PROPUESTA DE CRITERIOS PARA DETERMINAR LOS VALORES DE NIVELES CARACTERÍSTICOS DE METALES PESADOS EN SUELOS Y SEDIMENTOS A PARTIR DE MÉTODOS ENTRÓPICOS

Memoria que presenta Rafael Bellver Mancheño
para optar al grado de Doctor en Ciencias
Geológicas por la Universidad de Granada

Fdo.: Rafael Bellver Mancheño

Granada, 31 de mayo de 2004

VºBº de los Directores de la Tesis

Fdo.: Dr. Juan A. Fernández García

Fdo.: Mariano J. Valderrama Bonnet

UNIVERSIDAD DE GRANADA

COMISIÓN DE DOCTORADO

Fecha: 16 de Junio de 2004
Su Refª.
Nuestra Refª.
Fecha de Salida:
Unidad de Origen: COMISIÓN DE DOCTORADO

UNIVERSIDAD DE GRANADA
Facultad de Ciencias
Fecha 22/6/04
ENTRADA NUM. 2057

UNIVERSIDAD DE GRANADA
COMISION DE DOCTORADO
18 JUN. 2004
REGISTRO
SALIDA Nº 602

Destinatario:

Iltmo. Sr. DECANO DE LA Facultad de Ciencias

En cumplimiento del artículo 36 de las Normas Reguladoras de los estudios de Tercer Ciclo de esta Universidad, adjunto se remite un ejemplar de la Tesis Doctoral presentada por el Doctorando D.RAFAEL JUAN BELLVER MANCHEÑO titulada PROPUESTA DE CRITERIOS PARA DETERMINAR LOS VALORES DE NIVELES CARACTERISTICOS DE METALES PESADOS EN SUELOS Y SEDIMENTOS A PARTIR DE METODOS ENTROPICOS y dirigida por el Profesor/es Dr/es JUAN ANTONIO FERNANDEZ GARCIA, MARIANO JOSE VALDERRAMA BONNET con objeto de mantenerla depositada desde el día 16 de Junio de 2004 hasta el día 16 de Julio de 2004 para que pueda ser examinada por cualquier Doctor que así lo desee.

Granada, a 16 de Junio de 2004.

LA SECRETARIA DE LA COMISIÓN

DE DOCTORADO

UNIVERSIDAD DE GRANADA
Comisión de Doctorado

Fdo: María López Jurado Romero de la Cruz

TESIS DOCTORAL

PROPUESTA DE CRITERIOS PARA DETERMINAR LOS VALORES DE NIVELES CARACTERÍSTICOS DE METALES PESADOS EN SUELOS Y SEDIMENTOS A PARTIR DE MÉTODOS ENTRÓPICOS

Rafael Bellver Mancheño

Granada, 31 de mayo de 2004

PREAMBULO

El presente trabajo pretende profundizar en el estudio de la distribución de los llamados metales pesados en suelos y sedimentos y en la detección de zonas sometidas a procesos de contaminación. Se busca determinar los valores de concentración de estos elementos que se deben a factores naturales de las debidas a aportes antrópicos, que se reflejan en la aparición de los mismos en lugares que no deberían estar o en un aumento de las concentraciones de *origen natural*.

A fecha de hoy, no existe ninguna metodología precisa que permita establecer los criterios necesarios para calificar una zona como contaminada o no, especialmente cuando los valores de concentraciones de elementos contaminantes no alcanzan un umbral suficiente como para quedar reflejados en los llamados bioindicadores, es decir, que se halla alcanzado una concentración lo suficientemente alta como para que se empiecen a producir efectos adversos (tóxicos), en la cubierta vegetal principalmente, de las zonas sometidas a procesos de contaminación.

Probablemente, el hecho de que, a día de hoy, no exista tal metodología, se deba a condicionantes de tipo político-social, que han hecho que los problemas relativos a la contaminación se enfoquen más desde un punto de vista administrativo que científico. Así, se definen términos como **nivel base** o ***baseline*** (internacionalmente), con el que se quiere indicar el umbral a partir del cual un suelo se puede o debe considerar contaminado, o **niveles guía**, término que hace referencia a valores fijados por distintas instituciones que pretenden indicar el grado de contaminación de los suelos y para qué uso son adecuados, o si debe realizarse algún tipo de actuación sobre ellos. Por lo general, se fijan tres niveles con grado de exigencia creciente:

1.- **Nivel de referencia:** corresponde a la máxima concentración de un determinado elemento que puede encontrarse en un suelo sin que llegue a producir efectos no deseables, es el límite de seguridad, por debajo del cual se supone que se le puede dar al suelo cualquier uso. En la literatura

internacional, se utilizan como términos equivalentes: ***reference, guide, target, trigger, baseline, screning*** o ***preventive levels*** o ***values***.

2.- **Nivel de investigación**: delimita la concentración máxima de contaminantes en un suelo, a partir de la cual cabe se puedan producir efectos nocivos para ciertos usos, siendo irrelevante para otros. En la literatura internacional, se utilizan como términos equivalentes: ***investigation, signal*** o ***threshold values.***

3.- **Nivel de intervención**: corresponde al valor a partir del cual el riesgo deja de ser tal y pasa a convertirse en la seguridad de que se están produciendo daños medioambientales y es peligroso para la salud humana realizar cualquier tipo de actividad sobre él, se considera que el suelo está muy contaminado y que deben tomarse medidas de recuperación. En la literatura internacional, se utilizan como términos equivalentes: ***intervention, remediation, action*** o ***clean up values*** o ***levels***.

Resumiendo, en la actualidad los estudios de contaminación de suelos, se vienen enfocando desde una perspectiva de planificación y gestión medioambiental, se busca diferenciar áreas contaminadas de otras que no lo están, y el grado de *peligrosidad* de éstas en función de criterios como riesgo para la salud humana o deterioro del medioambiente. Y la importancia que se concede a este tipo de estudios, queda reflejada en la abundante, y relativamente reciente, legislación que existe sobre el tema, tanto a nivel nacional como internacional.

En España, la Ley de Residuos de 1998 (10/98), recoge la normativa fundamental de las Directivas de la comunidad Europea en materia de medioambiente, no existiendo hasta el momento una regulación que fije, a nivel nacional, los criterios para considerar un suelo como contaminado, si bien el artículo 3 en su apartado p, se define lo que se entiende por suelo contaminado: *Todo aquél cuyas características físicas, químicas o biológicas han sido alteradas negativamente por la presencia de componentes de carácter peligroso de origen humano, en concentración tal que comporte un riesgo para*

la salud humana o el medioambiente. Según este artículo, hay que destacar dos hechos importantes **a**. Sólo se considera contaminación, cuando se produce por actividad humana y **b**. Para calificar un suelo como contaminado, la contaminación tiene que ser lo suficientemente intensa como para que comporte riesgo para la salud humana o el medioambiente. En el artículo 27 de esta misma ley se delega en las Comunidades Autónomas el declarar, delimitar e inventariar los suelos contaminados de cada comunidad, así como a fijar las prioridades de actuación.

Como se deduce de lo anterior, los planteamientos actuales son más una declaración de intenciones o la exposición de una serie de necesidades que los que corresponderían a un planteamiento científico clásico, entre otras razones, porque las espectativas abiertas o las necesidades de conocimiento para una buena gestión del Medioambiente no son acordes con los conocimientos o los datos existentes.

Por ello, en este trabajo se parte de la base de que el planteamiento a seguir en los estudios de distribución y acumulación de elementos contaminantes debe en consonancia con los datos e información disponibles, para lo cual, se recurre a la definición de un nuevo concepto, que se ha dado en llamar Nivel Característico o niveles característicos, que puede definirse como *el valor o el intervalo de valores correspondientes a las concentraciones de origen natural de un elemento contaminante bajo unas condiciones determinadas de factor condicionante o patrón explicativo*, así tendríamos un valor característico para cada tipo de suelo, que además mejora su precisión al añadir factores condicionantes o patrones explicativos, como tipo de sustrato, grado de pendiente, etc. Pudiendo obtener así de forma más precisa, valores característicos para un suelo desarrollado sobre distintos sustratos, o situados en diversas condiciones de pendiente entre otros.

Esto ha obligado a desarrollar una metodología de cálculo, basada en la Teoría de la Información de Shanon (1948), que permite agrupar los datos siguiendo simultáneamente dos criterios, uno de similitud y otro de continuidad, para así delimitar los valores correspondientes a concentraciones de elementos

de origen natural, y por su diferencia con los encontrados en un muestreo, poder calificar la zona de estudio como *normal*, y por tanto libre de contaminación o como anómala y por consiguiente *sospechosa* de estar sometida a procesos de contaminación.

INDICE

1. INTRODUCCIÓN Y OBETIVOS

1.1. Consideraciones generales.

Según los expertos en Ecología, las mayores cotas de éxito evolutivo, pueden conseguirse siguiendo dos estrategias diferentes, una de *adaptabilidad*, que consiste en ser capaz de subsistir en diferentes nichos ecológicos junto con otras especies, en principio mejor adaptadas, y una segunda, de *especialización*, que consiste a su vez en la adaptación máxima al medio y por tanto en obtener el mejor aprovechamiento de sus recursos.

En principio, la primera presenta más posibilidades para la supervivencia en general, pues las fuentes de subsistencia están más diversificadas y por tanto las especies que sigan esta estrategia se verán menos afectadas por los cambios que se puedan producir en los sistemas a lo largo del tiempo. Y la segunda, supone una mayor facilidad para sobrevivir, pues requiere menos esfuerzo por parte de las especies mejor adaptadas para cubrir sus necesidades básicas, siempre y cuando se mantengan las condiciones del medio. Según lo anterior, el factor tiempo, en cuya dimensión se van a producir los cambios del sistema, es determinante a la hora de atribuir un mayor o menor éxito a la estrategia seguida. Generalizando, a corto plazo, y en tanto se mantengan las condiciones, el éxito obtenido es mayor a través de estrategias de adaptación, y a largo plazo, si cambian las condiciones, triunfan las de adaptabilidad.

En su lucha por la supervivencia, la humanidad ha hecho uso de las dos estrategias, en principio su éxito se debía a su adaptabilidad, más tarde, debido a su capacidad de construir herramientas, mejoró su adaptación, y por último y a diferencia de otras especies desarrolla e impone, una tercera estrategia, consistente en intentar adaptar el medio a sus necesidades. Estos intentos de adaptar el *sistema planeta* a nuestras necesidades, son los que producen los cambios en las *relaciones históricas* entre especies y con el medio, generando desequilibrios. Si hacemos un balance de las consecuencias de estos desequilibrios, desde el comienzo de la Historia hasta ahora y desde un punto de vista estricto de éxito de la especie frente a otras competidoras, obviando cualquier consideración de tipo sociológico, en general ha resultado positivo,

pues ha aumentado la población de manera exponencial a la vez que ha aumentado el tiempo de vida de los individuos y, al menos los países desarrollados, disponen actualmente de recursos suficientes para mantener e incluso aumentar su ritmo de desarrollo.

Mención aparte merece la pregunta: ¿En el futuro se va a poder mantener el nivel de éxito alcanzado? La respuesta sería negativa si seguimos rigiéndonos por las mismas normas que hasta ahora, que se resumen en hacer *caso omiso* de cualquier consideración que no afecte directamente y de forma reconocible a la salud humana o al medio de producción de recursos, *Sistema Tierra*. De manera, que el reto al que nos enfrentamos consiste esencialmente en poder poner a disposición de la especie los recursos suficientes para que mantenga o aumente su éxito.

La lógica indica que la cantidad de recursos presentes en un sistema debe de disminuir de manera proporcional al consumo que se haga de ellos, llegando a agotarse si no se renuevan con la suficiente rapidez, por lo que llegado un cierto momento y si la renovación no se realiza a una velocidad acorde con el ritmo de consumo, ocurre qué, para mantener el nivel de recursos es necesario importarlos.

Dado que no parece que a corto plazo puedan importarse recursos de otros planetas, la respuesta más prudente a la pregunta anterior sería no agotar o inutilizar los del nuestro, a la vez que se trabaja en el desarrollo y uso de otras energías y tecnologías que permitan generar nuevos recursos.

La tendencia de la humanidad, tanto a nivel particular como colectivo, de intentar alcanzar la mayor cota posible de *satisfacción personal* a lo largo de su existencia, plantea, por lo subjetivo del término en ambos niveles, gran cantidad de problemas a la hora de elegir criterios que permitan elaborar estrategias a medio y largo plazo para el uso, manejo y conservación de los nichos ecológicos que ocupamos o de otros que pueden llegar a ocuparse en el futuro, por lo que sería conveniente profundizar en los conocimientos sobre nuestro sistema de producción de recursos.

El hecho de que ciertas actividades antrópicas puedan ser lesivas para el Medio Ambiente o la salud humana, se conoce desde muy antiguo, de hecho ya existía una *normativa* en la Roma Republicana para la construcción de cloacas en las ciudades y la disposición de puntos potencialmente contaminantes en los campamentos de las legiones, en los cuales la construcción y situación de los vertederos y las letrinas de los legionarios debía de cumplir una serie de requisitos, pues se sabía la existencia de una relación entre la localización de estos *puntos negros* y una serie de infecciones y enfermedades. Este tipo de prácticas, cayó en desuso a raíz del hundimiento del Imperio, aunque se han ido recuperando a lo largo del tiempo. Pero ha sido a partir de la llamada Revolución Industrial (1760) cuando las actividades humanas han ido afectando de manera determinante el *comportamiento normal* de los *ciclos de la Naturaleza*, ya que en épocas anteriores, con una población inferior y una industria rudimentaria, las *acciones del hombre sobre el medio* se reducían a las derivadas de la agricultura, ganadería y minería a una escala relativamente pequeña, a partir de entonces estas actividades han tenido un crecimiento exponencial, y además se ha unido a ellas la industrial, junto con un aumento espectacular de la población, lo que ha multiplicado lo que podríamos llamar *efectos antrópicos* sobre los ecosistemas.

Para tener una idea de la intensidad de la actuación antrópica a lo largo de los últimos siglos, podemos recurrir a representar la evolución del PIB mundial (Fig. 1.) – recordemos que el Producto Interior Bruto (PIB), es la suma del valor de todos los bienes y servicios puestos en el mercado o registrados oficialmente, por lo que refleja de alguna manera, la actividad agrícola, ganadera e industrial -. Aunque algunos autores, como McNeill (2001), hacen notar que se trata de un procedimiento muy imperfecto desde el punto de vista económico, pues en algunos momentos y lugares ciertas producciones y prestaciones de servicios significativas, se realizan fuera de los controles, lo que supone una infravaloración del PIB calculado sobre el real. En los aspectos que conciernen a este trabajo la diferencia entre los valores de PIB oficiales y los reales, nos indica que la *presión sobre los ecosistemas* es probable que haya sido superior a la expuesta.

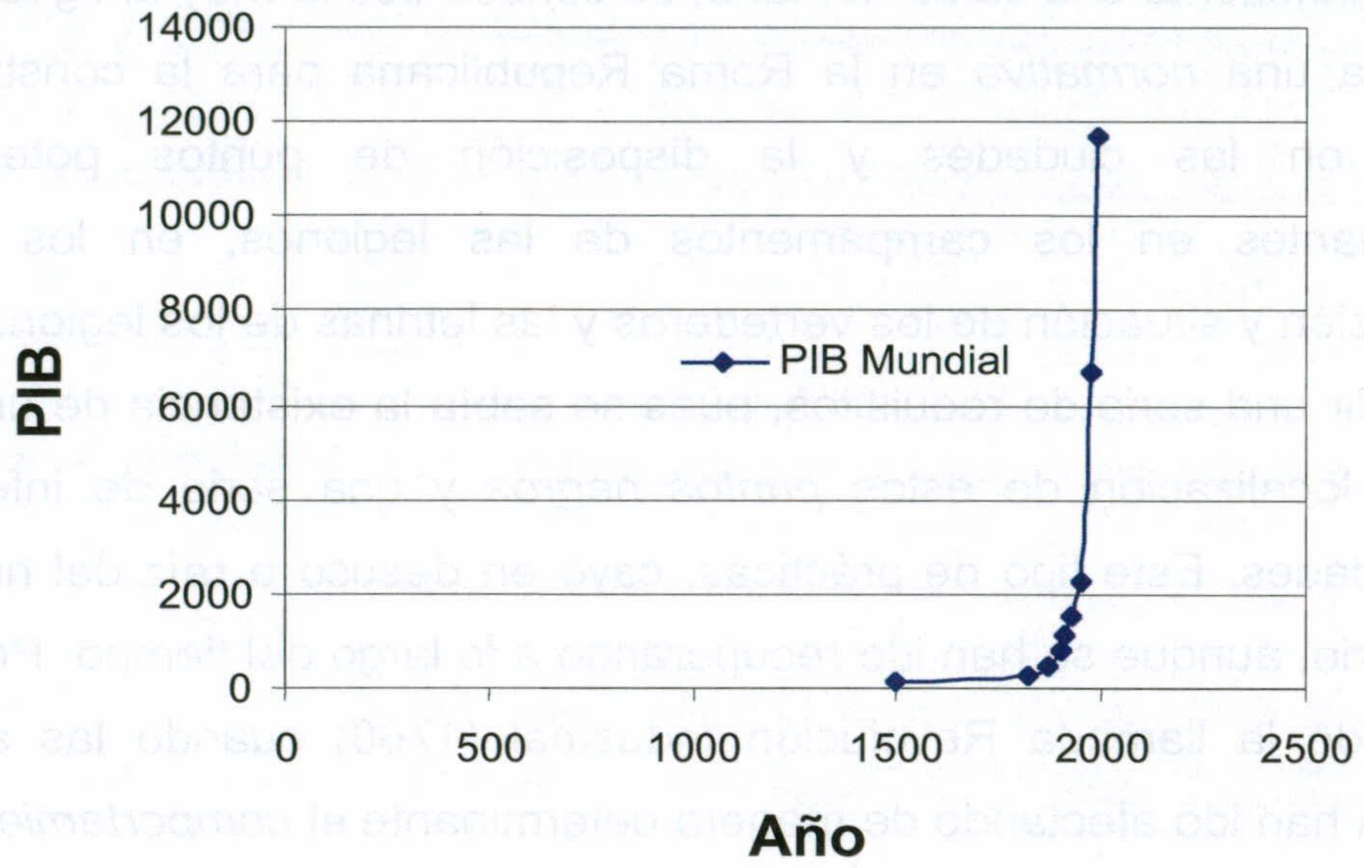


Fig. 1. Evolución del PIB mundial desde 1.500 hasta 1992, en valores referidos a 1.500. (Maddison, 1995, in McNeill, 2001).

En el gráfico, podemos ver, que la situación económica actual es del orden de 1.200 veces superior a la que existía en 1.500, lo cual, aunque no presente una relación lineal con el grado de presión al que se ha sometido al medio ambiente, es indicativo del aumento producido en los últimos tiempos, esto, unido al comportamiento no lineal que presentan algunos sistemas al sobrepasar ciertos umbrales, como pueden ser la formación de huracanes en el océano Atlántico tropical, una vez superado el umbral de 26ºC de temperatura, o la desaparición de bancos pesqueros cuando la tasa de capturas supera a la de reproducción, debe hacer que nos planteemos *si la humanidad ha comenzado a jugar con el planeta sin conocer todas las reglas del juego* (McNeill, 2001), con todos los riesgos que ello supone.

A partir de 1820 comienza a diverger el PIB y la población, lo que implica un cambio de costumbres y por lo tanto de tipo de contaminantes y forma de los mismos.

1.2. Aspectos sociales y legislativos.

Hace ya algunas décadas que la sociedad empezó a preocuparse por los temas relacionados con el estudio y la conservación del medio ambiente, pero no fue hasta la Conferencia de Río de Janeiro de 1992, propiciada por las Naciones Unidas, cuando las *altas esferas políticas* decidieron que era tan urgente como necesario, el aumentar el conocimiento de como las actividades humanas influyen en el medio ambiente y de cómo equilibrar el balance entre crecimiento económico y protección del medio ambiente para asegurarnos un desarrollo sostenible, entendiendo como desarrollo sostenible aquel que *satisface las necesidades actuales sin poner en peligro la capacidad de las generaciones futuras de satisfacer sus propias necesidades*.(EEU, 2002).

Anteriormente a esta conferencia - aunque solo dos décadas -, una serie de organismos internacionales, y algunos países de forma independiente ya habían comenzado a financiar y realizar proyectos de investigación encaminados hacia este fin.

En el concierto mundial la UNEP, comienza en 1972 el United Nations Environment Programme, cuyos avances se reflejan en los proyectos GEO (Global Environmental Outlook), en los que participan más de 171 paises y 40 instituciones internacionales, iniciados en 1999 (GEO 2000) y que actualmente se encuentra en su versión GEO-3 (2002), intentan proporcionar a los responsables de política medioambiental toda la información posible para poder tomar las decisiones necesarias para el futuro del Medio Ambiente.

En Europa, durante la cumbre de la CEE de París de 1972, se reconoció que, en el contexto de la expansión económica y la mejora de la calidad de vida, debía concederse una especial atención al medio ambiente y a raíz de ello se adoptó el primer programa de medio ambiente para el periodo 1973-1976, al cual siguieron una serie de programas plurianuales del mismo tipo que desembocaron en la adopción de diversas directivas. Pero no fue hasta la entrada en vigor del Acta Única de 1987 cuando se añadió al tratado un título específico sobre el medio ambiente (artículos 130R hasta 130T). El tratado

también introduce la idea de que *las exigencias de la protección del medio ambiente serán un componente de las demás políticas de la comunidad* (Dirección General de Medio Ambiente de la CE, 2002).

Pero fue a partir de 1993, con la entrada en vigor del Tratado de la Unión Europea, cuando se produjeron los mayores avances en estos aspectos, pues se introdujo entre los objetivos de la Unión Europea el concepto de *crecimiento sostenible y no inflacionista que respete el medio ambiente*, y se esbozó el principio de cautela en el artículo referente al medio ambiente, con lo que las medidas relacionadas con el medio ambiente adquirieron el rango de *política* por derecho propio. El Tratado de Amsterdam de 1999, recoge el principio de desarrollo sostenible en su artículo 2 y la cláusula del artículo 130 R del Acta Única de 1987, según la cual las exigencias de la protección del medio ambiente deben ser un componente de las demás políticas, pasó a convertirse en el artículo 6.

Todos estos aspectos, han sido considerados de nuevo en otras cumbres internacionales, como las de Kiev (2002) y Johanesburgo (2003), en las que se han hecho nuevas propuestas, debidas a la mejora de la información disponible, que siguen las mismas líneas de actuación.

Hoy día, prácticamente todos los países disponen de una legislación propia en lo relativo a conservación del Medio Ambiente (Ferguson, 1999), que generalmente sigue las directivas propuestas por los distintos organismos internacionales, pero resulta bastante imprecisa, en ciertos aspectos, por no disponer de criterios o modelos de comparación claros para determinar cuando y cuanto están afectando al medio ambiente las distintas actividades antrópicas, ello ha conducido a promover y financiar proyectos de investigación que reduzcan, en la medida de lo posible, esta imprecisión, y permitan por tanto resolver de manera efectiva los problemas derivados de la poco adecuada gestión del medio ambiente que, en general, se ha llevado a lo largo de la historia.

Generalmente se acepta, que los efectos negativos sobre el medio ambiente que pueden derivarse de la actuación humana, se centran principalmente en dos aspectos, uno debido a la sobreexplotación del medio en forma de agricultura, ganadería, pesca, talas de bosques, etc., que dan como resultado una degradación de tipo físico principalmente y el otro relacionado con las actividades industriales y mineras y con los modos de vida de la población (tendencia a concentrarse en núcleos de población cada vez mayores), cuyas emisiones y residuos afectan al medio siguiendo una componente esencialmente química.

El presente trabajo, intenta contribuir al segundo aspecto, específicamente en lo relativo a la situación en la que se encuentran los suelos y la distribución de los elementos llamados metales pesados, presentes en ellos, buscando un sistema que permita discriminar entre concentraciones *normales* o *naturales* y otras debidas a la actividad humana o derivadas de ésta, ante la ausencia, en la actualidad, de criterios precisos para conocer los valores correspondientes a la *concentración natural* de los distintos elementos o compuestos.

Las recomendaciones y directivas de los distintos organismos nacionales e internacionales, van encaminadas sobre todo hacia el control de situaciones que se consideran, con toda seguridad, real o potencialmente de riesgo tanto para la salud humana como para el medio ambiente, como son la vigilancia y control de emisiones de gases nocivos o la recuperación de zonas (suelos, sedimentos y acuíferos) cuyo contenido en éstos elementos o compuestos nocivos es con toda seguridad lo suficientemente elevada para que produzcan a corto y medio plazo efectos negativos en los aspectos citados de salud humana y medio ambiente, a la vez que se potencian las líneas de investigación dirigidas a un mejor conocimiento de las acciones y procesos que dan lugar al deterioro medioambiental.

Esta estrategia de trabajo de países y organismos internacionales, queda reflejada en publicaciones de distinto tipo, como guías, informes o comunicaciones, de distintos niveles, de las cuales, las que se han considerado

más relevantes, se pueden encontrar sus referencias en el apartado de Bibliografía, aunque no se citen en el texto de forma específica.

1.3.1. Necesidad de establecer niveles de referencia.

El hecho de que la detección de puntos o zonas contaminadas por metales pesados no resulte una tarea fácil, se debe principalmente a que, a diferencia de otros compuestos tóxicos, generalmente de naturaleza orgánica, su detección no es indicativa *per se* de la existencia de procesos contaminantes, ya que puede tener un origen natural debido a su presencia en los materiales que dan lugar al suelo como elementos minoritarios o trazas, de manera que además de por causas antrópicas, las acumulaciones de metales pesados pueden deberse a la actuación de fenómenos naturales sobre los materiales originarios que dan lugar al suelo como son los procesos de edafización a los que se superponen los fenómenos geológicos externos clásicos de erosión, transporte y sedimentación, que a su vez actúan de distinta forma según el clima, la topografía, etc. (Adriano, 2001; Matschullat *et al.*, 2000; Plant *et al.*, 2001; Salminen y Tarvainen, 1997; Salminen y Gregorauskienè, 2000; Tarvainen y Kallio, 2002, ...).

Gran cantidad de autores (Darnley *et al.*, 1995; Darnley, 1997; Plant *et al.*, 2001; Salminen y Tarvainen, 1997; Salminen y Gregorauskienè, 2000; Tidball y Ebens, 1976; ...), proponen el establecimiento de niveles de concentración o umbrales para cada elemento que permitan discriminar entre áreas contaminadas y no contaminadas, como solución al problema, es decir, delimitar un nivel de referencia o *background natural* a partir del cual una zona pueda etiquetarse como contaminada o no, e incluso indicar el grado de contaminación al que está o ha estado sometida.

De esta manera, queda resuelto el problema conceptual de la delimitación de zonas contaminadas, dejando en el aire, como cuestión a resolver, cual es el valor de concentración que hay que tomar como referencia.

Algunos autores, proponen como valor de referencia general, el correspondiente a las concentraciones medias de estos elementos en la corteza terrestre, como las publicadas por Turekian y Wedephol (1961), Taylor (1964), Taylor y McLennan (1985) o Rudnick y Fontain (1995), pero la mayoría

los desestiman precisamente por ser valores excesivamente genéricos, que no reflejan ni las fuentes locales de estos elementos, ni los procesos de acumulación, es decir, las litologías concretas sobre las que se desarrolla un suelo y los distintos procesos de alteración de rocas y sedimentos y procesos de edafización, que dan lugar a ese suelo. Baste considerar que, desde el punto de vista de la geoquímica, el término *background* o nivel base en un área determinada, se entiende como el valor o intervalo de valores de concentración de un elemento o compuesto a partir de los cuales, se supone que existe una anomalía positiva, en tanto que, desde el punto de vista ambiental lo que se quiere indicar es el umbral que separa las concentraciones de elementos o compuestos de origen natural, de las producidas como resultado de las actividades antrópicas. Pese a que, en principio, no parece que exista ningún problema para la conjunción de los dos criterios, realmente no es así, pues lo que se considera una anomalía positiva desde el punto de vista geoquímico no tiene por qué serlo desde el punto de vista medioambiental, de manera que los valores correspondientes a los *backgrounds* serían diferentes, por lo que lo normal es que se sobrevalore el *background ambiental* sobre el *geoquímico* si se trabaja a la misma escala, aunque tenderían a converger a medida que esta se fuese reduciendo.

Como se deduce de lo anterior, al problema de definir un *valor límite* de concentración natural, se añade el de poder generalizarlo al área que se esté estudiando.

Existen propuestas basadas en el uso de bioindicadores, que reflejen, tanto cuantitativa como arealmente, de alguna manera las concentraciones anómalas de elementos y compuestos tóxicos para la solución de ambos problemas conjuntamente (Caritat *et al*, 2001; Eberling *et al.*, 2003; Kooistra *et al.*, 2004; Ötvös *et al.*, 2003; ...), pero, en primer lugar, debido a su propia naturaleza, detectan los contaminantes cuando éstos han alcanzado ya un nivel de concentración lo suficientemente alto como para que queden reflejados en organismos vivos, y en segundo lugar la forma química en la que se encuentre el elemento es determinante a la hora de ser asimilado (Adriano, 2001; Cancès *et al.*, 2003; Thorton, 2002; Wang *et al.*, 2002; Su y Wong, 2003;

...), con lo que pueden quedar enmascarados los procesos que se quieren detectar. Este tipo de métodos son adecuados para detectar zonas contaminadas, pero no para ser usados en tareas de establecimiento de valores de concentración natural ni de prevención de la contaminación, pues sobrevalorarían los niveles de referencia.

Visto lo cual, a la hora de acometer un trabajo para el establecimiento de niveles de referencia y prevención de la contaminación, seguimos manteniendo las incógnitas de valor de referencia a tomar y de su representatividad, y por tanto de la fiabilidad del mismo, al extenderlo a toda el área de estudio

En cuanto al primer aspecto, se han propuesto una serie de soluciones a través de métodos de normalización, que como su propio nombre indica, consisten en estandarizar los valores de concentración del elemento bajo estudio frente a otro elemento o un isótopo estable del mismo, del que se esté seguro que no tiene enriquecimiento antrópico, para, a partir de los *datos normalizados*, construir un índice de enriquecimiento o acumulación por medio de la relación entre los valores obtenidos y los considerados naturales (Ciavola y Covelli, 1994; Covelli y Fontolan, 1995; Grant y Middleton, 1990; ...). En principio, parece que esta metodología resulta la más adecuada para definir valores de referencia, pero desgraciadamente, en la mayoría de las ocasiones suele ocurrir que no se dispone de los datos necesarios para ponerla en práctica, bien sea por falta del elemento de comparación, porque su concentración queda por debajo del límite de detección, porque la forma química del contaminante en cuestión lo haga más móvil y falsee los resultados o que la actuación de procesos geológicos externos o antrópicos, superpuestos a los de contaminación, produzca remociones o acumulaciones de materiales que proporcionen valores engañosos. No obstante, de cumplirse los requisitos necesarios, como son disponer de los datos e identificar los procesos que dan lugar al fraccionamiento, es probablemente el método que da los resultados más precisos, aunque su obtención suponga una mayor complejidad en la interpretación de los resultados, así como una mayor carga económica (Mason *et al.*, 2003; Meng y Maynard, 2001; Turer *et al.*, 2003; Weber y Hryñczuk, 2000; Yohn et al., 2002; ...).

Además de los métodos anteriores, para obtener los valores de referencia se han propuesto otros, basados en las relaciones entre las concentraciones de los contaminantes obtenidas en muestras a diferentes profundidades en suelos y en el material originario (Cobelo-García y Prego, 2003; Preda y Cox, 2002; Reimann *et al.* 2000; Reimann *et al.* 2001; Yohn *et al.*, 2002; ...), pero los valores obtenidos son sólo indicativos del comportamiento y evolución de la concentración del elemento en el suelo, no son lo suficientemente concretos en cuanto a la *concentración natural* del mismo, puesto que los valores de referencia se asignan subjetivamente en función de las variables consideradas (geología, topografía, etc.).

La complejidad del problema, ha llevado a algunos autores a preguntarse si realmente se puede obtener un valor real que refleje cual es la *concentración natural* de los metales pesados en una zona determinada. Matschullat *et al.* (2000), llegan incluso a publicar un artículo titulado *"Geochemical background – can we calculate it?"* , en el que ponen de manifiesto sus reservas a la hora de reconocer como nivel de referencia cualquier valor obtenido a través de los "métodos convencionales".

Desde un punto de vista más pragmático, Tidball y Ebens en 1976 resuelven este problema acuñando el término nivel base regional (*regional baseline*), al estudiar una zona fuertemente antropizada y sin grandes anomalías naturales, de manera que pudieron estimar un valor de fondo en el que la variabilidad de los datos se debía más al factor humano que a los naturales, por lo que el establecimiento de éste nivel de referencia, *baseline*, resultaba *más real* que si se hubiese realizado en otra zona en la que las fuentes de variación (geología, clima, topografía, etc.) hubieran sido mayores. Considera como valor de referencia o *regional baseline* al intervalo comprendido entre la media geométrica ± una desviación geométrica del 95% central de la distribución obtenida, este método se sigue utilizando actualmente, a pesar de sus consideraciones de partida (Chen *et al.*, 1988; Chen *et al.*, 1999; Ferreira *et al.*, 2001; Salminen y Gregorauskienè, 2000; ...).

El primer intento, a nivel internacional y escala global, de obtener unos niveles de referencia para elementos con el fin de proporcionar a las autoridades la información necesaria para establecer la legislación oportuna y las líneas de actuación futuras, se produce en 1996. Auspiciado por la IUGS (*International Union of Geological Sciences*) y la UNESCO, se establece formalmente el grupo de trabajo *Global Geochemical Baselines*, para coordinar las actividades científicas que se desarrollen en este campo. El punto de partida fue la planificación de la *Global Geochemical Database for Environmental and Resource Management* (Darnley *et al*, 1995), en la que se establecen los criterios de muestreo, y tratamiento de datos necesarios para establecer un nivel de referencia de elementos y compuestos, que pueda ser utilizado en estudios de contaminación (Darnley, 1997; Plant et al, 1997; Plant et al, 2001), también establecen el término *geochemical baseline*, definiéndolo como el valor de concentración de un elemento, obtenido siguiendo una serie de criterios estrictos: tamaño de cuadrícula (160x160 km), nº de muestras por celdilla (5) y metodología analítica a utilizar. Además indican el protocolo a seguir cuando se quieran integrar datos que hayan sido tomados con mayor detalle.

A partir del establecimiento del *Geochemical baseline*, se extiende su uso y se amplía su contenido, deja de ser un concepto estricto y pasa a considerarse como *el umbral de concentración de un cierto elemento a partir del cual se considera que ha habido aportes de origen no natural y por tanto existe contaminación* (Darnley, 1997; Plant et al. 2001; Salminen y Tarvainen, 1997; Salminen y Gregorauskienè, 2000; ...), de manera que pasa a considerarse como un equivalente al *background* desde el punto de vista medioambiental, con sus mismas ventajas e inconvenientes.

De esta manera, el problema de definir el nivel de referencia, esta vez bajo el apelativo de *geochemical baseline*, retorna: obtención de un valor de este umbral, que cumpla con la definición de *baseline* y sea representativo del área bajo estudio, aunque ahora queda delimitado a una superficie concreta y con una estrategia de muestreo e interpretación de datos sistemática, lo que deja vía libre a los investigadores para ensayar otras técnicas de análisis de

datos, cuyos resultados puedan mejorar la interpretación y el establecimiento de los niveles de referencia, *baselines* o niveles base.

La tendencia general hasta hace unos años, era representar el nivel base de una región por un solo valor o intervalo de valores, que se pretendía fuese representativo de la concentración de un elemento en dicha región, este valor o intervalo de valores solía ser la media aritmética de las concentraciones presentes en las muestras tomadas, en el caso de que se ajustasen a una distribución de tipo normal o la media geométrica si se ajustasen a una lognormal, pero dado que esta es una situación muy poco frecuente, pues las distribuciones normales o lognormales en la naturaleza, son más una excepción que una regla (Reimann y Filzmoser, 1999), algunos autores proponen, como mal menor, el uso de otros parámetros más robustos, en el sentido estadístico de término, como la mediana ± un margen de variación, que podría corresponder a los intervalos de confianza (Salminen y Tarvainen, 1997; Salminen y Gregorauskiene, 2000; ...).

Al planteamiento excesivamente simplista anterior, se oponen muchos autores que han puesto de manifiesto la incoherencia de querer representar por un solo valor el nivel de fondo para los suelos de una región, que normalmente presenta una distribución de datos tal, que hace que parámetros como la media o la mediana sean muy poco representativos de la misma, de hecho Salminen y Tarvainen (1997) y Tarvainen y Kallio (2002), ponen de manifiesto que existen áreas en el Norte de Finlandia, que consideran libres de contaminación, en las que las concentraciones de metales pesados son superiores a las que corresponden al baseline que marca la ley finesa, por lo que proponen la conveniencia de utilizar no un sólo nivel base, sino varios valores de fondo *más locales* para caracterizar a una región (Gregorauskienè y Kardunas, 1997; Salminen y Tarvainen, 1997; Salminen y Gregorauskiene, 2000, Tarvainen y Kallio, 2002; De Miguel *et al.*, 2002; ...). Estos *valores locales*, pueden estar referidos a distintas formaciones geológicas (De Miguel *et al.*, 2002; Salminen y Gregorauskiene, 2000; ...), a la litología, no sólo desde el punto de vista mineralógico sino también desde el granulométrico (Salminen y Tarvainen, 1998), o cualquier otro factor que se considere. En suma estos métodos se

basan en la subdivisión de la región bajo estudio siguiendo los criterios tradicionales en la línea del análisis de varianza, es decir, creando variables de control que determinen grupos y estableciendo si existen o no diferencias significativas entre ellos, para terminar asignando los valores de fondo a un parámetro característico, como pueden ser las medias o las medianas de los grupos.

A pesar de que el uso de varios niveles base supone una mejora en la exactitud y representatividad de los mismos, no deja de ser un método artificioso, en el que, por lo arbitrario y subjetivo de la elección de las variables de control *a priori*, puede inducir a plantear conclusiones sesgadas y por tanto incorrectas. Por ello, otros autores, hacen notar que el uso de otras metodologías basadas en el establecimiento de variables de control siguiendo criterios de probabilidad pueden dar mejor resultado (Dorronsoro *et al.*, 2003), y planteando que, probablemente, la mejor manera de representar los niveles de fondo de una región sería a través de un mapa temático para cada elemento, realizado a través de técnicas de interpolación o curvas de Bezier.

En la actualidad, gran parte de los trabajos que se realizan sobre contaminación de suelos por metales pesados, están enfocados desde un punto de vista de distribución, comportamiento y relaciones entre elementos, en ellos se hace uso de técnicas clásicas en geoestadística. Así, pueden ser estimadas las dependencias espaciales (ej. a través de variogramas o semivariogramas) y la distribución espacial (ej. por *kriging*), además de calcular y representar los errores estandar (Grosz *et al.*, 2004; Haapanen *et al.*, 2004; Lark, 2000; McGrath *et al.*, 2004; Juang *et al.*,2004;...), pero como hace notar Filzmoser (1999) el encontrar cuales son las razones que producen la distribución regional de una variable y en que medida lo hacen, es otra cuestión.

