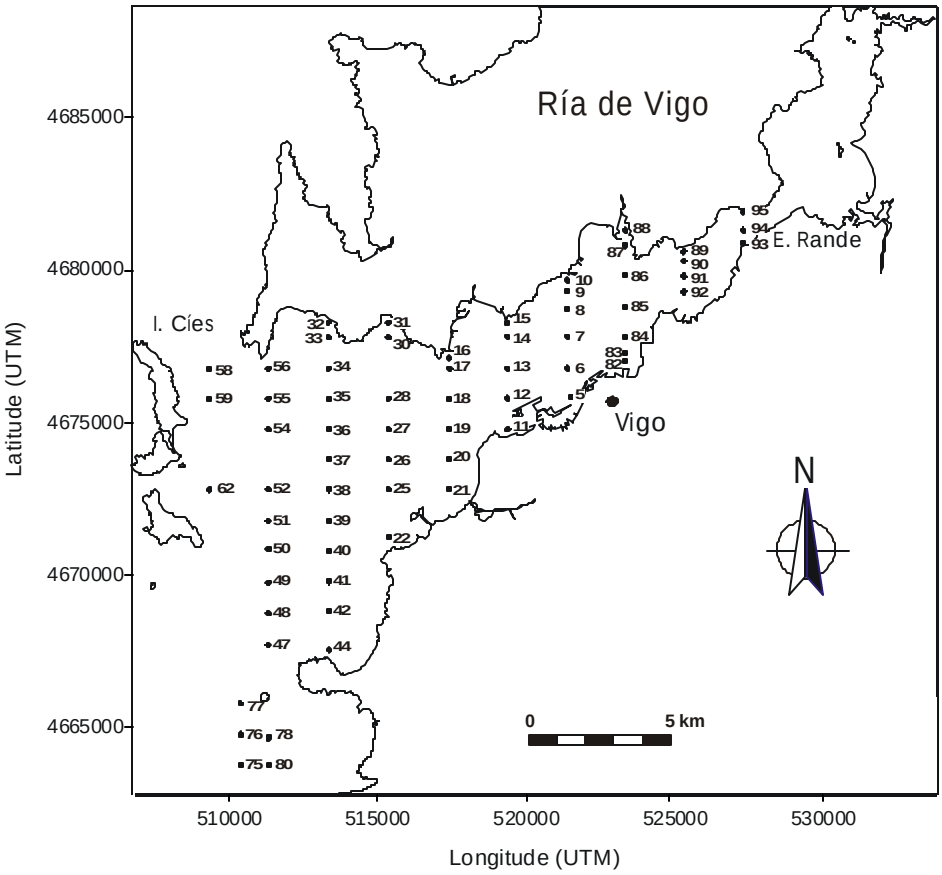

Figure 1. Study area and sampling locations.
Figura 1. Área de estudio y localización de las muestras.

tidal range of 2.2 m. The geology of the area is dominated by igneous and metamorphic rocks of Precambrian to Paleozoic age. The only younger rocks comprise Miocene and Quaternary continental sediments. There are no anomalous heavy metals in the metamorphic rocks (Guitián, 1992), i.e., the only important source of trace metals can be anthropogenic due to the increased establishment in the last decades of important centres of population and industrial activity along the ria's margins, as usually occurs in these shoreline environments.

aproximadamente 33 km de largo y se estrecha gradualmente hacia tierra, desde 10 km cerca de la boca a unos 0.6 km cerca del Estrecho de Rande, estando su zona externa protegida de mar abierto por las Islas Cíes (fig. 1). El área experimenta un clima templado-húmedo, con temperaturas medias en verano de 18.5°C y en invierno de 9.5°C, y precipitaciones cercanas a los 1000 mm año^{-1}, siendo más intensas en invierno. La ría es mesomareal, con un intervalo mareal medio de 2.2 m. La geología del área está dominada por rocas ígneas y metamórficas de edades comprendidas desde el

The floor of the outer Ría de Vigo is covered with mixed siliciclastic and skeletal gravels (with a sandy or muddy component) that, along the axis of the ria, pass landwards into sands and clayey silts. Towards the shoreline the sediments become coarser, grading through various intermediate sediment types into clean carbonate skeletal sands or mixed carbonate siliciclastic sands (Vilas *et al.*, 1995). In the inner parts, fine-grained and, usually, organic-rich sediments persist up to the shoreline. This is the result of a progressive change in hydrodynamic conditions along the ria. The outer parts are affected by severe storms in winter and by upwelling processes in summer, whilst the inner parts have an estuarine character throughout the year. The upwelling produces a marked increase in the biological productivity in the ria and, consequently, these sediments typically have very high contents of organic matter.

Field sampling and laboratory analyses

Sixty-seven samples of surficial subtidal sediments were collected with a Shipeck drag from the floor of the main Ría de Vigo (fig. 1). Table 1 lists mean values and the range for the main sedimentological characteristics. Organic carbon was determined by wet digestion, calcium carbonate with a Bernard calcimeter and particle-size distribution by dry sieving and the pipette method. All determinations were performed using the procedures described by Guitián and Carballas (1976). Sediment properties are in concordance with the sedimentological setting described previously in detail by Vilas *et al.* (1995). Nevertheless, the spatial variation in surficial grain size is shown in figure 2 to facilitate the understanding of the data. In order to simplify the granulometrical nomenclature we have divided the samples into two groups: group M (muddy samples), composed of samples that have a predominance of

Precámbrico al Paleozoico. Los únicos materiales más jóvenes se corresponden con sedimentos continentales Miocenos y Cuaternarios. Las rocas metamórficas no tienen concentraciones anómalas de metales pesados (Guitián, 1992), por lo que la única fuente de metales traza puede ser de tipo antropogénico debido a un incremento en el establecimiento de importantes centros de población y actividad industrial en los márgenes de la ría, tal como generalmente ocurre en estos ambientes costeros.

El fondo de la parte externa de la ría de Vigo está cubierto por una mezcla de gravas siliciclásticas y bioclásticas (con una componente arenosa o fangosa), que hacia tierra, y siguiendo el eje de la ría, pasan a arenas y limos arcillosos. Hacia la línea de costa los sedimentos se hacen más gruesos, pasando por varios tipos de sedimentos intermedios hasta arenas limpias bioclásticas o mezcla de carbonatadas-siliclásticas (Vilas *et al.*, 1995). En las partes internas, los sedimentos de grano fino, generalmente ricos en materia orgánica, persisten hasta la línea de costa. Éste es el resultado de un cambio progresivo en las condiciones hidrodinámicas en la ría. Las partes externas están afectadas por intensas tormentas en invierno y por procesos de afloramiento en verano, mientras que las partes internas tienen un carácter estuarino durante todo el año. El afloramiento produce un marcado incremento en la productividad biológica de la ría y, en consecuencia, estos sedimentos tienen contenidos muy elevados de materia orgánica.

Muestreo y análisis de laboratorio

Se recogieron 67 muestras de sedimentos superficiales submareales del fondo de la ría de Vigo (fig. 1), usando una draga tipo Shipeck. En la tabla 1 se presentan los valores medios y extremos para las principales características de los sedimentos. El carbono orgánico se

Table 1. Mean values and range of the main properties of the samples.
Tabla 1. Valores medios e intervalo de las principales propiedades de las muestras.

Properties	Mean	Range
% Gravel	8.00	0.00–86.15
% Sand	46.54	0.90–98.71
% Mud	45.75	0.92–98.23
% Organic matter	3.41	0.00–9.65
% $CaCO_3$	37.53	1.37–94.26

fine sediments (<63 μm), corresponding to samples located in the inner and axial parts of the ria; and group S (sandy and biogenic calcium carbonate-rich samples), including the rest of the samples located mainly in the outer and external parts of the ria. This classification is used in the dendrograms referred to later in the paper and reflects the two main environments in the ria.

Thirteen elements (Al, Fe, Mn, Ti, Sr, Zn, Cu, Pb, Cr, Co, Ni, As and Cd) were determined by inductively coupled plasma atomic emission spectroscopy (ICP-AES) after triacid total digestion (HNO_3, HF and $HClO_4$). The digestion procedures were tested using standard reference materials (CRM 277 of the BCR, estuarine sediment), and the recovery was practically complete for most of the elements (92–98.8%); only in the case of Ni was the recovery 83%. ICP-AES has the advantage of simultaneously analysing all the metals in a single sample at low detection levels. The detection limits for trace elements were 10 μg g^{-1} for Pb and As; 2.5 μg g^{-1} for Zn, Ni, Co and Cr; 1.8 μg g^{-1} for Cd; and 5 μg g^{-1} for Cu on a sediment basis. To calibrate the equipment, standard solutions with the same acidic matrix were used. Suspensions were duplicated with results expressed as mean values in μg g^{-1} of dry sediment.

determinó por digestión húmeda, el carbonato cálcico usando un calcímetro de Bernard y la distribución granulométrica por tamizado en seco y el método de la pipeta. Todas las determinaciones se llevaron a cabo siguiendo los procedimientos descritos por Guitián y Carballas (1976). Las propiedades del sedimento están en concordancia con el contexto sedimentológico descrito previamente en detalle por Vilas *et al*. (1995). No obstante, en la figura 2 se muestra la variación espacial del tamaño de grano en superficie para facilitar la comprensión de los datos. Para simplificar la nomenclatura granulométrica, hemos dividido las muestras en dos grupos: grupo M (muestras fangosas), compuesto de muestras que tienen un predominio de sedimentos finos (<63 μm), que corresponde a las muestras de la parte interna y axial de la ría; y grupo S (muestras arenosas y ricas en carbonato cálcico biogénico), que incluye el resto de las muestras localizadas en las partes más exteriores y externas de la ría. Esta clasificación se usa luego en los dendrogramas y refleja los dos principales ambientes en la ría.

Se determinaron 13 elementos (Al, Fe, Mn, Ti, Sr, Zn, Cu, Pb, Cr, Co, Ni, As y Cd) por espectroscopía de emisión por plasma (ICP-AES) tras digestión total triácida (HNO_3, HF y $HClO_4$). Los procedimientos de digestión

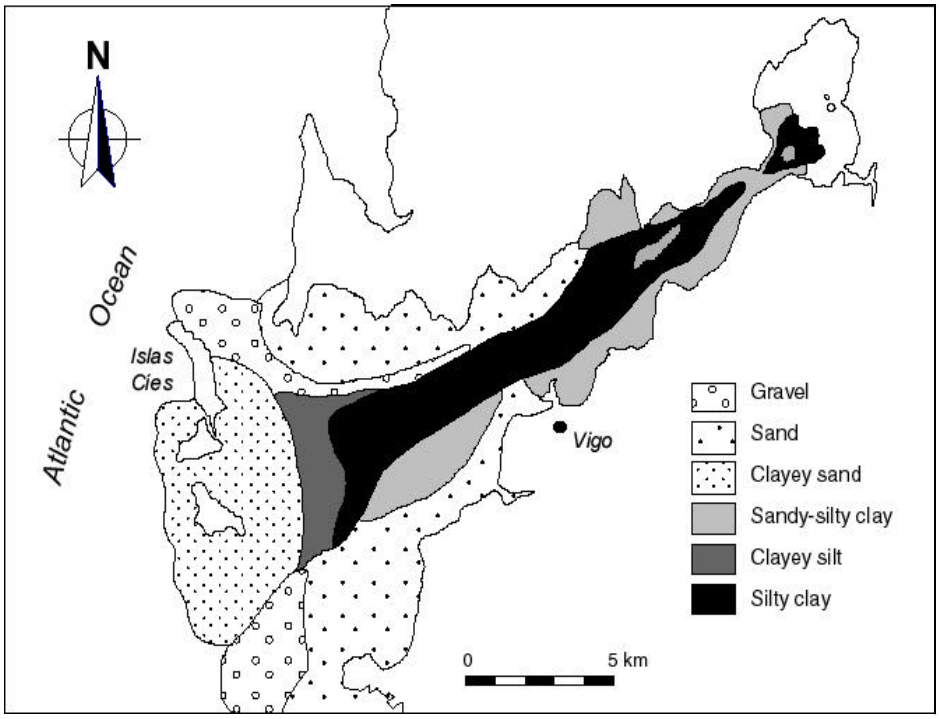

Figure 2. Surficial grain-size sediment distribution (modified from Vilas *et al*., 1995).
Figura 2. Distribución superficial de la textura del sedimento (modificado de Vilas *et al.*, 1995).