Para el estudio de las causas que dan lugar a una determinada distribución regional de una variable y estimar su valor de fondo, se vienen utilizando técnicas de análisis de datos, cada vez más complejas desde el punto de vista matemático, que van desde las correlaciones múltiples uni y

multivariantes (Huisman *et al.* 1997; Gordeev *et al.*, 2004; Reimann *et al.* 2001; Preda y Cox, 2002; ...), hasta el análisis discriminante, el análisis de factores (De Vivo *et al.*, 1997; Fachinelli *et al.*, 2001; Filzmoser, 1999; Gallego *et al.* 2002) o el análisis de *clusters* (Bandyopadhadyay, 2004; Buurman *et al.*, 2004; Farber y Kadmon, 2003; Globocanin *et al.*, 2004; Lacassie *et al.*, 2004; Meng y Maynard, 2001; Meyer *et al.*, 2004; Motelay-Massei *et al.*, 2004; Tayfur *et al.*, 2003; Therfeld *et al.*, 2003; Yücer y Demil, 2004; Zhang, Y.G. *et al.*, 2004; Zhang, B. *et al.*, 2004;...) e incluso técnicas de análisis de Fourier (Gallego *et al.*, 2002). En estos trabajos se suele buscar la identificación de las fuentes de contaminantes y el modo en el que se distribuyen, a partir de las relaciones existentes entre ellos y con otras variables explicativas (geología, topografía, etc.), de forma que se toman como *puntos fuente* de contaminantes los correspondientes a los valores extremos y su distribución se explica a partir del comportamiento general. Este tipo de metodologías resultan muy útiles a la hora de reducir la dimensionalidad de variables o datos, es decir, estimar el *comportamiento general* y las relaciones entre las variables, y por tanto de la mayoría de los datos, de forma que resulte más fácil la interpretación de los resultados, pero es aquí precisamente donde éstos métodos tienen su punto débil, pues la reducción de dimensionalidad conlleva una pérdida de información, de manera que los valores locales pierden importancia frente a los generales, por lo que al aplicarlos a estudios sobre contaminación encontraríamos que se les resta influencia a los *valores discordantes o puntos díscolos*, que corresponderían a los puntos clave y tienden a uniformizar los demás.

1.3.2. Estado actual y justificación del desarrollo de nuevas metodologías.

Tal y como se ha podido observar en el capítulo anterior, el hecho de utilizar con profusión los métodos estadísticos en estudios de contaminación de suelos, se debe a que, en muchos casos, el comportamiento del fenómeno objeto de estudio es demasiado complejo, o no se dispone de la suficiente información, para explicarlo mediante una sucesión de procesos físico-químicos, es decir que no se conocen con el suficiente detalle o profundidad, cómo actúan los procesos que dan lugar a los resultados finales ni qué grado de influencia tienen en ellos. Por ello, se ha incrementado el uso de técnicas de creciente sofisticación estadístico-matemática, en algunos casos con el fin de establecer las pautas de comportamiento generales, y a partir de ellas, clarificar los criterios de actuación para abordar problemas concretos (Abonyi y Szeifert, 2003; De Vivo *et al.*, 1997; Fachinelli *et al.*, 2001; Globocanin *et al.*, 2004; Farber y Kadmon, 2003; Filzmoser, 1999; Lacassie *et al.*, 2004); Lin y Chen, 2004; Meng y Maynard, 2001; Meyer *et al.*, 2004; Motelay-Massei *et al.*, 2004; Thierfeld *et al.*, 2003; Yücel y Demir, 2004; ...), y en otros aplicando directamente la nueva información para inferir comportamientos y establecer conclusiones (Gallego *et al.*, 2002; Gordeev *et al.*, 2004; Huisman *et al.* 1997; Preda y Cox, 2002; Reimann *et al.* 2001; ...).

Se debe tener en cuenta que, a diferencia de otros trabajos, en los que se suelen medir las respuestas de una variable frente a los cambios de un factor conocido, en este caso, de lo que se dispone es de una serie de datos o valores que son el resultado de la acción o acciones, en algunos casos superpuestas, de varios fenómenos o factores, de los que no se tiene una idea exacta de cómo actúan ni de con qué intensidad lo hacen, por lo que las técnicas de reducción de dimensionalidad y de análisis de datos que se vienen utilizando, en general, no están produciendo resultados satisfactorios, ya que la falta de conocimiento sobre cómo actúan de forma cuantitativa los fenómenos implicados, sus interrelaciones y su posible superposición, obliga a hacer uso de una serie de supuestos sobre el comportamiento de las variables y los factores de variación, que, en el mejor de los casos, hacen imprecisa la interpretación de los resultados.

Si exponemos de manera esquemática el problema en cuestión, identificando las fuentes de contaminantes, los procesos que pueden dar lugar a su concentración y cuales son los factores que pueden influir en su redistribución obtenemos el cuadro siguiente:

1º.- El origen de los elementos contaminantes puede tener dos fuentes primarias:

a) Sustrato geológico.

b) Aportaciones de origen antrópico.

2º.- Los elementos contaminantes pueden concentrarse en determinadas zonas por:

a) Causas naturales, por procesos de sedimentación o por procesos de edafización.

b) Causas antrópicas, como cercanía a la fuente de producción del contaminante o vertidos.

Además, pueden producirse redistribuciones superficiales de contaminantes debido a la actuación de procesos, los de tipo geológico externo fundamentalmente, y en profundidad, por procesos edáficos.

3º.- Debido a la evolución del desarrollo social y la mejora de los rendimientos industriales, el tipo, la *calidad* y la distribución areal de los contaminantes puede variar con el tiempo.

De la observación del cuadro anterior, se puede concluir que a la hora de estudiar los procesos de contaminación:

1º.- Las relaciones entre las concentraciones de distintos contaminantes en conjunto, no tienen por qué existir, y si existieran, ser lineales, debido a la variabilidad del sustrato y a la del origen y distribución de los contaminantes.

2º.- La única información directa y cuantitativa disponible sobre la condición de punto contaminado o no y sobre la distribución de la

contaminación se encuentra incluida en los datos de concentraciones de los elementos contaminantes y en la localización geográfica de los puntos muestrales.

Lo cual invalida el uso generalizado de las técnicas estadísticas o geomatemáticas uni o multivariantes convencionales en este tipo de trabajos, pues éstas consideran, en mayor o menor medida, que debe existir una relación en el comportamiento de las variables en su conjunto, así:

Las técnicas univariantes exigen un buen ajuste de los datos a un modelo conocido, pues de otra forma los parámetros clásicos que definen la función de distribución y de probabilidad (media, mediana, desviación típica u otros momentos), no son representativos de la misma, además, otros parámetros de uso común en otras areas, básicamente en econometría, como la mediala o los índices de concentración (Gini, Theil, etc.), nos indican sólo la existencia o no de homogeneidad en el reparto de los valores, es decir, el grado de agrupación que presentan los datos en la distribución.

Las técnicas multivariantes presuponen la existencia de algún tipo de relación entre las variables, o que estas se ajustan bien a un modelo de distribución clásico (normal, lognormal, etc.)

El análisis discriminante busca una combinación lineal entre variables que produzca las máximas diferencias entre grupos previamente definidos, lo cual obliga reducir la dimensionalidad de las variables y a elegir de forma subjetiva los grupos, lo que conlleva una pérdida de información por la reducción de dimensionalidad de las variables y porque la función discriminante obtenida persigue el objetivo de adaptarse a la mayor parte de los datos, con lo cual pierde *precisión local* y da más influencia a los datos que se ajusten al modelo previo que a los que no lo hagan.

El análisis de factores intenta localizar grupos homogéneos de variables a partir de un grupo mayor, los grupos se forman con las variables muy correlacionadas entre sí, y establece la estructura de los factores a partir de

matrices de correlación y covarianza, lo que implica la presuposición de existencia de algún tipo de relación entre las variables, así como una pérdida de información al reducir la dimensionalidad al igual que el análisis discriminante. además, a través de ésta técnica, sólo se distinguen como valores anómalos los *outliers*, no los *inliers* (datos anómalos cuyo valor no es lo suficientemente alto como para producir cambios detectables en los resultados del análisis, pero que lo desvirtúan o le hacen perder exactitud).

Mención aparte merece el análisis de *clusters*, que busca la división de los datos en subgrupos en base a criterios de similaridad entre todos los puntos muestrales y todas las variables consideradas. Las medidas de similaridad se hacen en base a correlaciones, distancias medias, euclidianas, etc.. Éste método parece el más adecuado, desde el punto de vista conceptual, para la detección de puntos contaminados, pues a partir de su aplicación, al menos en teoría, deberían distribuirse los datos ,y por tanto agruparse, de distinta manera los puntos que están contaminados y los que no lo están, dando lugar a agrupaciones *homogéneas naturales* a un cierto nivel, y encontrándose en niveles sucesivos, las agrupaciones correspondientes a los puntos afectados por un cierto grado de contaminación, acorde con la distancia de separación de los subgrupos previos.

Pese a lo expuesto, el uso de éste método desde un punto de vista estricto, presenta algunos inconvenientes serios, el primero consiste en la exigencia de utilización conjunta de más de una variable (todas o parte de las disponibles) para su cálculo, y algunas de ellas, como el tipo de suelo o la litología subyacente, son de tipo cualitativo, y su transformación a cuantitativa va acompañada de un error que interfiere en la interpretación de los resultados. Y un segundo relacionado con el método de agrupación utilizado, resultando:

a) Si se utilizan los algoritmos de k-medias se está obligado a definir un número máximo de agrupaciones de forma subjetiva

b) Los algoritmos de tipo jerárquico, funcionan de forma iterativa, van produciendo agrupaciones de mayor similitud hasta terminar dando lugar a un

solo grupo, presentándose entonces la necesidad de elegir subjetivamente el nivel o niveles que determinan los *grupos homogéneos* que se buscan.

Desde el punto de vista de este trabajo, estos métodos, deberían usarse únicamente para reducción de la dimensionalidad de datos o de variables previo al planteamiento del problema de forma definitiva, en el estudio de fenómenos complejos como análisis exploratorio, para inferir cuales son los procesos que influyen en mayor medida en el fenómeno o, como confirmación de hipótesis.

Como se deduce de lo anterior, para el estudio de fenómenos como el citado, u otros de problemática similar, se hace necesario el desarrollo de nuevas metodologías que, al menos en principio, no hagan uso de supuestos que puedan ser erróneos y desvirtúen las interpretaciones, así como el planteamiento de expectativas y el establecimiento de conceptos más acordes con la información disponible.

1.4. Objetivos concretos.

Una vez identificados los problemas que se presentan a la hora de identificar suelos contaminados, estimar el grado de contaminación y generalizar los valores de referencia. En este trabajo se van a proponer tres objetivos que se supone mejoren el conocimiento en las tareas citadas:

1º Definir el concepto de *niveles característicos* como *los intervalos de valores de las concentraciones de metales pesados que guardan una mayor similitud en agrupaciones de puntos muestrales realizadas en base a variables explicativas* (ej. Litología, topografía, clima, etc.).

2º Desarrollar la metodología de análisis de datos necesaria para poder agrupar de forma objetiva las muestras y eventualmente realizar una clasificación.

3º Una vez caracterizadas las *concentraciones normales* de los distintos elementos, *niveles característicos*, en base a ellos localizar zonas de concentración anómala, que permita aseverar si se trata de puntos contaminados o no y en caso de que resulten positivos inferir en que medida lo están.

4º Ajustar un sistema de umbralización que permita representar y extender los niveles de referencia (*baselines*) a otras escalas de trabajo, siguiendo criterios de maximización de su probabilidad de aparición.

2. METODOLOGÍA

3. METODOLOGÍA

2.2. Condiciones a cumplir por una nueva metodología.

Tal y como se ha planteado el problema en el capítulo anterior, la metodología a desarrollar debe de cumplir una serie de condiciones:

a) Estar libre de supuestos de comportamiento previos, pues no se puede afirmar *a priori* que las variables sigan una distribución determinada, ni que existan relaciones funcionales entre ellas.

b) El establecimiento de grupos de *datos homogéneos* debe hacerse con el mínimo de pérdida de información, siguiendo criterios de similaridad, que permitan su explicación a partir de la distribución de los factores condicionantes o patrones explicativos. Además debe asumir una cierta heterogeneidad tanto en el factor explicativo como en el muestreo.

c) Ha ser sensible tanto a los *outliers* (valores extremos) como a los *inliers* (valores anómalos situados en el interior de la distribución), no detectables por métodos convencionales.

d) Debe de ser *compatible*, en la medida de lo posible, con los métodos que se han venido utilizando hasta ahora.

Por lo tanto, el cumplimiento de estas condiciones, implica:

1º) Estudiar las variables por métodos libres de cualquier supuesto de comportamiento *a priori*.

2º) Desarrollar un método de agrupación que siga simultáneamente criterios de similaridad y continuidad.

3º) Elaborar un sistema de detección de *outliers* e *inliers*.

4º) Proponer un método de integración de datos que haga *compatibles* los resultados con otras metodologías.

En cuanto al punto 1º, se puede afirmar que el estudio de las variables una a una y por separado mediante el uso de técnicas no paramétricas, parece ser la solución más idónea, pues está libre de cualquier supuesto de ajuste a un tipo de distribución predeterminada así como de cualquier supuesto de relación funcional entre ellas.

La propuesta de estudiar las variables por separado, puede parecer equivocada desde el punto de vista de los trabajos tradicionales, pues se trata de técnicas univariantes que no aportan nada al conocimiento sobre cuales son las relaciones existentes entre los distintos elementos contaminantes y de su *comportamiento en conjunto*, pero hay que tener en cuenta que el uso de las técnicas de análisis multivariante, tanto las tradicionales (De Vivo *et al.*, 1997; Fachinelli *et al.*, 2001; Filzmoser, 1999; Gallego *et al.* 2002; Gordeev *et al.*, 2004; Huisman *et al.* 1997; Reimann *et al.* 2001; Preda y Cox, 2002; ...) como las modificaciones y desarrollos recientes (Buurman *et al.* 2004; Farber y Kadmon, 2003; Globocanin *et al*, 2004; Lacassie *et al.* 2004; Meng y Maynard, 2001; Meyer *et al.*, 2004; Motelay-Massei *et al.*, 2004; Tayfur *et al.*, 2003; Therfeld *et al.*, 2003; Yücer y Demil, 2004; ...), suponen en definitiva, el establecimiento de una *estructura general de comportamiento* de las variables en base a las relaciones existentes entre ellas, lo que además de producir una pérdida de información, que puede ser crucial, debida a la reducción de dimensionalidad de datos y variables, a los efectos de este trabajo se traducen en una serie de supuestos que están en contra de las observaciones y el conocimiento previo disponible:

a) Continuidad de las variables, no considerando de esta manera la existencia de umbrales, geoquímicos, como los que se pueden producir en contactos entre materiales de composición muy diferente como pueden ser una roca ígnea básica, una metamórfica o una sedimentaria, o de rupturas y cambios de pendiente entre otros.

b) Existencia de relaciones funcionales en el conjunto de los datos, ya que la estructura obtenida, es la resultante de la combinación de todos los

valores disponibles, considerando así, de manera implícita, que las relaciones entre las concentraciones de los distintos elementos contaminantes es parecida en todos los tipos de sustrato o roca, que la alteración de distintos minerales, y por tanto la liberación de elementos contaminantes, sigue una dinámica similar, etc.

Aunque, en aquellos casos en los que se den las condiciones adecuadas de homogeneidad de sustrato, como de condiciones del medio, las técnicas de análisis multivariante se puedan aplicar con buenos resultados.

Los tres puntos siguientes: 2º) Desarrollo de un método de agrupación que siga simultáneamente criterios de similitud y continuidad. 3º) Elaboración de un sistema de detección de *outliers* e *inliers*. y 4º) Propuesta de un método de integración de datos que haga *compatibles* los resultados con otras metodologías. No presentan una solución tan fácil como el primero pues exigen el desarrollo de nuevas metodologías de cálculo, que se abordan el los capítulos siguientes.

2.3. Propuesta de un sistema de agrupación.

Tal y como se refleja en los objetivos de este trabajo, la definición de niveles característicos, pasa por conseguir un método de agrupación de datos tal, que los subgrupos de datos resultantes cumplan la propiedad de que cada elemento que pertenezca a un grupo se parezca más al resto de los componentes de ese grupo que a los de otro, por próximo que se encuentre, y sin hacer uso de condiciones *apriorísticas*, que supongan un comportamiento predeterminado de las variables y que por tanto puedan introducir un sesgo en la interpretación de los valores resultantes. Es decir, se busca un método que establezca una serie de agrupaciones de valores que por su similitud individual y su *independencia* con respecto a grupos adyacentes permita considerar *homogéneos* a los valores que lo constituyen.

Normalmente, cuando se dispone de un conjunto de datos en los que se desea establecer subgrupos o clasificarlos, se suele hacer uso de criterios o medidas de semejanza, que en el caso de variables cuantitativas se basan en un criterio de minimización de las diferencias o distancias existentes entre los valores entre sí o con respecto a un parámetro. Para la obtención de éstas medidas de distancias o semejanzas, hay multitud de sistemas propuestos: distancias de Cook, euclidianas, de Mahalanobis, de Matushita, etc., a partir de las cuales se pueden agrupar los datos en función de su parecido, desde el punto de vista de que existan menos diferencias entre ellos, siguiendo un método específico de cálculo para cada una de ellas. Las agrupaciones resultantes, se realizan mediante iteraciones, de manera que se van estableciendo agrupaciones a distintos niveles en cada iteración, quedando reflejado el *factor de semejanza* a través de la distancia que separa a un nuevo punto o agrupación de un subgrupo previamente establecido. En tanto que, cuando se trata de variables cualitativas, se utilizan las llamadas distancias de Hamming, que consisten en el contaje y comparación de de atributos diferentes para cada muestra.

Estos métodos que, aparentemente suponen la solución al problema que se plantea, de hecho su uso es cada vez más común tanto en el campo en el

que nos movemos como en otros similares (Bandyopadhadyay, 2004; Buurman *et al.*, 2004; Farber y Kadmon, 2003; Globocanin *et al.*, 2004; Lacassie *et al.*, 2004; Meng y Maynard, 2001; Meyer *et al.*, 2004; Motelay-Massei *et al.*, 2004; Tayfur *et al.*, 2003; Therfeld *et al.*, 2003; Yücer y Demil, 2004; Zhang, Y.G. *et al.*, 2004; Zhang, B. *et al.*, 2004;...) se incluyen en los llamados métodos de análisis de *clusters*., que como ya se ha visto anteriormente presentan varios inconvenientes para su uso generalizado.

1º.- El tipo de medida de similaridad a utilizar, ya que según se tome una u otra, ej. euclidianas o de Mahalanobis, los resultados pueden ser diferentes, debido a que cada una de ellas sigue un método de cálculo diferente, asumiendo por tanto un comportamiento predeterminado de las variables, lo cual está en contra del planteamiento inicial.

2º.- Independientemente del sistema de obtención de similaridades (distancias), todas ellas parten del supuesto de que se está trabajando con una distribución continua y en la mayoría de los casos multivariante, lo que presupone un cierto grado de relación *a priori* de las variables (colinearidad o covariación), de manera que si no existe tal relación obtenemos unos resultados sesgados de muy difícil interpretación.

3º.- La conveniencia de estandarizar o normalizar las variables, para que en el cálculo de las similitudes no se les conceda más importancia a unas que a otras, pues si se realiza la estandarización se pierde precisión.

4º.- El orden de introducción de las variables puede influir en el resultado, ya que el cálculo de las similaridades al añadir una nueva variable, está condicionado por los cálculos previos.

5º.- Sea cual sea el método de *clustering* utilizado, k-medias o jerárquico, obliga a decidir de forma subjetiva los *grupos homogéneos* que se persiguen, el primero por definir desde el inicio el número de grupos resultantes y el segundo por obligar a decidir subjetivamente en qué iteración se considera que están formados los *grupos homogéneos*.

Pese a los inconvenientes citados, los métodos que produzcan *agrupaciones homogéneas*, conceptualmente parecen los más adecuados para trabajos como el presente, hay que recordar que se parte de la hipótesis de que *fenómenos iguales que actúan de idéntica manera, con la misma intensidad y en el mismo orden dan lugar a resultados idénticos*, y por tanto este comportamiento debe quedar reflejado en una cierta similitud en los datos obtenidos, lo que obliga a hacer una serie de correcciones en los planteamientos clásicos de los análisis de *clusters*:

1º.- No presuponer la existencia de una relación estadística o funcional entre las distintas variables.

2º.- No considerar *a priori* que las variables se ajustan a un tipo de distribución definida ni que son continuas.

Estas correcciones obligan a:

1º.- Estudiar las variables una a una y por separado, dejando las consideraciones sobre relaciones entre ellas para más tarde.

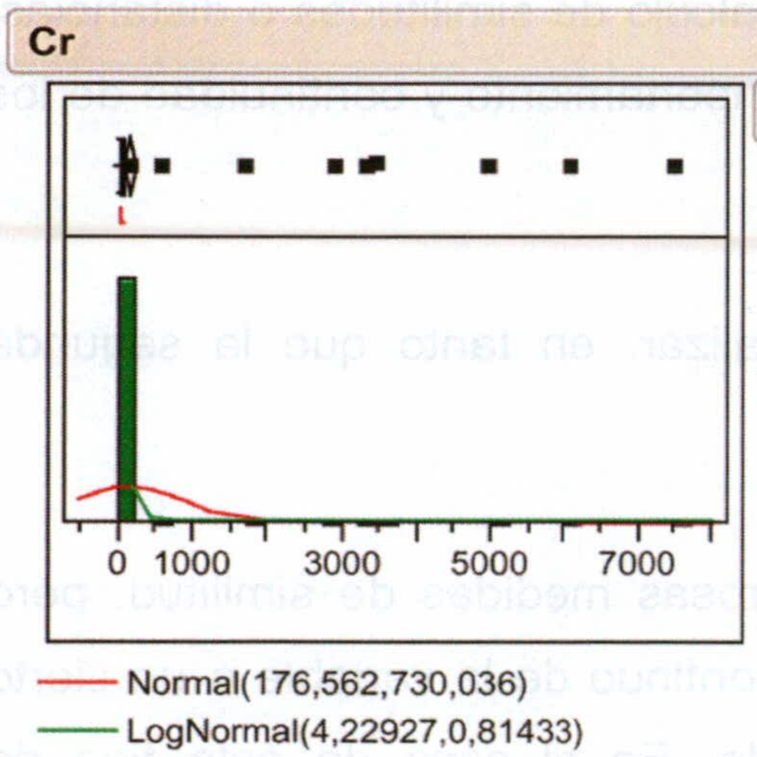
2º.- Desarrollar un sistema nuevo de cálculo de similitudes o distancias, que no se vea afectado por supuestos de comportamiento y continuidad de los datos.

La primera corrección es fácil de realizar, en tanto que la segunda requiere un esfuerzo un poco mayor.

Como ya se ha citado, existen numerosas medidas de similitud, pero todas ellas presuponen un comportamiento continuo de la variable o un cierto grado de ajuste a una distribución conocida. En el caso de éste tipo de trabajos, en el que podemos encontrar una variabilidad de datos tal que sugiere en muchos casos la existencia de varias poblaciones mezcladas, -hay que recordar que algunos autores como Salminen y Tarvainen (1997), Salminen y

Gregorauskienè (2000), etc., proponen el uso de varios niveles de referencia-, la *medida de similitud* que se proponga debe seguir otros derroteros, de forma que el criterio a seguir no suponga necesariamente un comportamiento continuo y más o menos lineal de las variables en cuestión.

Partiendo de la base de que la información de cualquier tipo contenida en una variable reside en la distribución de la misma, se deduce fácilmente que los parámetros clásicos, como la media, la desviación típica, otros momentos, la mediana o la moda, aportan muy poca información si los datos no se ajustan a una distribución conocida, normal, lognormal, etc., pues pierden tanto sus propiedades de inferencia, como su significado conceptual, pasando a ser meros indicadores del comportamiento general de los datos que en el mejor de los casos solo aportan información cualitativa, como puede verse en la Figura 2, en la que puede verse cuales serían las distribuciones normales y lognormales correspondientes a las medias y desviaciones obtenidas, así como comparar los valores correspondientes a la media y a la mediana, resultando que el 50% de los datos tiene valores iguales o inferiores a 66.00 y el valor medio asignado es de 176.56.



Quantiles			Moments	
100.0%	maximum	7530,0	Mean	176,56162
99.5%		6970,0	Std Dev	730,0359
97.5%		1750,0	Std Err Mean	43,706106
90.0%		115,0	upper 95% Mean	262,59857
75.0%	quartile	89,0	lower 95% Mean	90,524665
50.0%	median	66,0	N	279
25.0%	quartile	46,0	Sum Wgt	279
10.0%		31,0	Sum	49260,692
2.5%		21,0	Variance	532952,42
0.5%		5,6	Skewness	7,602706
0.0%	minimum	0,0	Kurtosis	61,911104
			CV	413,47372
			N Missing	0

Fig. 2. Estadística descriptiva del Cr.

En otras disciplinas como la econometría frecuentemente se añaden a las tablas de la estadística descriptiva otros parámetros e índices de concentración, como:

- Mediala (MI), valor de la variable tal que la suma de las observaciones mayores que él es igual a la suma de las observaciones menores que él.

- Índice de Gini, alcanza valores entre 0 y 1, el valor 0 indica concentración mínima, por tanto la muestra está uniformemente repartida a lo largo de todo el rango, y el 1 concentración máxima, de forma que un sólo valor acumula el 100% de los datos. Se calcula a partir de la Curva de Lorentz.

- Índice de Theil, que se basa en la entropía (medida de desorden) de la distribución de frecuencias, obteniéndose a partir de la fórmula:

$$T = 1 + (\sum pi*\log(pi)/\log(n))$$

Donde:

pi es el procentaje que representa el valor i frente a la muestra total.

Al igual que el índice de Gini, alcanza valores entre 0 y 1, el 0 indica concentración mínima y el 1 concentración máxima.

Éstas medidas de concentración o agrupamiento, proporcionan más información sobre la distribución de los datos, indicando en el caso de la mediala dónde se encuentran situados los valores más altos y más bajos, y en el caso de los índices de Gini y Theil el grado de concentración o dispersión general, pero no dan más que una información vaga de cuales son las zonas de mayor densidad de puntos, aunque algunos autores proponen métodos de segmentación o agrupación de los datos para delimitar las agrupaciones de concentraciones similares, Akita (2003) utiliza criterios apriorísticos similares a

los clásicos de análisis de varianza, Dikhanov (1996) utiliza los cuantiles, Kapur y Kesawan (1992) y Reesor y McLeish (2002) proponen métodos multivariantes y transformaciones de las variables, etc.. Pero en definitiva, estas propuestas, en lo que a este trabajo se refiere, se diferencian muy poco de las técnicas ya comentadas que parten de supuestos apriorísticos.

Visto lo anterior, se confirma que no existe ningún método que permita crear agrupaciones de datos de manera objetiva, sin hacer ningún tipo de suposición *a prioi*, por lo que para su desarrollo hay que partir de cero.

Para ello, se considera:

a) Las agrupaciones homogéneas, están refrendadas matemáticamente por las teorías de puntos autoconsistentes y puntos principales de Flury (1993 y 1995) y Tarpey (1998 y 2000), que definen puntos principales como el conjunto de k puntos que representa óptimamente una distribución en términos de error cuadrático medio y puntos autoconsistentes al conjunto de k puntos que coinciden con su media condicionada a los dominios de atracción que generan de acuerdo con la distancia minimal (González Caballero y Peralta Sáez, 2003). Es decir, que los conjuntos de puntos resultantes de una eventual agrupación, contienen un punto principal y que sus límites están fijados por los dominios de atracción que éste genera o Dominios de Voronoi, lo que en definitiva supone una serie de agrupaciones de puntos o valores que cumplen la condición de parecerse más entre sí que a cualquier otro punto o valor que pertenezca a un grupo distinto.

b) Las agrupaciones que se obtengan no deben sufrir pérdida de información o distorsión de la misma, lo que equivale a que la entropía correspondiente a la distribución completa calculada de forma discreta, sea igual a la de la suma de las entropías de

los grupos resultantes, pudiendo medir la entropía a través de los índices de Theil.

De esta manera se dispone, desde el punto de vista matemático, de los dos criterios necesarios para establecer las agrupaciones, uno derivado de los índices de concentración de Theil, pues las agrupaciones que se obtengan deben cumplir que la entropía de la distribución no varía al realizar su cálculo en forma discreta (punto a punto) o en intervalos, con lo cual se demuestra que no hay pérdida de información en los grupos y un segundo, basado en las teorías de puntos principales, que obliga a que los puntos o valores incluidos en un mismo grupo cumplan la condición de parecerse más entre sí que a otro punto perteneciente a un intervalo diferente.

Tanto Flury como Tarpey, destacan en sus trabajos la enorme dificultad que supone la obtención de los puntos principales de una distribución de tamaño medio, pues se basan en métodos de cálculo iterativos del tipo k-medias, que necesitan mucho tiempo de cálculo y un uso intensivo de ordenador.

Pero si se enfoca el problema de otra manera, haciendo uso de la filosofía de la teoría de la información (Shannon, 1948), se puede simplificar el problema del cálculo sensiblemente.

ENTROPÍA DE SHANON

El concepto de *entropía* de un sistema físico fue introducido por primera vez por Clausius en 1864 como una medida del desorden del sistema, que puede expresarse en términos de otras macrocoordenadas que sí pueden medirse de forma directa. No obstante, este concepto clásico de entropía de Clausius no era de naturaleza probabilística, y fue Bolzmann en 1896 quien definió la entropía en termodinámica estadística como una medida de las posiciones y velocidades de todas las partículas incluidas en el sistema físico, enfatizando así su significado probabilística.

En base al desarrollo de la teoría de la probabilidad durante la primera mitad del siglo XX, Shannon introdujo en 1948 la entropía en abstracto, por analogía con la expresión de Bolzmann, como una medida de la incertidumbre de experimentos probabilísticas arbitrarios, es decir una medida cuantitativa sobre la cantidad de información proporcionada por el experimento. Esto permitió a Kullback y Leibler (1951) definir la *divergencia* como una medida de la distancia entre dos poblaciones.

Consideremos un experimento aleatorio cuyos posibles resultados son a_1, a_2, …, a_n con probabilidades respectivas p_1, p_2, …, p_n ($pi \geq 0$, $1=1,2,\ldots,n$, $p_1+p_2+\cdots+p_n=1$). Es evidente que dicho experimento puede representarse mediante una variable aleatoria X con función de masa $P(X= a_i)=p_i$. Se denomina entropía de la variable X a la expresión:

$$H(X) = -\sum_{i=1}^{N} p_i \log p_i$$

Los logaritmos se pueden tomar respecto a cualquier base que sea mayor que la unidad. De hecho, si se toma en base dos, a la unidad correspondiente se le denomina *BIT* (Binary digit) y puede definirse como la entropía correspondiente a una variable aleatoria con dos resultados equiprobables:

$$H(X) = -\frac{1}{2}\log_2 \frac{1}{2} - \frac{1}{2}\log_2 \frac{1}{2} = \log_2 2 = 1\ BIT$$

Si se toman en base 10, como es usual, la unidad correspondiente se denomina *DIT* (Decimal digit o unidad de Hartley) y se define como la entropía de una variable aleatoria con diez resultados equiprobables:

$$H(X) = -\frac{1}{10}\log_{10} \frac{1}{10} - \cdots - \frac{1}{10}\log_{10} \frac{1}{10} = \log_{10} 10 = 1\ DIT$$

Si se toman logaritmos neperianos, la unidad correspondiente se denomina *NAT* y su interpretación encaja entro de la entropía como medida de la incertidumbre asociada a variables aleatorias continuas.

ESTIMACIÓN DE ENTROPÍAS (IGUALDAD DE DIVERSIDADES)

El concepto de diversidad aparece en numerosos campos de investigación asociado siempre a la idea de variabilidad de los elementos de una determinada población. Así, la entropía de Shannon permite cuantificar esta diversidad desde un punto de vista teórico, pero puede suceder que no se conozca su valor más que a nivel muestral, siendo necesario desarrollar expresiones para estimar de forma óptima la entropía de toda la población. En nuestra situación, dado que los tamaños muestrales son suficientemente grandes, el estadístico para el contraste

H_0: $H(\Theta) = D_0$

H_1: $H(\Theta) \neq D_0$

viene dado por:

$$Z = \frac{\sqrt{n}[H(\theta) - D_0]}{\hat{\sigma}}$$

siendo

$$\hat{\sigma}^2 = \sum_{i=1}^{n} p_i (\log p_i)^2 - H(\theta)^2$$

la estimación de la varianza. Su distribución asintótica se ajusta por el teorema de Slutsky (Ferguson, 1996) a un modelo Gausiano tipificado.

Igualmente, una vez calculadas las entropías para diferentes elementos, el objetivo siguiente consiste en averiguar si existen diferencias significativas entre ellas, es decir si la pérdida de información consecuencia de los distintos métodos de agrupación son dignas de tenerse en cuenta. Para ello contrastaremos la hipótesis nula

H_0: $H(\Theta_1) = H(\Theta_2)$

frente a la alternativa

H_1: $H(\Theta_1) \neq H(\Theta_2)$

Dado que el tamaño muestral es en todos los casos grande (n=278), se utilizará el siguiente estadístico de contraste (Pardo, 1997):

$$Z = \frac{\sqrt{n_1 n_2}[H(\theta_1) - H(\theta_2)]}{\sqrt{n_2 \hat{\sigma}_1^2 + n_1 \hat{\sigma}_2^2}}$$

donde n_i denota el tamaño muestral y s_i^2 es la cuasivarianza muestral asociada a la entropía $H(\Theta_i)$, que se distribuye asintóticamente según una normal N(0,1). Como $n_1 = n_2 = 278$, la expresión del estadístico queda reducida a:

$$Z = 16,67333 \frac{H(\theta_1) - H(\theta_2)}{\sqrt{\hat{\sigma}_1^2 + \hat{\sigma}_2^2}}$$

Así:

1º.- Si se ordenan los datos en forma creciente, obtenemos una sucesión de valores en la que los datos quedan situados de forma que los más parecidos, es decir los que guardan mayor similitud, se encuentran más próximos.

2º.- Si en vez de utilizar el criterio tradicional de calcular las diferencias/similitudes entre puntos y estudiar la distribución de los dominios obtenidos sobre los valores, se estudia el comportamiento de las distancias entre puntos consecutivos, se observa que los valores quedan agrupados automáticamente entre máximos relativos de las distancias calculadas.

Siguiendo este planteamiento, se obtienen agrupaciones que cumplen las condiciones impuestas, en primer lugar no hay pérdida de información, lo que queda demostrado a los efectos de este trabajo en que no hay variación en los índices de Theil y por tanto en la entropía de la distribución, y además los valores quedan agrupados siguiendo un criterio de máxima similitud puesto que el valor medio de cada una de estas agrupaciones constituye un punto principal de la distribución de los mismos, resultando *agrupaciones autoconsistentes* univariantes, lo que en definitiva resulta ser un método de obtención de *clusters* univariante, que además presenta una serie de ventajas sobre los métodos tradicionales, que esencialmente consisten en:

a) Obtener directamente las agrupaciones finales sin necesidad de cálculos iterativos.
b) Conocer de manera cuantitativa la pérdida de información que se produce en el proceso de clasificación comparando el índice de Theil de la distribución en forma discreta (original) y el obtenido al calcularlo con los intervalos resultantes.
c) En el caso de resultar un número de grupos demasiado elevado para interpretarlo con facilidad, se pueden recalcular las agrupaciones sabiendo la información que se pierde en cada iteración.

2.4. Propuesta de un sistema de detección de *outliers* e *inliers*.

Una vez agrupados los valores según la metodología propuesta en el apartado anterior, resta el problema de detectar valores anómalos que puedan haber sido incluidos en un grupos y que su presencia en ellos se deba más a la condición de continuidad que a la de cercanía o similitud, lo cual traería consigo el que pasaran desapercibidos a pesar de ser valores anómalos. Para evitar esto se hace necesario afrontar el problema sólo desde el punto de vista de la similitud pero esta vez dentro de grupos ya establecidos.

Podría pensarse que una solución aceptable sería el tomar como referencia alguna de las distancias intragrupo, bien la distancia media entre puntos consecutivos o bien la distancia mediana, pero su uso rompería la condición de continuidad, además de hacer muy difícil su interpretación. Para verlo, bastaría considerar que la distribución de datos en un grupo se ajustase a una distribución conocida (normal, lognormal, etc.), encontrando entonces que los valores correspondientes a las colas de la distribución quedarían marcados como puntos anómalos, lo cual invalidaría el criterio.

Sin embargo, si en vez de utilizar las distancias intragrupo utilizamos las intergrupo, se tiene que, la distancia media supone un indicador robusto de valores anómalos, ya que se ve afectada por los valores extremos, pero su precisión es baja por esa misma razón, además de no tener una interpretación clara. En tanto qué, si se utiliza la distancia mediana, obtendríamos un indicador menos robusto, pero más preciso y de interpretación sencilla, pues los valores que quedasen señalados como anómalos supondrían que ello se debe a que se encuentran situados a una distancia mayor de la que en la mayoría de los casos da lugar a un grupo independiente.

2.5. Compatibilización de la metodología propuesta con los *baseline* .

Tal y como queda reflejado en los capítulos anteriores, los métodos que se están utilizando en la actualidad para la detección de puntos o zonas contaminadas se basan en el concepto de *baseline*, es decir, en la obtención de un valor o conjunto de valores, en el caso de que se considere más de un *baseline*, de concentración de *origen natural*, que se supone representativo de la zona de estudio, y que permite diferenciar las zonas contaminadas de las que no lo están, y éste valor se suele obtener a través de la asignación directa de un parámetro estadístico.

En esencia, a través del concepto de *baseline*, se trata de extender una serie de valores puntuales, correspondientes a los puntos muestrales, a una superficie acotada, de forma que el valor obtenido sea representativo de la misma, lo cual, desde un punto de vista práctico, lo convierte en un proceso de integración areal de datos puntuales, que como tal puede realizarse de varias formas, que van desde la asignación de su valor a partir de un parámetro estadístico (Gregorauskienè y Kardunas, 1997; Salminen y Tarvainen, 1997; Salminen y Gregorauskiene, 2000, Tarvainen y Kallio, 2002; De Miguel *et al.*, 2002; ...), al uso de criterios de tipo más probabilístico como el rango intercuartílico (Canadian Soil Survey).

Desde el punto de vista de este trabajo, parece más lógica la utilización de criterios de tipo probabilístico como el rango intercuartílico, pues asumiendo riesgo que supone el uso de distribuciones de frecuencias como distribuciones de probabilidad en un muestreo único, a la hora de determinar un valor de concentración como referencia, si se toma el rango intercuartílico, se estará considerando un intervalo de valores que representan al menos al 50% de los puntos muestreados, lo cual le confiere robustez, desde el punto de vista estadístico y representatividad desde el punto de vista conceptual. Sin embargo su precisión es menor que la de métodos paramétricos como aquéllos que consideran un intervalo de confianza para la media o la mediana, pues su recorrido, es decir la diferencia entre los valores que definen el intervalo (25-75%), es mayor.

A tenor de lo visto, el ideal sería alcanzar un equilibrio entre robustez y precisión, de manera que se pudiera obtener un intervalo de valores que fuese representativo de al menos el 50% de los datos y que a la vez tuviese un recorrido menor. Esta idea no es nueva, de hecho autores como Rousseeuw y Leroy (1988) y Rousseew y Croux (1993), proponen lo que llaman *shortest half*, literalmente la mitad más corta, que consiste en el intervalo de valores que contiene al 50% de los datos con recorrido menor, como estimador robusto de escala. Así, se puede disponer de un intervalo de valores tan representativo como el rango intercuartílico, pues al igual que éste, contiene al 50% de los datos, que simultáneamente es más preciso puesto que su recorrido es menor, y que presenta la ventaja adicional de estar menos influenciado por valores extremos.