Data treatment

Statistical analyses were carried out using the SPSS package (v. 9.0). By convention, in the literature of multivariate analysis, a technique for investigating the association among objects is called a *Q*-mode and for the association of variables, an *R*-mode (Sneath and Sokal, 1973). The aim that we pursue, i.e., to identify contaminated areas, requires a *Q*-mode analysis. However, we also use *R*-mode analysis in order to compare with other multivariate techniques, such as PCA.

CA, unlike other multivariant tecniques, offers the advantage that it does not require data to be normally distributed, because it is

se validaron usando material estándar de referencia (sedimento estuarino, CRM 277 del BCR), siendo la recuperación prácticamente total (92–98.8%) para la mayor parte de los elementos, a excepción del Ni con un 83% de recuperación. ICP-AES tiene la ventaja de analizar en cada muestra simultáneamente todos los metales con límites de detección bajos. Los límites de detección para los elementos traza fueron 10 μg g^{-1} para Pb y As; 2.5 μg g^{-1} para Zn, Ni, Co y Cr; 1.8 μg g^{-1} para Cd; y 5 μg g^{-1} para Cu. Para calibrar el equipo se prepararon soluciones estándar con la misma matriz ácida; las muestras se digirieron por duplicado y los resultados corresponden a valores medios en μg g^{-1} de sedimento seco.

simply a descriptive method that does not make inferences about the background population. In order to simplify, all the *Q*-mode analysis were run using Euclidean distance as similarity coefficient, and for the *R*-mode analysis, correlation coefficients were used. Regarding the association methods, SLINK, CLINK and UPGMA are compared. The operations are performed on a similarity matrix, and all begin with the formation of the initial cluster by linkage of two objects with the greatest similarity (or least dissimilarity), and at each iteration the most similar pair of objects/clusters is linked. Once clusters have formed, they cannot be divided and can only amalgamate with others; consequently, the result is purely hierarchical. For the SLINK method, the similarity between an object and a new cluster is equal to the similarity between that object and the most similar object in the cluster. This is the minimum reasonable measure of distance for most of the objects contained. The CLINK method is exactly converse to the previous method; i.e., the apparent intercluster distances are maximized, because the similiarity between an object and a cluster (or between two clusters) is taken to be the least of all the possible similarities. Finally, for the UPGMA, the criterion for similarity recalculation is intermediate between the previous two.

An alternative approach to the analysis of the geochemical data involves the use of DA. Unlike CA, DA is a powerful statistical tool where we can test the significance of the separation between groups, Wilk's lambda being a measure of the goodness of the fit. DA is used to assign objects to one, two or more pre-established groups. The measure of difference between the means of two multivariate groups is called Mahalanobis' distance. The crucial aspect of discriminant functions is that the user chooses the group/s to be discriminated (training group). The discriminant function emphasizes the geological property of

Tratamiento de datos

Los análisis estadísticos se realizaron usando el paquete estadístico SPSS (v. 9.0). En la literatura de análisis multivariante, por convenio, una técnica para investigar la asociación entre objetos se llama modo *Q* y para la asociación entre variables se llama modo *R* (Sneath y Sokal, 1973). El objetivo que perseguimos, i.e., identificar áreas contaminadas, precisa de un análisis en modo *Q*. No obstante, en la comparación con otras técnicas de análisis multivariante, como el PCA, se han incluido también algunos análisis en modo *R.*

El CA es simplemente un método descriptivo que no hace deducciones sobre la población y, por ello, a diferencia de otras técnicas multivariantes, presenta la ventaja de que no es necesario que los datos se distribuyan normalmente. Para simplificar, los coeficientes de similitud usados han sido distancias euclidianas en todos los análisis en modo *Q* y coeficientes de correlación en modo *R*. En relación con los métodos de asociación, se comparan los siguientes: SLINK, CLINK y UPGMA. En cualquiera de los casos, las operaciones se llevan a cabo sobre una matriz de semejanza y comienzan con la formación de un grupo inicial por unión de los dos objetos con la máxima semejanza, y en cada iteración se agrupa el par de objetos más similar. El resultado es puramente jerárquico, ya que, una vez que los grupos o *clusters* se han formado, no pueden separarse y tan sólo pueden amalgamarse con otros. Con SLINK, la similitud entre un objeto y un nuevo *cluster* es igual a la similitud entre ese objeto y el objeto más similar en el *cluster*. Ésta es la medida mínima de distancia entre un objeto y un *cluster* (o entre dos *clusters*) y es, por tanto, menor que la verdadera distancia para la mayoría de los objetos. El método de CLINK es exactamente el opuesto del SLINK, i.e., maximiza las distancias aparentes entre los *clusters*, ya que la similitud entre un objeto

interest. Initially, training group data are needed for each of the categories that we wish to separate, so an initial representative datum percent is needed from each of the groups. Furthermore, no prior information is required about the variables having discriminate ability, so that variables can be included on a trial basis. The aim is to find discriminant functions: the linear equation which best separates two (or more) user-defined subgroups within the data set and allows allocation of new data to one or other of the groups on this basis.

Finally, PCA was applied in order to compare with CA in *R*-mode. PCA is a technique to find the directions of maximum variance in the data, i.e., linear compounds of correlated variables. These linear compounds are called principal components that are interpreted as factors influencing the data. More detailed information about multivariate analyses can be found in Davis, (1986), Cuadras (1991), Reyment and Savazzi (1999), among others.

RESULTS AND DISCUSSION

Metal concentrations (mean, standard deviation and extreme raw values) obtained for surface sediments of the Ría de Vigo are presented in table 2. To calculate these general statistical parameters, samples with values below detection limits were excluded from the data. The first column of the table shows the number of samples included for each element. Cd and As have high percentages of the samples, 77.6% and 58.2%, respectively, with values below the detection limit. A high degree of variability is also evident in the data, related to the wide variation in grain size of the samples.

All sediment samples were hierarchically clustered, using CA applied in *Q*-mode (grouping the stations by metal concentrations, i.e., 13 variables, 67 samples) to the raw data. Values below the limit of detection were assigned a value of half of this limit in order to

y un *cluster* (o entre dos *clusters*) se toma como la mínima similitud de las posibles. Finalmente, para el UPGMA, el criterio para el cálculo de la similitud es intermedio entre los dos anteriores.

El DA constituye una alternativa al CA empleada para el análisis de datos geoquímicos. A diferencia del CA, ésta sí es una poderosa herramienta estadística, donde podemos probar la significación de la separación entre grupos, siendo la lambda de Wilk el estadístico de la bondad del ajuste. Este análisis se usa para asignar objetos a uno, dos o más grupos establecidos. La medida de la diferencia entre las medias de dos grupos se llama distancia de Mahalanobis. El aspecto crucial de las funciones discriminantes es que el usuario elige el grupo que va a ser discriminado, enfatizándose la propiedad geológica de interés. En primer lugar, se necesita un grupo de datos para cada una de las categorías que se desea separar, por lo que se requiere que exista un porcentaje representativo de datos para cada uno de los grupos. Se van haciendo pruebas incluyendo nuevas variables, ya que no se necesita información inicial sobre cuáles son las variables que tienen capacidad discriminante. El objetivo consiste en encontrar funciones discriminantes, es decir, la ecuación lineal que mejor separa dos (o más) subgrupos del conjunto de datos definidos por el usuario y basándose en dicha ecuación permite la adjudicación de datos nuevos a uno u otro de los grupos.

Finalmente, se aplicó un PCA para comparar con los datos del CA en modo *R.* PCA es una técnica para encontrar las direcciones de máxima varianza de los datos, es decir, componentes lineales de variables correlacionadas. Estas componentes lineales se llaman componentes principales que se interpretan como factores que influyen en los datos. Información más detallada sobre las técnicas de análisis multivariante puede verse en Davis (1986), Cuadras (1991), Reyment y Savazzi (1999), entre otros.

Table 2. Metal concentrations in surface sediments from Ría de Vigo.
Tabla 2. Concentraciones de metales en sedimentos superficiales de la ría de Vigo.

	Number of samples	Minimum	Maximum	Mean	Standard deviation
Al	67	0.07	8.45	3.79	2.86
Fe	67	0.12	8.05	2.34	2.09
Ti	67	0.001	0.47	0.19	0.16
Mn	67	5.70	318.00	174.92	74.13
Sr	67	73.70	1990.00	786.16	651.06
Zn	67	6.40	274.00	91.40	74.15
Pb	58	10.40	180.00	62.31	45.64
Cu	64	2.40	85.00	27.02	22.42
Cr	61	4.20	108.00	35.08	30.44
Ni	57	2.30	67.40	24.51	14.85
Co	42	1.80	14.10	10.23	3.34
As	28	9.30	89.30	32.03	15.69
Cd	15	1.60	6.90	3.13	1.67

Units are % for Al, Fe and Ti, and μg g^{-1} for remaining elements.

avoid the exclusion of a high number of cases in the data matrix (Albani *et al.*, 1989). In the application of the CA, firstly the UPGMA method was used with squared Euclidean distance as the similarity coefficient. Figure 3 shows the dendrogram obtained for this analysis. In a first level of the hierarchy, sediments were clearly divided into two main clusters (M and S). These clusters coincide with the previous granulometrical classification. In figure 3, the number of the sample location is followed by the letter M to indicate that a sample belongs to group M, and when followed by the letter S, that it belongs to group S. A phenon line at the extreme left of the dendrogram (very high similarity at value 1 of the rescaled distance) allowed us to distinguish

RESULTADOS Y DISCUSIÓN

En la tabla 2 se presentan las concentraciones obtenidas para los metales (media, desviación estándar y valores extremos) en los sedimentos superficiales de la ría de Vigo. Para el cálculo de estos parámetros estadísticos generales, se excluyeron las muestras con valores por debajo del límite de detección. La primera columna de la tabla muestra el número de muestras incluidas para cada elemento, siendo el Cd y As los que tienen mayor porcentaje de muestras con valores por debajo del límite de detección, 77.6% y 58.2%, respectivamente. Dada la gran variabilidad en el tamaño de grano de las muestras, se evidencia un alto grado de variabilidad en los datos.

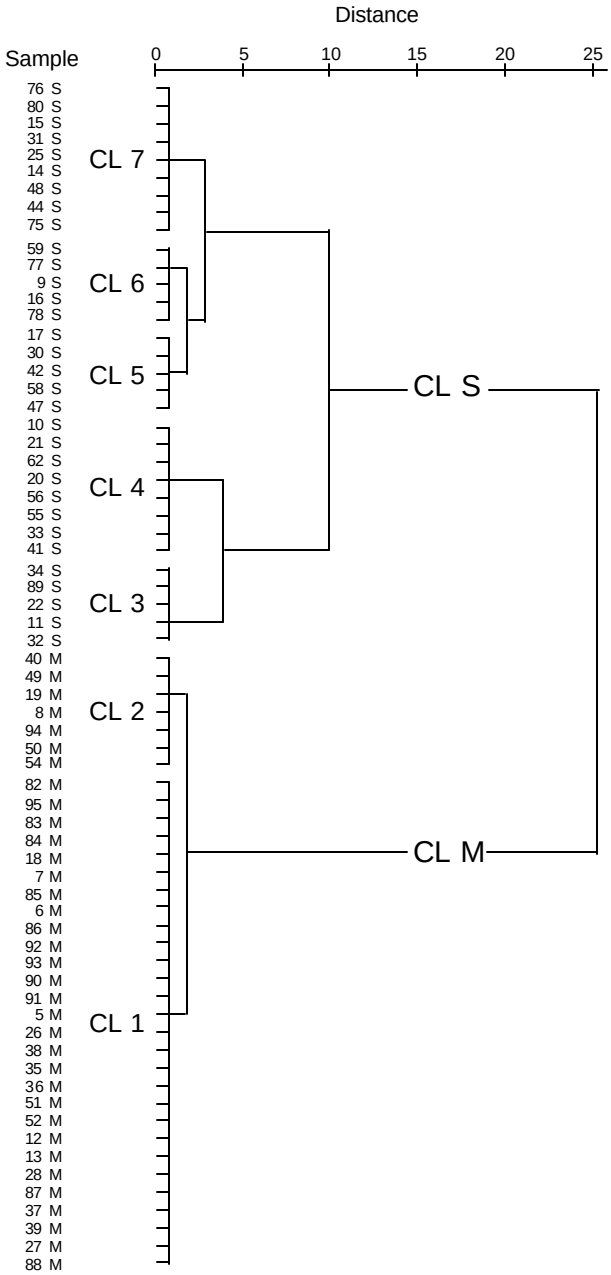

Figure 3. *Q*-mode dendrogram of the cluster analysis with the UPGMA method (non-standardized data).
Figura 3. Dendrograma en modo *Q* del análisis *cluster* realizado con el método UPGMA (datos no estandarizados).

seven clusters, designed by numbers 1 to 7. Cluster M (CL M) includes clusters 1 (CL 1) and 2 (CL 2). CL 1 comprises samples from the axial part of the ria; CL 2 also includes samples from the axial part of the ria, but more seaward than the preceding ones. Cluster S (CL S) comprises the other five clusters (CL 3, CL 4, CL 5, CL 6 and CL 7), in which the outer and external samples of the ria are included. CL3merges sandy and carbonate-rich samples from inner and middle areas in the ria, such as samples 89 and 11, respectively. CL 4 and 5 include samples located in the mid-outer part of the ria and samples located towards the shoreline. CL 6 and 7 include the outermost and external samples. Table 3 shows the mean and standard deviation for the heavy metals in these seven clusters. For most elements, the data reveal a marked decrease from CL 1 to CL7, with a strong difference for the muddy samples (CL 1 and CL 2), showing the well-known association between fine particles and heavy metal content.

In order to compare clustering algorithms, CA was again applied to the raw data, using CLINK and SLINK. The dendrograms obtained are shown in figures 4 and 5, respectively. In all the phenograms, Euclidean distance is presented rescaled to provide a better comparison. The dendrogram obtained with the CLINK method (fig. 4) is quite similar to the UPGMA method (fig. 3). Again we found two main clusters, denoted M and S as before. Although the order of the samples and the clusters has changed, the nomenclature is preserved as before to facilitate the comparison. In phenon line 1 we can also identify seven clusters and the grouping of samples is very similar. Only samples 41 and 80 appear in different clusters. The difference, however, is that now M comprises CL 1, CL 2 and CL 3, while S comprises CL 4, CL 5, CL 6 and CL 7. Some of the samples included in CL 3 have significant percentages of mud, specifically

Usando CA aplicado en modo *Q* a los datos sin transformar (agrupando las muestras por concentraciones de metales, i.e., 13 variables, 67 muestras), todas las muestras de sedimento se agruparon jerárquicamente. Para evitar la exclusión de un elevado número de casos de la matriz de datos, a los valores por debajo del límite de detección se les asignó el valor medio de dicho límite (Albani *et al*., 1989). En la aplicación del CA, en primer lugar se usó el método UPGMA con distancias euclidianas al cuadrado como coeficiente de similitud. La figura 3 muestra el dendrograma obtenido para este análisis. En un primer nivel de la jerarquía, los sedimentos se dividieron claramente en dos *clusters* principales (M y S), coincidentes con la clasificación granulométrica previa. En la figura 3, el número de la localización de la muestra seguido por la letra M indica que la muestra pertenece al grupo M y seguido por la letra S que pertenece al grupo S. Una línea (*phenon line*) en el extremo izquierdo del dendrograma (similitud muy elevada al valor 1 de la distancia) nos permitió distinguir siete clases o *clusters*, designados por números del 1 al 7. El *cluster* M (CL M) incluye las clases 1 y 2 (CL 1 y CL 2). CL 1 engloba las muestras de la parte axial de la ría y CL 2 también incluye muestras de la parte axial de la ría, pero más hacia mar adentro que las precedentes. El *cluster* S (CL S) agrupa los otros cinco *clusters* (CL 3, CL 4, CL 5, CL 6 y CL 7), en los que están incluidas las muestras más exteriores y externas de la ría. CL 3 une las muestras arenosas y carbonatadas de las zonas interna y media de la ría, como por ejemplo las muestras 89 y 11, respectivamente. CL 4 y CL 5 incluyen muestras localizadas en la parte media/externa de la ría y aquellas localizadas más próximas hacia la línea de costa. CL 6 y CL 7 incluyen las muestras más exteriores y externas. En la tabla 3 se muestra la media junto con la desviación estándar de los metales pesados para los siete *clusters*. Para la mayoría

Table 3. Comparison of means (± SD) of heavy-metal concentrations among clusters obtained using the UPGMA method (non-standardized data).
Tabla 3. Comparación de los valores medios (± DE) de las concentraciones de metales pesados entre los *clusters* obtenidos usando el método UPGMA (datos no estandarizados).

Variable	CL 1	CL 2	CL 3	CL 4	CL 5	CL 6	CL 7
Al	6.18 ± 1.46	6.30 ± 1.03	3.07 ± 1.27	1.68 ± 1.14	0.44 ± 0.23	0.56 ± 0.35	0.32 ± 0.25
Fe	4.19 ± 1.68	3.05 ± 0.37	1.09 ± 0.89	0.88 ± 0.82	0.33 ± 0.09	0.19 ± 0.07	0.35 ± 0.22
Ti	0.33 ± 0.05	0.30 ± 0.03	0.10 ± 0.05	0.06 ± 0.05	0.01 ± 0.01	0.01 ± 0.01	0.06 ± 0.10
Mn	220.61 ± 27.21	232.00 ± 15.76	157.80 ± 77.01	95.84 ± 59.15	153.60 ± 100.71	49.44 ± 20.31	149.72 ± 62.60
Sr	182.85 ± 57.84	450.57 ± 61.87	707.80 ± 81.36	1131.88 ± 106.13	1616.00 ± 49.30	1393.20 ± 48.64	1862.11 ± 91.58
Zn	150.94 ± 58.53	130.77 ± 16.87	79.28 ± 72.11	33.19 ± 23.16	17.48 ± 3.09	15.86 ± 5.13	17.02 ± 6.63
Pb	94.25 ± 42.13	62.99 ± 23.56	35.82 ± 19.41	25.55 ± 9.63	18.76 ± 9.36	17.95 ± 7.14	14.44 ± 3.46
Cu	45.64 ± 18.00	31.67 ± 5.63	15.34 ± 15.45	10.40 ± 10.60	4.22 ± 0.86	4.88 ± 2.34	4.08 ± 1.13
Cr	55.41 ± 29.51	40.48 ± 25.88	19.54 ± 13.13	14.34 ± 11.07	6.15 ± 1.53	10.80 ± 8.13	8.45 ± 4.82
Ni	34.02 ± 6.92	29.81 ± 5.16	10.26 ± 6.48	8.80 ± 8.06	20.55 ± 31.27	6.73 ± 3.69	7.33 ± 2.73
Co	11.61 ± 2.41	10.49 ± 0.59	3.24 ± 2.12	2.20 ± 1.82	1.64 ± 0.87	---	1.56 ± 0.92
As	23.23 ± 21.28	24.07 ± 13.99	---	7.01 ± 5.69	7.04 ± 4.56	8.52 ± 4.85	12.64 ± 9.99
Cd	1.25 ± 0.64	2.86 ± 2.55	---	1.10 ± 0.57	1.72 ± 1.47	1.78 ± 1.97	---

--- Less than the detection limit or constant.
Units are % for Al, Fe and Ti, and μg g^{-1} for remaining elements.

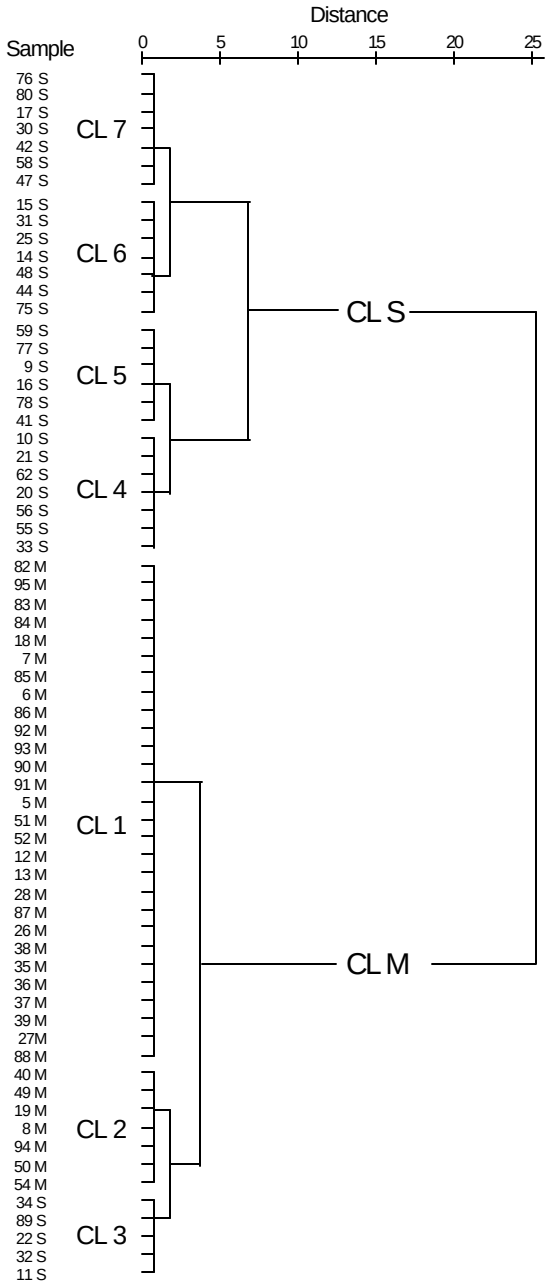

Figure 4. *Q*-mode dendrogram of the cluster analysis with the CLINK method (non-standardized data).
Figura 4. Dendrograma en modo *Q* del análisis *cluster* con el método CLINK (datos no estandarizados).