Dado que para la obtención, tanto del rango intercuartílico como del *shortest half*, que podríamos traducir como el 50% más denso, se hace uso únicamente de la posición los valores en la distribución de frecuencias, en distribuciones correspondientes a variables como las que se tratan es este trabajo, en las que frecuentemente encontramos valores repetidos, resultaría bastante útil definir un parámetro de *cobertura*, tomándolo como el porcentaje de datos, incluido entre los valores que delimitan el intervalo correspondiente al 50% más denso, de forma que además de disponer de un intervalo más preciso que el rango intercuartílico, se mejoraría su representatividad, ya que puede incluir un porcentaje mayor de las muestras.

Además, el hecho de utilizar un intervalo frente a un sólo valor presenta una serie de ventajas:

1º

Las variables se caracterizan por una serie de parámetros (media, mediana, moda, recorrido, rango intercuartílico, etc.). No hay que olvidar que en este caso, las variables están referidas a coordenadas geográficas (x,y), lo que permite dividir el área de estudio en casillas según una cuadrícula, y

asignar valores a cada casilla, de forma que se obtiene la distribución areal del o los *baselines*.

Si dibujamos un mapa de isolíneas, veremos unas concentraciones que corresponden a máximos y mínimos (que salvo que estén cerradas, pueden sufrir variaciones con las dimensiones del área considerada), y que además, en el caso de los máximos corresponden a modas areales de la distribución de la variable, su número, situación, valor y forma, son los que definen las pautas de comportamiento de la variable en la zona de estudio.

Se considera que el valor o intervalo de valores más representativo (para una variable) de la zona estudiada es el que ocupa mayor superficie, que por extensión es el que se da con mayor frecuencia y trasladándolo a probabilidad es el que es más probable encontrar.

Dentro de los intervalos clásicos, el rango intercuartílico representa el 50% de los datos y además por debajo del valor que alcance la variable en su límite superior se encuentra el 75% de los valores, se trata de un método "seguro" pero poco preciso. Si ese 50% de los valores lo hacemos correr por la escala hasta reducir al máximo su amplitud, tenemos el 50% más denso, que supone el intervalo más pequeño con mayor probabilidad de aparición, por lo tanto es el más preciso, pues no existe otro intervalo de las mismas dimensiones que se pueda presentar con un grado de probabilidad mayor. Este intervalo es tan representativo como el rango intercuartílico (50% de la muestra) pero más preciso, pues su amplitud es menor. Si la distribución es de tipo normal coinciden. Además, tiene la ventaja de que está menos influenciado por valores extremos. Como valor más representativo de este intervalo, podríamos tomar su mediana (50% de los valores arriba y abajo) que sigue la línea del planteamiento, o la media que representa, desde el punto de vista geométrico el "centro de gravedad" del intervalo. La mediana presenta la ventaja de ser el centroide desde el punto de vista del rango de los datos una vez ordenados de menor a mayor, pero a la hora de tomar un valor característico que represente a todo el intervalo, parece que la media es más

adecuado, pues es sensible a los valores extremos que también cuentan, además de dejar la puerta abierta a lagunas en el muestreo.

3. RESULTADOS Y DISCUSIÓN

3. Resultados y Discusión

Resultados y discusión.

En primer lugar, se ha procedido a realizar una serie de cálculos preliminares siguiendo los criterios tradicionales, con el fin de demostrar de manera fehaciente que los valores obtenidos son muy poco indicativos del estado y evolución de los procesos de contaminación, a la vez que se justifica la necesidad de abordar estos problemas desde otro punto de vista y por tanto de desarrollar otras metodologías que permitan avanzar en su estudio e interpretación

Se han realizado los siguientes cálculos y estadísticos:

1º.- Estadísticos descriptivos univariantes de todas las variables y contrastes de bondad de ajuste a las distribuciones normal y lognormal.

2º.- Análisis de regresión y correlaciones entre variables dos a dos.

3º. Análisis de componentes principales.

4º.- Análisis de factores (correlación y covarianza).

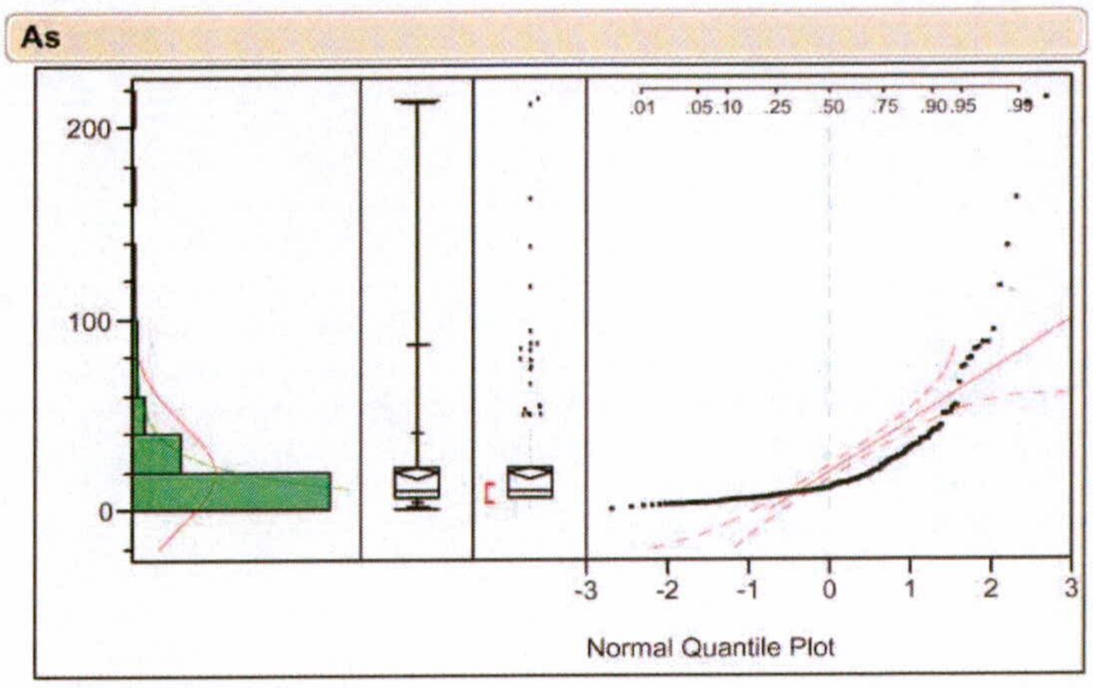
5.-º Análisis de *cluster* por varios métodos.

Obteniendo los resultados que se muestran a continuación.

3.1. Resultados análisis univariante.

En las tablas siguientes se muestran los resultados correspondientes a los estadísticos descriptivos de todos los elementos utilizados, analizados uno a uno, que tal como se suponía desde un principio, basta observar los valores correspondientes a los coeficientes de ajuste a las distribuciones consideradas (normal y lognormal) para desestimar su uso, pues tienen un ajuste bajo o nulo.

As



Quantiles

100.0%	maximum	214.00
99.5%		212.80
97.5%		86.50
90.0%		40.50
75.0%	quartile	22.70
50.0%	median	10.90
25.0%	quartile	6.70
10.0%		4.40
2.5%		2.40
0.5%		0.45
0.0%	minimum	0.00

Moments

Mean	19.75624
Std Dev	26.569535
Std Err Mean	1.5906764
upper 95% Mean	22.88754
lower 95% Mean	16.624939
N	279
Sum Wgt	279
Sum	5511.9909
Variance	705.94017
Skewness	4.2504398
Kurtosis	23.561501
CV	134.4868
N Missing	0

Fitted Normal

Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	Mu	19.75624	16.62494	22.88754
Dispersion	Sigma	26.56953	24.53266	28.97813

Goodness-of-Fit Test

Shapiro-Wilk W Test

W	Prob<W
0.561863	0.0000

Fitted LogNormal

Parameter Estimates

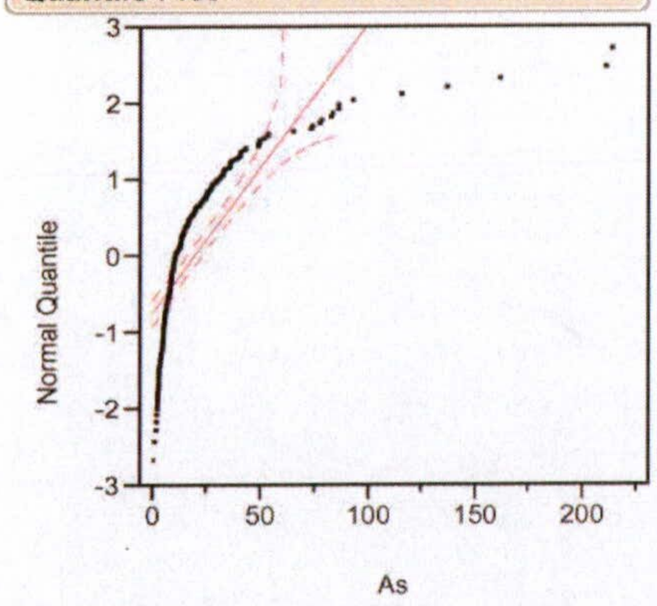
Type	Parameter	Estimate	Lower 95%	Upper 95%
Scale	Mu	2.525093	2.418317	2.631868
Shape	Sigma	0.904366	0.845627	0.972685

Goodness-of-Fit Test

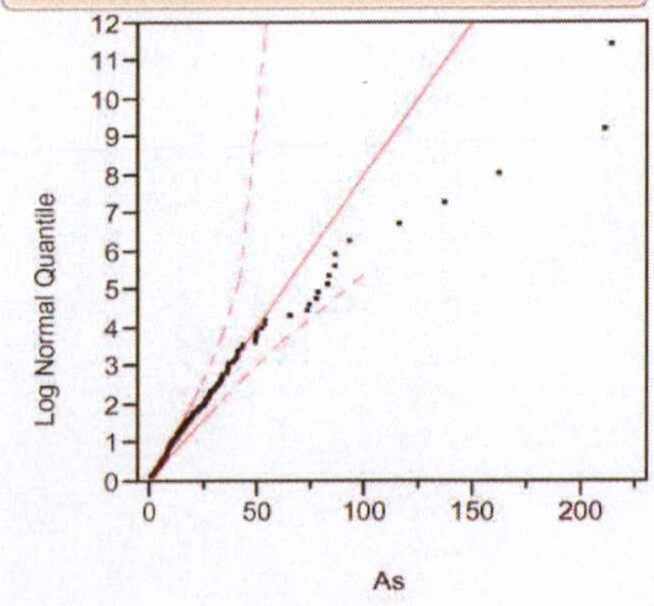
KSL Test

D		Prob>D
0.066184	<	0.0100

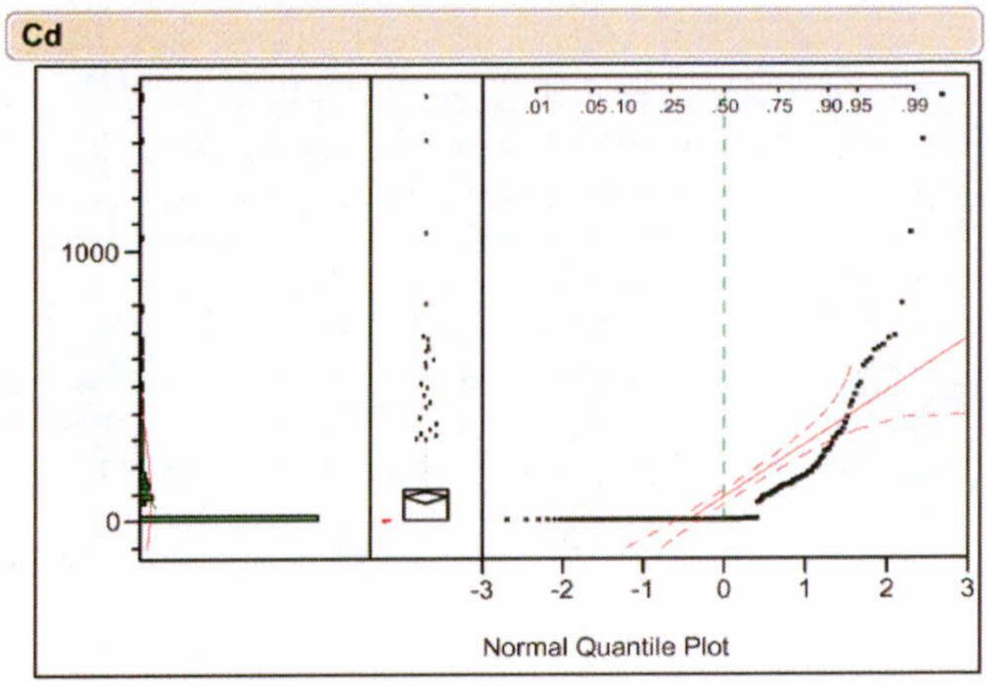
Quantile Plot



Quantile Plot



Cd



Normal(87.3379,195.16)
LogNormal(3.44837,2.68247)

Quantiles

100.0%	maximum	1565.0
99.5%		1499.7
97.5%		642.2
90.0%		259.3
75.0%	quartile	111.4
50.0%	median	0.32236
25.0%	quartile	0.0
10.0%		0.0
2.5%		0.0
0.5%		0.0
0.0%	minimum	0.0

Moments

Mean	87.337925
Std Dev	195.15976
Std Err Mean	11.683909
upper 95% Mean	110.3381
lower 95% Mean	64.337753
N	279
Sum Wgt	279
Sum	24367.281
Variance	38087.331
Skewness	4.0537993
Kurtosis	21.590882
CV	223.45362
N Missing	0

Fitted Normal

Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	Mu	87.3379	64.3378	110.3381
Dispersion	Sigma	195.1598	180.1984	212.8515

Goodness-of-Fit Test

Shapiro-Wilk W Test

W		Prob<W
0.502636		0.0000

Fitted LogNormal

Parameter Estimates

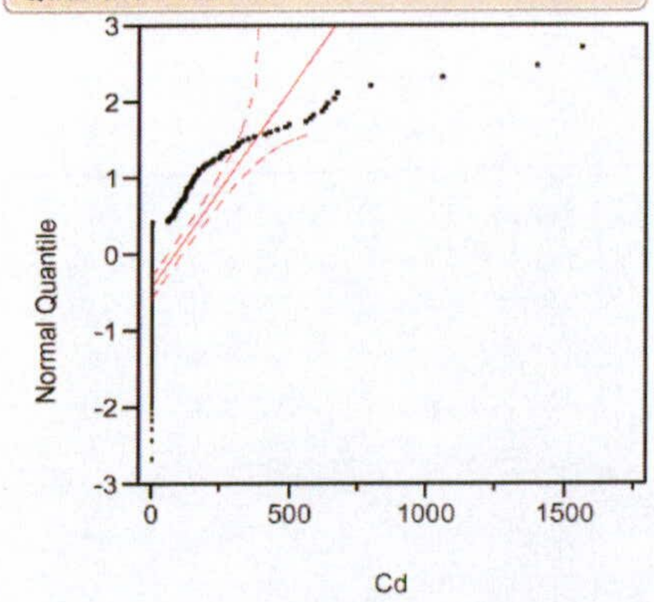
Type	Parameter	Estimate	Lower 95%	Upper 95%
Scale	Mu	3.448372	3.000126	3.896619
Shape	Sigma	2.682471	2.443523	2.978302

Goodness-of-Fit Test

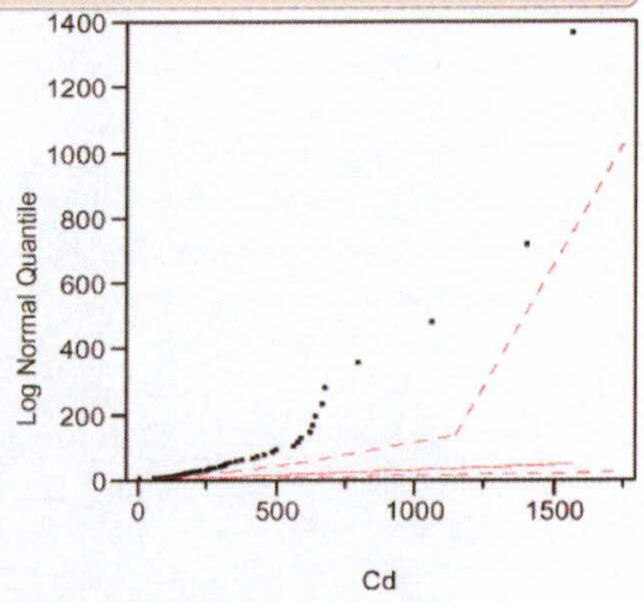
KSL Test

D		Prob>D
0.522615	<	0.0100

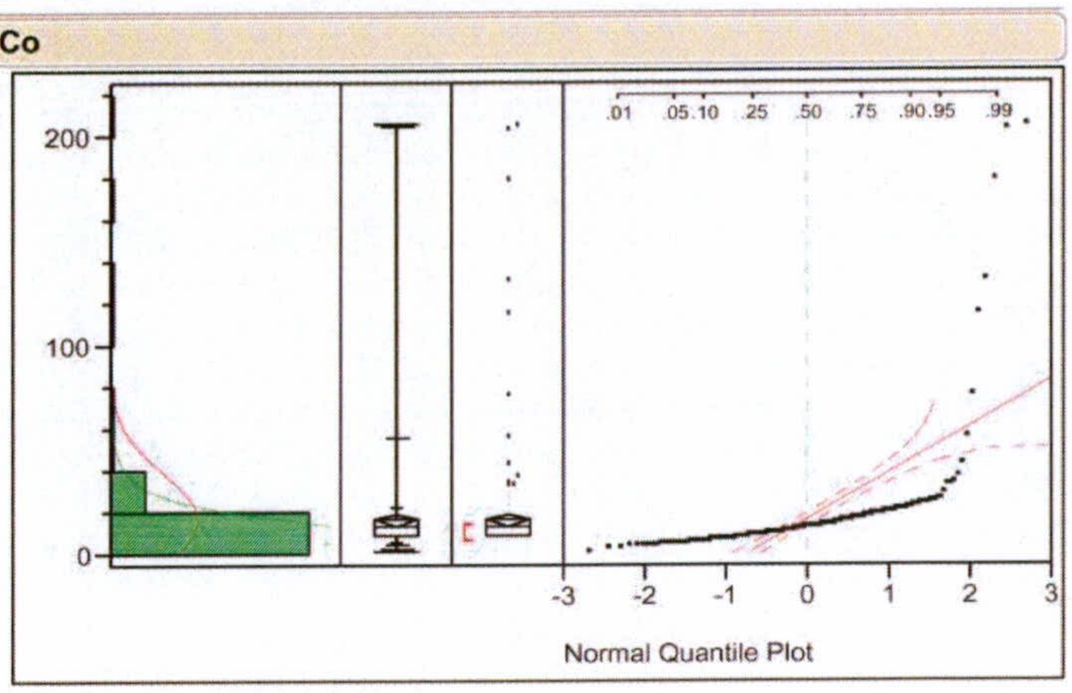
Quantile Plot



Quantile Plot



Co



Normal(16.3405,22.3097)
LogNormal(2.52343,0.62649)

Quantiles

100.0%	maximum	205.00
99.5%		204.20
97.5%		56.00
90.0%		22.00
75.0%	quartile	17.00
50.0%	median	13.00
25.0%	quartile	9.00
10.0%		6.00
2.5%		4.00
0.5%		1.80
0.0%	minimum	1.00

Moments

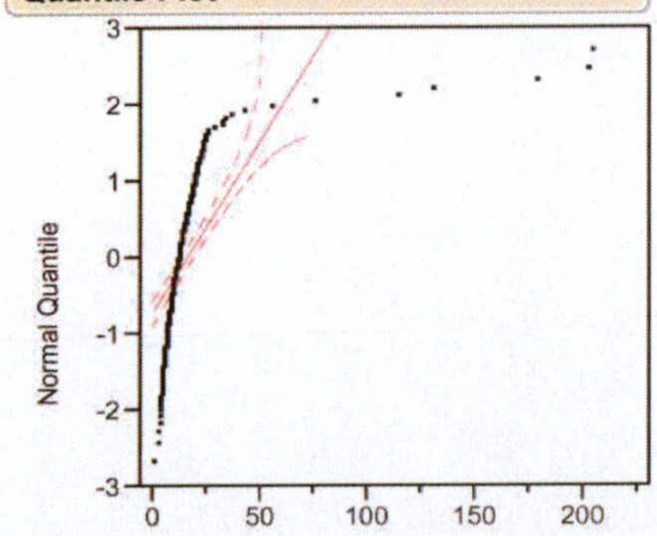
Mean	16.340502
Std Dev	22.309679
Std Err Mean	1.3356455
upper 95% Mean	18.969765
lower 95% Mean	13.711238
N	279
Sum Wgt	279
Sum	4559
Variance	497.72177
Skewness	6.6000021
Kurtosis	48.385571
CV	136.52995
N Missing	0

Fitted Normal

Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	Mu	16.34050	13.71124	18.96977
Dispersion	Sigma	22.30968	20.59937	24.33211

Quantile Plot



Goodness-of-Fit Test

Shapiro-Wilk W Test

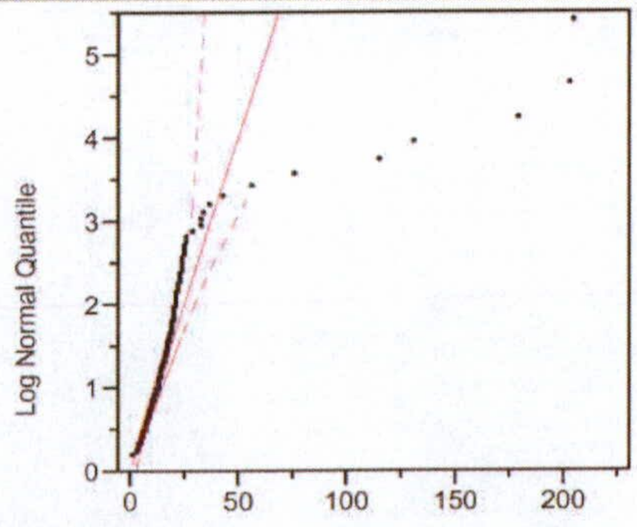
W	Prob<W
0.351194	0.0000

Fitted LogNormal

Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Scale	Mu	2.523430	2.449597	2.597263
Shape	Sigma	0.626486	0.585864	0.673722

Quantile Plot

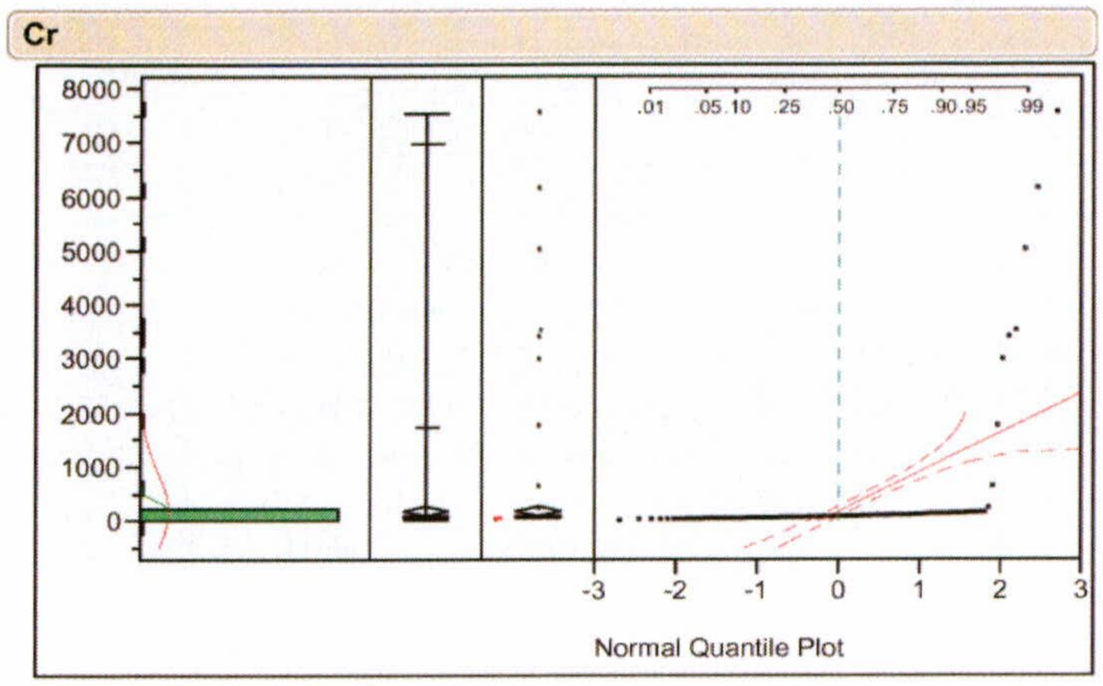


Goodness-of-Fit Test

KSL Test

D		Prob>D
0.091658	<	0.0100

Cr



Normal(176.544,730.04)
LogNormal(4.22927,0.81433)

Quantiles

100.0%	maximum	7530
99.5%		6970
97.5%		1750
90.0%		115
75.0%	quartile	89
50.0%	median	66
25.0%	quartile	46
10.0%		31
2.5%		21
0.5%		2.6
0.0%	minimum	0

Moments

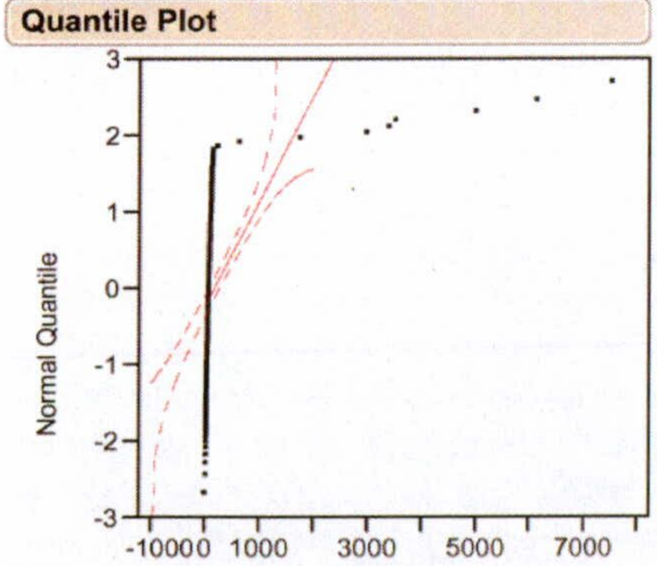
Mean	176.5437
Std Dev	730.04031
Std Err Mean	43.70637
upper 95% Mean	262.58117
lower 95% Mean	90.506224
N	279
Sum Wgt	279
Sum	49255.692
Variance	532958.86
Skewness	7.6026378
Kurtosis	61.910293
CV	413.51819
N Missing	0

Fitted Normal

Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	Mu	176.5437	90.5062	262.5812
Dispersion	Sigma	730.0403	674.0739	796.2202

Quantile Plot



Goodness-of-Fit Test

Shapiro-Wilk W Test

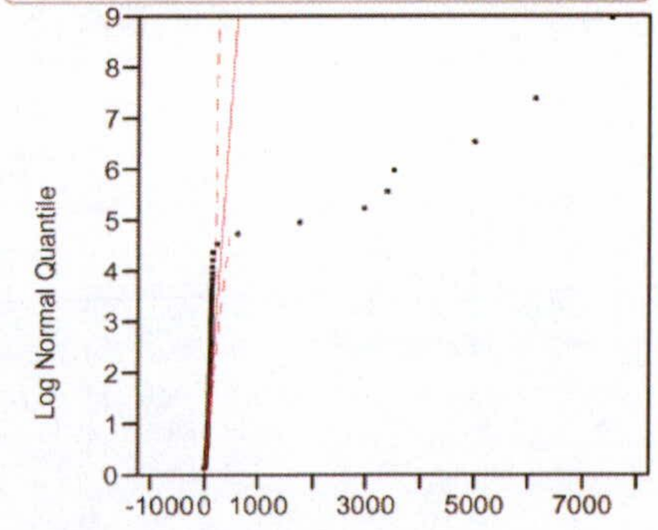
W		Prob<W
0.164861		0.0000

Fitted LogNormal

Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Scale	Mu	4.229273	4.133128	4.325417
Shape	Sigma	0.814326	0.761435	0.875843

Quantile Plot

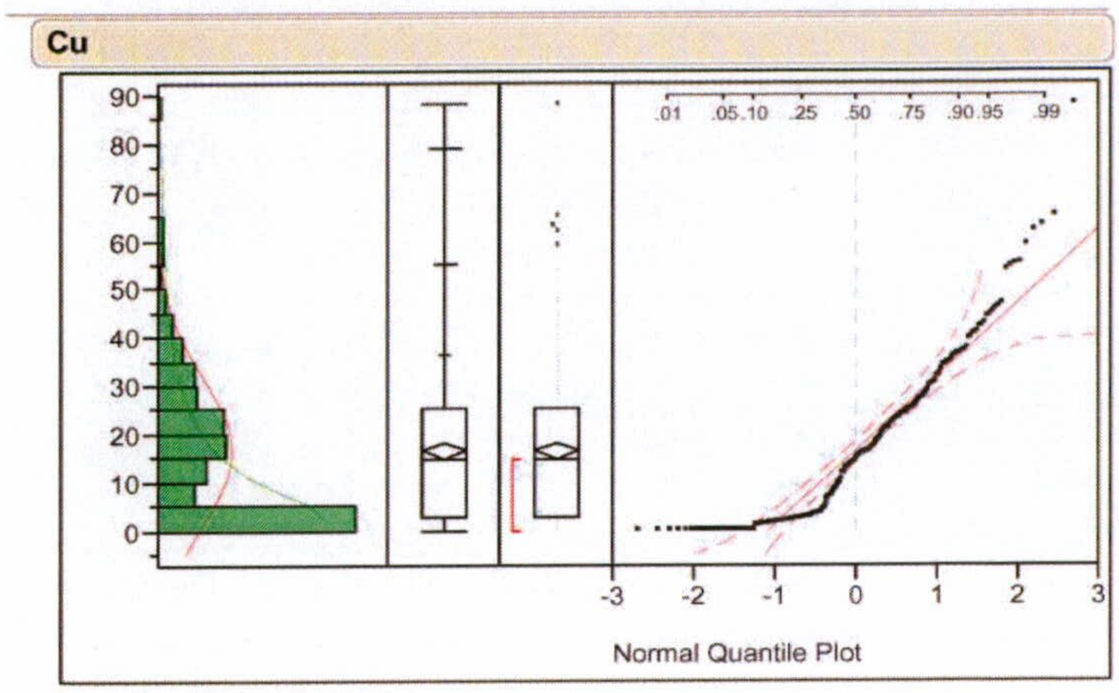


Goodness-of-Fit Test

KSL Test

D		Prob>D
1.000000	<	0.0100

Cu



Quantiles

100.0%	maximum	88.000
99.5%		78.798
97.5%		55.000
90.0%		36.379
75.0%	quartile	25.000
50.0%	median	14.747
25.0%	quartile	2.479
10.0%		0.000
2.5%		0.000
0.5%		0.000
0.0%	minimum	0.000

Moments

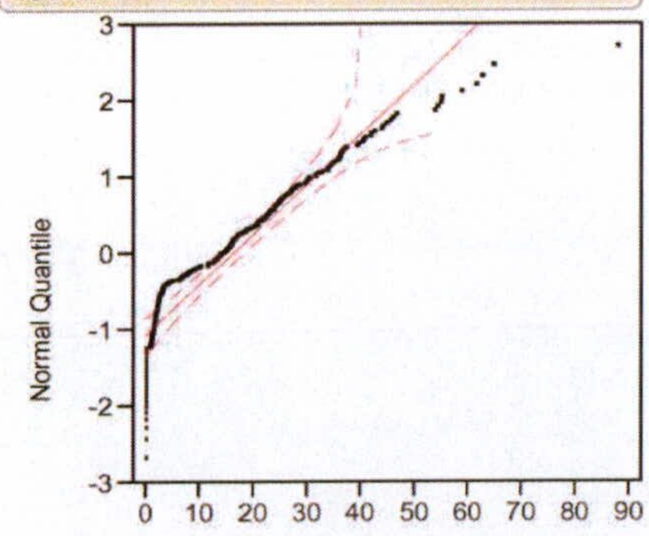
Mean	16.429974
Std Dev	15.265986
Std Err Mean	0.9139507
upper 95% Mean	18.229117
lower 95% Mean	14.630831
N	279
Sum Wgt	279
Sum	4583.9627
Variance	233.05032
Skewness	1.1118163
Kurtosis	1.562412
CV	92.915459
N Missing	0

Fitted Normal

Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	Mu	16.42997	14.63083	18.22912
Dispersion	Sigma	15.26599	14.09566	16.64988

Quantile Plot



Goodness-of-Fit Test

Shapiro-Wilk W Test

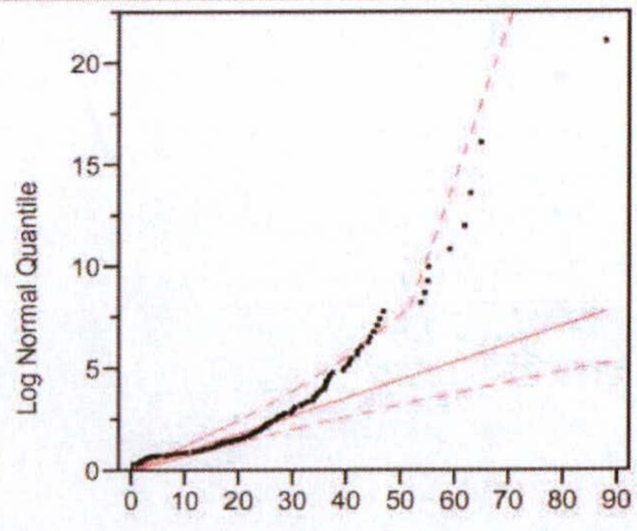
W	Prob<W
0.892520	<.0001

Fitted LogNormal

Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Scale	Mu	2.429775	2.288768	2.570783
Shape	Sigma	1.132006	1.054758	1.222601

Quantile Plot

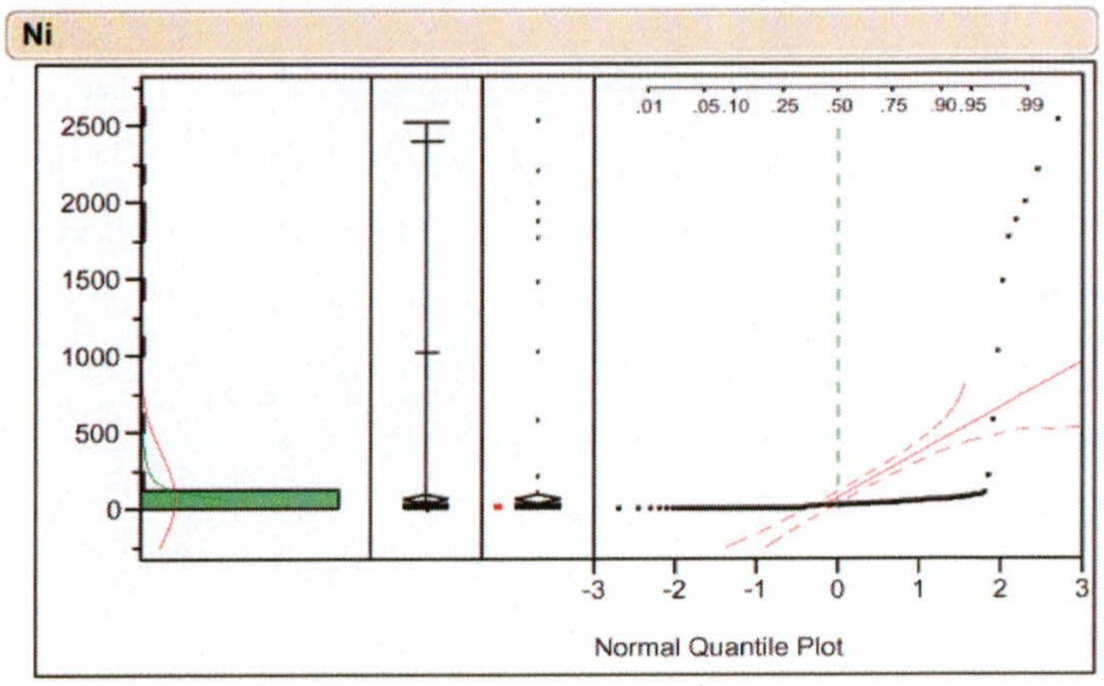


Goodness-of-Fit Test

KSL Test

D		Prob>D
0.185500	<	0.0100

Ni



Normal(69.2103,295.533)
LogNormal(3.13266,1.49196)

Quantiles

100.0%	maximum	2526.5
99.5%		2395.9
97.5%		1019.8
90.0%		55.9
75.0%	quartile	34.9
50.0%	median	19.6
25.0%	quartile	0.0
10.0%		0.0
2.5%		0.0
0.5%		0.0
0.0%	minimum	0.0

Moments

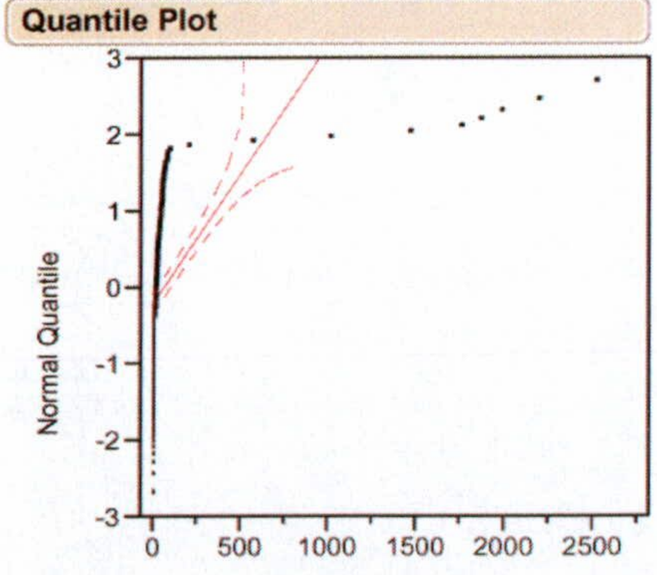
Mean	69.210282
Std Dev	295.53292
Std Err Mean	17.693093
upper 95% Mean	104.03974
lower 95% Mean	34.380827
N	279
Sum Wgt	279
Sum	19309.669
Variance	87339.706
Skewness	6.4611995
Kurtosis	42.381511
CV	427.00724
N Missing	0

Fitted Normal

Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	Mu	69.2103	34.3808	104.0397
Dispersion	Sigma	295.5329	272.8768	322.3237

Quantile Plot



Goodness-of-Fit Test

Shapiro-Wilk W Test

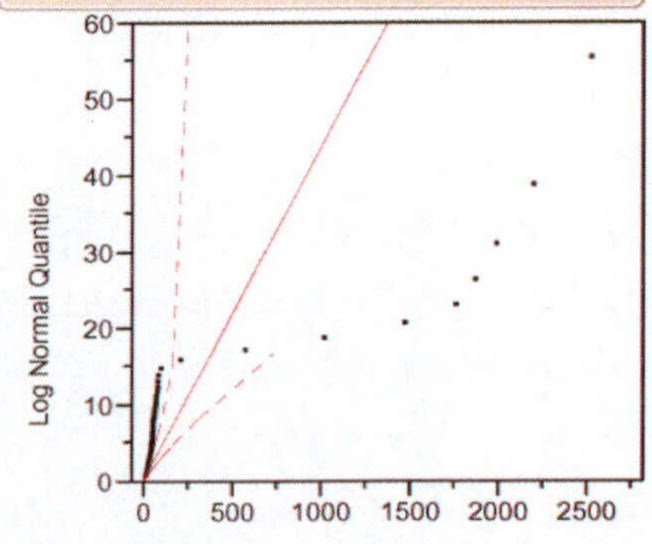
W	Prob<W
0.200853	0.0000

Fitted LogNormal

Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Scale	Mu	3.132656	2.927203	3.338109
Shape	Sigma	1.491959	1.380352	1.625062

Quantile Plot

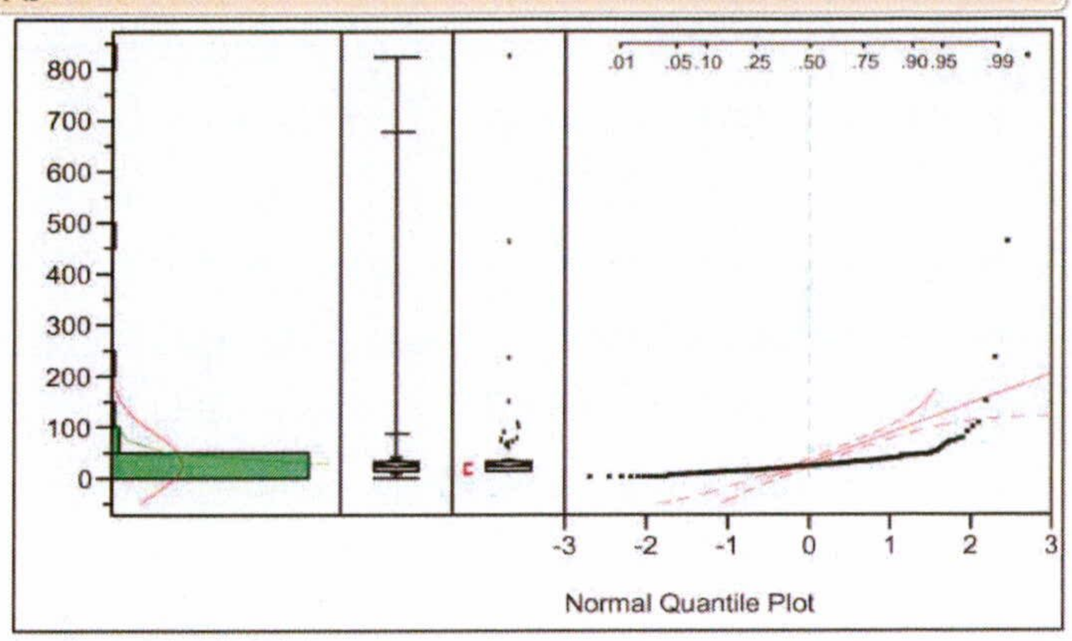


Goodness-of-Fit Test

KSL Test

D		Prob>D
0.306597	<	0.0100

Pb



Normal(28.673,58.31)
LogNormal(3.04591,0.6971)

Quantiles

100.0%	maximum	821.84
99.5%		677.20
97.5%		87.43
90.0%		41.55
75.0%	quartile	30.01
50.0%	median	21.48
25.0%	quartile	12.63
10.0%		7.15
2.5%		0.00
0.5%		0.00
0.0%	minimum	0.00

Moments

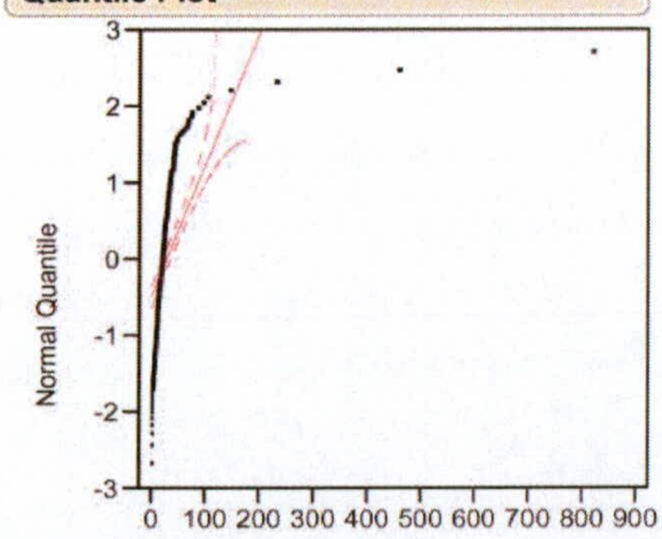
Mean	28.673005
Std Dev	58.310034
Std Err Mean	3.4909304
upper 95% Mean	35.54502
lower 95% Mean	21.80099
N	279
Sum Wgt	279
Sum	7999.7685
Variance	3400.0601
Skewness	10.786078
Kurtosis	134.48892
CV	203.36213
N Missing	0

Fitted Normal

Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	Mu	28.67301	21.80099	35.54502
Dispersion	Sigma	58.31003	53.83987	63.59598

Quantile Plot



Goodness-of-Fit Test

Shapiro-Wilk W Test

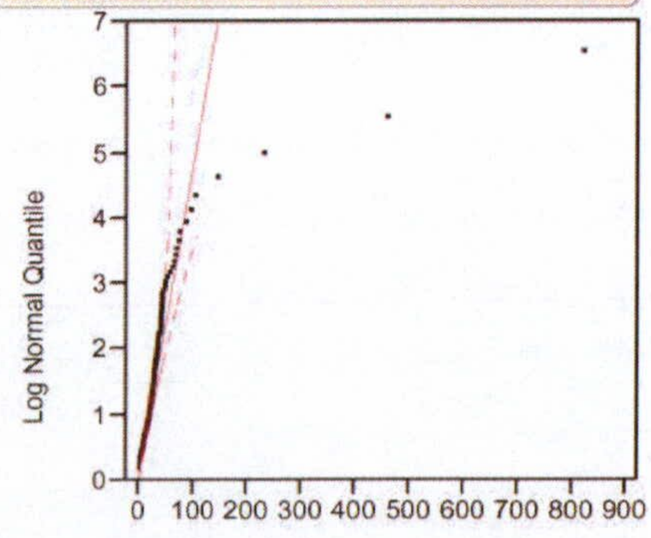
W		Prob<W
0.245817		0.0000

Fitted LogNormal

Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Scale	Mu	3.045907	2.962068	3.129747
Shape	Sigma	0.697100	0.651044	0.750819

Quantile Plot

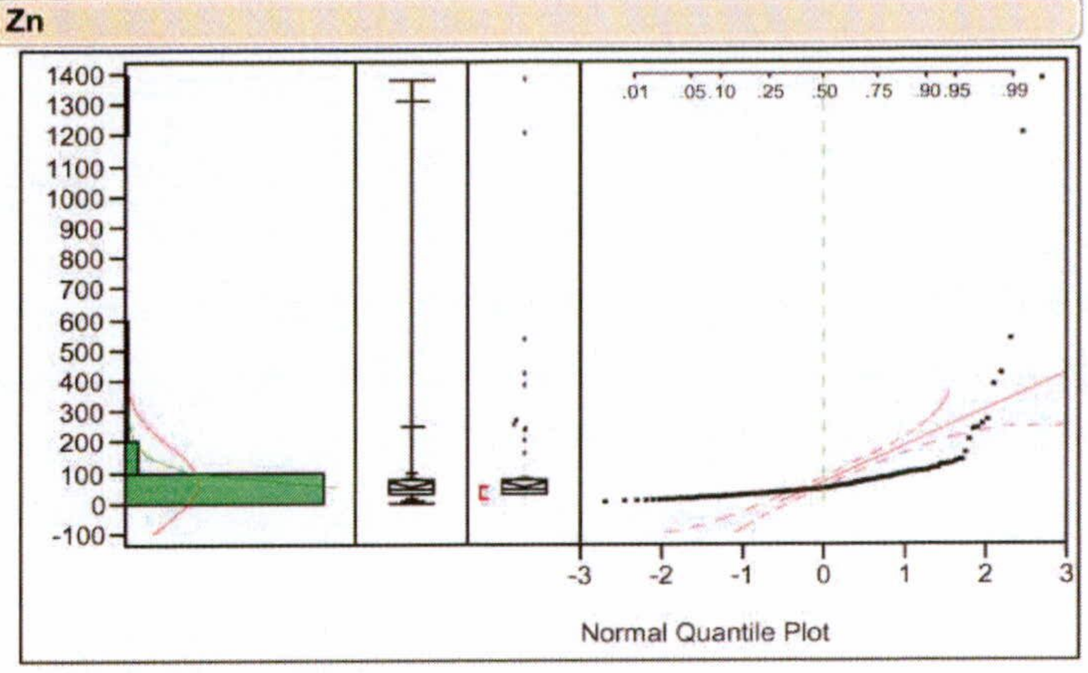


Goodness-of-Fit Test

KSL Test

D		Prob>D
0.080123	<	0.0100

Zn



Normal(65.2202,118.726)
LogNormal(3.75997,0.82101)

Quantiles

100.0%	maximum	1376.3
99.5%		1306.0
97.5%		251.3
90.0%		99.1
75.0%	quartile	72.8
50.0%	median	41.3
25.0%	quartile	26.1
10.0%		15.7
2.5%		7.2
0.5%		1.2
0.0%	minimum	0.0

Moments

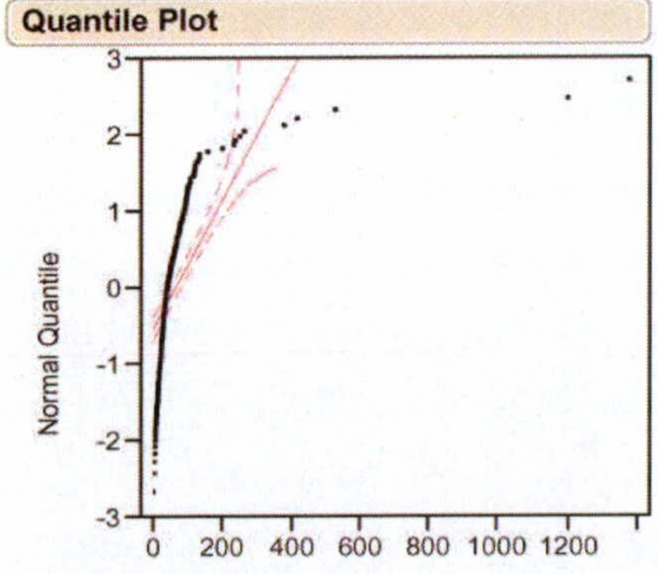
Mean	65.22024
Std Dev	118.72637
Std Err Mean	7.1079617
upper 95% Mean	79.212505
lower 95% Mean	51.227976
N	279
Sum Wgt	279
Sum	18196.447
Variance	14095.95
Skewness	8.4676701
Kurtosis	83.774879
CV	182.03914
N Missing	0

Fitted Normal

Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	Mu	65.2202	51.2280	79.2125
Dispersion	Sigma	118.7264	109.6246	129.4892

Quantile Plot



Goodness-of-Fit Test

Shapiro-Wilk W Test

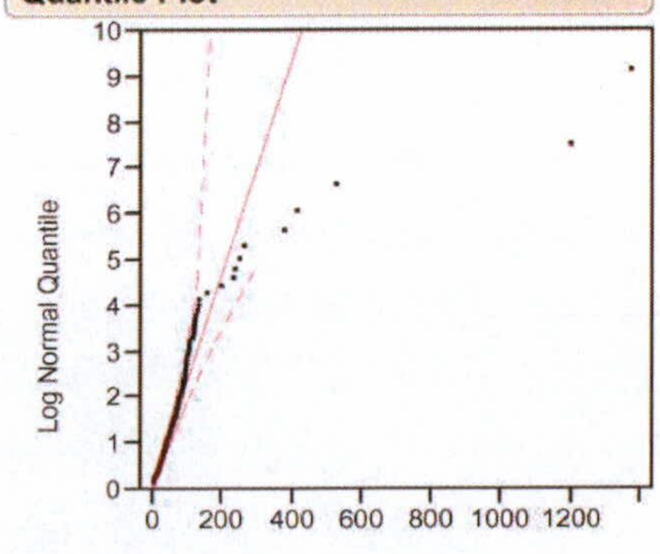
W	Prob<W
0.317372	0.0000

Fitted LogNormal

Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Scale	Mu	3.759972	3.663038	3.856906
Shape	Sigma	0.821012	0.767687	0.883035

Quantile Plot



Goodness-of-Fit Test

KSL Test

D	Prob>D
0.057853	0.0316

3.2 Análisis de correlación y regresión bivariante.

Tal y como se muestra en la matriz de correlación adjunta, las correlaciones entre los distintos elementos son altas sólo en Co-Cr-Ni, elementos que suelen estar relacionados desde el punto de vista mineralógico, ($r > 0.9$), pero no son significativas, la razón de que esto ocurra, se debe a la distribución de los datos, basta observar los gráficos de ajuste correspondientes, para notar unas agrupaciones de gran densidad de puntos en las partes inferiores, que dan como resultado un r elevado y una serie de puntos dispersos, que parecen ser los causantes de la falta de significación.

Otros elementos que suelen estar relacionados desde el punto de vista mineralógico, como el Pb y el Zn, dan coeficientes de correlación significativos, pero sensiblemente inferiores a los habituales en mineralogía, que suelen encontrarse en torno a 0.9, en este caso el r calculado es de 0.56, que probablemente se deba a a dos causas, una primera relacionada con procesos de meteorización y remoción posterior de materiales y una segunda a efectos antrópicos.

El resto de los elementos entre sí y con los ya citados presenta coeficientes de correlación bajos y no significativos.

Además del coeficiente de correlación clásico (r de Pearson), se han calculado los coeficientes de correlación Rho de Spearman, recordemos que es similar al de Pearson, pero utilizando métodos no paramétricos, obteniéndose resultados similares, como se muestra en la tabla.

	As 1S	Co 1S	Cr 1S	Fe 1S	Hg 1S	Mo 1S	Sb 1S	Sn 1S	Cu 2S	Pb 2S	Zn 2S	Ag 2S	Ni 2S	Mn 2S	Cd 2S	V 2S	Ca 2S	Mg 2S	Ti 2S	Al 2S
As 1S				8,9	2,0	8,5	51,7			9,9	21,1	8,1					3,3			
Co 1S			91,4	41,6									89,1	22,9		3,2		30,0		
Cr 1S				25,1									93,1	18,4		1,4		35,6		
Fe 1S							1,5				2,3	1,9	25,1	7,9		5,9	7,0	2,3	1,5	4,0
Hg 1S																				
Mo 1S										6,1	3,5							3,9		
Sb 1S									1,5 (-)								3,8			
Sn 1S																	2,9			
Cu 2S										3,8	3,0	7,2 (-)		52,6	18,7 (-)	61,6	38,1 (-)	3,3	13,7 (-)	51,6
Pb 2S											33,3			4,1		6,6		3,1		1,9
Zn 2S												33,3		1,7		8,4		3,9		1,7
Ag 2S														6,6 (-)	3,7	8,4 (-)	12,2	2,0 (-)	7,9	7,5 (-)
Ni 2S														21,9		3,1	2,9 (-)	40,3		
Mn 2S															17,6 (-)	53,0	35,9 (-)	17,0	13,9 (-)	43,0
Cd 2S																20,9 (-)	15,9	5,2 (-)	20,3	20,5 (-)
V 2S																	48,4 (-)	9,2	13,7	75,5
Ca 2S																		10,7 (-)	20,8	41,7 (-)
Mg 2S																			6,7 (-)	2,5
Ti 2S																				13,6 (-)
Al 2S																				

3.3 Análisis multivariante.

Los métodos multivariantes ensayados, han sido un análisis de factores/componentes principales según correlaciones y según covarianzas y un análisis de *clusters* por varios métodos.

Como puede verse en la tabla adjunta, los análisis de componentes principales realizados por los dos métodos arrojan unos buenos resultados, tanto por el método de correlaciones como por el de covarianzas, dando pocos factores y un alto porcentaje de varianza explicada.

En el caso de componentes principales por covarianza, resultan tres factores que explican el 98% de la varianza, dominados por los elementos Cr, Cd y Zn.

Otro tanto ocurre al utilizar las correlaciones, donde tres factores explican el 78.4%, y que en este caso están comandados por Co-Cr, Pb-Zn y Cd.

Pese a los, aparentemente, buenos resultados que proporcionan los análisis de factores/componentes principales, desde el punto de vista de este trabajo no son demasiado fiables, pues basta observar la matriz de correlación del apartado anterior para ver que las correlaciones entre elementos son muy bajas en general, y por tanto los resultados que puedan obtenerse tomándolas como base de partida son como poco dudosos. Éste parece ser uno de los casos a los que Davis (1978) se refiere como *ajustes fantasmas*, debido a la sensación de *falsa seguridad* que proporciona un buen ajuste que no se ve apoyado por resultados preliminares.

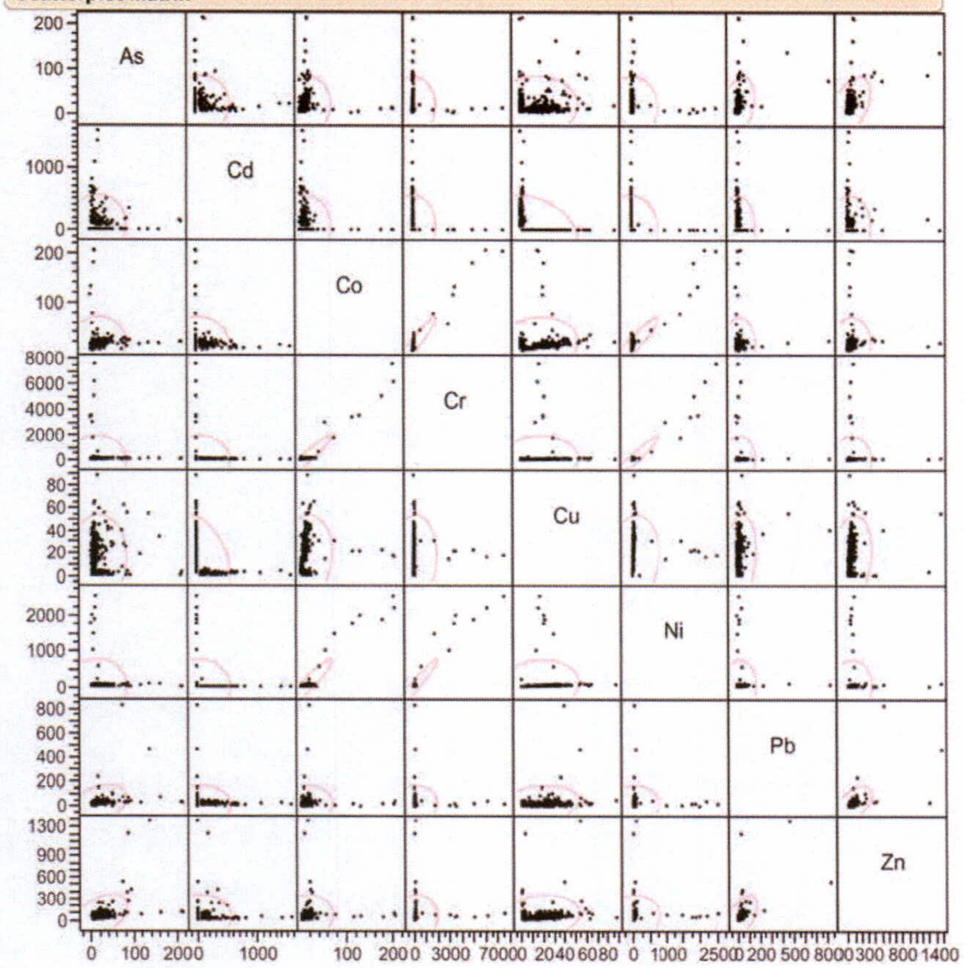
En cuanto a los análisis de *clusters* jerárquicos, realizados por distintos métodos, basta ver la tabla correspondiente para notar que proporcionan resultados diferentes según el método utilizado, lo que en general los invalida, pues en el mejor de los casos se puede utilizar el método que más nos convenga, que no tiene por qué ser el más objetivo

3. Resultados y discusión

Correlations

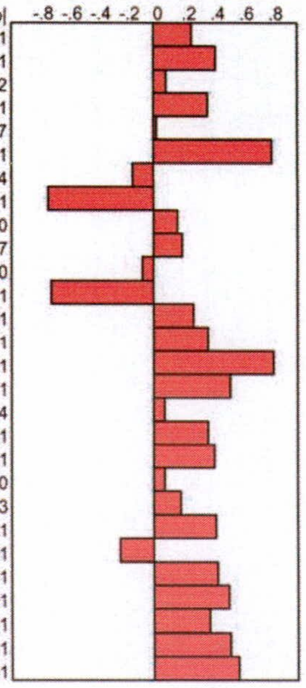
	As	Cd	Co	Cr	Cu	Ni	Pb	Zn
As	1.0000	0.0151	0.0218	-0.0552	0.0012	-0.0691	0.3140	0.4585
Cd	0.0151	1.0000	-0.0703	-0.0694	-0.4319	-0.1032	-0.0584	-0.0304
Co	0.0218	-0.0703	1.0000	0.9561	0.1007	0.9435	0.0204	0.0446
Cr	-0.0552	-0.0694	0.9561	1.0000	0.0385	0.9649	-0.0152	0.0091
Cu	0.0012	-0.4319	0.1007	0.0385	1.0000	0.0970	0.1939	0.1732
Ni	-0.0691	-0.1032	0.9435	0.9649	0.0970	1.0000	-0.0112	0.0142
Pb	0.3140	-0.0584	0.0204	-0.0152	0.1939	-0.0112	1.0000	0.5573
Zn	0.4585	-0.0304	0.0446	0.0091	0.1732	0.0142	0.5573	1.0000

Scatterplot Matrix



Nonparametric: Spearman's Rho

Variable	by Variable	Spearman Rho	Prob>|Rho|
Cd	As	0.2715	<.0001
Co	As	0.4326	<.0001
Co	Cd	0.0845	0.1592
Cr	As	0.3843	<.0001
Cr	Cd	0.0171	0.7757
Cr	Co	0.8288	<.0001
Cu	As	-0.1435	0.0164
Cu	Cd	-0.7365	<.0001
Cu	Co	0.1643	0.0060
Cu	Cr	0.2021	0.0007
Ni	As	-0.0818	0.1730
Ni	Cd	-0.7099	<.0001
Ni	Co	0.2805	<.0001
Ni	Cr	0.3801	<.0001
Ni	Cu	0.8333	<.0001
Pb	As	0.5363	<.0001
Pb	Cd	0.0757	0.2074
Pb	Co	0.3760	<.0001
Pb	Cr	0.4260	<.0001
Pb	Cu	0.0769	0.2000
Pb	Ni	0.1918	0.0013
Zn	As	0.4306	<.0001
Zn	Cd	-0.2313	<.0001
Zn	Co	0.4482	<.0001
Zn	Cr	0.5213	<.0001
Zn	Cu	0.3963	<.0001
Zn	Ni	0.5319	<.0001
Zn	Pb	0.5868	<.0001



Principal Components / Factor Analysis

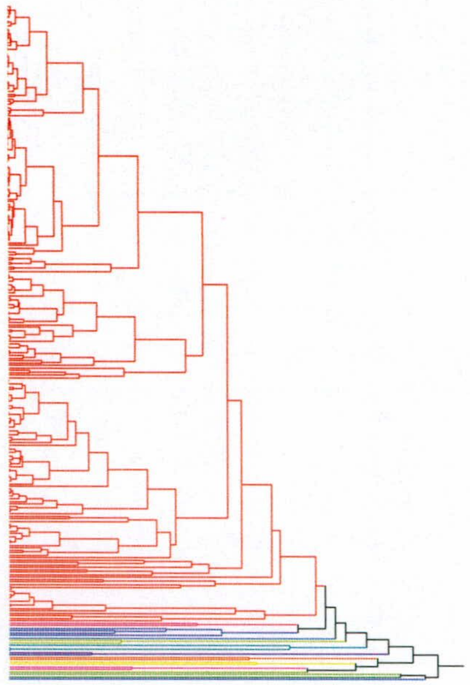
Principal Components: on Covariances

Eigenvalue	615762.8	38057.05	15474.52	5120.624	2148.661	546.1834	174.5069	34.3194
Percent	90.9118	5.6188	2.2847	0.7560	0.3172	0.0806	0.0258	0.0051
Cum Percent	90.9118	96.5306	98.8153	99.5713	99.8885	99.9692	99.9949	100.0000
Eigenvectors								
As	-0.00194	0.00005	0.10283	-0.02787	0.01044	0.98977	0.05491	-0.07663
Cd	-0.01972	0.99668	0.03991	0.05909	0.01336	-0.00475	0.03060	-0.00436
Co	0.02725	0.00041	0.00696	0.01406	0.00739	0.06985	0.11647	0.99021
Cr	0.92969	0.03976	-0.00394	-0.36529	-0.00070	-0.01024	0.01036	-0.02087
Cu	0.00095	-0.03438	0.01973	0.03474	0.02745	-0.06499	0.98972	-0.11268
Ni	0.36680	-0.04706	0.00903	0.92776	0.00600	0.02619	-0.03557	-0.02102
Pb	-0.00107	-0.02346	0.30452	-0.01016	0.95064	-0.04012	-0.03576	-0.00202
Zn	0.00154	-0.03318	0.94582	-0.00740	-0.30846	-0.09394	-0.01687	0.00434

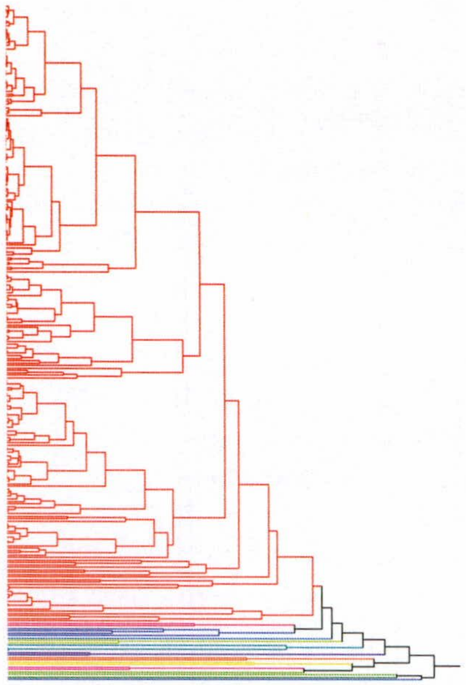
Principal Components / Factor Analysis

Principal Components: on Correlations

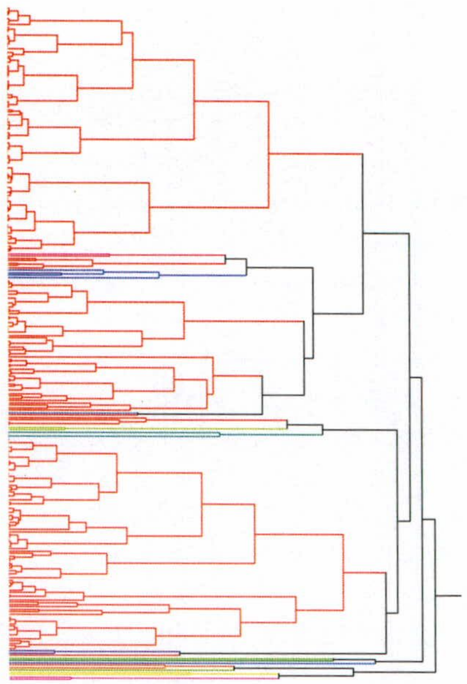
Eigenvalue	2.9384	1.9714	1.3660	0.6998	0.5347	0.4080	0.0515	0.0301
Percent	36.7298	24.6425	17.0756	8.7474	6.6839	5.1000	0.6443	0.3765
Cum Percent	36.7298	61.3723	78.4478	87.1953	93.8792	98.9792	99.6235	100.0000
Eigenvectors								
As	-0.02153	0.46663	0.30408	-0.73645	0.21467	0.30741	0.07767	0.01872
Cd	-0.09389	-0.16271	0.67823	0.39021	0.58644	0.08613	0.03309	0.00499
Co	0.57004	-0.00440	0.08382	-0.03377	0.05257	0.05496	-0.75254	-0.30785
Cr	0.57223	-0.05413	0.08966	0.00346	-0.04161	-0.00616	0.11108	0.80465
Cu	0.09473	0.28428	-0.62638	0.16819	0.68894	0.10650	0.02677	0.05357
Ni	0.57294	-0.04327	0.04188	0.01852	-0.01073	-0.01539	0.64258	-0.50448
Pb	0.01542	0.56077	0.08742	0.51004	-0.36159	0.53541	0.00955	-0.00110
Zn	0.02929	0.59639	0.17528	0.12433	-0.00054	-0.77251	-0.02233	-0.00364



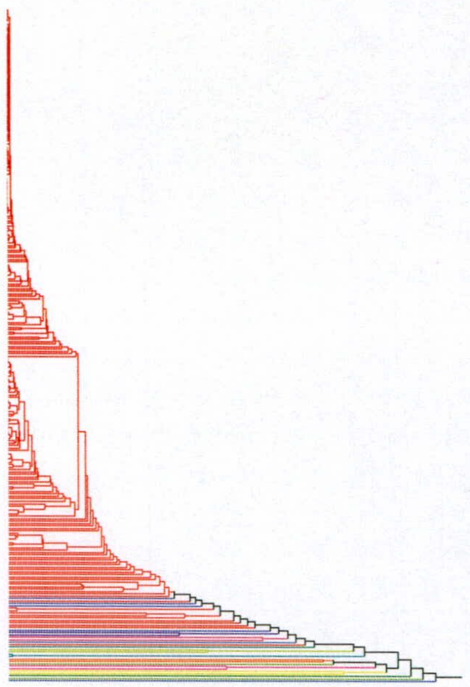
Average



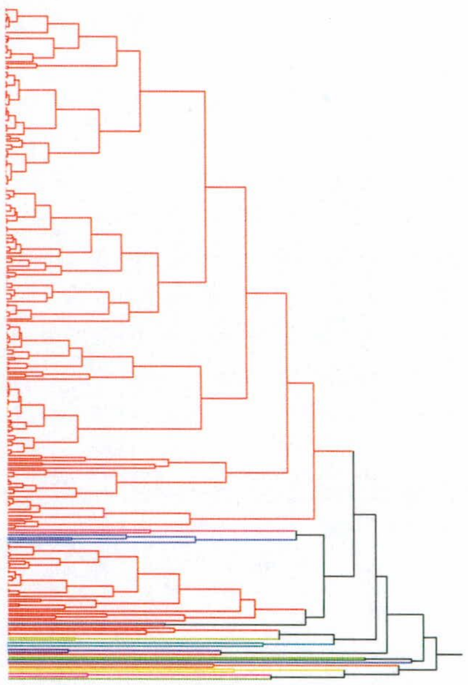
Centroid



Ward



Single



Complete

Resultados de un análisis *cluster* por varios métodos

Resumiendo, tal y como se refleja en los resultados mostrados, las metodologías tradicionales son muy poco precisas a la hora de ser aplicadas a estudios sobre contaminación, pues:

1º.- Los contrastes de bondad de ajuste de todas las variables muestran que no existe un buen ajuste a ningún tipo de distribución de las que se utilizan habitualmente, normal o lognormal, y por tanto los parámetros estadísticos son muy poco representativos del conjunto de los datos.

2º.- La correlación entre las distintas variables es muy baja, y en la mayoría de los casos no es siquiera significativa.

3º.- Los métodos multivariantes ensayados proporcionan, en el mejor de los casos, unos resultados muy vagos.

4º.- Los análisis de *cluster* realizados, muestran distintos resultados en función del método de cálculo/agrupación utilizado.

3. Resultados y discusión

3.4. Obtención de niveles característicos.

Tal y como se ha explicado en el capítulo anterior, la metodología que se propone para la obtención de niveles característicos consiste, en esencia, en la clasificación de los datos en *grupos homogéneos*, tras lo cual se obtienen, por comparación con los patrones explicativos, los valores menores y con respecto a ellos se obtienen las anomalías.

Los pasos seguidos han sido:

1º.- Obtención de agrupaciones siguiendo seis métodos:

a) Distancia media
b) Distancia Mediana
c) Máximos relativos
d) Máximos relativos con detección de inliers/outliers
e) Máximos relativos 2
f) Máximos relativos 2 con detección de inliers/outliers

2º.- Comparación de las agrupaciones obtenidas con los patrones explicativos primarios considerados, en este caso unidades estructurales, litología grado de pendiente e índice de torrencialidad, para obtener los valores característicos.

3º.- Resta de los valores característicos de los datos para obtener la distribución de anomalías.

4º.- Representación de las anomalías obtenidas.

Sistemas de agrupación.

Los sistemas de agrupación propuestos proporcionan unos buenos resultados como puede verse en las tablas resumen adjuntas (Tablas 1 a 8).

En ellas se observa que, independientemente del número de agrupaciones resultante, la pérdida de información, medida a través de la variación de entropía, es mínima, por lo que las agrupaciones resultantes parecen un buen sistema de agrupación de datos sin pérdida de información, por lo que puede decirse que los resultados obtenidos son mejores que los que proporcionan los métodos convencionales.

En cuanto a los resultados que proporcionan los distintos métodos de agrupación, como puede verse en la tabla resumen, se observa que no existen diferencias significativas en cuanto a variación de entropía se refiere, entre los distintos métodos ensayados, por lo que la elección de alguno de ellos, desde un punto de vista general, debe hacerse por el número de grupos que genere cada uno de ellos, si bien la elección del de máximos relativos 2 (MR2) y máximos relativos 2 con detección de inliers/outliers, son los que presentan una mejor relación nº grupos/*ganancia* de información, y por tanto una menor distorsión de la misma, lo que resulta crucial a la hora de interptetar los resultados.

En las figuras y mapas que se acompañan puede observarse la distribución de anomalías obtenidas a partir de cada uno de los métodos utilizados.

No obstante se hace una descripción somera de como se estima la variación de la cantidad de información en el apartado siguiente.

ESTIMACIÓN DE ENTROPÍAS (IGUALDAD DE DIVERSIDADES)

El concepto de diversidad aparece en numerosos campos de investigación asociado siempre a la idea de variabilidad de los elementos de

una determinada población. Así, la entropía de Shannon permite cuantificar esta diversidad desde un punto de vista teórico, pero puede suceder que no se conozca su valor más que a nivel muestral, siendo necesario desarrollar expresiones para estimar de forma óptima la entropía de toda la población. En nuestra situación, dado que los tamaños muestrales son suficientemente grandes, el estadístico para el contraste

H_0: $H(\Theta) = D_0$

H_1: $H(\Theta) \neq D_0$

viene dado por:

$$Z = \frac{\sqrt{n}[H(\theta) - D_0]}{\hat{\sigma}}$$

siendo

$$\hat{\sigma}^2 = \sum_{i=1} p_i (\log p_i)^2 - H(\theta)^2$$

la estimación de la varianza. Su distribución asintótica se ajusta por el teorema de Slutsky (Ferguson, 1996) a un modelo Gausiano tipificado.

Igualmente, una vez calculadas las entropías para diferentes elementos, el objetivo siguiente consiste en averiguar si existen diferencias significativas entre ellas, es decir si la pérdida de información consecuencia de los distintos métodos de agrupación son dignas de tenerse en cuenta. Para ello contrastaremos la hipótesis nula

H_0: $H(\Theta_1) = H(\Theta_2)$

frente a la alternativa

H_1: $H(\Theta_1) \neq H(\Theta_2)$

Dado que el tamaño muestral es en todos los casos grande (n=278), se utilizará el siguiente estadístico de contraste (Pardo, 1997):

$$Z = \frac{\sqrt{n_1 n_2}[H(\theta_1) - H(\theta_2)]}{\sqrt{n_2 \hat{\sigma}_1^2 + n_1 \hat{\sigma}_2^2}}$$

donde n_i denota el tamaño muestral y s_i^2 es la cuasivarianza muestral asociada a la entropía $H(\Theta_i)$, que se distribuye asintóticamente según una normal N(0,1). Como $n_1 = n_2 = 278$, la expresión del estadístico queda reducida a:

$$Z = 16,67333 \frac{H(\theta_1) - H(\theta_2)}{\sqrt{\hat{\sigma}_1^2 + \hat{\sigma}_2^2}}$$

En la Tabla 9 se muestran los distintos valores de entropía, así como los resultados de los tests de hipótesis resultantes de contrastar la entropía de cada método de agrupación frente a la variable que representa los elementos sin agrupar.