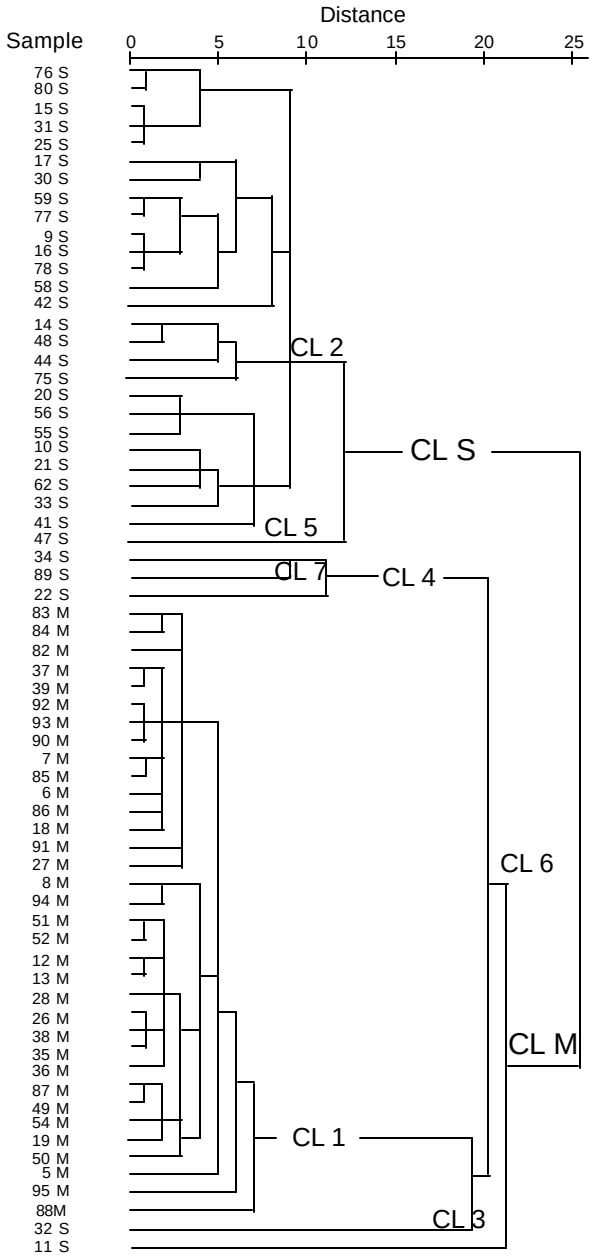

Figure 5. *Q*-mode dendrogram of the cluster analysis with the SLINK method (non-standardized data).
Figura 5. Dendrograma en modo *Q* del análisis *cluster* con el método SLINK (datos no estandarizados).

samples 34 and 89, with 28.8% and 33.7%, respectively. It is possible that the CA reflects grain size and these samples were erroneously classified as S. Nevertheless, the other three samples in CL 3 (samples 11, 22 and 32) do not exceed 3.5% of mud.

The dendrogram obtained using the SLINK method (fig. 5) is clearly different to that obtained with the UPGMA and CLINK methods. More than 50% of the samples are now classified in different clusters. Furthermore, it is more difficult to establish a single phenon line. We attempted to identify the same clusters as previously. CL M merges CL 1, CL3, CL 4, CL 6 and CL 7, while CL S merges CL 2 and CL 5. Several samples now form isolated clusters. Samples from the previous CL 3 are now included in CL 3, 4, 6 and 7. Also, sample number 47 forms CL 5. CL1 comprises all the samples considered as muddy and CL 2 the samples considered as sandy, excepting those previously mentioned. However, this classification, contrary to the previous UPGMA, does not seem to show a clear relation with polluted or non-polluted areas. The CA using this method does not actually differ from the others as much as it apparently does. The difference in appearance is due to an effect known as chaining that frequently occurs with SLINK. Chaining is a term that describes the following situation: at the first clustering step, two objects merge to form one cluster, and throughout the remaining clustering steps this cluster grows progressively larger through the annexation of lone objects that have not yet clustered. Because SLINK embodies the nearest neighbour philosophy, it is the most likely to cause chaining and tends to result in unnatural-looking hierarchical systems. Using this procedure, it is easy for points to link on to the ends of straggly, elongated clusters. Points at opposite ends of such clusters will be substantially different. For these reasons, the nearest-neighbour approach

de los elementos los datos revelan un marcado descenso desde CL 1 a CL 7, con una gran diferencia para las muestras de CL 1 y CL 2 que se corresponden con las muestras más fangosas, reflejando la ya conocida asociación entre las partículas finas y el contenido metálico.

Para comparar los algoritmos de agrupamiento (*clustering*), CA se aplicó de nuevo a los datos sin transformar, usando CLINK y SLINK. Los dendrogramas obtenidos se muestran en las figuras 4 y 5, respectivamente. En todos los dendrogramas la distancia euclidiana se presenta reescalada para facilitar la comparación. El dendrograma obtenido con el método CLINK (fig. 4) es bastante similar al del método UPGMA (fig. 3). De nuevo, encontramos dos *clusters* principales, denominados M y S al igual que antes. Aunque el orden de las muestras y los *clusters* han cambiado se mantiene la misma nomenclatura para facilitar la comparación. A la distancia fenotípica 1 podemos identificar siete *clusters* en los que las muestras se agrupan de modo similar; sólo las muestras 41 y 80 aparecen en *clusters* diferentes. La diferencia, sin embargo, es que ahora el *cluster* M engloba los *clusters* 1, 2 y 3, mientras que S comprende los *clusters* 4, 5, 6 y7. Algunas de las muestras incluidas en CL 3 tienen importantes porcentajes de fango, en concreto, las muestras 34 y 89, con un 28.8% y 33.7%, respectivamente. Es posible que el CA refleje diferencias en tamaño de grano y estas muestras se hubieran clasificado erróneamente como S. No obstante, las otras tres muestras en CL 3 (muestras 11, 22 y 32) no superan el 3.5% de fango.

El dendrograma obtenido con el método SLINK (fig. 5) es claramente diferente a los obtenidos con los métodos CLINK y UPGMA. Más del 50% de las muestras se clasifican ahora en diferentes *clusters*. Además, es más difícil establecer una única distancia fenotípica (*phenon line*). Intentamos identificar los

is among the least favoured method, even though several researchers have recently used it in their studies (Nombela *et al.*, 1998; Segovia-Zavala *et al*., 1998). The criterion for similarity calculation is not unreasonable and it should not be rejected unless there is a special reason for doubting the validity of elongated clusters. Chaining does not occur with the CLINK and UPGMA methods, producing neat, tight clusters. When choosing between the three clustering methods, most researchers select UPGMA. They consider that UPGMA tends to give higher values of the cophenetic correlation coefficient (Farris, 1969, cited by Romesburg, 1990). This means that, on average, UPGMA produces less distortion in transforming the similarities between objects into a tree.

Hughes (1979) argued that since SLINK and CLINK are based on opposite philosophies, if they eventually give the same trees for a given data matrix, then the clusters are well-defined in their attributed space and, in a sense, are real. Conversely, if the two clustering methods give entirely different trees, the clusters are only weakly defined and perhaps are artifacts of the clustering method.

Until now, the clustering of the sampling stations satisfactorily reflects the properties of the separate zones and the conditions they suggest for sedimentation and diagenesis. CA reveals information about sediment characteristics and grain size effects are predominant over contaminant effects. Obviously, physical properties and chemical characteristics of the sediments are intimately linked. These relationships and a geochemical analysis of the data are widely treated in Rubio *et al.* (2000).

In order to eliminate the effect of differences in magnitude and variance of the data, a linear transformation know as *z*-transformation was utilized for scaling the raw analytical data of each element to zero mean and unit variance (Simeonov and Andreev, 1989; Huang *et al.*,

mismos *clusters* que anteriormente. CL M une CL 1, CL 3, CL 4, CL 6 y CL 7, mientras que CL S une CL 2 y CL 5. Varias muestras ahora forman *clusters* aislados. Las muestras del CL3 anterior ahora están incluidas en CL 3, 4, 6 y 7. También, la muestra 47 forma el CL 5. CL 1 incluye todas las muestras consideradas fangosas y CL 2 las muestras consideradas arenosas, exceptuando las previamente mencionadas. Sin embargo, esta clasificación, a diferencia de la primera obtenida, no parece mostrar una clara relación con áreas contaminadas o no contaminadas. En realidad, el CA usando este método no difiere de los otros tanto como aparenta. La diferente apariencia es debida a un efecto conocido como encadenamiento (*chaining*) que ocurre con frecuencia con SLINK. El encadenamiento es un término que describe la siguiente situación: en el primer paso del agrupamiento, dos objetos se unen para formar un *cluster*; en los siguientes pasos del *cluster*, éste se va haciendo progresivamente mayor por unión de objetos aislados que todavía no habían sido agrupados. Dado que el método del SLINK incorpora la filosofía del vecino más cercano, es el más probable de los métodos que cause encadenamiento y tiende a dar como resultado sistemas jerárquicos que no parecen naturales. Con este procedimiento es fácil que determinados puntos estén ligados en los extremos de *clusters* desordenados y alargados. Los puntos de los extremos opuestos de tales *clusters* serán sustancialmente diferentes. Por estas razones, el método del vecino más cercano está entre los más desfavorecidos, aun cuando varios investigadores lo han usado recientemente en sus estudios (Nombela *et al.*, 1998; Segovia-Zavala *et al*., 1998). El criterio para el cálculo de la similitud es razonable y no debería ser rechazado a menos que haya una razón especial para dudar de la validez de los *clusters* alargados. Con los métodos CLINK y UPGMA no se produce el encadenamiento, dando *clusters* más ordenados y cerrados. En la

1994). Standardization is normally applied prior to distance calculation. This is in order to avoid interpoint distances being dominated by one variable, which happens to be measured in parts per million when the rest are percentages. CA was again applied to the standardized data,and the tree obtained (UPGMA method and squared Euclidean distance) is shown in figure 6. When standardizing, we also distinguish between two main clusters (M and S), but the difference now is that cluster S includes samples from CL 2, besides samples considered sandy. This cluster is almost the same as CL 2 in the previous analyses. Average values for heavy metals in these clusters are presented in table 4. Average values for CL 2 are closer to the ones for CL 4 and CL 5 than for CL 3, especially for typical anthropogenic metals, such as Zn, Pb, Cu and Cr. In granulometrical terms, samples for CL 2 are intermediate between the two simple classifications made (muddy or sandy). These are classified as sandy silty clay in Shepard's (1954) textural diagram. Again, the CA shows a granulometrical signal rather than an anthropogenic enrichment signal.