Tabla resumen As

General

Σ:	Σpi:	Σ pi*log(pi):
5511.99085	1.00000015	-2.2223564
	I. Theil:	0.09070554
	T*Nd:	25.3068449

DM

Σpi:	Σ pi*log(pi):
1.00000015	-2.23805274
I. Theil:	0.08428326
Nd/ng:	r:
5.96666667	0.99109766
IS:	D:
0.00021408	2.53888169
Ganancia:	G*Ng:
0.07080363	2.1241089
T*Nd:	2.52849768

DMd

Σpi:	Σ pi*log(pi):
1.00008633	-2.22259547
I. Theil:	0.09060772
Nd/ng:	r:
1.55652174	0.99997101
IS:	D:
8.5059E-07	0.10816701
Ganancia:	G*Ng:
0.00107842	0.12401792
T*Nd:	10.4198877

MR

Σpi:	Σ pi*log(pi):
0.99947159	-2.22166242
I. Theil:	0.09098949
Nd/ng:	r:
2.03409091	0.99704841
IS:	D:
-3.2267E-06	0.36146599
Ganancia:	G*Ng:
-0.00313044	-0.27547863
T*Nd:	8.00707472

IO MR

Σpi:	Σ pi*log(pi):
0.99978415	-2.22199004
I. Theil:	0.09085544
Nd/ng:	r:
1.73786408	0.99998988
IS:	D:
-0.00088209	0.06707005
Ganancia:	G*Ng:
-0.0016526	-0.17021764
T*Nd:	9.35811003

MR2

Σpi:	Σ pi*log(pi):
1.00000015	-2.22976798
I. Theil:	0.08767303
Nd/ng:	r:
6.88461538	0.93533452
IS:	D:
0.00011663	2.04222314
Ganancia:	G*Ng:
0.03343242	0.86924286
T*Nd:	2.27949883

IO MR2

Σpi:	Σ pi*log(pi):
1.00000088	-2.22254553
I. Theil:	0.09062815
Nd/ng:	r:
4.06818182	0.99982246
IS:	D:
-0.00205973	19.7562539
Ganancia:	G*Ng:
0.00085314	0.0375383
T*Nd:	3.98763871

Tabla resumen Cd

General

Σpi:	Σ pi*log(pi):
1.00000004	-1.8266599
I. Theil:	0.25260785

DM

Σpi:	Σ pi*log(pi):
1.00004079	-1.82809579
I. Theil:	0.25202034
Nd/ng:	r:
4.83783784	0.99986577
IS:	D:
1.5879E-05	1.41315373
Ganancia:	G*Ng:
0.00232576	0.08605329
T*Nd:	9.3247526

DMd

Σpi:	Σ pi*log(pi):
1.00000004	-1.8266599
I. Theil:	0.25260785
Nd/ng:	r:
1.31617647	1
IS:	D:
1.3878E-16	2.6847E-07
Ganancia:	G*Ng:
7.4716E-14	1.0161E-11
T*Nd:	34.3546672

MR

Σpi:	Σ pi*log(pi):
1.00000004	-1.8279388
I. Theil:	0.25208458
Nd/ng:	r:
3.80851064	0.99438757
IS:	D:
1.1133E-05	3.60219844
Ganancia:	G*Ng:
0.00207148	0.0973595
T*Nd:	11.847975

IO MR

Σpi:	Σ pi*log(pi):
1.00032563	-1.82714327
I. Theil:	0.25241007
Nd/ng:	r:
2.84126984	0.99999352
IS:	D:
-0.00400651	0.24238864
Ganancia:	G*Ng:
0.00078292	0.04932404
T*Nd:	15.9018347

MR2

Σpi:	Σ pi*log(pi):
1.00000004	-1.82955731
I. Theil:	0.25142235
Nd/ng:	r:
10.5294118	0.98940468
IS:	D:
6.9735E-05	6.49270771
Ganancia:	G*Ng:
0.00469303	0.07978147
T*Nd:	4.27417997

IO MR2

Σpi:	Σ pi*log(pi):
1.00004065	-1.82700199
I. Theil:	0.25246788
Nd/ng:	r:
4.97222222	0.99981696
IS:	D:
-0.007013	1.42818163
Ganancia:	G*Ng:
0.0005541	0.01994752
T*Nd:	9.08884359

Tabla resumen Co

General

Σpi:	Σ pi*log(pi):
1	-2.27717145
I. Theil:	0.06827753

DM

Σpi:	Σ pi*log(pi):
1	-2.27636889
I. Theil:	0.06860591
Nd/ng:	r:
4.83783784	1
IS:	D:
-8.875E-06	0
Ganancia:	G*Ng:
-0.00480939	-0.17794758
T*Nd:	2.53841856

DMd

Σpi:	Σ pi*log(pi):
1	-2.27717145
I. Theil:	0.06827753
Nd/ng:	r:
4.83783784	1
IS:	D:
0	0
Ganancia:	G*Ng:
0	0
T*Nd:	2.52626874

MR

Σpi:	Σ pi*log(pi):
1.00365577	-2.28222337
I. Theil:	0.0662105
Nd/ng:	r:
6.39285714	0.99308936
IS:	D:
7.3823E-05	0.43864141
Ganancia:	G*Ng:
0.03027403	0.84767282
T*Nd:	1.85389393

IO MR

Σpi:	Σ pi*log(pi):
1	-2.2771724
I. Theil:	0.06827714
Nd/ng:	r:
4.97222222	0.99999759
IS:	D:
1.0791E-08	0.00477897
Ganancia:	G*Ng:
5.6898E-06	0.00020483
T*Nd:	2.45797722

MR2

Σpi:	Σ pi*log(pi):
0.99741467	-2.32653887
I. Theil:	0.04807847
Nd/ng:	r:
59.6666667	0.75344717
IS:	D:
0.00673302	6.84016619
Ganancia:	G*Ng:
0.29583766	0.88751297
T*Nd:	0.1442354

IO MR2

Σpi:	Σ pi*log(pi):
0.99999781	-2.30614349
I. Theil:	0.05642339
Nd/ng:	r:
14.9166667	0.97183056
IS:	D:
0.00098784	4.24243525
Ganancia:	G*Ng:
0.17361699	2.08340391
T*Nd:	0.67708072

Tabla resumen Cr

General

Σpi:	Σ pi*log(pi):
1.00000004	-1.69914205
I. Theil:	0.30478277

DM

Σpi:	Σ pi*log(pi):
1.00000004	-1.71431461
I. Theil:	0.2985748
Nd/ng:	r:
17.9	0.9992348
IS:	D:
0.0006208	22.8628321
Ganancia:	G*Ng:
0.02036851	0.20368507
T*Nd:	2.985748

DMd

Σpi:	Σ pi*log(pi):
1.00205036	-1.70655182
I. Theil:	0.30175101
Nd/ng:	r:
1.34586466	0.997858
IS:	D:
2.2795E-05	4.39784946
Ganancia:	G*Ng:
0.00994729	1.32298977
T*Nd:	40.1328839

MR

Σpi:	Σ pi*log(pi):
0.99972374	-1.71304371
I. Theil:	0.2990948
Nd/ng:	r:
2.41891892	0.9821259
IS:	D:
7.6864E-05	20.7009045
Ganancia:	G*Ng:
0.01866237	1.38101565
T*Nd:	22.1330152

IO MR

Σpi:	Σ pi*log(pi):
1.00000004	-1.69914469
I. Theil:	0.30478169
Nd/ng:	r:
2.08139535	0.99999986
IS:	D:
1.2522E-08	0.16734836
Ganancia:	G*Ng:
3.5332E-06	0.00030386
T*Nd:	26.2112256

MR2

Σpi:	Σ pi*log(pi):
1.00002034	-1.72971263
I. Theil:	0.29227458
Nd/ng:	r:
10.5294118	0.93583314
IS:	D:
0.00073578	44.7513808
Ganancia:	G*Ng:
0.04103969	0.69767465
T*Nd:	4.96866787

IO MR2

Σpi:	Σ pi*log(pi):
1.00000004	-1.69952502
I. Theil:	0.30462607
Nd/ng:	r:
6.62962963	0.9999875
IS:	D:
5.8035E-06	2.67395009
Ganancia:	G*Ng:
0.00051412	0.01388128
T*Nd:	8.22490401

Tabla resumen Cu

General

Σpi:	Σ pi*log(pi):
1.00000059	-2.25205801
I. Theil:	0.07855289

DM

Σpi:	Σ pi*log(pi):
1.00011019	-2.25317055
I. Theil:	0.07809769
Nd/ng:	r:
3.72916667	0.99939561
IS:	D:
9.4834E-06	0.36417688
Ganancia:	G*Ng:
0.00579487	0.27815375
T*Nd:	3.74868899

DMd

Σpi:	Σ pi*log(pi):
1.00000059	-2.25209229
I. Theil:	0.07853887
Nd/ng:	r:
1.26950355	0.99998573
IS:	D:
9.9469E-08	0.03484868
Ganancia:	G*Ng:
0.00017854	0.02517477
T*Nd:	11.0739801

MR

Σpi:	Σ pi*log(pi):
1.00000059	-2.25239549
I. Theil:	0.07841481
Nd/ng:	r:
2.45205479	0.99623116
IS:	D:
1.8915E-06	0.25476327
Ganancia:	G*Ng:
0.0017578	0.12831913
T*Nd:	5.72428121

IO MR

Σpi:	Σ pi*log(pi):
1.0000464	-2.25217538
I. Theil:	0.07850487
Nd/ng:	r:
1.98888889	0.99997004
IS:	D:
5.3357E-07	0.07325875
Ganancia:	G*Ng:
0.00061132	0.05501884
T*Nd:	7.06543831

MR2

Σpi:	Σ pi*log(pi):
1.00000059	-2.25292675
I. Theil:	0.07819744
Nd/ng:	r:
6.62962963	0.99244383
IS:	D:
1.3165E-05	0.54590096
Ganancia:	G*Ng:
0.00452497	0.12217428
T*Nd:	2.11133092

IO MR2

Σpi:	Σ pi*log(pi):
1.00013465	-2.25246096
I. Theil:	0.07838802
Nd/ng:	r:
4.36585366	0.99965031
IS:	D:
4.0212E-06	0.28132535
Ganancia:	G*Ng:
0.00209885	0.08605303
T*Nd:	3.21390882

Tabla resumen Ni

General

Σpi:	Σ pi*log(pi):
0.99999993	-1.53830527
I. Theil:	0.3705904

DM

Σpi:	Σ pi*log(pi):
0.99999993	-1.60406725
I. Theil:	0.34368337
Nd/ng:	r:
5.96666667	0.99772397
IS:	D:
0.0008969	15.9188132
Ganancia:	G*Ng:
0.07260585	2.17817545
T*Nd:	10.310501

DMd

Σpi:	Σ pi*log(pi):
1.00000007	-1.53830767
I. Theil:	0.37058941
Nd/ng:	r:
1.25174825	1
IS:	D:
6.8834E-09	0.01071928
Ganancia:	G*Ng:
2.6561E-06	0.00037982
T*Nd:	52.994286

MR

Σpi:	Σ pi*log(pi):
0.99999958	-1.54870453
I. Theil:	0.36633546
Nd/ng:	r:
2.84126984	0.97935842
IS:	D:
6.7539E-05	9.00495675
Ganancia:	G*Ng:
0.01148151	0.72333536
T*Nd:	23.0791338

IO MR

Σpi:	Σ pi*log(pi):
0.99999993	-1.53830836
I. Theil:	0.37058913
Nd/ng:	r:
2.03409091	0.99999991
IS:	D:
1.4402E-08	0.06652714
Ganancia:	G*Ng:
3.42E-06	0.00030096
T*Nd:	32.6118434

MR2

Σpi:	Σ pi*log(pi):
0.99999993	-1.78758598
I. Theil:	0.26859525
Nd/ng:	r:
9.42105263	0.59172728
IS:	D:
0.00536817	59.3637043
Ganancia:	G*Ng:
0.2752234	5.22924458
T*Nd:	5.10330971

IO MR2

Σpi:	Σ pi*log(pi):
0.99771343	-1.5324119
I. Theil:	0.37300171
Nd/ng:	r:
3.89130435	0.99999595
IS:	D:
-5.242E-05	0.42789327
Ganancia:	G*Ng:
-0.00650669	-0.29930763
T*Nd:	17.1580788

Tabla resumen Pb

General

Σpi:	Σ pi*log(pi):
0.99999981	-2.19043722
I. Theil:	0.10376552

DM

Σpi:	Σ pi*log(pi):
0.99999981	-2.21590221
I. Theil:	0.09334632
Nd/ng:	r:
10.5294118	0.99059266
IS:	D:
0.00061289	6.24487194
Ganancia:	G*Ng:
0.10041102	1.70698731
T*Nd:	1.5868874

DMd

Σpi:	Σ pi*log(pi):
0.99996423	-2.19036015
I. Theil:	0.10379705
Nd/ng:	r:
1.18543046	0.99999969
IS:	D:
-2.0883E-07	0.02586046
Ganancia:	G*Ng:
-0.00030389	-0.0458874
T*Nd:	15.6733549

MR

Σpi:	Σ pi*log(pi):
1.02865347	-2.2814646
I. Theil:	0.06652095
Nd/ng:	r:
2.15662651	0.71818768
IS:	D:
0.00044873	6.21152111
Ganancia:	G*Ng:
0.35893007	29.7911955
T*Nd:	5.52123923

IO MR

Σpi:	Σ pi*log(pi):
0.99940166	-2.18898577
I. Theil:	0.10435939
Nd/ng:	r:
1.75490196	0.9999974
IS:	D:
-0.00102313	0.07646745
Ganancia:	G*Ng:
-0.00572321	-0.58376736
T*Nd:	10.6446579

MR2

Σpi:	Σ pi*log(pi):
0.99999981	-2.25065741
I. Theil:	0.07912596
Nd/ng:	r:
6.17241379	0.61395742
IS:	D:
0.00084964	7.37886635
Ganancia:	G*Ng:
0.23745424	6.8861731
T*Nd:	2.29465273

IO MR2

Σpi:	Σ pi*log(pi):
0.99999981	-2.19048607
I. Theil:	0.10374553
Nd/ng:	r:
3.25454545	0.99998423
IS:	D:
-0.00188628	0.24705602
Ganancia:	G*Ng:
0.00019263	0.01059473
T*Nd:	5.70600419

Tabla resumen Zn

General

Σpi:	Σ pi*log(pi):
0.99999984	-2.19642048
I. Theil:	0.10131742

DM

Σpi:	Σ pi*log(pi):
0.99999984	-2.24331616
I. Theil:	0.08212969
Nd/ng:	r:
14.9166667	0.97715087
IS:	D:
0.00159898	20.2742653
Ganancia:	G*Ng:
0.1893824	2.27258879
T*Nd:	0.98555623

DMd

Σpi:	Σ pi*log(pi):
0.99999984	-2.19642858
I. Theil:	0.10131411
Nd/ng:	r:
1.34586466	0.99999773
IS:	D:
2.492E-08	0.14094168
Ganancia:	G*Ng:
3.2713E-05	0.00435086
T*Nd:	13.4747763

MR

Σpi:	Σ pi*log(pi):
0.99986017	-2.1967584
I. Theil:	0.10117916
Nd/ng:	r:
2.03409091	0.99610637
IS:	D:
1.5712E-06	1.92736343
Ganancia:	G*Ng:
0.00136464	0.12008859
T*Nd:	8.90376609

IO MR

Σpi:	Σ pi*log(pi):
0.99995834	-2.19631133
I. Theil:	0.10136208
Nd/ng:	r:
1.73786408	0.99999848
IS:	D:
-0.0009841	0.14178958
Ganancia:	G*Ng:
-0.00044079	-0.04540144
T*Nd:	10.4402945

MR2

Σpi:	Σ pi*log(pi):
1.00049903	-2.26023468
I. Theil:	0.07520735
Nd/ng:	r:
6.39285714	0.62394504
IS:	D:
0.0009325	17.7277641
Ganancia:	G*Ng:
0.2577057	7.21575956
T*Nd:	2.10580567

IO MR2

Σpi:	Σ pi*log(pi):
1.01051927	-2.2123705
I. Theil:	0.09479135
Nd/ng:	r:
3.50980392	0.99529925
IS:	D:
-0.00185865	1.28708832
Ganancia:	G*Ng:
0.06441217	3.28502056
T*Nd:	4.83435872

As	Sin agrup.	DM	DMd	MR	MR2	MR dep	MR2 dep
H(Θ)	2,22236	2,23805	2,22260	2,22166	2,22977	2,22199	2,22254
s	0,45281	0,45829	0,45234	0,45505	0,43610	0,45402	0,45285
test		Z=0,4061 p=0,6849	z=0,0063 p=0,9950	z=0,0182 p=0,9855	z=0,1965 p=0,8443	z=0,0096 p=0,9923	z=0,0047 p=0,9963

Cd	Sin agrup.	DM	DMd	MR	MR2	MR dep	MR2 dep
H(Θ)	1,82666	1,82810	1,82666	1,82794	1,82956	1,82714	1,82700
s	0,39150	0,40020	0,39150	0,38901	0,38691	0,39013	0,39170
test		z=0,0304 p=0,9757	z=0,0000 p=0,9999	z=0,0274 p=0,9781	z=0,0623 p=0,9503	z=0,0103 p=0,9918	z=0,0073 p=0,9942

Co	Sin agrup.	DM	DMd	MR	MR2	MR dep	MR2 dep
H(Θ)	2,27717	2,27637	2,27717	2,28222	2,32654	2,27717	2,30614
s	0,45385	0,45463	0,45385	0,43633	0,39695	0,45386	0,44399
test		z=0,0208 p=0,9834	z=0,0000 p=0,9999	z=0,1337 p=0,8937	z=1,3652 p=0,1727	z=0,0000 p=0,9999	z=0,7608 p=0,4471

Cr	Sin agrup.	DM	DMd	MR	MR2	MR dep	MR2 dep
H(Θ)	1,69914	1,71431	1,70655	1,71304	1,72971	1,69914	1,69952
s	0,88640	0,89929	0,88210	0,88014	0,86453	0,88643	0,88684
test		z=0,2003 p=0,8413	z=0,0988 p=0,9213	z=0,1855 p=0,8529	z=0,4117 p=0,6808	z=0,0000 p=0,9999	z=0,0051 p=0,9960

Cu	Sin agrup.	DM	DMd	MR	MR2	MR dep	MR2 dep
H(Θ)	2,25206	2,25317	2,25209	2,25240	2,25293	2,25218	2,25246
s	0,29142	0,29210	0,29157	0,29046	0,28958	0,29101	0,29040
test		z=0,0425 p=0,9661	z=0,0012 p=0,9991	z=0,0131 p=0,9896	z=0,0335 p=0,9733	z=0,0046 p=0,9963	z=0,0154 p=0,9877

Ni	Sin agrup.	DM	DMd	MR	MR2	MR dep	MR2 dep
H(Θ)	1,53831	1,60407	1,53831	1,54870	1,78759	1,53831	1,53241
s	0,79072	0,88693	0,79074	0,78255	0,57815	0,79074	0,79273
test		z=0,7924 p=0,4286	z=0,0000 p=0,9999	z=0,1337 p=0,8937	z=3,6437 p=0,0003	z=0,0000 p=0,9999	z=0,0754 p=0,9399

Pb	Sin agrup.	DM	DMd	MR	MR2	MR dep	MR2 dep
H(Θ)	2,19044	2,21589	2,19036	2,28147	2,25066	2,18899	2,19049
s	0,56456	0,56496	0,55201	0,28666	0,42877	0,56714	0,56456
test		z=0,5216 p=0,6021	z=0,0017 p=0,9987	z=2,3536 p=0,0189	z=1,3906 p=0,1649	z=0,0297 p=0,9763	z=0,0010 p=0,9992

Zn	Sin agrup.	DM	DMd	MR	MR2	MR dep	MR2 dep
H(Θ)	2,19642	2,24331	2,19643	2,19676	2,26023	2,19631	2,21237
s	0,54142	0,54029	0,54142	0,54087	0,40490	0,54164	0,49528
test		z=1,0221 p=0,3072	z=0,0002 p=0,9898	z=0,0074 p=0,9941	z=1,5739 p=0,1161	z=0,0024 p=0,9981	z=0,3624 p=0,7172

Tabla XX: Resultados del contraste de hipótesis de igualdad entre las entropías

Obtención de valores característicos.

Para la obtención de valores característicos se recurre a la ordenación de los datos según patrones explicativos primarios (estructura geológica, litología, pendiente y torrencialidad) y el estudio de los mismos, asignando los valores característicos a los valores de menor concentración encontrados bajo las mismas condiciones de patrones explicativos. En las tablas resumen siguientes se muestran los valores característicos asignados.

Cálculo y localización de anomalías.

El cálculo de valores anómalos, se realiza de una manera sencilla, restando el valor del nivel característico al de los datos de concentración, obteniendo así un nuevo valor que indica la anomalía presente en función del patrón teórico obtenido para los niveles característicos. Los valores obtenidos para las anomalías de los distintos elementos, se muestran el los mapas siguientes.

As	Unidades Estructurales	N. característico	Litología	N. característico
	Complejo Alpujárride	g1-g26/g1-g46	Arcillas	g16-g20/g17-g24
	Complejo Alpujárride		Calizas-Mármoles	g2-g26/g2/g46
	Complejo Alpujárride		Esquistos	g3-g24/g32
	Complejo Alpujárride		Filitas	g4-g16/g5-g17
	Complejo Alpujárride		Gneiss	g4-g20/g5-g22
	Complejo Alpujárride		Peridotitas	g1-g11/g1-g12
	Complejo Maláguide	g6-g25/g7-g38	Areniscas	g6-g21/g7-g27
	Complejo Maláguide		Esquistos	g16-g25/g17-g38
	Complejo Maláguide		Filitas	g9/g10
	Complejo Nevado-Filábride	g3-g26/g4-g48	Areniscas	g26/g48
	Complejo Nevado-Filábride		Calizas-Mármoles	g11-g21/g12-g27
	Complejo Nevado-Filábride		Esquistos	g3-g11/g4-g12
	Complejo Nevado-Filábride		Filitas	g26/g47
	Complejo Nevado-Filábride		Filitas-Cuarcitas	g21/g27
	Depresiones Postorogénicas	g2-g26/g3-g44	Arcillas-limos	g3-g26/g4-g43
	Depresiones Postorogénicas		Arenas	g4-g19/g5-g20
	Depresiones Postorogénicas		Areniscas y margas	g5-g13/g6-g14
	Depresiones Postorogénicas		Calizas	g9-g25/g10-g41
	Depresiones Postorogénicas		Conglomerado	g3-26/g4-g44
	Depresiones Postorogénicas		Esquistos	g11-g26/g12-g40
	Depresiones Postorogénicas		Filitas	g13/g14
	Depresiones Postorogénicas		Margas	g4-g19/g5-g20
	Depresiones Postorogénicas		Margocalizas	g2-g10/g4-g11
	Depresiones Postorogénicas		Mat.fluviales	g4-g20/g5-g21
	Rocas Volcánicas	g10-g22/g11-g28	Andesita	g10-g22/g11-g28
	Rocas Volcánicas		Dacitas	
	Rocas Volcánicas		Lamproita	
	U. Campo de Gibraltar	g2-g19/g3-g20	Arcillas	g10/g11
	U. Campo de Gibraltar		Areniscas	g10-g13/g11-g14
	U. Campo de Gibraltar		Calizas	g2-g17/g3-g18
	U. Campo de Gibraltar		Conglomerado	g10/g11
	U. Campo de Gibraltar		Margas	g12/g13
	U. Campo de Gibraltar		Margocalizas	g19/g20
	Zona Prebética	g2-g21/g3-g27	Calizas	g4-g21/g5-g27
	Zona Prebética		Margas	g2-g46/g3-g5
	Zona Subbética	g2-g26/g3-g42	Arcillas	g4-g11/g5-g12
	Zona Subbética		Areniscas	g6/g7
	Zona Subbética		Calizas	g2-g26/g3-g42
	Zona Subbética		Conglomerado	g7-g16/g8-g18
	Zona Subbética		Margas	g2-g14/g3-g15
	Zona Subbética		Margocalizas	g3-g10/g4-g11

Cd	Unidades Estructurales	N. característico
	Complejo Alpujárride	g1-g16/g1-g28
	Complejo Alpujárride	
	Complejo Alpujárride	
	Complejo Alpujárride	
	Complejo Alpujárride	
	Complejo Alpujárride	
	Complejo Maláguide	g1-g3
	Complejo Maláguide	g1-g3
	Complejo Maláguide	
	Complejo Nevado-Filábride	
	Complejo Nevado-Filábride	g1-g14
	Complejo Nevado-Filábride	
	Complejo Nevado-Filábride	
	Complejo Nevado-Filábride	
	Depresiones Postorogénicas	g1-g17
	Depresiones Postorogénicas	
	Depresiones Postorogénicas	
	Depresiones Postorogénicas	
	Depresiones Postorogénicas	
	Depresiones Postorogénicas	
	Depresiones Postorogénicas	
	Depresiones Postorogénicas	
	Depresiones Postorogénicas	
	Rocas Volcánicas	g6-g36
	Rocas Volcánicas	g6-g15
	Rocas Volcánicas	
	U. Campo de Gibraltar	g1-g4
	U. Campo de Gibraltar	g1-g4
	U. Campo de Gibraltar	
	U. Campo de Gibraltar	
	U. Campo de Gibraltar	
	U. Campo de Gibraltar	
	Zona Prebética	g1-g4
	Zona Prebética	
	Zona Subbética	g1-g16
	Zona Subbética	
	Zona Subbética	
	Zona Subbética	
	Zona Subbética	
	Zona Subbética	

Litología	N. característico
Arcillas	g10-g14
Calizas-Mármoles	g1-g5
Esquistos	g1-g10
Filitas	g1-g16
Gneiss	g1
Peridotitas	g1-g4
Areniscas	g1-g2
Esquistos	g1-g3
Filitas	g1/g1
Areniscas	g8/g8
Calizas-Mármoles	g1-g13
Esquistos	g1-14
Filitas	g10/g10
Filitas-Cuarcitas	g1-g6
Arcillas-limos	g1-g14
Arenas	g1-g7
Areniscas y margas	g1-g16
Calizas	g1-g16
Conglomerado	g1-g17
Esquistos	g1-g13
Filitas	g14/g16
Margas	g1-g16
Margocalizas	g1-g16
Mat.fluviales	g1-g12
Andesita	g10/g10
Dacitas	g6/g6
Lamproita	g15/g23
Arcillas	g1/g1
Areniscas	g1/g1
Calizas	g1-g4
Conglomerado	g1/g1
Margas	g1/g1
Margocalizas	g1/g1
Calizas	g1-g4
Margas	g1-g3
Arcillas	g1-g16
Areniscas	g1/g1
Calizas	g1-g8
Conglomerado	g1/g1
Margas	g1-g4
Margocalizas	g1-g12

Co	Unidades Estructurales	N. característico
	Complejo Alpujárride	g1-g3
	Complejo Alpujárride	
	Complejo Alpujárride	
	Complejo Alpujárride	
	Complejo Alpujárride	
	Complejo Alpujárride	
	Complejo Maláguide	g2-g2
	Complejo Maláguide	
	Complejo Maláguide	
	Complejo Nevado-Filábride	
	Complejo Nevado-Filábride	
	Complejo Nevado-Filábride	
	Complejo Nevado-Filábride	
	Complejo Nevado-Filábride	
	Depresiones Postorogénicas	g2
	Depresiones Postorogénicas	
	Depresiones Postorogénicas	
	Depresiones Postorogénicas	
	Depresiones Postorogénicas	
	Depresiones Postorogénicas	
	Depresiones Postorogénicas	
	Depresiones Postorogénicas	
	Depresiones Postorogénicas	
	Rocas Volcánicas	
	Rocas Volcánicas	
	Rocas Volcánicas	g2/g2
	U. Campo de Gibraltar	g2
	U. Campo de Gibraltar	
	U. Campo de Gibraltar	g2
	U. Campo de Gibraltar	
	U. Campo de Gibraltar	
	U. Campo de Gibraltar	
	Zona Prebética	
	Zona Prebética	
	Zona Subbética	
	Zona Subbética	
	Zona Subbética	
	Zona Subbética	
	Zona Subbética	
	Zona Subbética	
	Zona Subbética	

Litología	N. característico
Arcillas	g2/g2
Calizas-Mármoles	g2
Esquistos	g2
Filitas	g2/g2
Gneiss	g2/g2
Peridotitas	g3
Areniscas	g2/g2
Esquistos	g2
Filitas	g2/g2
Areniscas	g2/g2
Calizas-Mármoles	g2/g2
Esquistos	g2-g3
Filitas	g2/g2
Filitas-Cuarcitas	g2/g2
Arcillas-limos	g2/g2
Arenas	g2/g2
Areniscas y margas	g2/g2
Calizas	g2/g2
Conglomerado	g2/g2
Esquistos	g2/g2
Filitas	g2/g2
Margas	g2/g2
Margocalizas	g2/g2
Mat.fluviales	g2
Andesita	g2/g2
Dacitas	g2/g2
Lamproita	g2/g2
Arcillas	g2/g2
Areniscas	g2
Calizas	g2/g2
Conglomerado	g2/g2
Margas	g2/g2
Margocalizas	g2/g2
Calizas	g2
Margas	g2/g2
Arcillas	g2/g2
Areniscas	g2/g2
Calizas	g2/g2
Conglomerado	g2/g2
Conglomerado	g2/g2
Margas	g2/g2
Margocalizas	g2

Cr	Unidades Estructurales	N. característico
	Complejo Alpujárride	
	Complejo Alpujárride	g1-g17
	Complejo Alpujárride	
	Complejo Alpujárride	
	Complejo Alpujárride	
	Complejo Alpujárride	
	Complejo Maláguide	g4-g16
	Complejo Maláguide	
	Complejo Maláguide	
	Complejo Nevado-Filábride	
	Complejo Nevado-Filábride	
	Complejo Nevado-Filábride	g5-g16
	Complejo Nevado-Filábride	
	Complejo Nevado-Filábride	
	Depresiones Postorogénicas	g2-g16
	Depresiones Postorogénicas	
	Depresiones Postorogénicas	
	Depresiones Postorogénicas	
	Depresiones Postorogénicas	
	Depresiones Postorogénicas	
	Depresiones Postorogénicas	
	Depresiones Postorogénicas	
	Depresiones Postorogénicas	
	Depresiones Postorogénicas	
	Rocas Volcánicas	g5-g15
	Rocas Volcánicas	g3-g13
	Rocas Volcánicas	
	U. Campo de Gibraltar	
	U. Campo de Gibraltar	
	U. Campo de Gibraltar	g6-g15
	U. Campo de Gibraltar	
	U. Campo de Gibraltar	
	U. Campo de Gibraltar	
	Zona Prebética	g5g13
	Zona Prebética	g3-g11
	Zona Subbética	
	Zona Subbética	
	Zona Subbética	g2-g16
	Zona Subbética	
	Zona Subbética	
	Zona Subbética	g2-g18

Litología	N. característico
Arcillas	g6-g7
Calizas-Mármoles	g1-g16
Esquistos	g3-g16
Filitas	g6-g9
Gneiss	g6-g16
Peridotitas	g16-g17
Areniscas	g2-g3
Esquistos	g7-g16
Filitas	g6/g8
Areniscas	g9/g11
Calizas-Mármoles	g8/g9
Esquistos	g3-g14
Filitas	g8/g10
Filitas-Cuarcitas	g8/g10
Arcillas-limos	g2-g7
Arenas	g4-g11
Areniscas y margas	g3-g8
Calizas	
Conglomerado	g2-g16
Esquistos	g4-g11
Filitas	g9/g11
Margas	g2-g15
Margocalizas	g2-g5
Mat.fluviales	g5-g9
Andesita	g8/g10
Dacitas	g3/g5
Lamproita	g13/g15
Arcillas	g6/g8
Areniscas	g10-g14
Calizas	g6-g15
Conglomerado	g12/g14
Margas	g7/g9
Margocalizas	g8/g10
Calizas	g3-g11
Margas	g3-g4
Arcillas	g4-g9
Areniscas	g3/g5
Calizas	g2-g16
Conglomerado	g3-g5
Margas	g2-g10
Margocalizas	g2-g5

Cu	Unidades Estructurales	N. característico	Litología	N. característico
	Complejo Alpujárride	g1-g27	Arcillas	g1-g5
	Complejo Alpujárride		Calizas-Mármoles	g1-g27
	Complejo Alpujárride	g1-g27	Esquistos	g1-g25
	Complejo Alpujárride		Filitas	g2-g24
	Complejo Alpujárride		Gneiss	g15-g29
	Complejo Alpujárride		Peridotitas	g12-g22
	Complejo Maláguide	g8-g41	Areniscas	g7/g8
	Complejo Maláguide		Esquistos	g21-g27
	Complejo Maláguide		Filitas	g16/g17
	Complejo Nevado-Filábride		Areniscas	g4/g4
	Complejo Nevado-Filábride		Calizas-Mármoles	g3-g15
	Complejo Nevado-Filábride	g1-g26	Esquistos	g1-g21
	Complejo Nevado-Filábride		Filitas	g1/g1
	Complejo Nevado-Filábride		Filitas-Cuarcitas	g2-g26
	Depresiones Postorogénicas		Arcillas-limos	g1-g24
	Depresiones Postorogénicas		Arenas	g3-g24
	Depresiones Postorogénicas		Areniscas y margas	g4/g4
	Depresiones Postorogénicas		Calizas	g3-g15
	Depresiones Postorogénicas	g1-g26	Conglomerado	g1-25
	Depresiones Postorogénicas		Esquistos	g3-g15
	Depresiones Postorogénicas		Filitas	g1/g1
	Depresiones Postorogénicas		Margas	g4/g4
	Depresiones Postorogénicas		Margocalizas	g3-g15
	Depresiones Postorogénicas		Mat.fluviales	g3-g15
	Rocas Volcánicas	g1-g5	Andesita	g1/g1
	Rocas Volcánicas		Dacitas	g5/g5
	Rocas Volcánicas	g1-g5	Lamproita	g1/g1
	U. Campo de Gibraltar		Arcillas	g17/g18
	U. Campo de Gibraltar		Areniscas	g14-g25
	U. Campo de Gibraltar	g13-g27	Calizas	g13-g27
	U. Campo de Gibraltar		Conglomerado	g27/g34
	U. Campo de Gibraltar		Margas	g20/g21
	U. Campo de Gibraltar		Margocalizas	g16/g17
	Zona Prebética	g7-g18	Calizas	g7-g16
	Zona Prebética		Margas	g15-g18
	Zona Prebética		Margas	g16-g19
	Zona Subbética		Arcillas	g5-g32
	Zona Subbética	g3-g40	Calizas	g4-g27
	Zona Subbética		Conglomerado	g9-g11
	Zona Subbética		Margas	g11-g25
	Zona Subbética		Margocalizas	

Ni	Unidades Estructurales	N. característico	Litología	N. característico
	Complejo Alpujárride	g1-g19/g1-g46	Arcillas	g1-g2/g1-g2
	Complejo Alpujárride		Calizas-Mármoles	g1-g19/g1-g39
	Complejo Alpujárride		Esquistos	g1-g19/g1-g38
	Complejo Alpujárride		Filitas	g1-g16/g1-g21
	Complejo Alpujárride		Gneiss	g11-g19/g14-g28
	Complejo Alpujárride		Peridotitas	g19/g40-g46
	Complejo Maláguide	g3-g19/g5-g35	Areniscas	g3-g6/g5-g9
	Complejo Maláguide		Esquistos	g15-g19/g19-g35
	Complejo Maláguide		Filitas	g12/g15
	Complejo Nevado-Filábride	g1-g19/g1-g27	Areniscas	g2/g2
	Complejo Nevado-Filábride		Calizas-Mármoles	g1/g13/g1-g16
	Complejo Nevado-Filábride		Esquistos	g1-g19/g1-g27
	Complejo Nevado-Filábride		Filitas	g2/g2
	Complejo Nevado-Filábride		Filitas-Cuarcitas	g2/g12/g2-g15
	Depresiones Postorogénicas	g1-g19/g1-g34	Arcillas-limos	g1-g11/g1-g14
	Depresiones Postorogénicas		Arenas	g1-g18/g1-g24
	Depresiones Postorogénicas		Areniscas y margas	g1-g4/g1-g7
	Depresiones Postorogénicas		Calizas	g1-g18/g1-g24
	Depresiones Postorogénicas		Conglomerado	g1-g2/g1-g2
	Depresiones Postorogénicas		Esquistos	g1-g15/g1-g20
	Depresiones Postorogénicas		Filitas	g1/g1
	Depresiones Postorogénicas		Margas	g1-g19/g1-g32
	Depresiones Postorogénicas		Margocalizas	g2-g12/g2-g15
	Depresiones Postorogénicas		Mat.fluviales	g1-g12/g1-g15
	Rocas Volcánicas	g1-g2/g1-g2	Andesita	g1/g1
	Rocas Volcánicas		Dacitas	g2/g2
	Rocas Volcánicas		Lamproita	g2/g2
	U. Campo de Gibraltar	g9-g19/g12-g30	Arcillas	g11/g14
	U. Campo de Gibraltar		Areniscas	g11-g18/g14-g24
	U. Campo de Gibraltar		Calizas	g9-g19/g12-g30
	U. Campo de Gibraltar		Conglomerado	g18/g24
	U. Campo de Gibraltar		Margas	g11/g14
	U. Campo de Gibraltar		Margocalizas	g11/g14
	Zona Prebética	g3-g14/g6-g17	Calizas	g5-g14/g8-g17
	Zona Prebética		Margas	g3-g9/g6-g12
	Zona Subbética	g1-g19/g1-g37	Arcillas	g1-g18/g1-g24
	Zona Subbética		Areniscas	g6/g9
	Zona Subbética		Calizas	g1-g19/g1-g37
	Zona Subbética		Conglomerado	g7-g13/g10-g16
	Zona Subbética		Margas	g4-g11/g7-g13
	Zona Subbética		Margocalizas	g2-g16/g2-g21

Pb

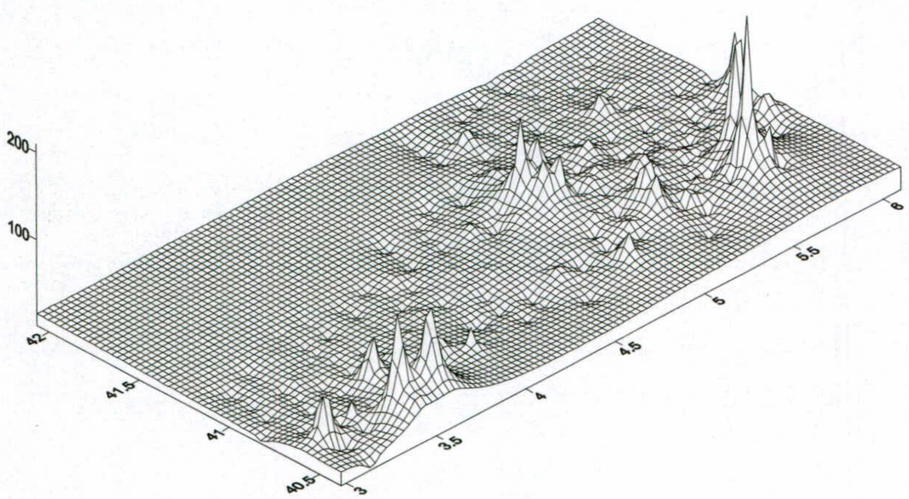
Unidades Estructurales	N. característico	Litología	N. característico
Complejo Alpujárride	g1-g29/g1-g54	Arcillas	g13-g16/g14-g17
Complejo Alpujárride		Calizas-Mármoles	g15-g29/g16-g54
Complejo Alpujárride		Esquistos	g8-g29/g9-g51
Complejo Alpujárride		Filitas	g2-g26/g3-g30
Complejo Alpujárride		Gneiss	g9-g22/g10-g24
Complejo Alpujárride		Peridotitas	g1-g27/g1-g33
Complejo Maláguide	g1-g28/g1-g40	Areniscas	g3-g28/g4-g35
Complejo Maláguide		Esquistos	g1-g28/g1-g40
Complejo Maláguide		Filitas	g5/g6
Complejo Nevado-Filábride	g6-g29/g7-g43	Areniscas	g23/g25
Complejo Nevado-Filábride		Calizas-Mármoles	g9-g12/g13
Complejo Nevado-Filábride		Esquistos	g6-g10/g7-g11
Complejo Nevado-Filábride		Filitas	g25/g29
Complejo Nevado-Filábride		Filitas-Cuarcitas	g16-g28/g38
Depresiones Postorogénicas	g1-g29/g1-g55	Arcillas-limos	g2-g29/g2-g53
Depresiones Postorogénicas		Arenas	g8-g27/g9-g32
Depresiones Postorogénicas		Areniscas y margas	g1-g17/g1-g19
Depresiones Postorogénicas		Calizas	g1-g28/g1-g39
Depresiones Postorogénicas		Conglomerado	g1-g29/g1-g55
Depresiones Postorogénicas		Esquistos	g14-g25/g15-g28
Depresiones Postorogénicas		Filitas	g3/g4
Depresiones Postorogénicas		Margas	g1-g29/g1-g45
Depresiones Postorogénicas		Margocalizas	g2-g15/g3-g16
Depresiones Postorogénicas		Mat.fluviales	g6-g28/g7-g36
Rocas Volcánicas	g10-g24/g11-g26	Andesita	g24/g26
Rocas Volcánicas		Dacitas	g10/g11
Rocas Volcánicas		Lamproita	g16/g18
U. Campo de Gibraltar	g1-g27/g1-g33	Arcillas	g1/g1
U. Campo de Gibraltar		Areniscas	g20-g26/g22-g31
U. Campo de Gibraltar		Calizas	g3-g27/g4-g33
U. Campo de Gibraltar		Conglomerado	g16/g17
U. Campo de Gibraltar		Margas	g8/g9
U. Campo de Gibraltar		Margocalizas	g11/g12
Zona Prebética	g10-g26/g11-g30	Calizas	g10-g26/g11-g30
Zona Prebética		Margas	g10-g11/g11-g12
Zona Subbética	g1-g29/g1-g46	Arcillas	g5-g27/g6-g34
Zona Subbética		Areniscas	g1/g1
Zona Subbética		Calizas	g1-g29/g1-g46
Zona Subbética		Conglomerado	g5-g12/g6-g13
Zona Subbética		Margas	g3-g28/g4-g35
Zona Subbética		Margocalizas	g1-g11/g1-g12

Zn	Unidades Estructurales	N. característico	Litología	N. característico
	Complejo Alpujárride	g4-g28	Arcillas	g4-g24
	Complejo Alpujárride		Calizas-Mármoles	g12-g51
	Complejo Alpujárride		Esquistos	g5-g28
	Complejo Alpujárride	g5-g51	Filitas	g5-g28
	Complejo Alpujárride		Gneiss	g14-g32
	Complejo Alpujárride		Peridotitas	g14-g25
	Complejo Maláguide		Areniscas	g19-g25
	Complejo Maláguide	g5-g29	Esquistos	g12-g27
	Complejo Maláguide		Filitas	g12/g13
	Complejo Nevado-Filábride		Areniscas	g27/g34
	Complejo Nevado-Filábride		Calizas-Mármoles	g14-g28
	Complejo Nevado-Filábride	g3-g40	Esquistos	g3-g11
	Complejo Nevado-Filábride		Filitas	g28/g35
	Complejo Nevado-Filábride		Filitas-Cuarcitas	g6-g20
	Depresiones Postorogénicas		Arcillas-limos	g3-g48
	Depresiones Postorogénicas		Arenas	g4-g18
	Depresiones Postorogénicas		Areniscas y margas	g4-g14
	Depresiones Postorogénicas	g2-g28	Calizas	g2-g28
	Depresiones Postorogénicas		Conglomerado	g4-g28
	Depresiones Postorogénicas		Esquistos	g7-g28
	Depresiones Postorogénicas		Filitas	g11/g12
	Depresiones Postorogénicas		Margas	g4-g28
	Depresiones Postorogénicas		Margocalizas	g2-g16
	Depresiones Postorogénicas		Mat.fluviales	g6-g26
	Rocas Volcánicas	g3-g28	Andesita	g3/g4
	Rocas Volcánicas	g4-g42	Dacitas	g24/g31
	Rocas Volcánicas		Lamproita	g28/g42
	U. Campo de Gibraltar		Arcillas	g22/g28
	U. Campo de Gibraltar		Areniscas	g21-g26
	U. Campo de Gibraltar	g9-g41	Calizas	g9-g41
	U. Campo de Gibraltar		Conglomerado	g25/g32
	U. Campo de Gibraltar		Margas	g20/g23
	U. Campo de Gibraltar		Margocalizas	g16/g18
	Zona Prebética	g2-g19	Calizas	g2-g19
	Zona Prebética	g3-g21	Margas	g4-g12
	Zona Subbética		Arcillas	g4-g26
	Zona Subbética		Areniscas	g9/g10
	Zona Subbética		Calizas	g2-g28
	Zona Subbética		Conglomerado	g8-g19
	Zona Subbética	g1-g46	Margas	g2-g23
	Zona Subbética	g1-g28	Margocalizas	g1-g21

Muestra	X	Y	Código Loc.	As	Unidades Estructurales	N. característico	Litología	N. característico	I Combinado	N. característico	Cod. MR2	Anomalía MR2	I/O MR2	Anomalía I/O MR2	
MA-38-1	2.973020	40.909940	1	9.60	U. Campo de Gibraltar		Areniscas	g10-g13/g11-g14	5	g10/g11	10	0.17	11	0.17	0.17
MA-37-1	2.983500	40.823500	2	5.20	Depresiones Postorogénicas		Margas	g4-g19/g5-g20	5		4	0.33	5	0.33	0.33
MA-8-1	3.087460	40.673450	3	11.20	Complejo Alpujárride		Pelitas		5		11	11.20	12	11.20	11.20
MA-9-1	3.098240	40.535590	4	12.60	Complejo Alpujárride		Gneis		5		13	7.73	14	7.73	7.73
MA-35-1	3.109970	40.921840	5	2.10	Zona Subbética		Calizas		3		2	0.00	3	0.00	0.00
MA-7-1	3.112850	40.615380	6	86.50	Zona Subbética		Calizas		5		28	84.36	42	84.24	84.24
MA-36-1	3.118220	40.820950	7	7.90	Depresiones Postorogénicas		Areniscas		3		9	3.03	10	3.03	3.03
MA-6-1	3.119290	40.721240	8	7.80	Depresiones Postorogénicas		Calizas		3		9	0.00	10	0.00	0.00
MA-39-1	3.125490	41.024350	9	3.20	Zona Subbética		Margocalizas		3	g3/g4	3	0.00	4	0.00	0.00
MA-40-1	3.127180	41.132810	10	2.40	Zona Subbética		Margas	g2-g14/g3-g15	6	g2/g3	2	0.26	3	0.14	0.14
MA-1-1	3.142550	40.381520	11	5.20	Complejo Alpujárride		Pelitas		5	g1/g1	4	5.20	5	5.20	5.20
MA-21-1	3.146600	40.530200	12	0.00	Complejo Alpujárride		Pelitas		5		1	0.00	1	0.00	0.00
MA-20-1	3.226200	40.725200	13	9.90	Zona Subbética		Calizas		4		10	7.76	11	7.64	7.64
MA-2-1	3.254870	40.476250	14	17.70	Complejo Alpujárride		Calizas-Mármoles		6		16	15.56	17	15.10	15.10
MA-19-1	3.255300	40.734700	15	25.80	U. Campo de Gibraltar		Margocalizas	g19/g20	4	g19/g20	19	0.00	20	0.00	0.00
MA-45-1	3.257070	40.834410	16	1.90	Zona Subbética		Calizas		3	g2/g3	2	0.00	3	0.00	0.00
MA-4-1	3.258730	40.660170	17	7.90	Complejo Alpujárride		Calizas-Mármoles		3	g9/g10	9	0.00	10	0.00	0.00
MA-46-1	3.275380	40.923810	18	4.10	U. Campo de Gibraltar		Calizas		3	g4/g5	4	0.00	5	0.00	0.00
MA-41-1	3.275700	41.108620	19	6.79	Zona Subbética		Calizas		6		7	4.65	8	4.53	4.53
MA-42-1	3.280980	41.196130	20	4.50	Zona Subbética		Calizas		5		4	2.36	5	2.24	2.24
MA-5-1	3.281890	40.740030	21	4.90	U. Campo de Gibraltar		Calizas		3		4	0.03	5	0.03	0.03
MA-34-1	3.282040	41.008320	22	3.20	Zona Subbética		Margas		3	g3/g4	3	0.00	4	0.00	0.00
MA-3-1	3.289650	40.559180	23	6.30	Complejo Alpujárride		Pelitas		6		6	3.04	7	3.04	3.04
MA-17-1	3.322300	40.667800	24	32.50	Complejo Maláguide		Esquistos		6		21	15.35	26	15.35	15.35
MA-18-1	3.324600	40.674700	25	3.00	Complejo Alpujárride		Pelitas	g1-g11/g1-g12	6	g3/g4	3	0.00	4	0.00	0.00
MA-5-1	3.360100	41.019500	26	13.90	Depresiones Postorogénicas		Margas		4		14	9.03	15	9.03	9.03
MA-6-1	3.367800	40.939300	27	9.10	U. Campo de Gibraltar	g2-g19/g3-g20	Arcillas	g10/g11	4	g10/g11	10	0.00	11	0.00	0.00
MA-7-1	3.397000	40.909500	28	5.70	Depresiones Postorogénicas		Conglomerado		4		5	2.44	6	2.44	2.44
MA-15-1	3.406200	40.805500	29	10.00	Complejo Alpujárride		Pelitas		4		10	0.57	11	0.57	0.57
MA-44-1	3.412320	41.104860	30	10.90	Complejo Nevado-Filábride		Calizas-Mármoles		4	g11/g12	11	0.05	12	0.05	0.05
MA-8-1	3.418400	40.913000	31	2.60	Zona Subbética		Calizas		4		2	0.00	3	0.00	0.00
MA-18-1	3.423450	40.637250	32	9.00	U. Campo de Gibraltar		Conglomerado	g10/g11	4	g10/g11	10	0.00	11	0.00	0.00
MA-43-1	3.428250	41.209950	33	13.30	Zona Subbética		Calizas		2	g14/g15	14	0.00	15	0.00	0.00
MA-12-1	3.430130	40.920410	34	2.50	U. Campo de Gibraltar		Calizas		4	g2-g3	2	0.36	3	0.24	0.24
MA-16-1	3.430400	40.826000	35	4.80	Complejo Alpujárride		Gneis		3	g4/g5	4	0.00	5	0.00	0.00
MA-14-1	3.431900	40.794300	36	9.00	Complejo Alpujárride		Pelitas		4	g10/g11	10	0.00	11	0.00	0.00
MA-21-1	3.434130	40.429010	37	53.60	Complejo Alpujárride		Esquistos		4		24	49.90	32	50.34	50.34
MA-13-1	3.434500	40.784800	38	8.10	Depresiones Postorogénicas		Margas		4		9	3.23	10	3.23	3.23
MA-14-1	3.437730	40.727420	39	12.90	U. Campo de Gibraltar		Areniscas		4	g13/g14	13	3.47	14	3.47	3.47
MA-13-1	3.442650	40.816880	40	6.20	Complejo Alpujárride		Gneis		4		6	1.33	7	1.33	1.33
MA-17-1	3.443800	40.532820	41	137.00	Complejo Alpujárride		Calizas-Mármoles		6		26	134.86	45	134.40	134.40
MA-4-1	3.456600	41.026200	42	9.60	Depresiones Postorogénicas		Areniscas		4		10	4.73	11	4.73	4.73
MA-9-1	3.481600	40.917200	43	12.10	U. Campo de Gibraltar		Margas	g12/g13	4	g12/g13	12	0.00	13	0.00	0.00
MA-12-1	3.500300	40.852500	44	14.00	Zona Subbética		Margas		4		14	11.86	15	11.74	11.74
MA-3-1	3.505400	41.087200	45	7.60	Depresiones Postorogénicas		Calizas		3	g9/g10	9	0.00	10	-0.52	-0.52
MA-19-1	3.547300	40.630440	46	5.70	Complejo Alpujárride		Calizas-Mármoles		4	g5/g6	5	0.00	6	0.00	0.00
MA-11-1	3.548900	40.930700	47	8.80	Zona Subbética		Margas		4		10	6.66	11	6.54	6.54
MA-31-1	3.563070	41.294370	48	9.00	Zona Subbética		Margocalizas		4		10	5.74	11	5.74	5.74
MA-15-1	3.565210	40.806870	49	78.50	Complejo Maláguide		Esquistos		4		25	58.52	38	58.52	58.52
MA-11-1	3.565520	40.916890	50	2.90	U. Campo de Gibraltar		Calizas		4		3	0.76	4	0.64	0.64
MA-16-1	3.565610	40.720900	51	26.80	Complejo Maláguide		Esquistos		4		20	6.82	21	6.82	6.82
MA-32-1	3.573040	41.180590	52	6.20	Zona Subbética		Calizas		3		6	4.06	7	3.94	3.94
MA-20-1	3.573100	40.526920	53	74.30	Complejo Alpujárride		Calizas-Mármoles		5		25	72.16	36	71.70	71.70
MA-47-1	3.575320	41.005220	54	6.00	Zona Subbética		Areniscas	g6/g7	4	g6/g7	6	0.00	7	0.00	0.00
MA-50-1	3.575930	41.104760	55	5.49	Zona Subbética		Arcillas		3	g4/g5	4	0.62	5	0.62	0.62
MA-10-1	3.577080	41.006310	56	8.80	Zona Subbética		Calizas		4		10	6.66	11	6.54	6.54
MA-10-1	3.582700	40.949900	57	8.90	Depresiones Postorogénicas		Conglomerado		4		10	5.64	11	5.64	5.64
MA-23-1	3.588100	40.507300	58	162.00	Complejo Alpujárride		Calizas-Mármoles		5		26	159.86	46	159.40	159.40
MA-2-1	3.604200	41.049000	59	11.90	Depresiones Postorogénicas		Conglomerado		3	g12/g13	12	0.00	13	0.00	0.00
MA-22-1	3.641600	40.532600	60	49.70	Depresiones Postorogénicas		Conglomerado		5		23	46.44	30	46.44	46.44
MA-24-1	3.646900	40.635200	61	13.30	Depresiones Postorogénicas		Margas		4		14	8.43	15	8.43	8.43
MA30-1	3.696940	41.290750	62	8.40	Zona Subbética		Margas		3		9	5.14	10	5.14	5.14
MA-33-1	3.701660	41.180230	63	4.10	Depresiones Postorogénicas		Margas		3	g4/g5	4	0.00	5	0.00	0.00
MA-49-1	3.704000	40.987000	64	8.18	Zona Subbética		Calizas		3		9	6.04	10	5.92	5.92
MA-24-1	3.714200	40.730260	65	8.40	Complejo Maláguide		Filitas	g9/g10	6	g9/g10	9	0.00	10	0.00	0.00
MA-48-1	3.717600	41.096960	66	9.80	Depresiones Postorogénicas		Calizas		3		10	1.68	11	1.68	1.68
MA-1-1	3.720450	41.051500	67	10.70	Zona Subbética		Arcillas		3		11	5.83	12	5.83	5.83
MA-23-1	3.723130	40.832700	68	33.90	Complejo Maláguide		Areniscas		4		21	27.82	27	27.82	27.82
MA-22-1	3.733940	40.914050	69	14.00	U. Campo de Gibraltar		Calizas	g2-g17/g3-g18	5		14	11.86	15	11.74	11.74
MA-25-1	3.809500	40.667200	70	16.30	Complejo Maláguide		Esquistos	g16-g25/g17-g38	6	g16/g17	16	0.00	17	0.00	0.00
GR-14-1	3.865000	41.103750	71	9.60	Depresiones Postorogénicas		Calizas		2	g10/g11	10	0.00	11	0.00	0.00
MA-27-1	3.866840	40.726540	72	41.30	Complejo Maláguide		Esquistos		6		22	24.15	29	24.15	24.15
MA-29-1	3.870210	41.277460	73	6.60	Zona Subbética		Margocalizas		2	g7/g8	7	0.00	8	0.00	0.00
MA-25-1	3.884350	40.914190	74	20.60	U. Campo de Gibraltar		Calizas		5		17	18.46	18	18.34	18.34
MA-26-1	3.886570	40.821600	75	21.00	Complejo Maláguide		Esquistos		4	g17/g18	17	1.02	18	1.02	1.02
GR-13-1	3.887250	41.207250	76	5.00	Depresiones Postorogénicas		Areniscas		3	g4/g5	4	0.13	5	0.13	0.13
GR-15-1	3.893750	41.002000	77	10.10	Zona Subbética		Calizas		3		10	7.96	11	7.84	7.84
MA-27-1	3.993900	40.668600	78	49.20	Complejo Maláguide		Esquistos		6		23	32.05	30	32.05	32.05
MA-52-1	3.999110	40.709030	79	16.74	Depresiones Postorogénicas		Conglomerado		6		16	13.48	17	13.48	13.48
GR-16-1	4.016250	40.923250	80	15.10	Depresiones Postorogénicas		Calizas		3		15	6.98	16	6.98	6.98
MA-26-1	4.018200	40.674600	81	9.20	Depresiones Postorogénicas		Arcillas-limos		3		10	5.94	11	5.94	5.94
GR-11-1	4.033750	41.306000	82	5.00	Zona Subbética		Calizas	g2-g26/g3-g42	6		4	2.86	5	2.74	2.74
GR-17-1	4.035750	41.012500	83	17.50	Zona Subbética		Calizas		3		16	15.36	17	15.24	15.24
GR-18-1	4.054000	41.117000	84	5.00	Depresiones Postorogénicas		Margocalizas		2	g4/g5	4	0.00	5	0.00	0.00
GR-12-1	4.090500	41.203500	85	7.90	Depresiones Postorogénicas		Mat.fluviales		2	g9/g10	9	0.00	10	0.00	0.00
MA-28-1	4.143300	40.658700	86	13.30	Depresiones Postorogénicas		Conglomerado		6		14	10.04	15	10.04	10.04
MA-29-1	4.147000	40.688000	87	29.00	Complejo Alpujárride		Gneis		6		20	24.13	22	24.13	24.13
MA-30-1	4.150100	40.692700	88	13.00	Complejo Alpujárride		Gneis	g4-g20/g5-g22	6		13	8.13	14	8.13	8.13
MA-51-1	4.152570	40.820930	89	1.12	Complejo Alpujárride		Calizas-Mármoles	g2-g26/g2/g46	6	g2/g2-g3	2	0.00	2	0.00	0.00
MA-31-1	4.153800	40.691800	90	9.70	Complejo Alpujárride		Gneis		6		10	4.83	11	4.83	4.83
MA-28-1	4.169530	40.717510	91	11.90	Complejo Alpujárride		Esquistos		6		12	8.20	13	8.64	8.64
GR-9-1	4.171500	41.190000	92	5.60	Depresiones Postorogénicas		Margas		2	g5/g6	5	0.00	6	0.00	0.00
GR-21-1	4.174820	40.905830	93	10.70	Depresiones Postorogénicas		Calizas		4		11	2.58	12	2.58	2.58
GR-10-1	4.175500	41.284000	94	16.30	Zona Subbética		Calizas		3		16	14.16	17	14.04	14.04
GR-135-1	4.180520	40.814580	95	1.72	Complejo Alpujárride		Calizas-Mármoles		6		2	0.00	3	0.00	0.00
GR-20-1	4.182190	41.007310	96	5.90	Depresiones Postorogénicas		Margas		4		6	1.03	7	1.03	1.03
GR-19-1	4.182240	41.099500	97	13.40	Depresiones Postorogénicas		Conglomerado		3		14	1.30	15	1.30	1.30
GR-2-1	4.291750	41.310750	98	4.90	Depresiones Postorogénicas		Mat.fluviales		3	g4/g5	4	0.00	5	0.00	0.00
GR-133-1	4.306300	40.831500	99	2.86	Complejo Alpujárride		Esquistos		6		3	0.00	4	0.00	0.00
GR-3-1	4.308500	41.384000	100	5.60	Zona Subbética		Margocalizas		3		5	2.34	6	2.34	2.34
GR-23-1	4.320000	41.010000	101	17.40	Depresiones Postorogénicas		Margas		4		16	12.53	17	12.53	12.53
GR-1(1)	4.324500	41.208000	102	14.80	Depresiones Postorogénicas		Arcillas-limos		3		15	11.54	16	11.54	11.54
GR-4-1	4.324750	41.473000	103	3.00	Zona Subbética		Margas		5		3	0.86	4	0.74	0.74
GR-22-1	4.326030	40.929940	104	6.70	Complejo Alpujárride		Calizas-Mármoles		4		7	1.00	8	1.10	1.10
GR-134-1	4.327530	40.711640	105	6.67	Complejo Alpujárride		Esquistos		6		7	2.97	8	3.41	3.41
GR-24-1	4.357400	41.133020	106	18.40	Depresiones Postorogénicas		Mat.fluviales		3		16	13.53	17	13.53	13.53
GR-131-1	4.451140	40.731700	107	7.05	Complejo Alpujárride		Calizas-Mármoles		4		7	1.35	8	1.45	1.45
GR-132-1	4.454930	40.806400	108	3.24	Complejo Alpujárride		Esquistos		5	g3/g4	3	0.00	4	0.00	0.00
GR-7-1	4.466750	41.295250	109	8.50	Depresiones Postorogénicas		Calizas		4		9	0.38	10	0.38	0.38
GR-8-1	4.470250	41.200750	110	14.70	Depresiones Postorogénicas		Mat.fluviales		3		15	9.83	16	9.83	9.83
GR-26-1	4.474870	41.024180	111	20.20	Depresiones Postorogénicas		Conglomerado		3		17	8.10	18	8.10	8.10
GR-6-1	4.475250	41.392250	112	8.20	Depresiones Postorogénicas		Margocalizas		4		9	6.06	10	5.94	5.94
GR-25-1	4.476530	41.104880	113	15.40	Depresiones Postorogénicas		Arcillas-limos		4		15	12.14	16	12.14	12.14
GR-5-1	4.520750	41.486750	114	6.70	Depresiones Postorogénicas		Margas		4		7	1.83	8	1.83	1.83
GR-130-1	4.603470	40.631430	115	6.77	Complejo Alpujárride		Esquistos		3	g7/g8	7	0.00	8	0.00	0.00
GR-37-1	4.609670	41.288120	116	26.20	Complejo Alpujárride		Calizas-Mármoles		6		19	24.06	20	23.60	23.60
GR-31-1	4.620830	41.388000	117	14.40	Zona Subbética		Calizas		4		14	12.26	15	12.14	12.14
GR33-1	4.620890	41.496320	118	8.70	Zona Subbética		Margocalizas		4		10	5.44	11	5.44	5.44
GR-127-1	4.624210	40.893600	119	16.64	Complejo Nevado-Filábride		Esquistos		6		16	11.77	17	11.77	11.77
GR-128-1	4.625420	40.821470	120	27.74	Complejo Alpujárride		Esquistos		6		20	24.04	21	24.48	24.48
GR32-1	4.634000	41.553290	121	5.40	Depresiones Postorogénicas		Margocalizas		4		4	3.26	5	3.14	3.14
GR28-1	4.638140	41.199020	122	14.30	Complejo Alpujárride		Calizas-Mármoles		3		14	6.18	15	6.18	6.18
GR-129-1	4.648020	40.715870	123	13.63	Complejo Alpujárride		Esquistos		6		14	9.93	15	10.37	10.37
GR29-1	4.697520	41.241880	124	5.70	Depresiones Postorogénicas		Areniscas y margas		3	g5/g6	5	0.04	6	0.04	0.04
GR-125-1	4.747720	40.808020	125	7.21	Complejo Alpujárride		Filitas		4		8	2.34	9	2.34	2.34
GR30-1	4.757570	41.291980	126	15.20	Depresiones Postorogénicas		Calizas		4		15	7.08	16	7.08	7.08
GR-126-1	4.762080	40.915970	127	23.76	Complejo Nevado-Filábride		Esquistos		6		18	18.89	19	18.89	18.89
GR-35-1	4.769890	41.490700	128	4.60	Zona Subbética		Margocalizas		4		4	1.34	5	1.34	1.34
GR-34-1	4.784380	41.549560	129	8.50	Zona Subbética		Margocalizas		4		9	5.24	10	5.24	5.24
GR-38-1	4.785200	41.193070	130	33.10	Complejo Nevado-Filábride		Esquistos		4		21	28.23	27	28.23	28.23
GR-124-1	4.813000	40.722250	131	18.40	Complejo Alpujárride		Esquistos	g3-g24/g32	8		16	16.49	17	15.14	15.14
GR-123-1	4.896160	40.703820	132	49.28	Complejo Alpujárride		Esquistos		7		23	45.58	30	46.02	46.02
GR-122-1	4.898250	40.816200	133	10.26	Complejo Alpujárride		Esquistos		5		10	6.56	11	7.00	7.00
GR-43-1	4.903150	41.302180	134	116.00	Depresiones Postorogénicas		Conglomerado		4		26	112.74	44	112.74	112.74

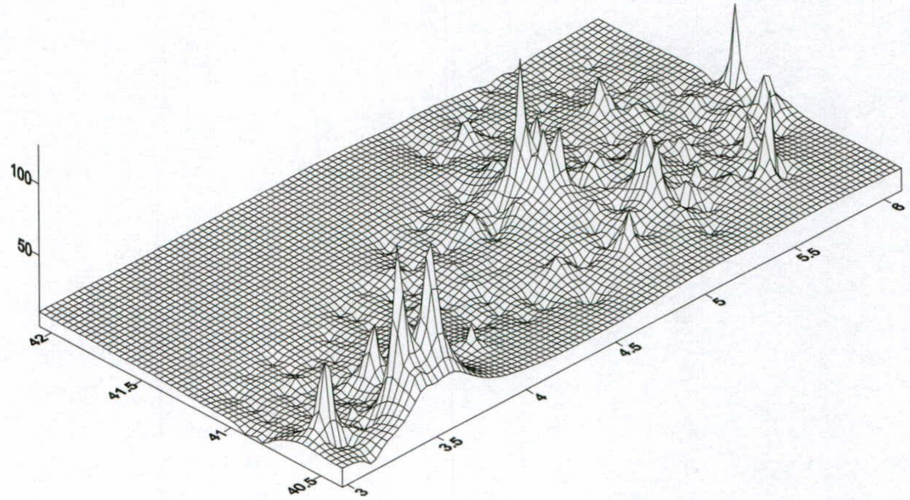
GR-121-1	4.921620	40.920290	135	13.31	Depresiones Postorogénicas		Conglomerado		6		14	10.05	15	10.05
GR-40-1	4.926480	41.107770	136	36.80	Complejo Nevado-Filábride		Esquistos		4		21	31.93	27	31.93
GR44-1	4.927220	41.361870	137	25.10	Depresiones Postorogénicas		Calizas		4		19	16.98	20	16.98
GR39-1	4.928240	41.198780	138	77.70	Depresiones Postorogénicas		Calizas		4		25	69.58	37	69.58
GR95-1	4.929600	41.771900	139	5.10	Zona Subbética		Arcillas		3		4	0.23	5	0.23
GR36-1	4.936000	41.583750	140	15.60	Depresiones Postorogénicas		Arcillas-limos		3		15	12.34	16	12.34
GR-45-1	4.937200	41.473590	141	35.70	Depresiones Postorogénicas		Calizas		3		21	27.58	27	27.58
GR96-1	4.939490	41.874610	142	2.30	Zona Prebética		Margas	g2-g46/g3-g5	3	g2/g3	2	0.16	3	0.04
GR97-1	4.940170	41.948860	143	5.00	Zona Prebética		Margas		3		4	2.86	5	2.74
A1-1057-1-1	4.949500	40.686000	144	25.80	Complejo Alpujárride		Esquistos		6		19	22.10	20	22.54
GR41-1	4.995790	41.187540	145	82.90	Depresiones Postorogénicas		Esquistos		4		26	72.05	39	72.05
1028-3(1)	5.001000	40.991000	146	31.70	Complejo Nevado-Filábride		Esquistos		6		21	26.83	25	26.83
1043-1(1)	5.008600	40.947000	147	22.70	Complejo Alpujárride		Esquistos		6		18	19.00	19	19.44
A1-1057-2-1	5.021300	40.760000	148	19.30	Depresiones Postorogénicas		Calizas		4		17	11.18	18	11.18
A1-1043-2-1	5.032000	40.833000	149	36.30	Depresiones Postorogénicas		Calizas		4		21	28.18	27	28.18
1028-2(1)	5.047000	40.984000	150	23.00	Complejo Nevado-Filábride		Esquistos		6		18	18.13	19	18.13
GR71-1	5.058500	41.652500	151	8.80	Depresiones Postorogénicas		Margocalizas		4		10	6.66	11	6.54
GR-46-1	5.061250	41.378520	152	73.40	Depresiones Postorogénicas		Conglomerado		3		25	61.30	35	61.30
GR-42-1	5.061500	41.173070	153	4.10	Complejo Nevado-Filábride		Esquistos		4	g4/g5	4	0.00	5	0.00
GR48-1	5.065780	41.659750	154	13.30	Depresiones Postorogénicas		Conglomerado		4		14	10.04	15	10.04
GR47-1	5.067800	41.472480	155	22.20	Depresiones Postorogénicas		Conglomerado		3		18	10.10	19	10.10
GR-72-1	5.071850	41.776560	156	33.10	Zona Prebética		Calizas		5		21	28.23	27	28.23
GR73-1	5.072740	41.844010	157	6.00	Zona Prebética		Calizas		6		6	1.13	7	1.13
1028-1(1)	5.086000	40.952000	158	15.30	Complejo Alpujárride		Filitas		4		15	10.43	16	10.43
A1-1057-3-1	5.107000	40.644000	159	25.60	Depresiones Postorogénicas		Arenas	g4-g19/g5-g20	4		19	20.73	20	20.73
A1-1012-2-1	5.145580	41.161600	160	36.50	Depresiones Postorogénicas		Conglomerado		4		21	33.24	27	33.24
A1-1029-1-1	5.152790	41.075200	161	22.20	Complejo Nevado-Filábride		Esquistos		4		18	17.33	19	17.33
A1-1058-3-1	5.201100	40.714300	162	29.90	Depresiones Postorogénicas		Conglomerado		4		20	26.64	23	26.64
GR49-1	5.206020	41.307080	163	10.60	Complejo Nevado-Filábride		Esquistos		4		11	5.73	12	5.73
1044-6-1	5.207000	40.860000	164	65.45	Complejo Alpujárride		Calizas-Mármoles		6		25	63.31	34	62.85
GR-53-1	5.213850	41.589520	165	22.30	Depresiones Postorogénicas		Margas		4		18	17.43	19	17.43
GR-59-1	5.214200	41.772940	166	10.10	Depresiones Postorogénicas		Calizas		3		10	1.98	11	1.98
GR-51-1	5.216120	41.474340	167	17.10	Depresiones Postorogénicas		Arcillas-limos		4		16	13.84	17	13.84
GR-56-1	5.219850	41.668940	168	10.20	Depresiones Postorogénicas		Arcillas-limos		4		10	6.94	11	6.94
GR65-1	5.220500	41.936100	169	5.30	Zona Prebética		Calizas		4	g4/g5	4	0.43	5	0.43
GR-62-1	5.222200	41.868600	170	13.00	Zona Prebética		Calizas		6		13	8.13	14	8.13
A1-1012-3-1	5.223700	41.197200	171	38.60	Complejo Nevado-Filábride		Esquistos		5		22	35.34	28	35.34
GR-50-1	5.223770	41.390040	172	14.80	Depresiones Postorogénicas		Margas		4		15	9.93	16	9.93
A1-1029-2-1	5.228600	41.099800	173	28.30	Complejo Nevado-Filábride		Esquistos		4		20	23.43	21	23.43
A1-1029-3-1	5.230560	41.044400	174	26.20	Complejo Nevado-Filábride		Esquistos		4		19	21.33	20	21.33
1044-4-1	5.232000	40.934000	175	86.40	Depresiones Postorogénicas		Calizas	g9-g25/g10-g41	6		26	78.28	41	78.28
A1-1058-2-1	5.259600	40.608300	176	11.50	Depresiones Postorogénicas		Conglomerado		4		11	8.24	12	8.24
1044-3(1)	5.303000	40.802000	177	16.50	Complejo Alpujárride		Filitas		3	g16/g17	16	0.00	17	0.00
A1-1058-4-1	5.313000	40.650000	178	16.10	Depresiones Postorogénicas		Conglomerado		4		16	12.84	17	12.84
1058-4-2	5.313000	40.650000	179	19.41	Complejo Alpujárride		Esquistos		4		17	15.71	18	16.15
994-2-1	5.334000	41.365000	180	36.10	Complejo Nevado-Filábride		Filitas-Cuarcitas	g21/g27	4	g21/g27	21	0.00	27	0.00
994-2-1	5.334000	41.365000	181	36.10	Complejo Nevado-Filábride		Filitas-Cuarcitas		4		21	0.00	27	0.00
GR66-1	5.337730	41.978410	182	2.60	Depresiones Postorogénicas		Margocalizas	g2-g10/g4-g11	6	g2/g3	2	0.00	3	0.00
1044-2(1)	5.350000	40.812000	183	43.40	Complejo Alpujárride		Calizas-Mármoles		3		22	35.28	29	35.28
GR60-1	5.356490	41.773010	184	8.30	Depresiones Postorogénicas		Arcillas-limos		4		9	5.04	10	5.04
A1-1029-4-1	5.360700	41.093000	185	53.80	Complejo Nevado-Filábride		Esquistos		6		24	48.93	33	48.93
GR-63-1	5.361620	41.874890	186	8.00	Zona Subbética		Calizas		6		9	5.86	10	5.74
GR52-1	5.362150	41.480790	187	12.30	Depresiones Postorogénicas		Calizas		4		12	4.18	13	4.18
GR-54-1	5.363060	41.577240	188	12.30	Depresiones Postorogénicas		Arcillas-limos		4		12	9.04	13	9.04
GR57-1	5.364470	41.666790	189	3.60	Depresiones Postorogénicas		Arcillas-limos		3	g3/g4	3	0.34	4	0.34
A1-1058-1-1	5.366500	40.733000	190	8.40	Depresiones Postorogénicas		Conglomerado		6		9	5.14	10	5.14
994-1(1)	5.370000	41.411000	191	17.20	Zona Subbética		Conglomerado		4		16	10.42	17	10.42
A1-1029-5-1	5.371970	41.028500	192	43.00	Complejo Nevado-Filábride		Esquistos		4		22	38.13	29	38.13
A1-1044-5-1	5.382400	40.926600	193	27.50	Depresiones Postorogénicas		Conglomerado		4		20	24.24	21	24.24
1044-1(1)	5.417000	40.872000	194	9.50	Complejo Alpujárride		Filitas		4		10	4.63	11	4.63
1012-1(1)	5.418000	41.283000	195	10.90	Depresiones Postorogénicas		Calizas		4		11	2.78	12	2.78
1013-5-1	5.428000	41.187000	196	31.90	Complejo Nevado-Filábride		Esquistos		6		21	27.03	25	27.03
1013-1-1	5.434000	41.265000	197	7.30	Complejo Alpujárride		Esquistos		4		8	3.60	9	4.04
1045-4(1)	5.434900	40.865000	198	30.30	Complejo Alpujárride		Arcillas		4		20	13.15	24	13.15
1045-3(1)	5.444500	40.884000	199	24.80	Depresiones Postorogénicas		Areniscas		4		19	19.93	20	19.93
1045-2(1)	5.462000	40.896000	200	25.10	Depresiones Postorogénicas		Conglomerado		4		19	21.84	20	21.84
1045-1(1)	5.464000	40.893000	201	31.30	Depresiones Postorogénicas		Conglomerado		4		21	28.04	25	28.04
1045-39	5.476600	40.922400	202	20.00	Depresiones Postorogénicas		Conglomerado		4		17	16.74	18	16.74
GR70-1	5.479710	42.040270	203	4.50	Zona Prebética	g2-g21/g3-g27	Calizas	g4-g21/g5-g27	6		4	0.00	5	0.00
GR-58-1	5.481150	41.655810	204	10.40	Zona Subbética		Calizas		5		11	8.26	12	8.14
1013-6-1	5.491000	41.242000	205	5.20	Complejo Nevado-Filábride		Esquistos		5		4	1.94	5	1.94
995-4(1)	5.503000	41.393000	206	9.40	Depresiones Postorogénicas		Calizas		4		10	1.28	11	1.28
GR61-1	5.503740	41.771270	207	8.40	Depresiones Postorogénicas		Calizas		4		9	0.28	10	0.28
GR67-1	5.504220	41.969200	208	4.40	Zona Subbética		Calizas		3		4	2.26	5	2.14
GR55-1	5.517350	41.581190	209	14.80	Complejo Alpujárride		Esquistos		5		15	11.10	16	11.54
1030-4(1)	5.518000	41.123000	210	33.60	Complejo Nevado-Filábride		Esquistos		6		21	28.73	27	28.73
1045-5-1	5.535000	40.894000	211	12.67	Depresiones Postorogénicas		Filitas	g13/g14	6	g13/g14	13	0.00	14	0.00
1045-1	5.537000	40.765000	212	27.50	Depresiones Postorogénicas		Conglomerado		5		20	24.24	21	24.24
1030-3(1)	5.545000	41.017000	213	20.20	Depresiones Postorogénicas		Esquistos	g11-g26/g12-g40	5		17	9.35	18	9.35
GR-64-1	5.554580	41.868440	214	5.20	Depresiones Postorogénicas		Margocalizas		4		4	3.06	5	2.94
995-1(1)	5.560000	41.478000	215	19.70	Depresiones Postorogénicas		Calizas		5		17	11.58	18	11.58
1030-1(1)	5.569000	41.067000	216	27.20	Depresiones Postorogénicas		Mat.fluviales	g4-g20/g5-g21	5		20	22.33	21	22.33
1045-35	5.573300	41.068800	217	40.50	Depresiones Postorogénicas		Conglomerado	g3-26/g4-g44	7		22	37.24	29	37.24
1045-37	5.588100	41.109000	218	23.40	Depresiones Postorogénicas		Conglomerado		6		18	20.14	19	20.14
973-2(1)	5.604000	41.554000	219	40.90	Complejo Alpujárride		Esquistos		4		22	37.20	29	37.64
973-2(1)	5.604000	41.554000	220	40.90	Complejo Alpujárride		Esquistos		4		22	37.20	29	37.64
1045-33	5.627800	41.045900	221	13.30	Depresiones Postorogénicas		Arcillas-limos		3		14	10.04	15	10.04
1013-3(1)	5.628000	41.177000	222	4.70	Complejo Nevado-Filábride		Esquistos	g3-g11/g4-g12	6	g4/g5	4	0.00	5	0.00
GR68-1	5.629070	41.963670	223	5.10	Zona Subbética		Calizas		4		4	2.96	5	2.84
1045-3	5.629600	40.764900	224	214.00	Complejo Nevado-Filábride	g3-g26/g4-g48	Areniscas	g26/g48	5	g26/g48	26	0.00	48	0.00
GR-69-1	5.630350	42.067360	225	6.90	Zona Prebética		Calizas		4		7	2.03	8	2.03
1013-2(1)	5.640000	41.261000	226	8.10	Complejo Nevado-Filábride		Esquistos		4		9	3.23	10	3.23
951-2(1)	5.640000	41.859000	227	6.70	Zona Subbética		Conglomerado	g7-g16/g8-g18	4	g7/g8	7	0.00	8	0.00
973-1(1)	5.641000	41.620000	228	9.80	Depresiones Postorogénicas		Calizas		4		10	1.68	11	1.68
973-1(1)	5.641000	41.620000	229	9.80	Complejo Alpujárride		Esquistos		5		10	6.10	11	6.54
951-4(1)	5.642000	41.730000	230	7.20	Zona Subbética		Calizas		4		8	5.06	9	4.94
951-3(1)	5.661000	41.785000	231	6.00	Complejo Maláguide	g6-g25/g7-g38	Areniscas	g6-g21/g7-g27	4	g6/g7	6	0.00	7	0.00
995-3(1)	5.666000	41.359000	232	6.20	Depresiones Postorogénicas		Areniscas y margas		4		6	0.54	7	0.54
1045-7	5.682900	40.709100	233	93.10	Depresiones Postorogénicas	g2-g26/g3-g44	Arcillas-limos	g3-g26/g4-g43	5		26	89.84	43	89.84
1045-5	5.684900	40.751200	234	23.70	Depresiones Postorogénicas		Conglomerado		5		18	20.44	19	20.44
930-1(1)	5.685600	41.923000	235	5.40	Depresiones Postorogénicas		Margas		4		4	0.53	5	0.53
995-2(1)	5.703000	41.463000	236	20.60	Depresiones Postorogénicas		Conglomerado		5		17	17.34	18	17.34
973-3(1)	5.707000	41.546000	237	4.70	Complejo Alpujárride		Filitas	g4-g16/g5-g17	4	g4/g5	4	0.00	5	0.00
973-3(1)	5.707000	41.546000	238	4.70	Complejo Alpujárride		Filitas		4		4	0.00	5	0.00
1030-2(1)	5.709000	41.021000	239	9.10	Depresiones Postorogénicas		Areniscas y margas	g5-g13/g6-g14	5		10	3.44	11	3.44
1013-4(1)	5.709000	41.173000	240	3.50	Complejo Nevado-Filábride		Esquistos		5	g3/g4	3	0.24	4	0.24
951-1(1)	5.712000	41.730000	241	9.30	Depresiones Postorogénicas		Conglomerado		4		10	6.04	11	6.04
1045-9	5.713300	40.668500	242	28.30	Rocas Volcánicas		Dacitas		6	g20/g21	20	0.00	21	0.00
952-6(1)	5.721000	41.806000	243	3.10	Depresiones Postorogénicas		Margocalizas		4		3	0.96	4	0.84
1045-13	5.737700	40.930100	244	211.00	Complejo Nevado-Filábride		Filitas	g26/g47	5	g26/g47	26	0.00	47	0.00
974-2(1)	5.741000	41.653000	245	6.90	Depresiones Postorogénicas		Areniscas		4		7	2.03	8	2.03
1014-4-1	5.743000	41.254000	246	8.50	Depresiones Postorogénicas		Conglomerado		4		9	5.24	10	5.24
1045-11	5.761200	40.891300	247	49.80	Depresiones Postorogénicas		Conglomerado		5		23	46.54	30	46.54
952-5(1)	5.773000	41.734000	248	3.20	Zona Subbética		Margocalizas	g3-g10/g4-g11	4		3	0.00	4	0.00
1031-1(1)	5.788000	41.061000	249	13.00	Depresiones Postorogénicas		Areniscas y margas		5		13	7.34	14	7.34
996-4(1)	5.791000	41.353000	250	9.70	Complejo Alpujárride		Esquistos		4		10	6.00	11	6.44
952-4(1)	5.791000	41.825000	251	3.70	Depresiones Postorogénicas		Margocalizas		4		3	1.56	4	1.44
996-3(1)	5.806000	41.445000	252	10.60	Depresiones Postorogénicas		Esquistos		3	g11/g12	11	0.00	12	0.00
952-3(1)	5.823000	41.789000	253	7.60	Zona Subbética		Calizas		4		9	5.46	10	5.34
A1-1031-2-1	5.828000	41.047000	254	8.30	Depresiones Postorogénicas		Margas		5		9	3.43	10	3.43
974-4-1	5.832000	41.545000	255	15.00	Complejo Alpujárride		Esquistos		4		15	11.30	16	11.74
974-4(2)	5.832000	41.545000	256	11.30	Complejo Alpujárride		Esquistos		4		11	7.60	12	8.04
974-1(1)	5.833000	41.659000	257	18.00	Complejo Alpujárride	g1-g26/g1-g46	Arcillas	g16-g20/g17-g24	4	g16/g17	16	0.00	17	0.00
1014-2(1)	5.836000	41.211000	258	8.80	Complejo Nevado-Filábride		Esquistos		3	g10/g11	10	0.00	11	0.00
1014-3(1)	5.848000	41.286000	259	5.90	Complejo Nevado-Filábride		Esquistos		6		6	1.03	7	1.03
1045-15	5.850300	40.935400	260	19.10	Depresiones Postorogénicas		Conglomerado		5		17	15.84	18	15.84
952-1(1)	5.863000	41.822800	261	7.30	Zona Subbética		Calizas		4		8	5.16	9	5.04
931-1(1)	5.868000	41.891000	262	3.60	Depresiones Postorogénicas		Conglomerado		4	g3/g4	3	0.34	4	0.34
931-1(1)	5.868000	41.891000	263	3.60	Depresiones Postorogénicas		Conglomerado		4		3	0.34	4	0.34
931-2(1)	5.875000	41.911000	264	10.50	Zona Subbética	g2-g26/g3-g42	Arcillas	g4-g11/g5-g12	4		11	5.63	12	5.63
974-3(1)	5.886000	41.563000	265	16.90	Complejo Alpujárride		Esquistos		4		16	13.20	17	13.64
1045-17	5.912400	40.944400	266	30.20	Depresiones Postorogénicas		Conglomerado		5		20	26.94	24	26.94
996-5(1)	5.915000	41.348000	267	19.70	Depresiones Postorogénicas		Conglomerado		5		17	16.44	18	16.44
1045-31	5.928100	41.128300	268	33.50	Complejo Nevado-Filábride		Calizas-Mármoles	g11-g21/g12-g27	6		21	22.65	27	22.65
1014-1(1)	5.940000	41.287000	269	10.00	Depresiones Postorogénicas		Conglomerado		4		10	6.74	11	6.74