CL M in figure 6 could be divided into several clusters, and several samples form isolated clusters, such as samples 88 and 95. In an attempt to follow the previous classification, we only single out CL 1, 4, 5, 6 and 7, but the similarities between samples in these clusters are questionable. CL 1, 6 and 7 presented the highest average values (table 4) of anthropogenic elements (Zn, Cu and Pb), although for Cr, Ni and Co, CL 4 and CL 5 also have high values.

Grain-size effects are clearly obvious, and these probably mask the degree of contamination. For this reason, they must be eliminated by normalizing the raw values using a grain-size proxy. Rubio *et al.* (2000) showed that Al is a good indicator of particle size for the sediments studied. These authors rejected the use

elección entre los tres métodos de agrupamiento, la mayoría de los investigadores eligen UPGMA. Ellos consideran que este último tiende a dar mayores valores del coeficiente de correlación cofenética (Farris, 1969; citado por Romesburg, 1990), lo cual quiere decir que por término medio, UPGMA produce menos distorsión a la hora de transformar en un árbol las similitudes entre objetos.

Hughes (1979) argumentó que puesto que SLINK y CLINK están basados en filosofías opuestas, si dan los mismos árboles para una matriz dada entonces los *clusters* están bien definidos y, en cierto sentido, son reales. Por el contrario, si los dos métodos de *cluster* dan árboles totalmente diferentes, entonces los *clusters* están mal definidos y quizás son artefactos del método de agrupamiento.

Hasta ahora, el agrupamiento de las estaciones de muestreo refleja bastante satisfactoriamente las propiedades de las distintas zonas y las condiciones que dichas zonas sugieren en cuanto a sedimentación y diagénesis. CA revela información sobre las características del sedimento, siendo los efectos de tamaño de grano predominantes sobre los efectos de contaminación. Obviamente, las propiedades físicas y las características químicas de los sedimentos están íntimamente ligadas. Estas relaciones, así como un análisis geoquímico de los datos se tratan ampliamente en Rubio *et al.* (2000).

Para eliminar los efectos de diferencias en magnitud y varianza de los datos, se utilizó una transformación lineal conocida como transformación *z* para escalar los datos analíticos de cada elemento a media cero y varianza uno (Simeonov y Andreev, 1989; Huang *et al.*, 1994). La estandarización normalmente se aplica antes de calcular las distancias y generalmente se usa para evitar que las distancias estén dominadas por una variable que por ejemplo esté medida en partes por millón cuando las restantes están en porcentaje. Se aplicó CA de

Table 4. Comparison of means (± SD) of heavy-metal concentrations among clusters obtained using the UPGMA method (standardized data).
Tabla 4. Comparación de los valores medios (± DE) de las concentraciones de metales pesados entre los *clusters* obtenidos usando el método UPGMA (datos estandarizados).

Variable	CL 1	CL 2	CL 3	CL 4	CL 5	CL 6	CL 7
Al	6.06 ± 2.14	5.57 ± 1.87	1.11 ± 1.25	6.52 ± 0.74	6.25 ± 0.77	3.96 ± 0.28	7.57 ± 0.98
Fe	4.94 ± 2.10	2.68 ± 0.97	0.57 ± 0.63	3.39 ± 0.51	3.66 ± 0.30	7.46 ± 0.63	3.81 ± 0.34
Ti	0.36 ± 0.04	0.26 ± 0.09	0.05 ± 0.09	0.34 ± 0.04	0.38 ± 0.09	0.32 ± 0.02	0.30 ± 0.02
Mn	227.67 ± 23.13	220.43 ± 27.45	121.01 ± 75.24	237.91 ± 18.91	224.33 ± 36.20	213.00 ± 9.85	197.00 ± 15.12
Sr	164.83 ± 27.84	488.43 ± 80.49	1411.84 ± 396.81	255.82 ± 53.88	141.00 ± 13.89	161.67 ± 25.89	147.75 ± 12.63
Zn	194.17 ± 18.92	139.91 ± 131.66	25.19 ± 18.81	108.86 ± 25.08	110.67 ± 22.00	231.33 ± 33.71	161.00 ± 26.20
Pb	109.60 ± 17.50	54.21 ± 28.05	23.53 ± 13.34	61.31 ± 14.36	54.67 ± 5.93	135.00 ± 24.27	154.25 ± 19.70
Cu	55.10 ± 16.22	28.10 ± 11.18	11.42 ± 9.23	31.53 ± 5.19	32.17 ± 1.85	69.80 ± 13.32	57.18 ± 6.18
Cr	84.53 ± 18.87	36.22 ± 26.68	7.99 ± 9.23	42.95 ± 22.86	74.21 ± 7.69	6.85 ± 1.57	71.53 ± 9.30
Ni	40.62 ± 6.97	25.74 ± 11.04	11.06 ± 14.00	31.84 ± 5.03	32.73 ± 2.53	36.03 ± 2.25	34.15 ± 1.59
Co	12.87 ± 1.22	10.35 ± 0.51	4.43 ± 1.21	11.83 ± 1.33	12.50 ± 0.20	12.03 ± 0.65	9.35 ± 5.05
As	53.20 ± 21.65	31.70 ± 6.26	19.08 ± 6.06	26.80 ± 15.48	36.50 ± 0.57	37.70 ± 1.11	---
Cd	2.13 ± 0.15	2.73 ± 2.63	1.23 ± 1.01	---	---	---	---

--- Less than the detection limit or constant.
Units are % for Al, Fe and Ti, and μg g^{-1} for remaining elements.

of both Fe and organic carbon as grain-size proxies, since they considered them as contaminants in the study area. However, the application of CA for metal/Al data (fig. 7) did not lead us to a conclusive result about the status of contamination; 82% of all the samples are now merged in CL 1, which comprises both sandy samples and muddy samples with different degree of contamination. Figure 7 shows CA with UPGMA; however, we again obtained different clusters depending on the algorithm used. The CA must be interpreted with caution and it should be combined with the calculation of other types of enrichment or contamination indexes or factors.

Other researchers have also combined the use of CA with other multivariate analysis. Bradfield and Orlóci (1975), Green and Vascotto (1978) and Emmerson *et al.* (1997) combined CA with DA. They first used CA to make a classification and then, with the same data, examined whether another method, such as DA, could identify the objects in the identified classes. They took the percentage of correct predictions as a measure of the validity of the classification.

In our case, the groups established were those mentioned previously for groups M and S. Prior to the DA, tests for equality of population centroids (Rao, 1965) were carried out. It was inferred from Wilk's lambda value (0.087) and the chi-square value (142.59) that there is a significant difference ($P < 0.0001$) in population centroids. The data and the histogram obtained from the DA are presented in table 5 and figure 8, respectively. During classification of the samples, 100% of the grouped cases were correctly classified. This indicates that it was possible to discriminate between the external and internal areas in the ria and the two main types of sediment that exist in the ria. This is also supported by the difference between the average values of group centroids (table 5). Nevertheless, in table 5 we

nuevo a los datos estandarizados y el árbol obtenido (método UPGMA y distancia euclidiana al cuadrado) se muestra en la figura 6. Al estandarizar, también distinguimos dos *clusters* principales (M y S), pero la diferencia ahora es que en el *cluster* S, además de las muestras consideradas como arenosas, están incluidas muestras de CL 2. Este *cluster* es casi el mismo que el CL 2 de análisis previos. En la tabla 4 se presentan los valores medios para los metales en estos *clusters*, siendo los valores para CL2más cercanos a los de CL 4 y CL 5 que a los de CL 3, especialmente para los típicos metales antropogénicos, tales como Zn, Pb, Cu y Cr. En términos granulométricos, las muestras para CL 2 son intermedias entre las dos simples clasificaciones hechas (fangosa o arenosa). Éstas se clasifican como arcilla limo-arenosa en el diagrama textural de Shepard (1954). De nuevo, CA muestra una señal granulométrica más que una señal de enriquecimiento antropogénico.

En la figura 6, CL M podría ser dividido en varios *clusters* y varias muestras formarían *clusters* aislados, tales como las muestras 88 y 95. En un intento por seguir la clasificación previa hemos sólo aislado CL 1, 4, 5, 6 y 7, pero las similitudes entre muestras en estos *clusters* son cuestionables. CL 1, 6 y 7 presentaron los valores medios más elevados (tabla 4) para los elementos antropogénicos (Zn, Cu y Pb), aunque CL 4 y CL 5 también tienen elevados valores para Cr, Ni y Co.

Los efectos de tamaño de grano son claramente visibles y éstos probablemente enmascaran el grado de contaminación. Por ello, deben eliminarse normalizando los valores usando un sustituto del tamaño de grano. Rubio *et al.* (2000) demostraron que para los sedimentos estudiados, el Al es un buen indicador del tamaño de partícula. Estos autores descartan tanto el uso del Fe como el del carbono orgánico como normalizadores debido a la contaminación existente por éstos en el área de

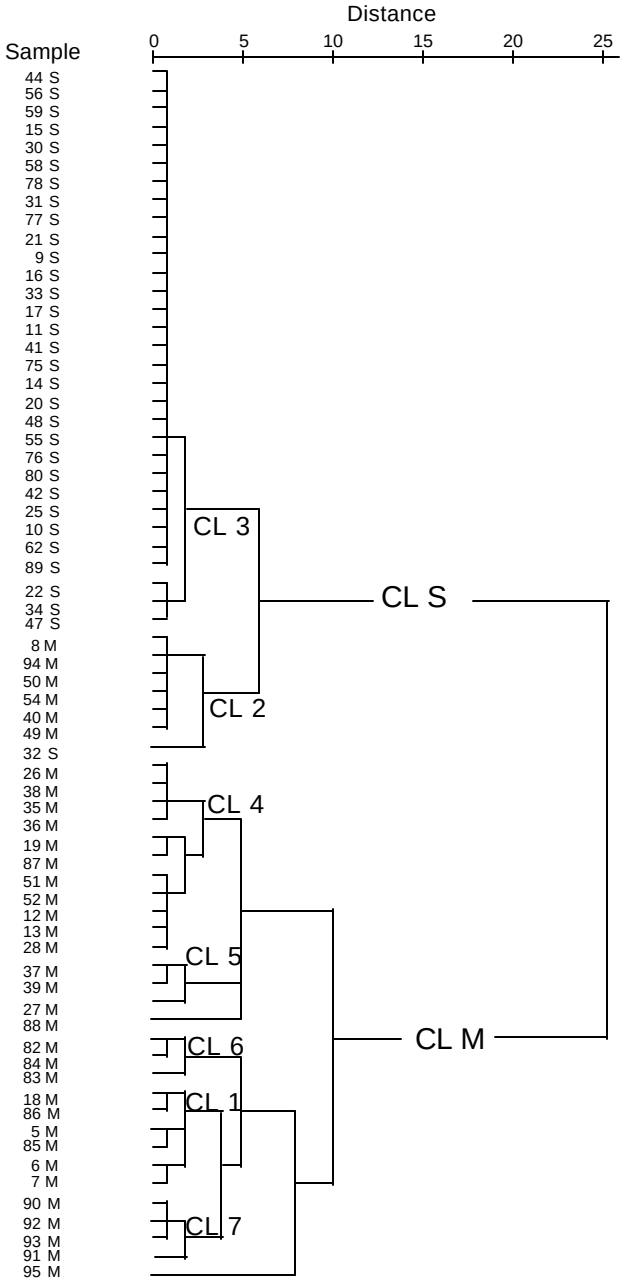

Figure 6. *Q*-mode dendrogram of the cluster analysis with the UPGMA method (standardized data).
Figura 6. Dendrograma en modo *Q* del análisis *cluster* con el método UPGMA (datos estandarizados).