1045-19	5.943900	40.938100	270	8.80	Rocas Volcánicas	g10-g22/g11-g28	Andesita	g10-g22/g11-g28	5	g10/g11	10	0.00	11	0.00
996-2(1)	5.948000	41.467000	271	7.60	Depresiones Postorogénicas		Mat.fluviales		3		9	2.73	10	2.73
996-1(1)	5.968000	41.403000	272	13.40	Depresiones Postorogénicas		Arcillas-limos		4		14	10.14	15	10.14
1045-21	6.017700	41.052600	273	52.00	Complejo Nevado-Filábride		Esquistos		6		24	47.13	31	47.13
1045-23	6.043800	41.169100	274	9.10	Depresiones Postorogénicas		Conglomerado		6		10	5.84	11	5.84
997-2(1)	6.060000	41.355000	275	13.00	Depresiones Postorogénicas		Conglomerado		5		13	9.74	14	9.74
1045-27	6.068200	41.289800	276	39.60	Rocas Volcánicas		Lamprobs		4	g22/g28	22	0.00	28	0.00
1045-25	6.080000	41.251900	277	25.30	Depresiones Postorogénicas		Margas		5		19	20.43	20	20.43
997-1(1)	6.083000	41.422000	278	9.60	Depresiones Postorogénicas		Margas		4		10	4.73	11	4.73
1045-29	6.099100	41.298900	279	83.70	Depresiones Postorogénicas		Esquistos		5		26	72.85	40	72.85

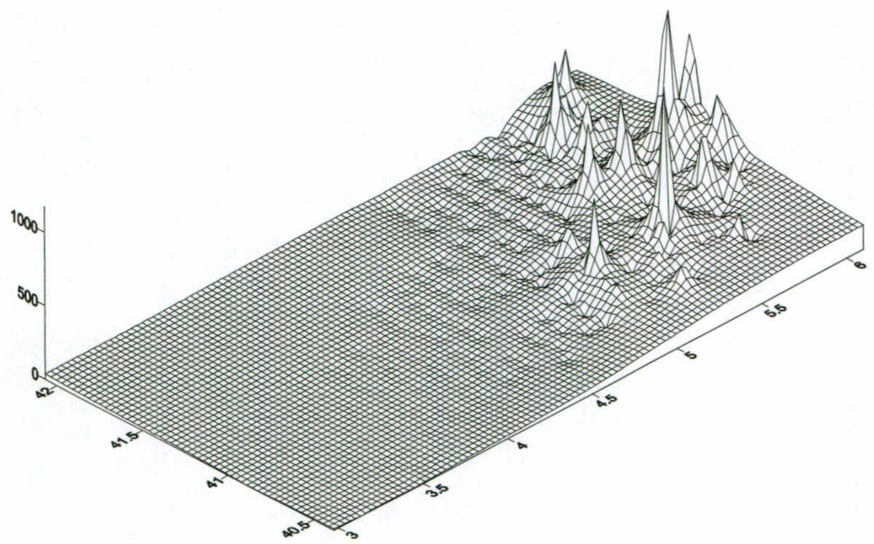


Valores de concentración

As

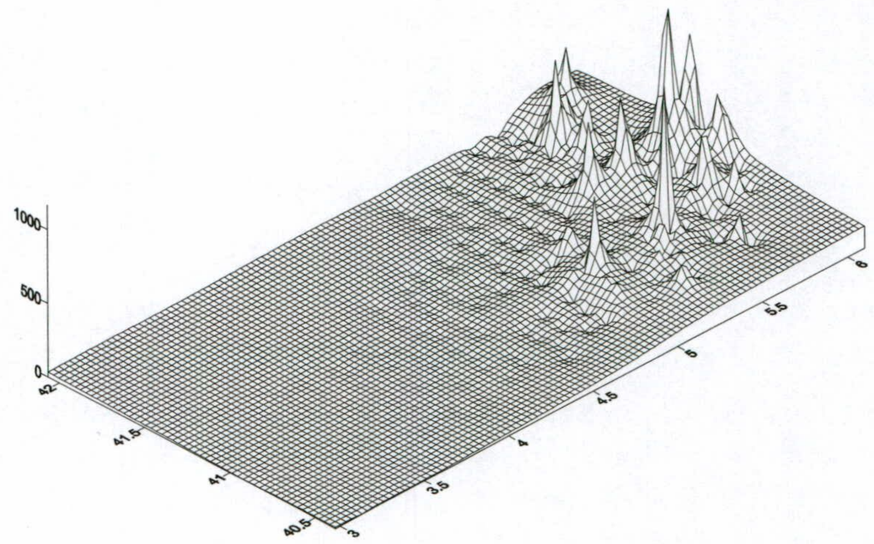


Anomalía calculada

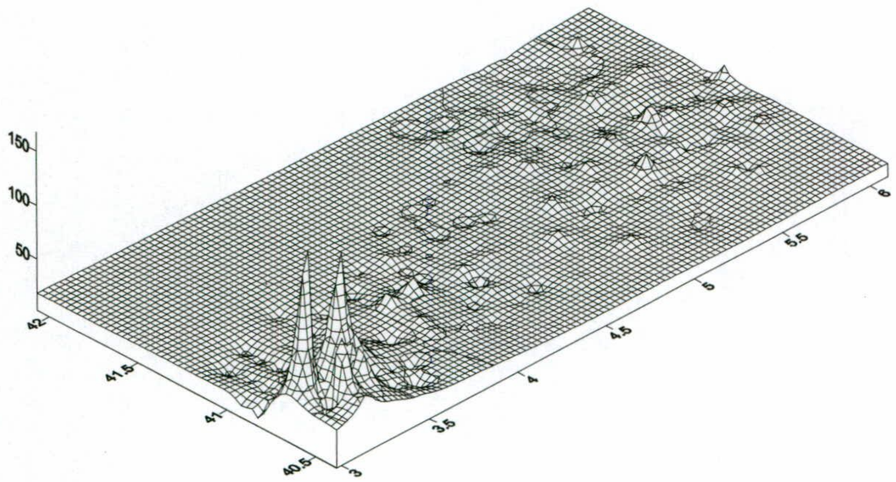


Valores de concentración

Cd

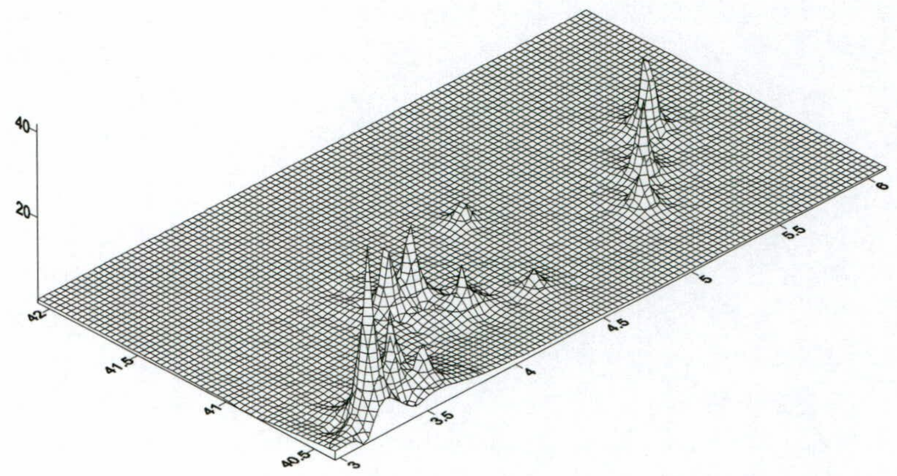


Anomalía calculada

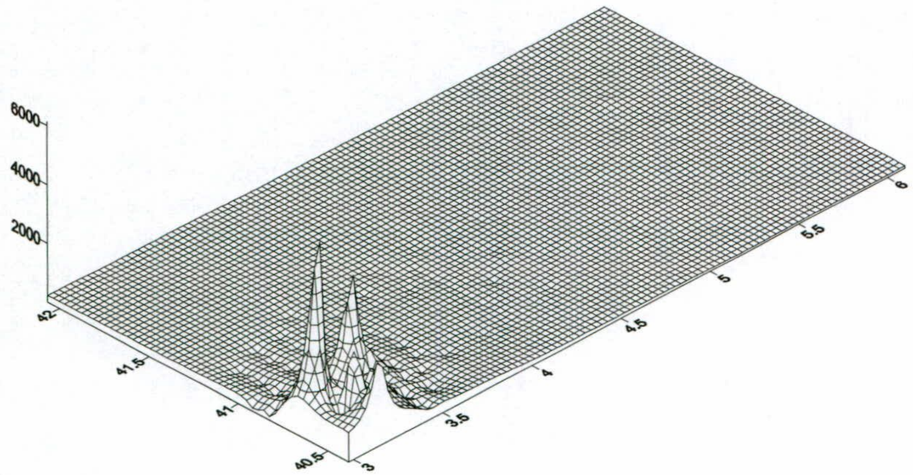


Valores de concentración

Co

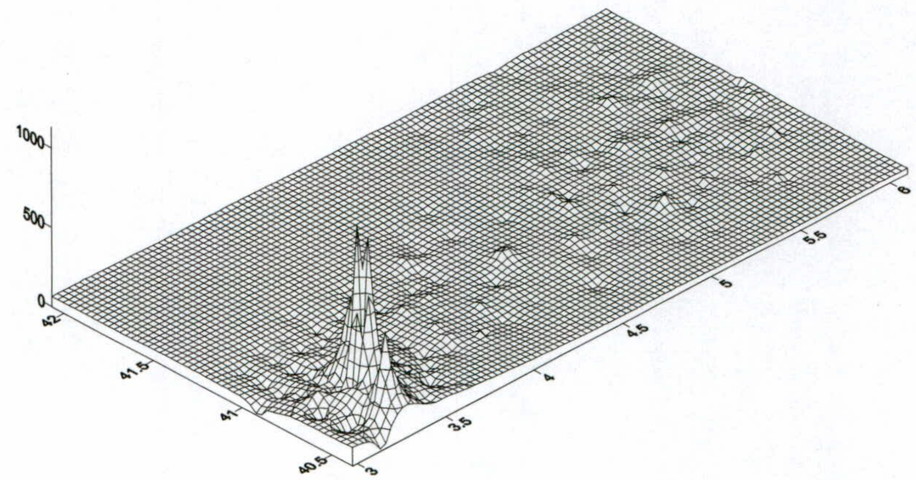


Anomalía calculada

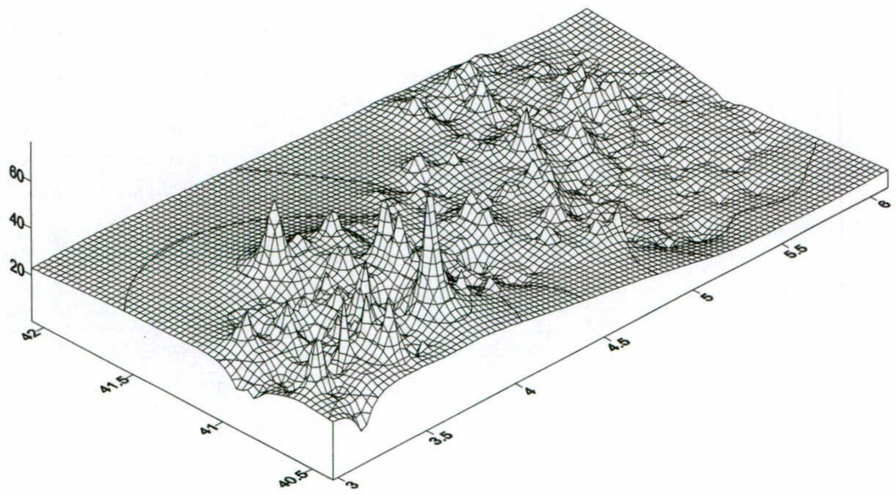


Valores de concentración

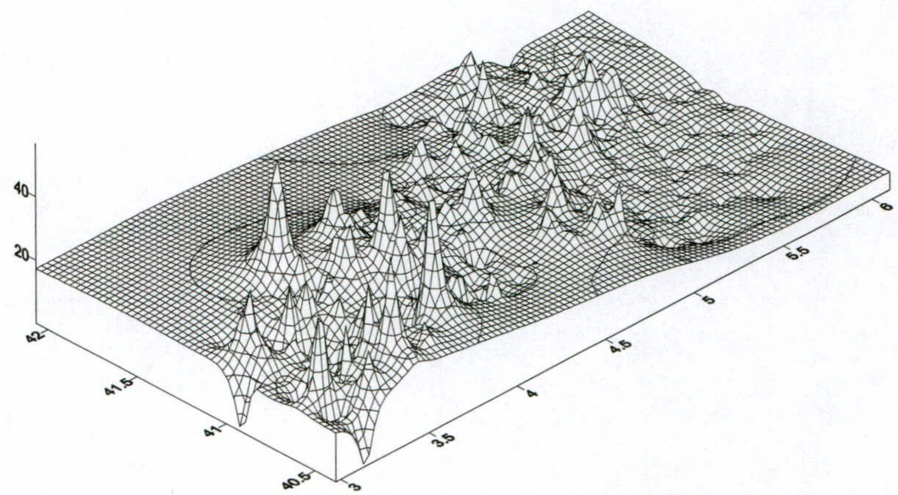
Cr



Anomalía calculada

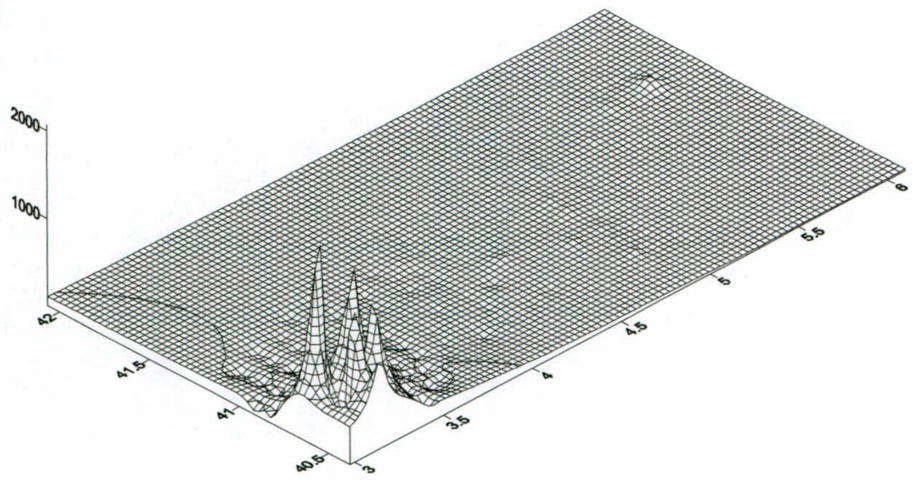


Valores de concentración



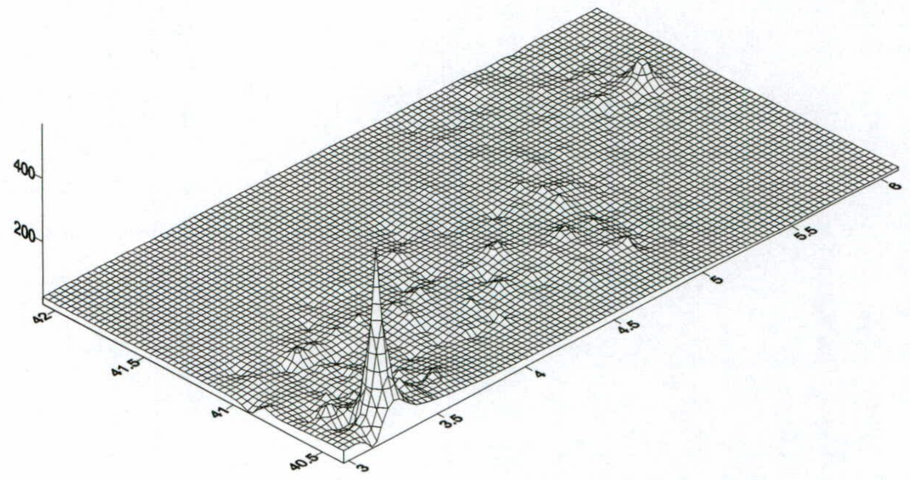
Anomalía calculada

Cu

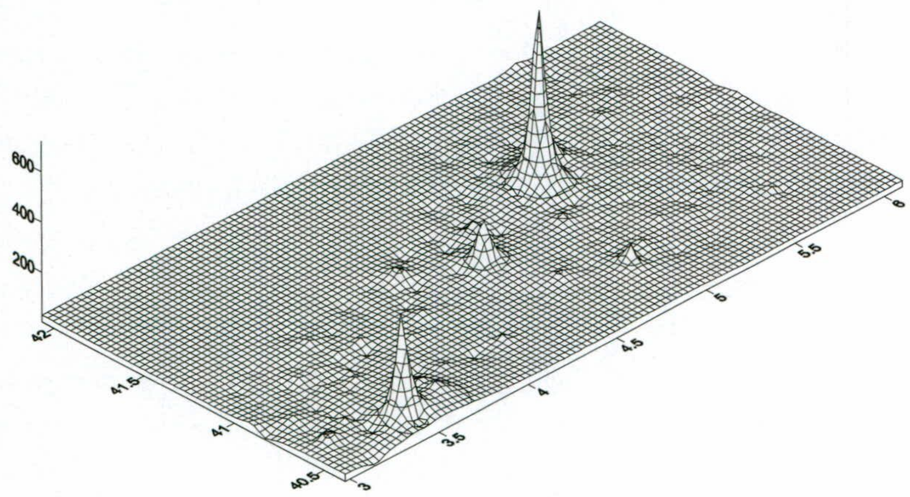


Valores de concentración

Ni

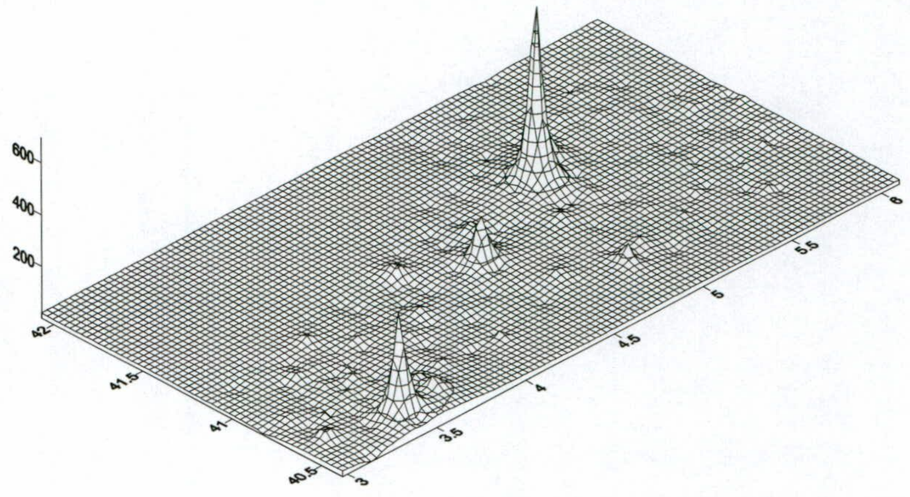


Anomalía calculada

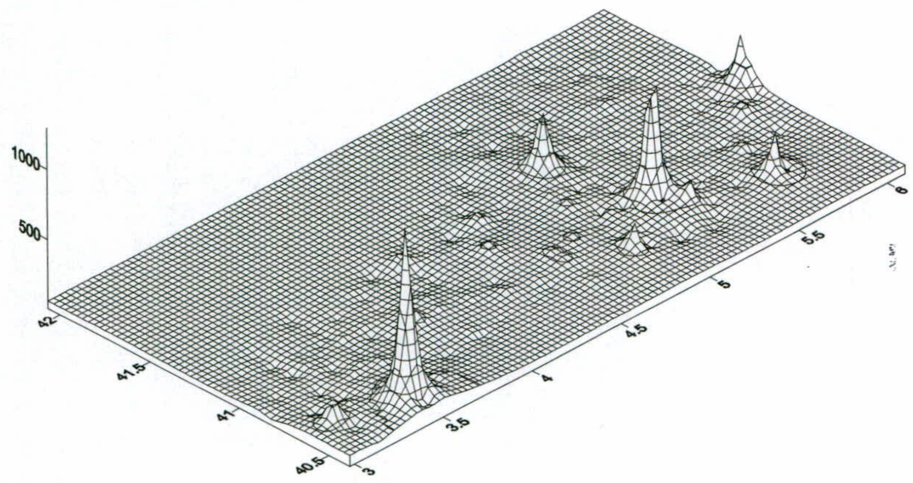


Valores de concentración

Pb

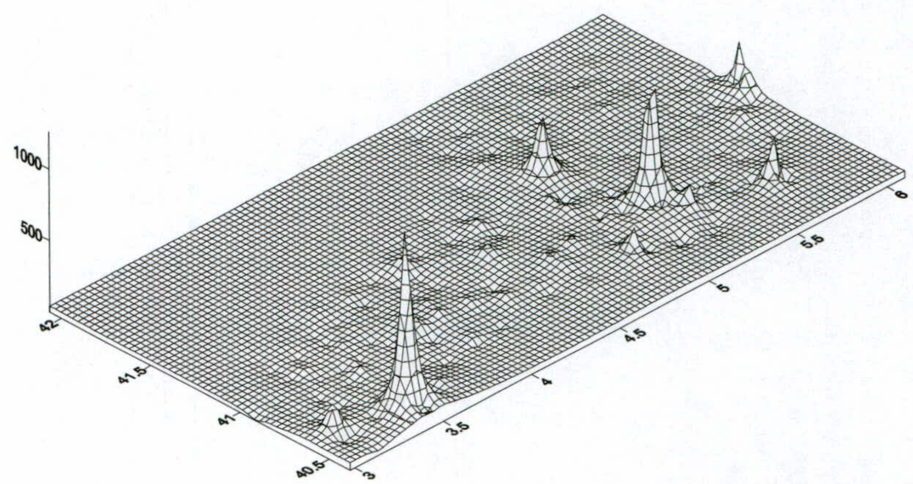


Anomalía calculada



Zn

Valores de concentración



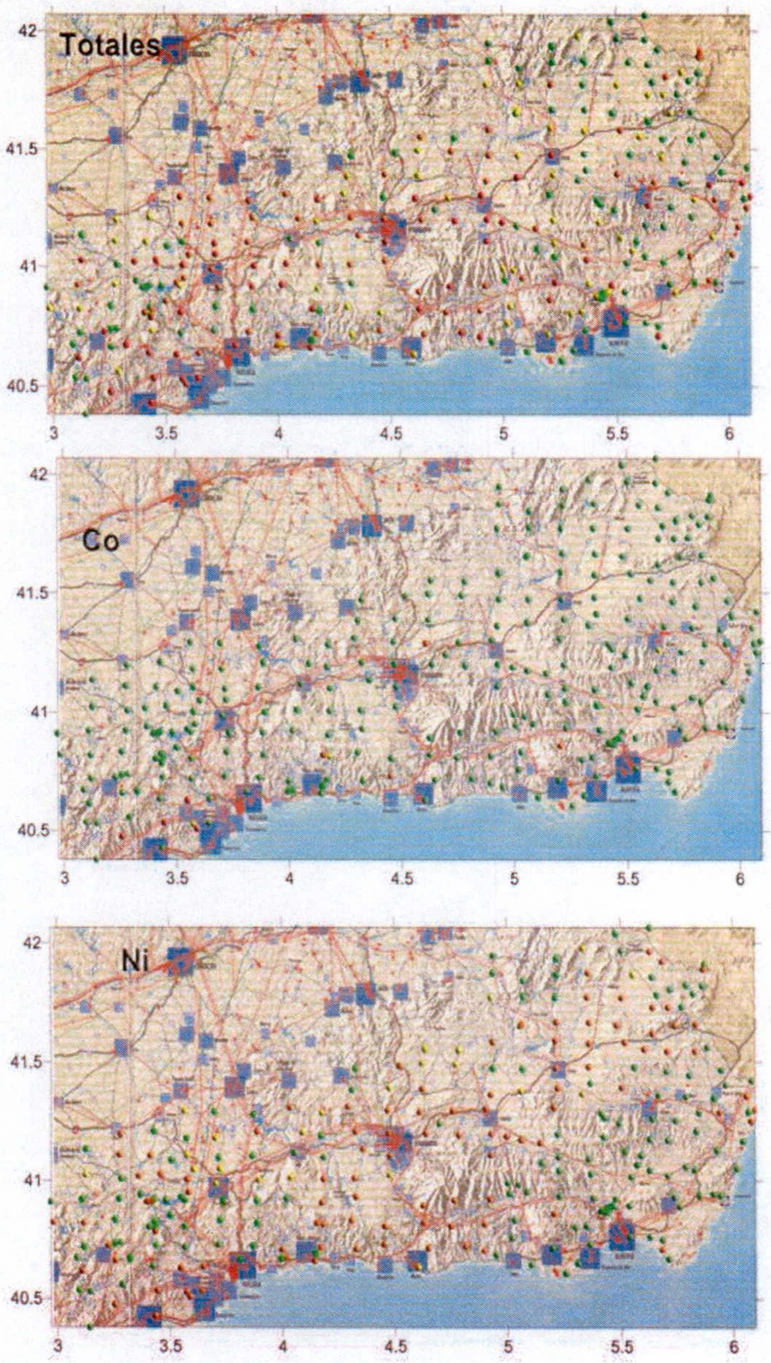
Anomalía calculada

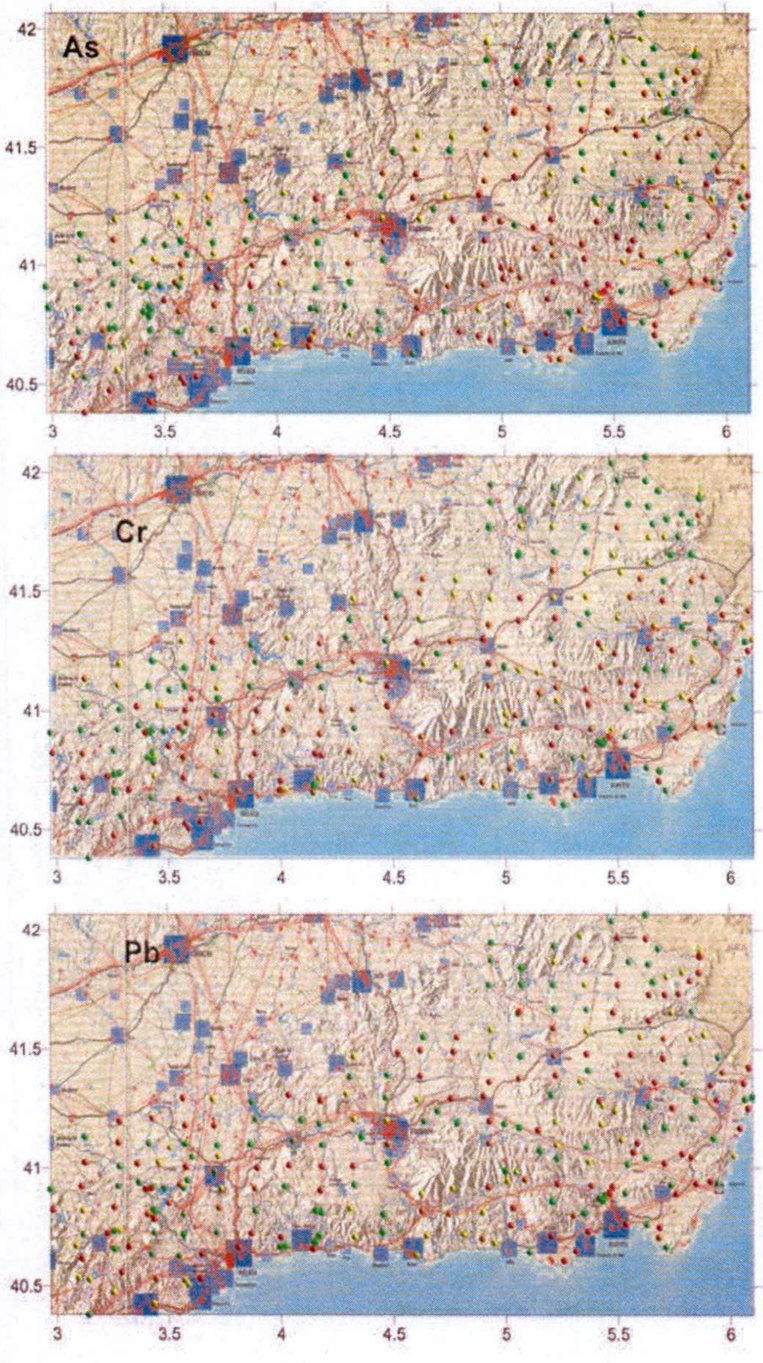
3.5. Discusión de resultados.

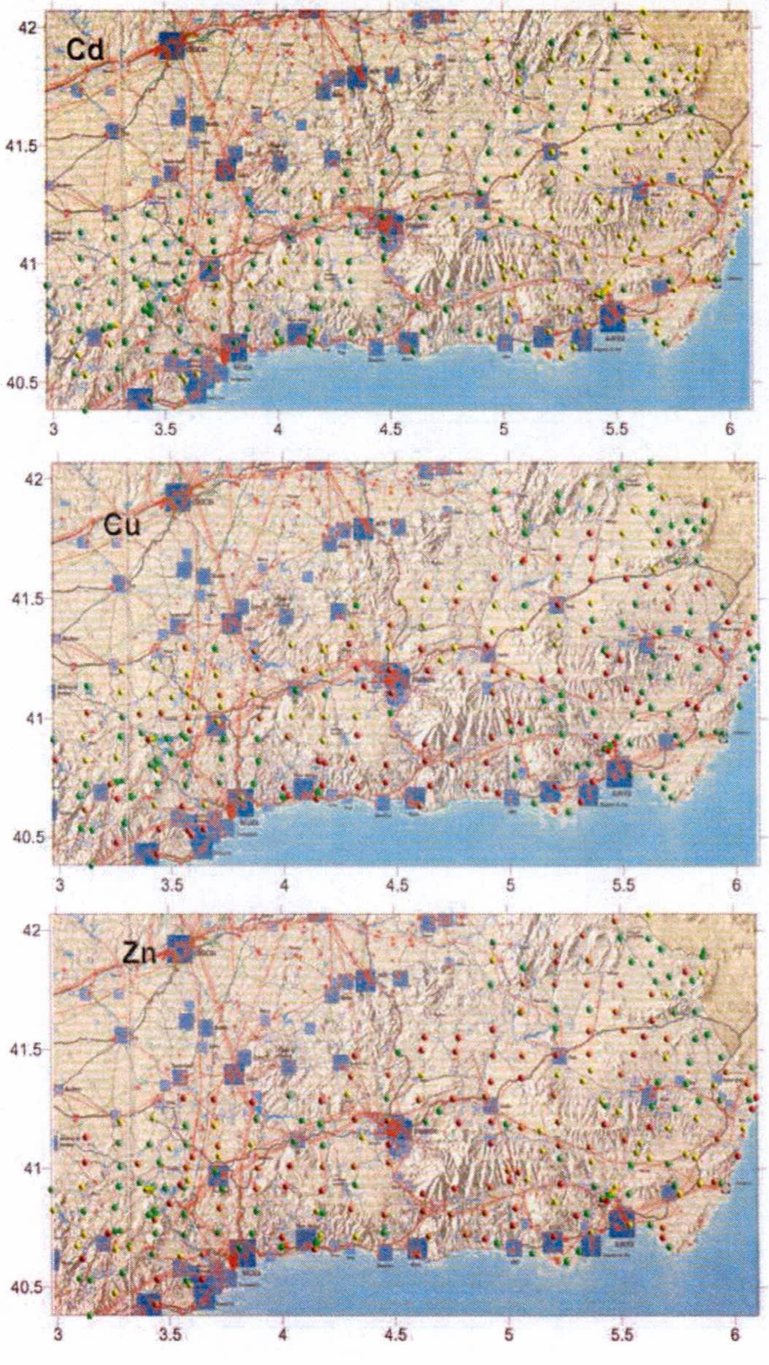
Como puede suponerse, la justificación de la bondad de la nueva metodología propuesta frente a la utilizada normalmente, en un tema como el que nos ocupa no resulta fácil, recordemos que la mayoría de los autores considera que no existe un método fiable para detectar anomalías correspondientes a puntos contaminados que no pase por la comparación con valores correspondientes a puntos contaminados con seguridad y que por tanto sobrevaloran los valores de referencia.

Para ello se ha desarrollado un índice que pretende maximizar los valores anómalos frente a los valores de fondo. Éste índice se consigue dividiendo el valor de la anomalía obtenida por el dato de concentración del metal bajo estudio, de forma que se maximiza la importancia de las anomalías en valores de concentración original bajo, obteniendo los mapas que se muestran a continuación.

Si recurrimos a métodos que se pueden calificar como poco ortodoxos y cualitativos, pero que su resultado sea coherente, obtenemos de forma cualitativa una estimación de la bondad del método, tal y como se muestra en los mapas adjuntos, en los que puede observarse la concentración de puntos potencialmente contaminados en torno a las areas industriales.