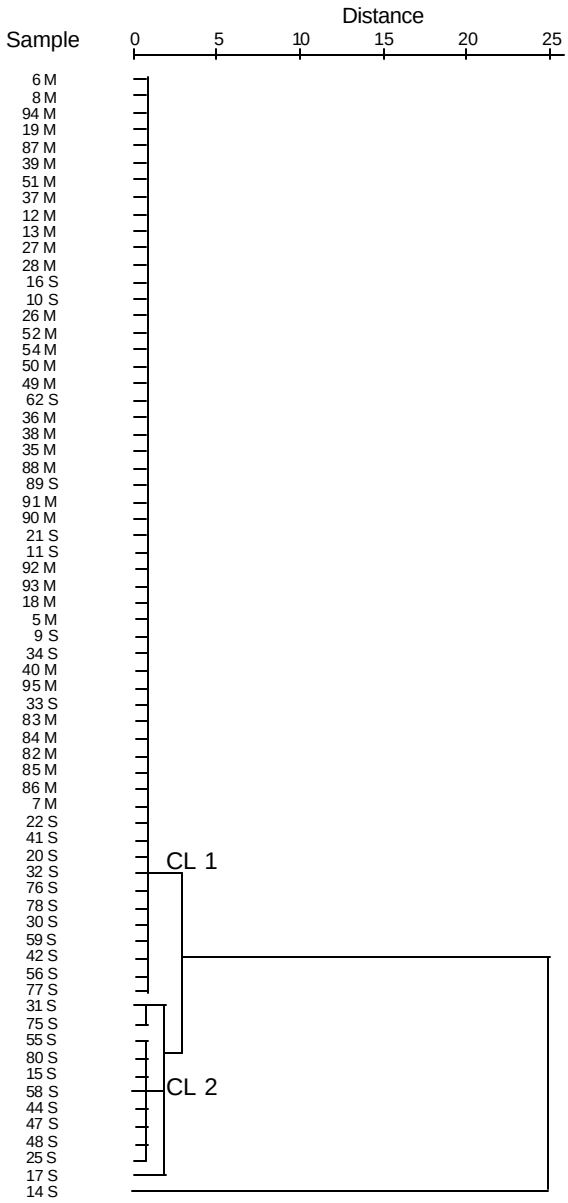

Figure 7. *Q*-mode dendrogram of the cluster analysis with the UPGMA method for normalized data with Al (metal/Al).
Figura 7. Dendrograma en modo *Q* del análisis *cluster* con el método UPGMA para los datos normalizados por Al (metal/Al).

Table 5. Discriminant analysis for sediments from Ría de Vigo.
Tabla 5. Análisis disciminante para los sedimentos de la ría de Vigo.

(a) Canonical discriminant function.

Function	Eigenvalue	% of variance	Canonical correlation	Wilk's lambda	Chi-square	Degrees of freedom	Significance
1	10.443	100.00	0.955	0.087	142.587	13	0.0000

(b) Standardized canonical discriminant function coefficients.

Variable	Function 1
Al	0.213
As	–0.115
Cd	0.188
Co	0.480
Cr	0.144
Cu	–0.137
Fe	0.239
Mn	0.087
Ni	–0.004
Pb	0.178
Sr	–0.268
Ti	0.235
Zn	0.041

(c) Canonical discriminant functions evaluated at group means (group centroids).

Group	Function 1
Muddy (M)	3.043
Sandy (S)	–3.329

Table 5 (Cont.)

(d) Classification results.

			Predicted		
		Group	Muddy	Sandy	Total
Actual group	Number of cases	Muddy	35	0	35
		Sandy	0	32	32
	%	Muddy	100.0	0.0	100.0
		Sandy	0.0	100.0	100.0

(e) Classification output.

			Highest group			2nd highest group			Discriminant scores
Sample	Actual group	Predicted group	P(D/G) probability	P(G/D)	Square Mahalanobis' distance	Group	P(G/D)	Square Mahalanobis' distance	Function 1
5	M	M	0.500	1.000	0.454	S	0.000	49.649	3.717
6	M	M	0.230	1.000	1.439	S	0.000	57.331	4.243
7	M	M	0.432	1.000	0.617	S	0.000	51.236	3.829
8	M	M	0.442	1.000	0.592	S	0.000	31.395	2.274
9	S	S	0.688	1.000	0.161	M	0.000	45.881	–3.730
10	S	S	0.230	1.000	1.442	M	0.000	26.744	–2.128
11	S	S	0.630	1.000	0.232	M	0.000	34.700	–2.847
12	M	M	0.826	1.000	0.048	S	0.000	43.460	3.264
13	M	M	0.913	1.000	0.012	S	0.000	39.230	2.935
14	S	S	0.319	1.000	0.994	M	0.000	54.310	–4.326
15	S	S	0.301	1.000	1.071	M	0.000	54.869	–4.364
16	S	S	0.547	1.000	0.364	M	0.000	48.655	–3.932
17	S	S	0.828	1.000	0.047	M	0.000	43.422	–3.546
18	M	M	0.278	1.000	1.178	S	0.000	55.617	4.129
19	M	M	0.881	1.000	0.022	S	0.000	42.540	3.193
20	S	S	0.776	1.000	0.081	M	0.000	44.321	–3.614

Table 5e (Cont.)

			Highest group			2nd highest group			Discriminant scores
Sample	Actual group	Predicted group	P(D/G) probability	P(G/D)	Square Mahalanobis' distance	Group	P(G/D)	Square Mahalanobis' distance	Function 1
21	S	S	0.891	1.000	0.019	M	0.000	38.878	–3.192
22	S	S	0.123	1.000	2.381	M	0.000	23.323	–1.786
25	S	S	0.309	1.000	1.033	M	0.000	54.595	–4.345
26	M	M	0.900	1.000	0.016	S	0.000	39.025	2.918
27	M	M	0.818	1.000	0.053	S	0.000	43.597	3.274
28	M	M	0.745	1.000	0.106	S	0.000	44.853	3.368
30	S	S	0.420	1.000	0.650	M	0.000	51.528	–4.135
31	S	S	0.236	1.000	1.404	M	0.000	57.111	–4.514
32	S	S	0.349	1.000	0.879	M	0.000	29.539	–2.391
33	S	S	0.929	1.000	0.008	M	0.000	41.745	–3.418
34	S	S	0.004	0.853	8.472	M	0.147	11.983	–0.418
35	M	M	0.648	1.000	0.208	S	0.000	46.634	3.500
36	M	M	0.750	1.000	0.102	S	0.000	44.770	3.362
37	M	M	0.627	1.000	0.236	S	0.000	47.035	3.529
38	M	M	0.822	1.000	0.051	S	0.000	43.531	3.269
39	M	M	0.509	1.000	0.436	S	0.000	49.454	3.704
40	M	M	0.189	1.000	1.727	S	0.000	25.584	1.729
41	S	S	0.548	1.000	0.360	M	0.000	33.315	–2.728
42	S	S	0.991	1.000	0.000	M	0.000	40.466	–3.318
44	S	S	0.211	1.000	1.567	M	0.000	58.126	–4.581
47	S	S	0.718	1.000	0.130	M	0.000	45.340	–3.690
48	S	S	0.855	1.000	0.033	M	0.000	42.966	–3.511
49	M	M	0.990	1.000	0.000	S	0.000	40.761	3.056
50	M	M	0.454	1.000	0.561	S	0.000	31.619	2.294
51	M	M	0.984	1.000	0.000	S	0.000	40.862	3.064
52	M	M	0.938	1.000	0.006	S	0.000	39.625	2.966

Table 5e (Cont.)

			Highest group			2nd highest group			Discriminant scores
Sample	Actual group	Predicted group	P(D/G) probability	P(G/D)	Square Mahalanobis' distance	Group	P(G/D)	Square Mahalanobis' distance	Function 1
54	M	M	0.301	1.000	1.069	S	0.000	28.500	2.010
55	S	S	0.732	1.000	0.117	M	0.000	45.089	–3.671
56	S	S	0.566	1.000	0.329	M	0.000	48.251	–3.903
58	S	S	0.411	1.000	0.675	M	0.000	51.753	–4.150
59	S	S	0.900	1.000	0.016	M	0.000	42.224	–3.454
62	S	S	0.011	0.984	6.473	M	0.016	14.654	–0.785
75	S	S	0.371	1.000	0.799	M	0.000	52.800	–4.223
76	S	S	0.404	1.000	0.697	M	0.000	30.663	–2.494
77	S	S	0.382	1.000	0.764	M	0.000	52.514	–4.203
78	S	S	0.470	1.000	0.522	M	0.000	50.334	–4.051
80	S	S	0.349	1.000	0.878	M	0.000	53.428	–4.266
82	M	M	0.632	1.000	0.229	S	0.000	46.940	3.522
83	M	M	0.694	1.000	0.155	S	0.000	35.746	2.650
84	M	M	0.736	1.000	0.114	S	0.000	45.018	3.381
85	M	M	0.414	1.000	0.668	S	0.000	51.689	3.861
86	M	M	0.786	1.000	0.074	S	0.000	44.139	3.315
87	M	M	0.080	1.000	3.058	S	0.000	21.379	1.295
88	M	M	0.022	0.997	5.259	S	0.003	16.639	0.750
89	S	S	0.012	0.986	6.353	M	0.014	14.836	–0.808
90	M	M	0.405	1.000	0.695	S	0.000	51.923	3.877
91	M	M	0.283	1.000	1.153	S	0.000	55.448	4.117
92	M	M	0.023	0.997	5.141	S	0.003	16.851	0.776
93	M	M	0.550	1.000	0.358	S	0.000	48.591	3.642
94	M	M	0.651	1.000	0.205	S	0.000	35.041	2.591
95	M	M	0.940	1.000	0.006	S	0.000	41.565	3.118

Shaded areas indicate samples with greater Mahalanobis' distances to the centroids.

have highlighted some of the samples with greater Mahalanobis' distances to the centroids, or in other words, samples that have the lowest probabilities of belonging to the group in which they are included. Some of these samples are those that in CA generally form isolated clusters, such as samples 34 and 89 in the sandy group, or samples 88 and 92 in the muddy group. This is also shown in the histogram in figure 8, where the discriminant scores are plotted along the discriminant function line for both groups. The samples far away from the centroid classes are numbers 34, 62 and 89 for group S, and 88, 92 and 97 for group M.

Another combination of multivariate techniques is CA with PCA (Huang *et al.*, 1994; Soares *et al.*, 1999), generally applied in order to interpret the association between variables. In this case the analysis must be run in *R*-mode. PCA was previously applied to these data by Rubio *et al.* (2000). They found three principal components explaining nearly 93% of the variance of the data (see fig. 10 and table 6 in Rubio *et al.*, 2000). These authors identified three groups of variables that they interpreted as different sources of the metals: lithogenic, anthropogenic and biogenic. Here, we run CA in *R*-mode using different algorithm clusterings (fig.9a,b). Despite slight differences between the algorithms, the results are nearly coincident with those of PCA from Rubio *et al.* (2000). A first cluster (CL 1) merges $CaCO_3$, Sr, sand and gravel, which is representative of the biogenic signal. The variables Cu, Pb, Zn, Fe and organic matter are grouped in CL 2, which is consistent with the mentioned anthropogenic signal. The rest of the variables that constituted the lithogenic group with PCA are not so clearly grouped with CA in only one cluster, although in CL 3, Co, Ti, Al, Ni and mud are forming one cluster with both algorithm methods used (fig. 9), whilst the rest of the variables are in different clusters.

estudio. La aplicación del CA para los datos normalizados con respecto a Al (fig. 7) tampoco llevó a un resultado concluyente acerca del grado de contaminación; así, el 82% de las muestras están ahora agrupadas en CL 1, donde se incluyen tanto muestras fangosas como arenosas y con distinto grado de contaminación. La figura 7 muestra el CA con UPGMA; sin embargo, nuevamente se obtuvieron ciertas diferencias en función del algoritmo usado. La utilización de CA para este tipo de estudios debe ser interpretada con precaución y debe combinarse con el cálculo de otro tipo de factores o índices de enriquecimiento o contaminación.

La mayor parte de los investigadores combinan el CA con otros análisis multivariantes. Bradfield y Orlóci (1975), Green y Vascotto (1978) y Emmerson *et al.* (1997) combinan el uso de CA con DA. Estos autores primeramente usan CA para hacer una clasificación y luego, con los mismos datos, examinan si con otro método, tal como DA, podrían identificar los objetos en las clases obtenidas. Como una medida de la validez de la clasificación tienen en cuenta el porcentaje de predicciones correctas.

En nuestro caso, los grupos establecidos fueron los mencionados grupos M y S. Antes del DA, se llevó a cabo una prueba para detectar igualdad de poblaciones centroides (Rao, 1965), deduciéndose del valor de la lamda de Wilk (0.087) y del valor de chi cuadrado (142.59) que hay una diferencia significativa ($P < 0.0001$) en las poblaciones centroides. Los datos y el histograma obtenido del DA se presentan en la tabla 5 y figura 8, respectivamente. En la clasificación de las muestras, el 100% de los casos agrupados se clasificaron correctamente. Esto indica que fue posible discriminar entre las áreas internas y externas de la ría y los dos tipos principales de sedimentos que existen en la misma. Además, esto está apoyado por la diferencia entre los

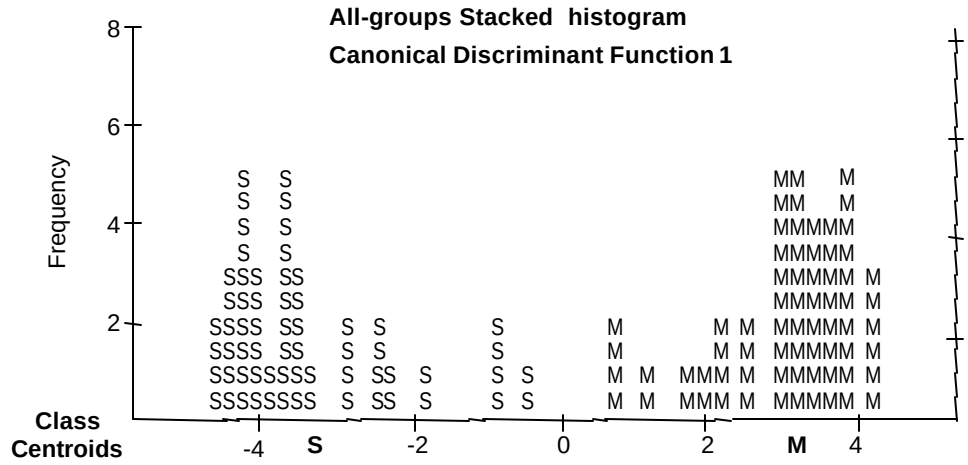

Figure 8. Plot of discriminant scores along the discriminant function line.
Figura 8. Diagrama de las puntuaciones discriminantes para la función lineal discriminante.

CONCLUSION

Choices need to be made at every stage of CA: the choice of similarity coefficient, choice of linkage method, choice of the phenon line and choice of standardization type, and there are very few recommendations or guidelines to help. As has been showed, it is therefore possible to obtain a substantial range of results from the same data set. CA is not a statistical procedure, and there is no easy way of deciding whether or not data are more clustered than would be expected from a random population. Our results from the Ría de Vigo show that clustering reflects quite satisfactorily the granulometrical properties of the two main environments in the ria. Sediments in the Ría de Vigo are variable in texture and chemical composition, because they accumulated in environments where physical conditions are themselves quite variable, so CA information can be used as provenance indicators. However, the ability of this form of multivariant analysis (*Q*-mode) to distinguish contaminated valores medios de los grupos centroides (tabla5). Sin embargo, en la tabla 5 hemos resaltado (gris) algunas de las muestras con mayores distancias de Mahalanobis a los centroides, o en otras palabras, las muestras que tienen las probabilidades más bajas de pertenecer al grupo en el que están incluidas. Algunas de estas muestras son aquellas que en CA generalmente formaban *clusters* aislados, tales como las muestras 34 y 89 en el grupo arenoso o las muestras 88 y 92 en el grupo fangoso. Esto también se muestra en el histograma de la figura 8, donde se representan las puntuaciones discriminantes para la función discriminante para ambos grupos. Las muestras alejadas de las clases centroides son los números 34, 62 y 89 para el grupo S y 88, 92 y 97 para el grupo M.

Otra combinación que suele hacerse es CA con PCA (Huang *et al*., 1994; Soares *et al.*, 1999), generalmente para ver la asociación entre variables, con lo que los análisis se realizan en modo *R.* PCA se aplicó previamente a estos datos por Rubio *et al.* (2000).

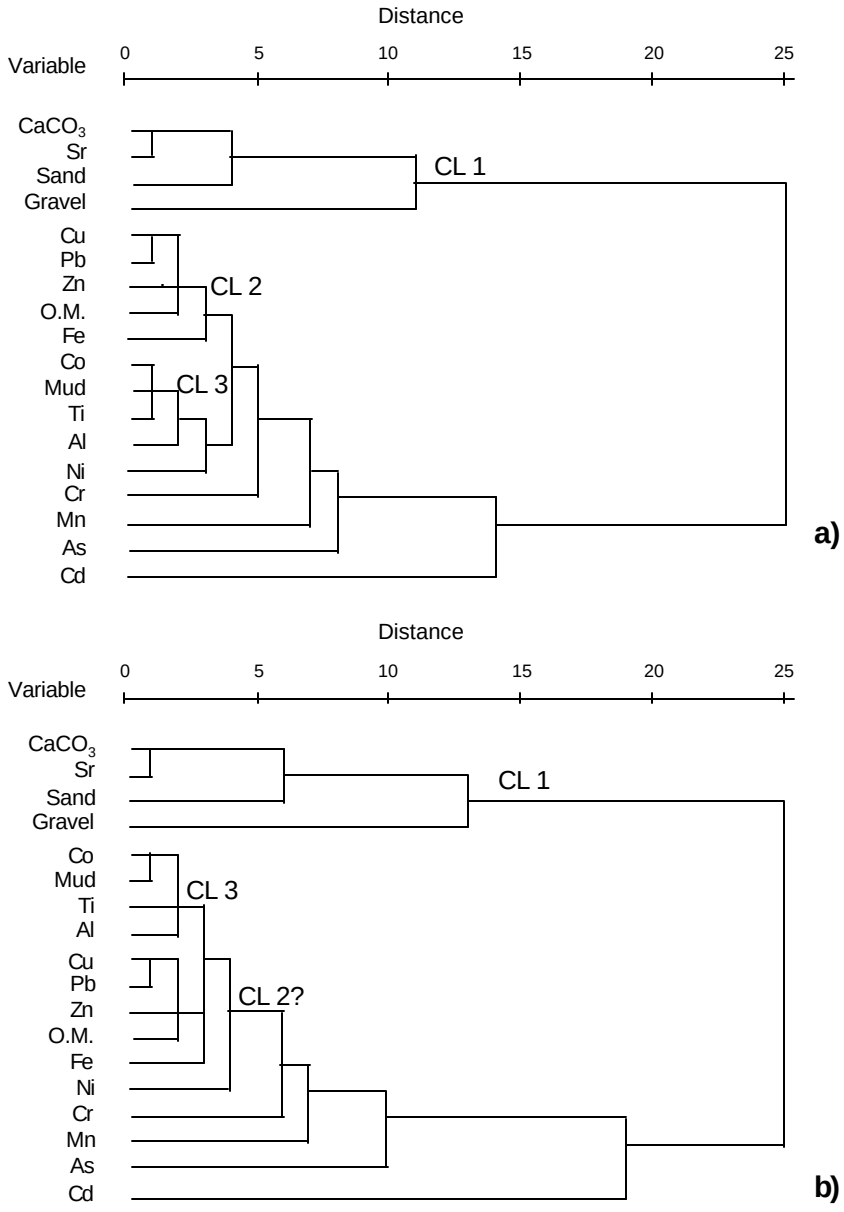

Figure 9. (**a**) *R*-mode dendrogram of the cluster analysis with the UPGMA method. (**b**) *R*-mode dendrogram of the cluster analysis with the SLINK method.
Figura 9. (**a**) Dendrograma en modo *R* del análisis *cluster* con el método UPGMA. (**b**) Dendrograma en modo *R* del análisis *cluster* con el método SLINK.

and uncontaminated sediments is questionable, since several interpretations can be drawn using different CA procedures. Information about the degree of contamination is mixed and masked by textural effects. Discriminant analysis provides a better basis on which to differentiate textural and geochemical environments in the ria. CA in *R*-mode has allowed us to obtain a clustering of variables similar to PCA, showing the biogenic, lithogenic and anthropogenic signals in the ria sediments.

ACKNOWLEDGEMENTS

We would like to thank three anonymous reviewers for very constructive reviews, which improved the paper. We are greatly indebited to K. Pye from the Department of Geology, Royal Holloway, University of London, for the helpful suggestions and critical review of the manuscript. This work is a contribution to projects PGIDT00MAR30103PR and PGIDT00PX 130105PR, of Xunta de Galicia, and REN2000-1102 MAR, of Plan Nacional de I+D+I.

REFERENCES

Albani, A.D., Rickwood, P.C., Favero, V.M. and Serandrei Barbero, R. (1989). The geochemical anomalies in sediments on the shelf near the Lagoon of Venice, Italy. Mar. Pollut. Bull., 20: 438–442.

Álvarez-Iglesias, P., Rubio, B. and Vilas, F. (2000). Especiación de plomo en sedimentos y niveles de concentración en organismos de la zona intermareal de la ensenada de San Simón (Ría de Vigo, NO España). Thalassas, 16: 81–94.

Angelidis, M.O. and Aloupi, M. (2000). Geochemical study of coastal sediments influenced by river-transported pollution: Southern Evoikos Gulf, Greece. Mar. Pollut. Bull., 40: 77–82.

Badarudeen, A., Damodaran, K.T., Sajan, K. and Padmalal, D. (1996). Texture and geochemistry of the sediments of a tropical mangrove

Estos autores encontraron tres componentes principales que explican casi un 93% de la varianza de los datos (ver fig. 10 y tabla 6 en Rubio *et al.*, 2000), agrupando las variables en tres grupos que interpretan como diferentes orígenes de los metales: litogénico, antropogénico y biogénico. Se realizó el CA en modo *R*, con distintos algoritmos de *cluster* (fig. 9a, b) y pese a las ligeras diferencias existentes entre éstos, la agrupación entre variables es bastante coincidente con los datos de PCA de Rubio *et al*. (2000). Así, en un primer *cluster* (CL 1) se unen $CaCO_3$, Sr, arena y grava, que marcaría la señal biogénica de los sedimentos. En CL 2 se agrupan Cu, Pb, Zn, Fe y materia orgánica, que coincide con la citada señal antropogénica. El resto de las variables constituye el grupo litogénico mediante PCA; sin embargo, con CA, el agrupamiento en un solo cluster no es tan claro, aunque en CL 3 se agrupan Co, Ti, Al, Ni y fango con los dos métodos de agrupamiento (fig. 9) y el resto de las variables forman *clusters* distintos.

CONCLUSIÓN

En cada paso del CA se necesita hacer varias elecciones: la elección del coeficiente de similitud, del método de agrupamiento, de la distancia fenotípica y del tipo de estandarización, y hay muy pocas recomendaciones o guías que ayuden. Como se ha demostrado es, por tanto, posible obtener un gran intervalo de resultados de los mismos datos. El CA no es un procedimiento estadístico y no es fácil decidir si los datos estarían agrupados más allá de lo que sería esperado de una población al azar. Nuestros resultados para la ría de Vigo muestran que el agrupamiento refleja bastante satisfactoriamente las propiedades granulométricas de los dos ambientes principales en la ría. Los sedimentos en la ría de Vigo son variables en textura y composición química porque se acumulan en ambientes donde las

ecosystem, southwest coast of India. Environ. Geol., 27: 164–169.

Barreiro-Lozano, R., Carballeira-Ocaña, A. and Real-Rodríguez, C. (1988). Metales pesados en los sedimentos de cinco sistemas de ria (Ferrol, Burgo, Arousa, Pontevedra y Vigo). Thalassas, 6: 61–70.

Bebianno, M.J. and Machado, L.M. (1997). Concentrations of metals and metallothioneins in *Mytilus galloprovincialis* along the south coast of Portugal. Mar. Pollut. Bull., 34: 666–671.

Bradfield, G.E. and Orlóci, L. (1975). Classification of vegetation data from an open beach environment in southwestern Ontario: Cluster analysis followed by generalized distance assignment. Can. J. Botany, 53: 495–502.

Carral, E., Villares, R., Puente, X. and Carballeira, A. (1995). Influence of watershed lithology on heavy metal levels in estuarine sediments and organisms in Galicia (north-west Spain). Mar. Pollut. Bull., 30: 604–608.

Chen, T.B., Wong, J.W.C., Zhou, H.Y. and Wong, M.H. (1997). Assessment of trace metal distribution and contamination in surface soils of Hong Kong. Environ. Pollut., 96: 61–68.

Cortesao, C. and Vale, C. (1995). Metals in sediments of the Sado estuary, Portugal. Mar. Pollut. Bull., 30: 34–37.

Cuadras, C.M. (1991). Métodos de Análisis Multivariante. Promociones y Publicaciones Universitarias (PPU), Barcelona, 644 pp.

Davis, J.C. (1986). Statistics and Data Analysis in Geology. John Wiley, New York, 646 pp.

DelValls, T.A. and Chapman, P.M. (1998). Site-specific sediment quality values for the Gulf of Cádiz (Spain) and San Francisco Bay (USA), using the sediment quality triad and multivariate analysis. Ciencias Marinas, 24(3): 313–336.

DelValls, T.A., Forja, J.M., González-Mazo, E., Gómez-Parra, A. and Blasco, J. (1998). Determining contamination sources in marine sediments using multivariate analysis. TRAC-Trend Anal. Chem., 17: 181–192.

Emmerson, R.H.C., O'Reilly-Wiese, S.B., Macleod, C.L. and Lester, J.N. (1997). A multivariate assessment of metal distribution in inter-tidal sediments of the Blackwater Estuary, UK. Mar. Pollut. Bull., 34: 960–968.

condiciones físicas en sí mismas son variables, así que la información del CA puede ser usada como un indicador de procedencia. Sin embargo, la capacidad de este tipo de análisis multivariante para distinguir sedimentos contaminados y no contaminados es cuestionable, puesto que se pueden hacer varias interpretaciones usando diferentes procedimientos de CA. La información sobre el grado de contaminación está mezclada y enmascarada por los efectos texturales. El análisis discriminante proporciona una base mejor sobre la cual diferenciar ambientes texturales y geoquímicos en la ría. Aunque los CA en modo *Q* no nos han permitido diferenciar grupos de muestras contaminadas o no contaminadas, el CA en modo *R* ha reflejado una asociación entre variables similar al PCA, indicando la influencia biogénica, litogénica y antropogénica en los sedimentos de la ría.

AGRADECIMIENTOS

Los autores agradecen a tres revisores anónimos sus sugerencias e indicaciones que han mejorado sensiblemente la presente contribución. Igualmente quieren agredecer a K. Pye, del Departamento de Geología, Royal Holloway, Universidad de Londres, su revisión crítica del manuscrito. Ésta es una contribución a los proyectos PGIDT00MAR30103PR y PGIDT00PX 130105PR, de la Xunta de Galicia, y REN2000-1102 MAR, del Plan Nacional de I+D+I.

Traducido al español por los autores.

Giordano, R., Musmeci, L., Ciaralli, L., Vernillo, I., Chirico, M., Piccioni, A. and Costantini, S. (1992). Total contents and sequential extractions of mercury, cadmium and lead in coastal sediments. Mar. Pollut. Bull., 24: 350–357.

Gónzalez, H. and Torres, I. (1990). Heavy metals in sediments around a sewage outfall at Havana, Cuba. Mar. Pollut. Bull., 21: 253–255.

Green, R.H. and Vascotto, G.L. (1978). A method for the analysis of environmental factors controlling patterns of species composition in aquatic communities. Water. Res., 12: 583–590.

Guitián, F. (ed.) (1992). Atlas Geoquímico de Galicia. Ed. Xunta de Galicia, Consellería de Industria y Comercio, Santiago, 29 pp.

Guitián, F. and Carballas, T. (1976). Técnicas de Análisis de Suelos. Ed. Pico Sacro, Santiago, 288 pp.

Huang, W., Campredon, R., Abrao, J.J., Bernat, M. and Latouche, C. (1994). Variation of heavy metals in recent sediments from Pirainginga Lagoon (Brazil): Interpretation of geochemical data with the aid of multivariate analysis. Environ. Geol., 23: 241–247.

Hughes, M.M. (1979). Exploration and play revisited: A hierarchical analysis. Int. J. Behav. Develop., 2: 225–233.

Karbassi, A.R. and Nadjafpour, S. (1996). Flocculation of dissolved Pb, Zn and Mn during estuarine mixing of river water with the Caspian Sea. Environ. Pollut., 93: 257–260.

Macías-Zamora, J.V., Villaescusa-Celaya, J.A., Muñoz-Barbosa, A. and Gold-Bouchot, G. (1999). Trace metals in sediment cores from the Campeche Shelf, Gulf of Mexico. Environ. Pollut., 104: 69–77.

Nombela, M.A., Vilas, F., García-Gil, S., García-Gil, E., Alejo, I., Rubio, B. and Pazos, O. (1994). Metales pesados en el registro sedimentario reciente en la ensenada de San Simón, parte interna de la Ría de Vigo (Galicia, España). Gaia, 8: 149–156.

Nombela, M.A., Rubio, B., Alejo, I. and Vilas, F. (1998). Distribución granulométrica en playas de las rías de Pontevedra y Arosa (NO España): Una comparación mediante técnicas multivariantes. Thalassas, 14: 81–88.

Rao, C.R. (1965). Linear Statistical Inference and its Application. John Wiley, New York, 522 pp.

Ratha, D.S. and Sahu, B.K. (1993). Source and distribution of metals in urban soil of Bombay, India, using multivariate statitiscal techniques. Environ. Geol., 22: 276–285.

Reyment, R.A. and Savazzi, E. (1999). Aspects of Multivariate Statistical Analysis in Geology. Elsevier, Amsterdam, 285 pp.

Richthofen, F. von (1886). Führer für Forchungsreisende. Oppenheim, Berlin.

Romesburg, H.C. (1990). Cluster Analysis for Researchers. Robert E. Krieger Publishing Co., Florida, 334 pp.

Rubio, B., Nombela, M.A., Vilas, F., Alejo, I., García-Gil, E., García-Gil, S. and Pazos, O. (1995). Distribución y enriquecimiento de metales en sedimentos actuales de la parte interna de la Ría de Pontevedra. Thalassas, 11: 35–45.

Rubio, B., Nombela, M.A. and Vilas, F. (2000). Geochemistry of major and trace elements in sediments of the Ria de Vigo (NW Spain): An assessment of metal pollution. Mar. Pollut. Bull., 40: 968–980.

Ruíz, F., González-Regalado, M.L., Borrego, J., Morales, J.A., Pendón, J.G. and Muñoz, J.M. (1998). Stratigraphic sequence, elemental concentrations and heavy metal pollution in Holocene sediments from the Tinto-Odiel estuary, southwestern Spain. Environ. Geol., 34:270–278.

Segovia-Zavala, J.A., Delgadillo-Hinojosa, F. and Álvarez-Borrego, S. (1998). Cadmium in the coastal upwelling area adjacent to the California-Mexico border. Estuar. Coast. Shelf Sci., 46: 475–481.

Shepard, F.P. (1954). Nomenclature based on sand-silt-clay ratios. J. Sed. Petrol., 24: 151–158.

Simeonov, V. and Andreev, G. (1989). Interpretation of Black Sea sediments analytical data by the clustering approach. Toxicol. Environ. Chem., 24: 233–240.

Soares, H.M.V.M., Boaventura, R.A.R., Machado, A.A.S.C. and Esteves da Silva, J.C.G. (1999). Sediments as monitors of heavy metal contamination in the Ave River basin (Portugal):

Multivariate analysis of data. Environ. Pollut., 105: 311–323.

Sneath, P.H.A. and Sokal, R.R. (1973). Numerical Taxonomy. W.H. Freeman, San Francisco, 573pp.

Swan, A.R.H. and Sandilands, M. (1995). Introduction to Geological Data Analysis. Blackwell, Oxford, 446 pp.

Vilas, F., Nombela, M.A., García-Gil, E., García-Gil, S., Alejo, I., Rubio, B. and Pazos, O. (1995). Cartografía de Sedimentos Submarinos. Ría de Vigo. Xunta de Galicia, Santiago de Compostela, 40 pp.