Totales
42
41.5
41
40.5
3
3.5
4
4.5
5
5.5
6


As
42
41.5
41
40.5
3
3.5
4
4.5
5
5.5
6


Cd
42
41.5
41
40.5
3
3.5
4
4.5
5
5.5
6

Co
42
41.5
41
40.5
3
3.5
4
4.5
5
5.5
6

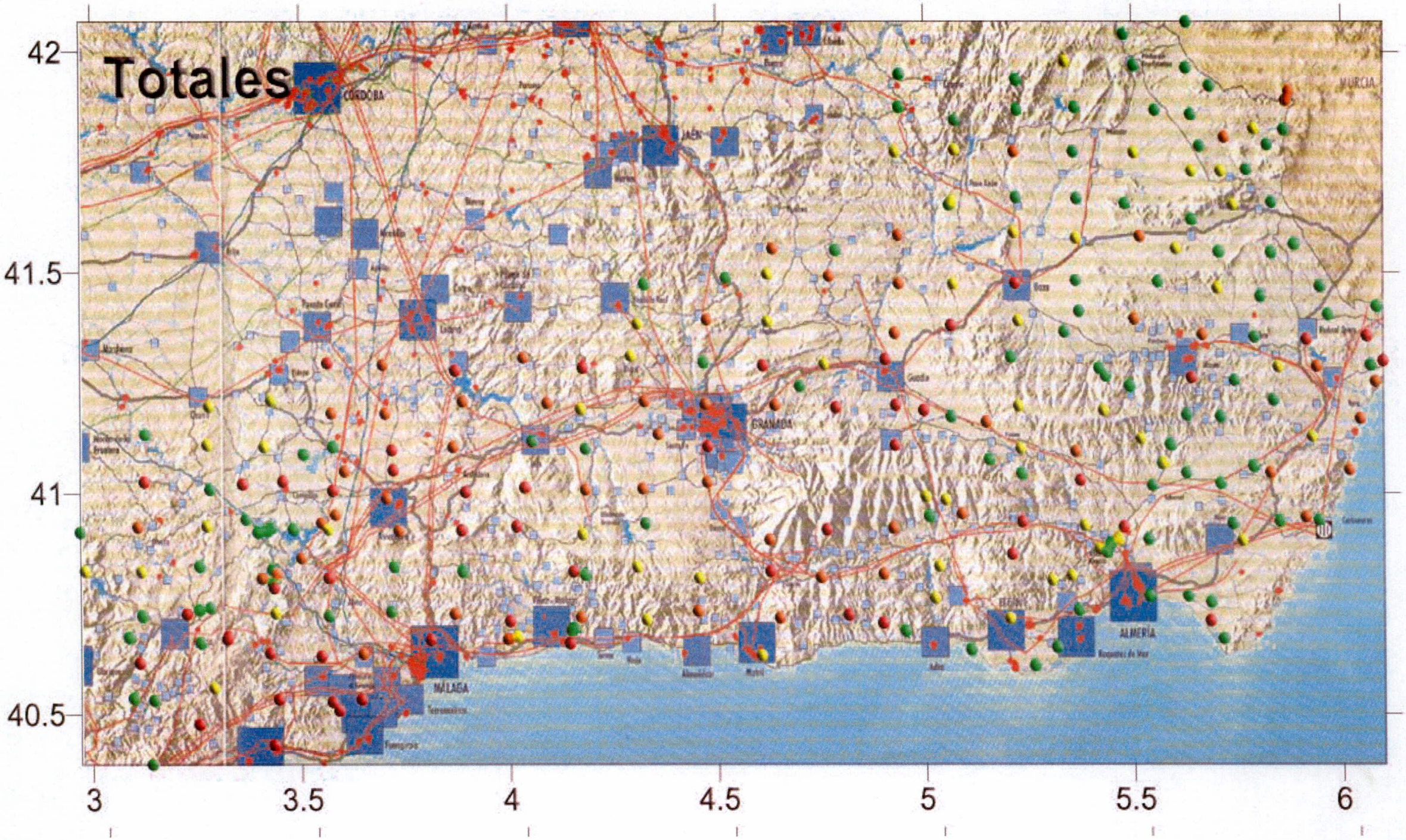
Cr
42
41.5
41
40.5
3
3.5
4
4.5
5
5.5
6

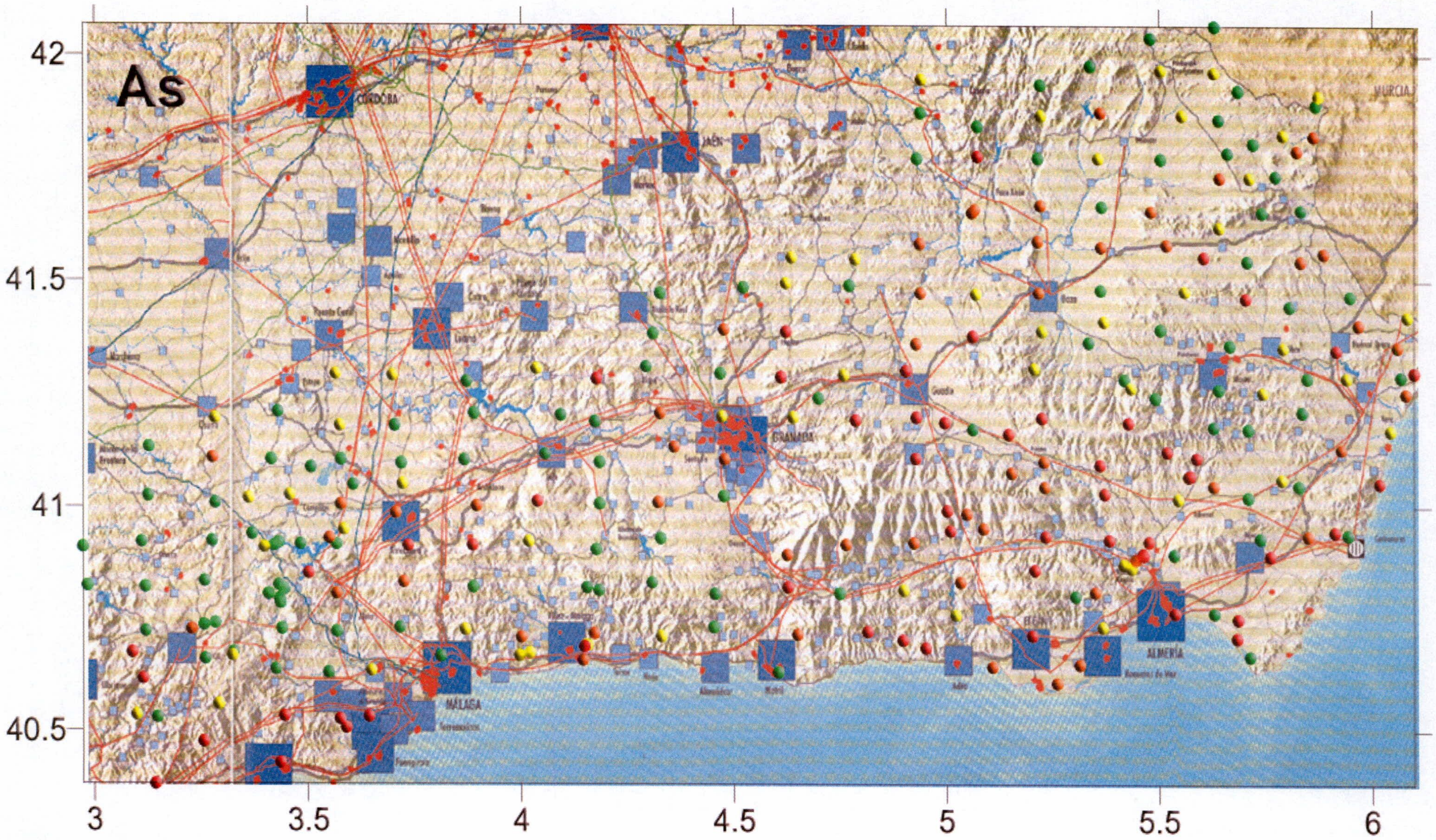
Cu
42
41.5
41
40.5
3
3.5
4
4.5
5
5.5
6

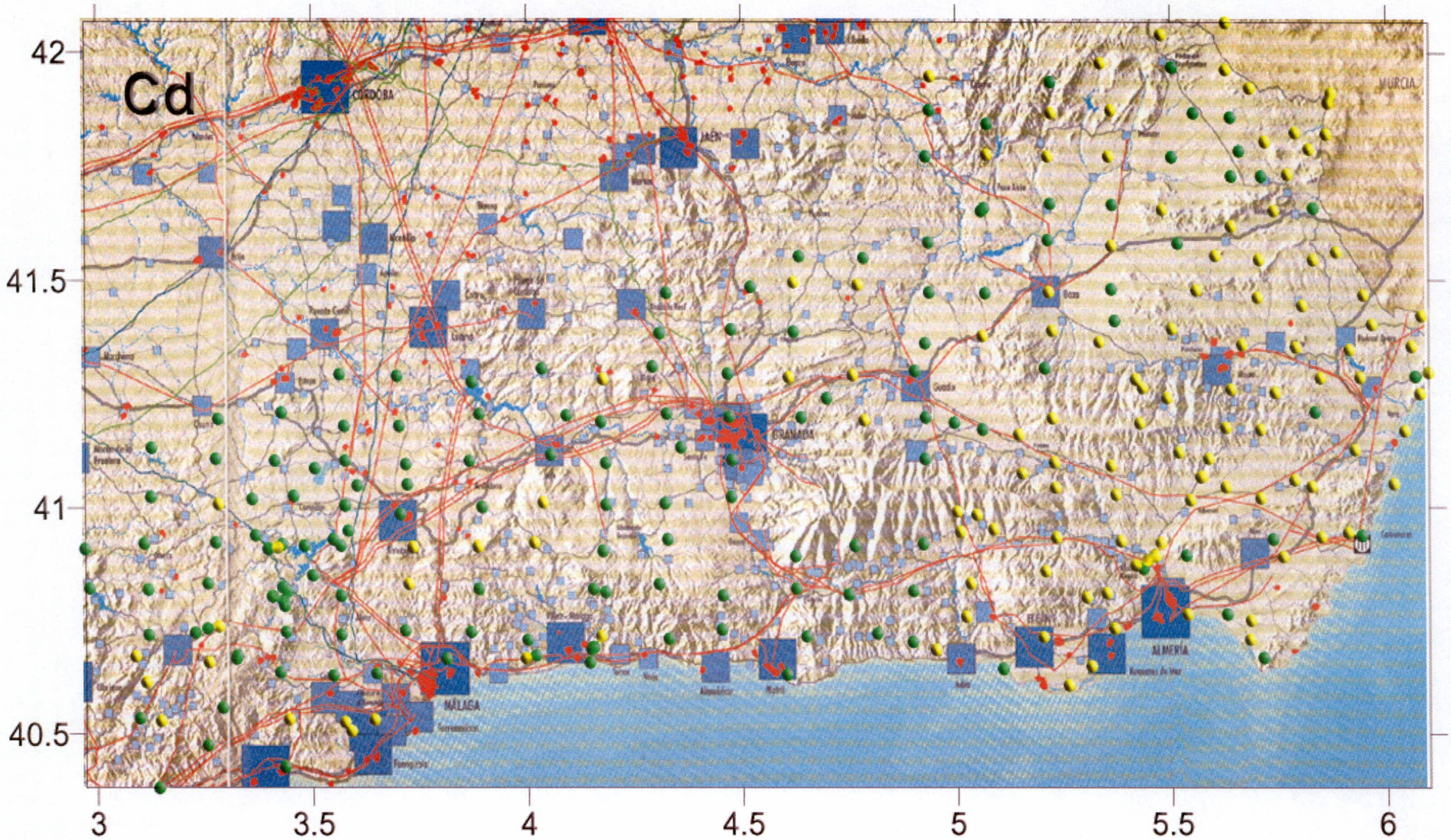
Ni
42
41.5
41
40.5
3
3.5
4
4.5
5
5.5
6

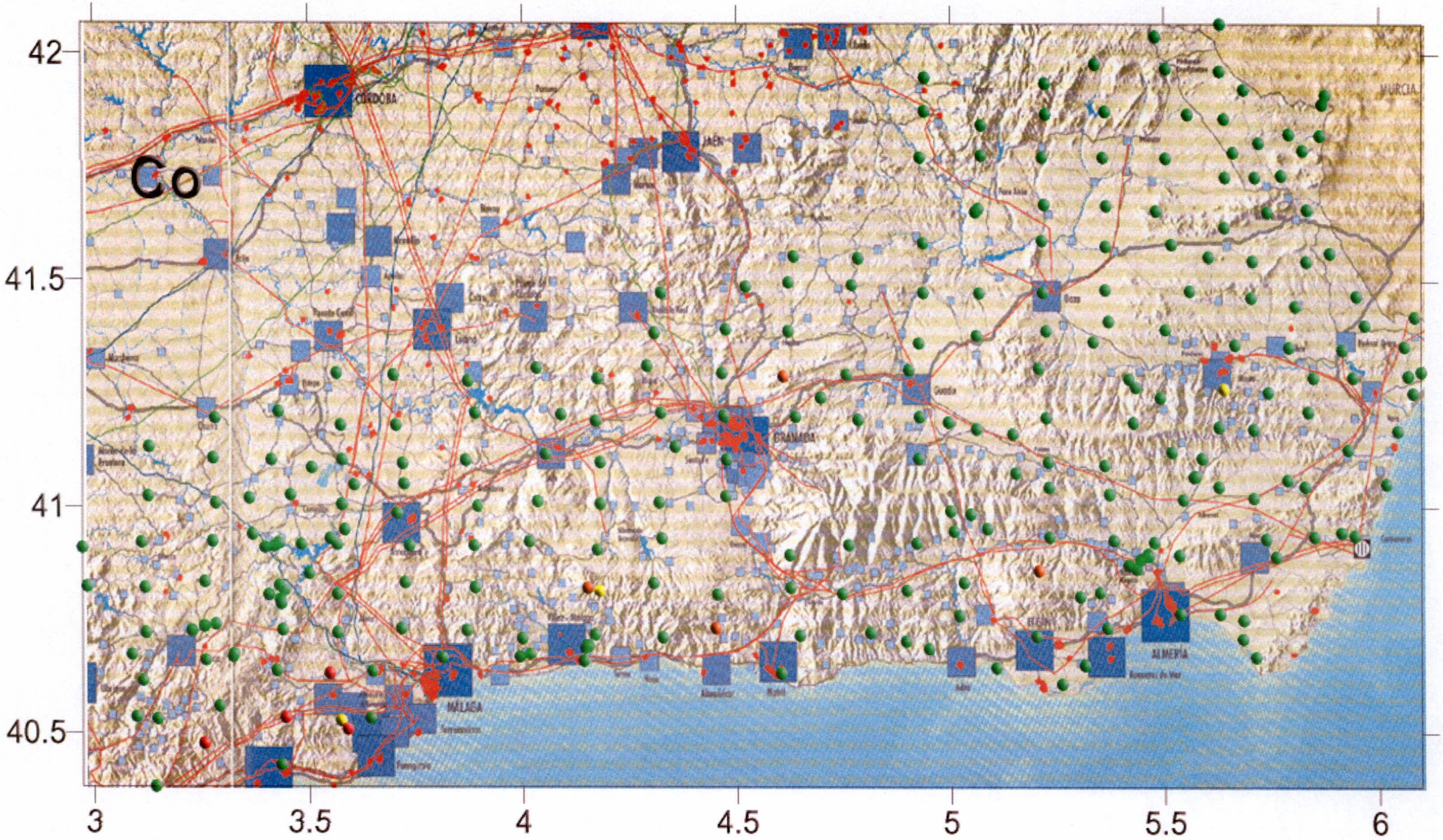
Pb
42
41.5
41
40.5
3
3.5
4
4.5
5
5.5
6

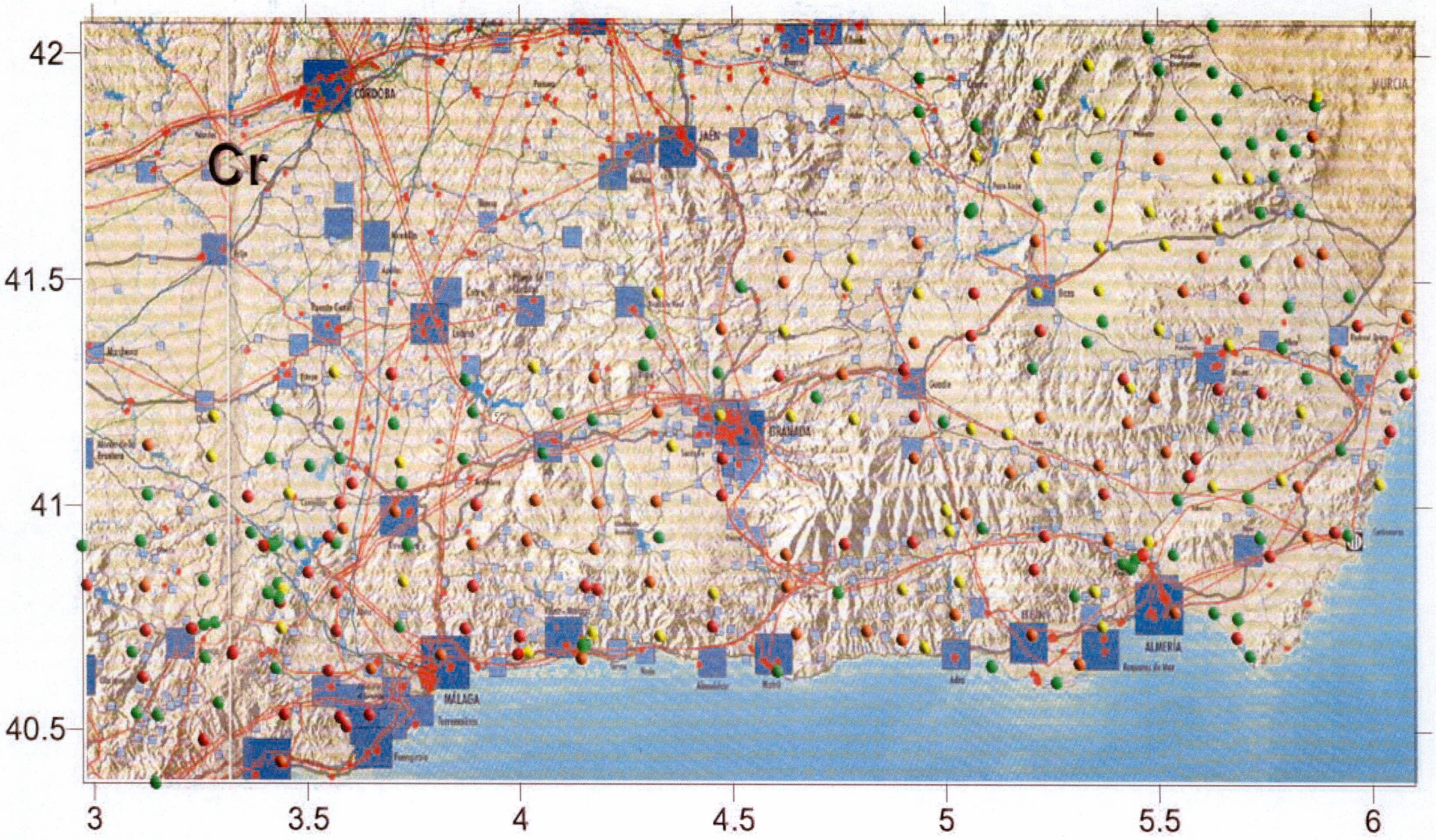
Zn
42
41.5
41
40.5
3
3.5
4
4.5
5
5.5
6

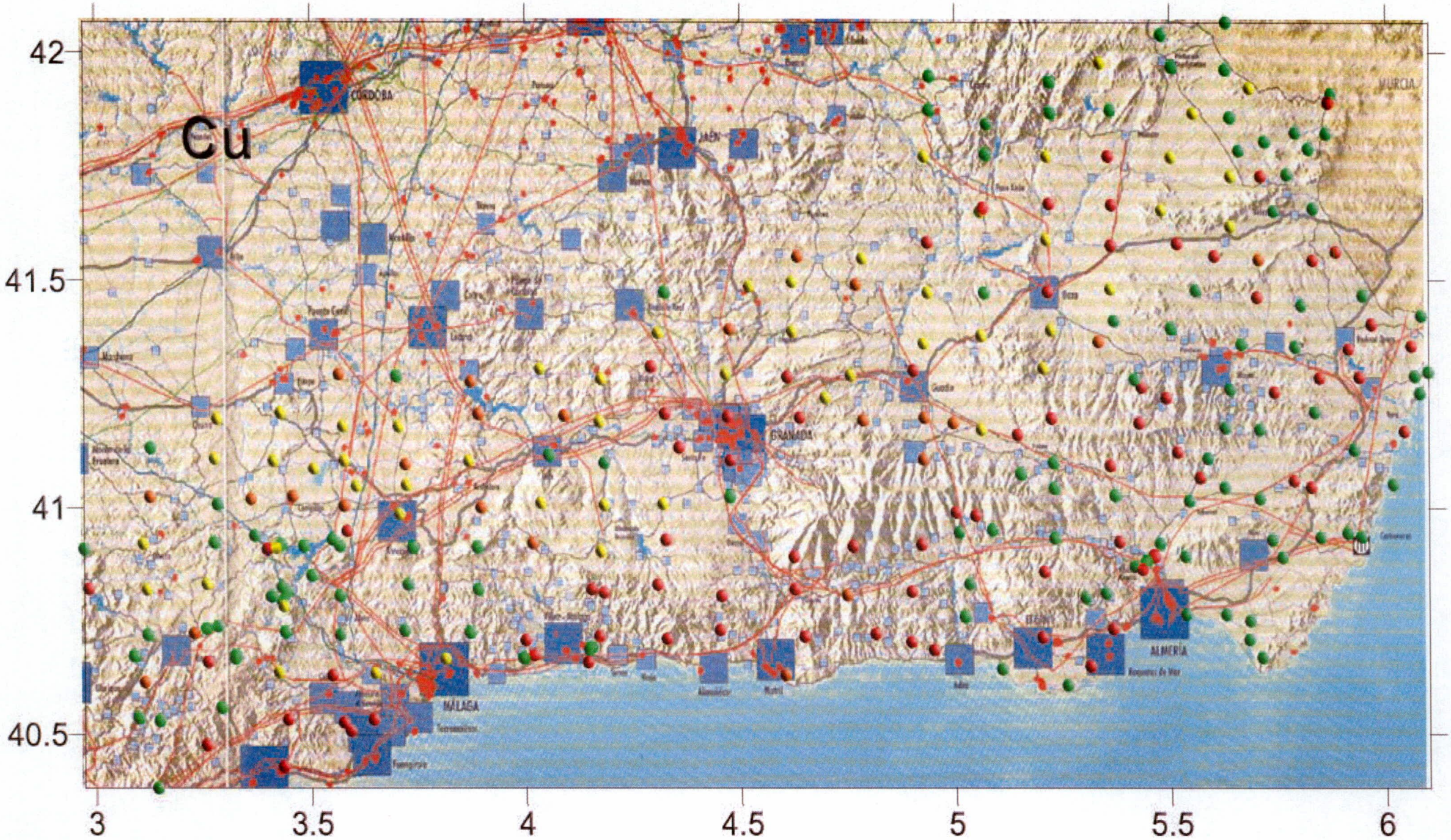

Totales
42
41.5
41
40.5
3
3.5
4
4.5
5
5.5
6
CÓRDOBA
JAÉN
GRANADA
MÁLAGA
ALMERÍA
MURCIA

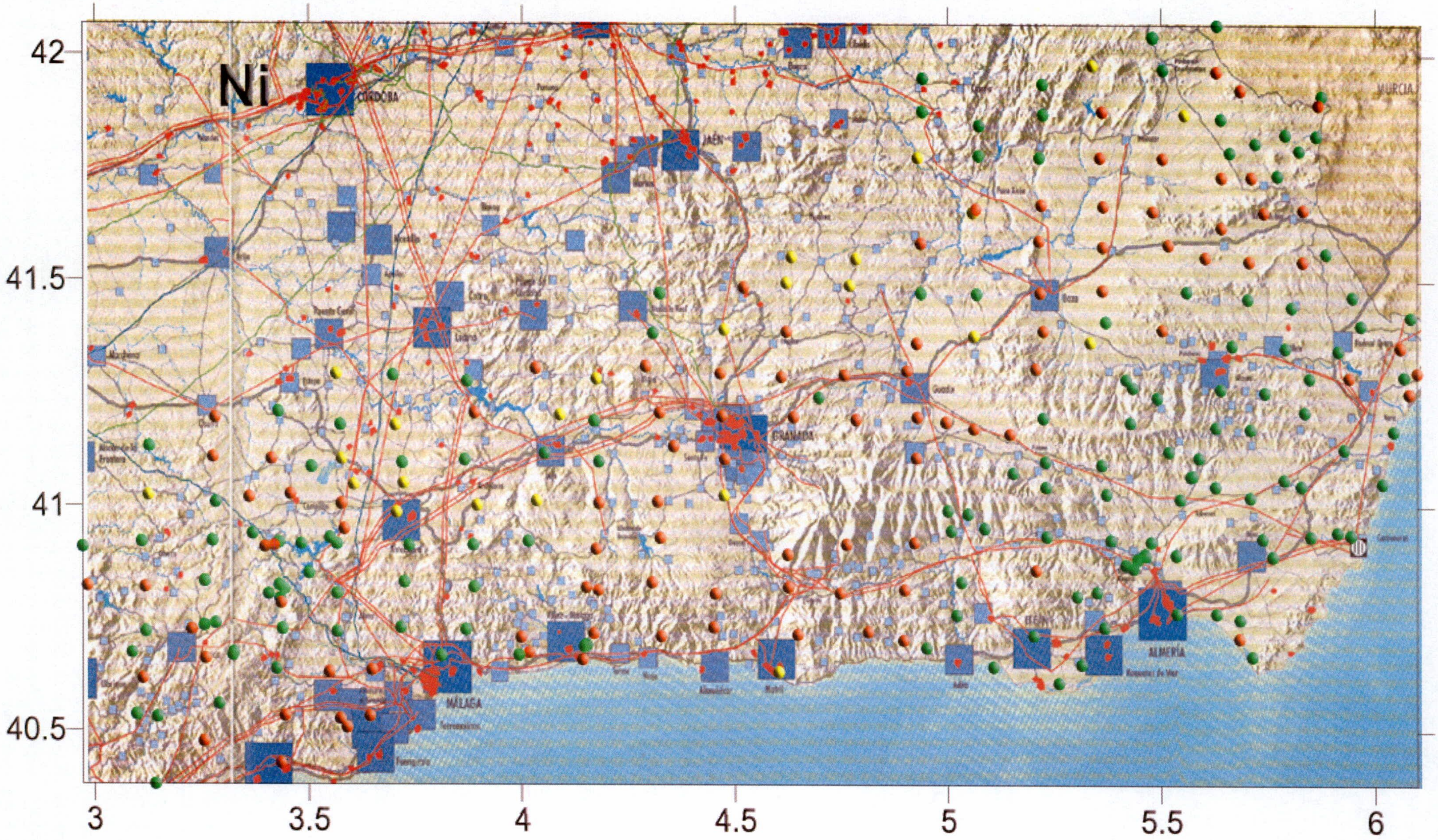

As
42
41.5
41
40.5
3
3.5
4
4.5
5
5.5
6
CÓRDOBA
JAÉN
GRANADA
MÁLAGA
ALMERÍA
MURCIA

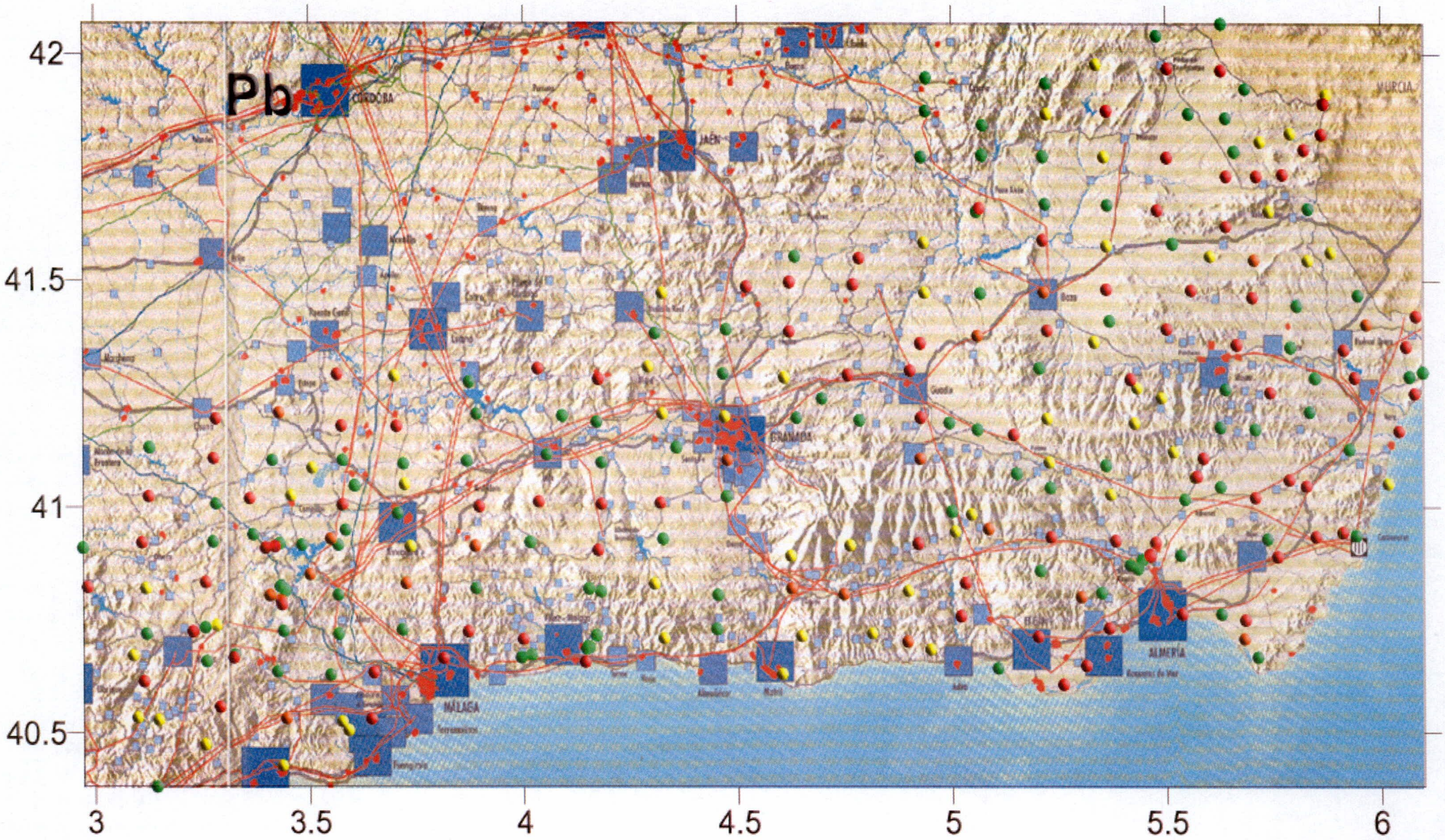

Cd
42
41.5
41
40.5
3
3.5
4
4.5
5
5.5
6
CÓRDOBA
JAÉN
GRANADA
MÁLAGA
ALMERÍA
MURCIA

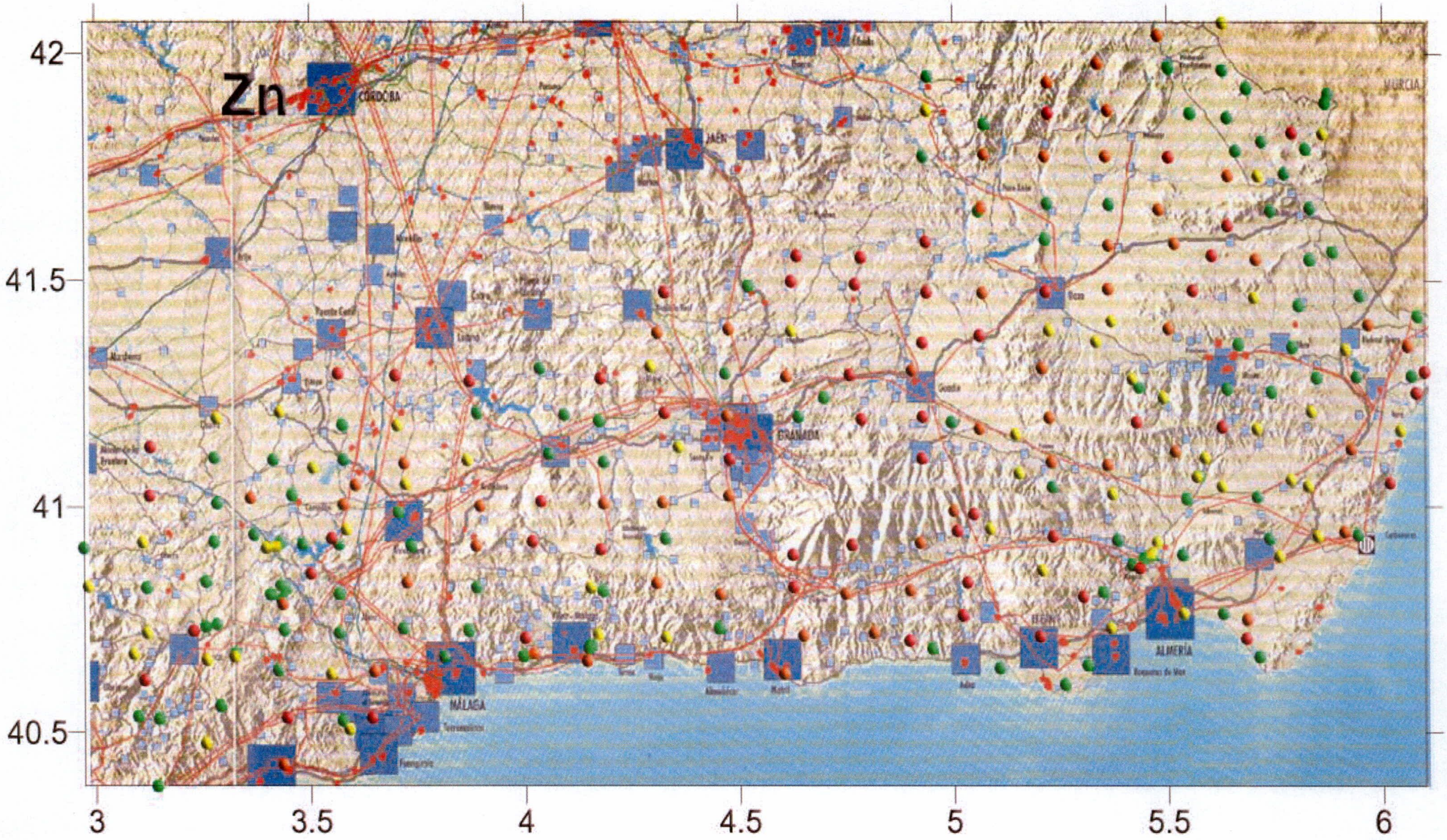

42
41.5
41
40.5
3
3.5
4
4.5
5
5.5
6
Co
CÓRDOBA
JAÉN
MURCIA
GRANADA
MÁLAGA
ALMERÍA


42
41.5
41
40.5
3
3.5
4
4.5
5
5.5
6
Cr
CÓRDOBA
JAÉN
MURCIA
GRANADA
MÁLAGA
ALMERÍA


Cu
42
41.5
41
40.5
3
3.5
4
4.5
5
5.5
6
CÓRDOBA
JAÉN
GRANADA
MÁLAGA
ALMERÍA
MURCIA


Ni
42
41.5
41
40.5
3
3.5
4
4.5
5
5.5
6
CÓRDOBA
JAÉN
GRANADA
MÁLAGA
ALMERÍA
MURCIA


Pb
42
41.5
41
40.5
3
3.5
4
4.5
5
5.5
6
CÓRDOBA
JAÉN
GRANADA
MÁLAGA
ALMERÍA
MURCIA


Zn
42
41.5
41
40.5
3
3.5
4
4.5
5
5.5
6
CÓRDOBA
JAÉN
GRANADA
MÁLAGA
ALMERÍA
MURCIA

Tal y como puede verse en los mapas anteriores, al menos desde un punto de vista cualitativo, los resultados obtenidos son mejores que los que se obtienen por métodos tradicionales.

4. CONCLUSIONES

4. CONCLUSIONES

1º.- La definición de Niveles característicos es más acorde con los datos disponibles que la de baseline, pues permite una mayor precisión, es independiente de la escala de trabajo y eventualmente se pueden hacer extrapolaciones o utilizarlos como referencia.

2º.- Los métodos de agrupación/segmentación desarrollados, en especial los de máximos relativos proporcionan unos buenos resultados, con una pérdida de información mínima.

3º.- La resta de los valores correspondientes a las agrupaciones obtenidas a los valores de concentración, da como resultado la obtención de zonas de valores anormalmente altos que según puedan ser explicados por patrones informacionales o no, se califican de contaminadas.

4º.- El método de compatibilización/integración de datos con los baseline, da mejores resultados que los utilizados tradicionalmente.

5. BIBLIOGRAFÍA

5. BIBLIOGRAPHY

Actander, E.; Solberg, H. 1999. Norway. En: Risk Assessment for Contaminated Sites in Europe. Vol 2. Policy Frameworks. Ed. Ferguson,C. y Kasamas, H. LQM Press. Nottingham NG7 2RD. UK.123-127.

Adriano, D.C.. 2001. Trace elements in terrestrial environments. II ed. Springer. New York. 865 pp.

Adriano, D.C.; Chlopecka, A. Kaplan, D.I.; Clisjters, H. and Vangronsveld, J.. 1997. Soil contamination and remediation: philosophy, science and technology. 3erd International conference on Biogeochemistry of the Trace Elements. Paris (France), May 15-19. R. Proust ed. Ed. INRA. pp. 465-504.

Aguilar, L.; Dorronsoro, C.; Galán, E.; Gómez, J.L. 1999. Criterios y estándares para declarar un suelo contaminado en Andalucía y la metodología y técnicas de toma de muestras y análisis para su investigación. Consejería de Medio Ambiente: Junta de Andalucía.

Allen, H.E.; Huang, C.P.; Bailey, G.W.; Bowers, A.R. (eds). 1995. Metal speciation and contamination of soil. Lewis Pub. Londres. UK.

Alloway BJ (ed). 1995. Heavy Metals in soils. Blackie Academic and Profesional. Londres.

Assmuth, T.; Seppanen, A. 1999. Finland. En: Risk Assessment for Contaminated Sites in Europe. Vol 2. Policy Frameworks. Ed. Ferguson,C. y Kasamas, H. LQM Press. Nottingham NG7 2RD. UK. 41-48.

Bachmann,G.; Freier, K.; Konietzka, R. 1995. Soil levels based on the german soil protection bill. Contaminated Soil´95. Kluwer Academic Pub. 711-719.

Bailey, R.A.; Clark, H.M.; Ferris, J.P.; Krause, S. and Strong, R.L..2002. Chemistry of the environment. 2ª Ed..Academic Press.

Baize D. 1997. Teneur totales en elements traces metalliques dans les sols (France). INRA editions. Paris. Francia.

Baize, D. and Stekerman, T.. 2001. Of the necessity of knowledge of the natural pedo-geochemical background contente in the evaluation of the contamination of soils by trace elements. Science of Total environment. 264: 127-139.

Barbizzi, S.; de Zorzi, P.; Belli, M.; Pati, A.; Sansone, U.; Stellato, L.; Barbina, M.; Deluisa, A.; Menegon, S. and Coletti, W.. 2004. Characterisation of a reference site for quantifying uncertainities related to soil sampling. Environmental pollution. 127: 131-135.

Behiels, G.; maes, F.; Vandermeulen, D. and Suetens, P.. 2002. Evaluation of image features and search strategies for segmentation of bone structures in radiographs using Active Shape Models. Medical Image Analist. 6: 47-62.

Bellotti, E.. 1998. Assesment of a soil quality criterion by means of a field survey. Applied Soil Ecology. 10: 51-63.

Bernard, A.M.. 1997. Effects of heavy metals in the environment on human health. Contaminated Soils. 3erd International conference on Biogeochemistry of the Trace Elements. Paris (France), May 15-19. R. Proust ed. Ed. INRA. pp. 21-34.

Bieber, A.; Francius, V.; Freier, K. 1999. Germany. En: Risk Assessment for Contaminated Sites in Europe. Vol 2. Policy Frameworks. Ed. Ferguson,C. y Kasamas, H. LQM Press. Nottingham NG7 2RD. UK. 61-76.

Bohn, H.I.; Brian M.L.; O´connor, G.A. 2001. Soil chemistry. John Wiley and sons. New York

Bowen, HIM. 1979. Environmental Chemistry of the Elements. Academic Press. Londres.

Brinkmann, R. 1998. Background concentration of metals in Florida soils. University of South Florida. State University System of Florida. Florida Center For Solid And Hazardous Waste Management. 2207 NW 13 Street, Suite D. Gainesville, FL 32609

Brogan, J.; Carty, G.; Crowe, M.; Leech, B. 1999. Ireland. En: Risk Assessment for Contaminated Sites in Europe. Vol 2. Policy Frameworks Ed. Ferguson,C. y Kasamas, H. LQM Press. Nottingham NG7 2RD. UK. 85-98.

Burt, R.; Wilson, M.A.; Keck, T.J.; Dougherty, B.D.; Strom, D.E. and Lindahl, J.A.. 2003. Trace element speciation in selected smelter-contaminated soils in Anaconda and Deer Lodge Valley, Montana, USA. Advances in Environmental Research. 8: 51-67.

Busquets, E. 1997. Elaborasio dels Criteris de Cualitat des Sols a Catalunya. Junta de Residus. Departament de Medi Ambient. Generalitat de Catalunya.

Caritat, P.; Reinman, C. Bogatyrev, I.; Chekushin, V.; Finne, T.E.; Halleraker, Jo H.; Kashulina, G.; Niskavaara, H.; Pavlov, V. and Äyräs, M., 2001. Regional distribution of Al, B, Ba, Ca, K, La, Mg, Mn, Na, P, Rb, Si, Sr, Th, U and Y in terrestrial moss within a 188,000km2 area on the central Barents region: influence of geology, seaspray and human activity. Applied Geochemistry (16): 137-159.

Chen, M.; Ma, L.Q.; Harris, W.G. 1999. Baseline concentration of 15 Trace elements in Florida surface soils. J. Environ. Qual. 28: 1173-1181.

Chiang, L.H.; Pell, R.J. and Seasholtz, M.B.. 2003. Exploring process data with the use of robust outlier detection algorithms. Journal of Process Control. 13: 437-449.

CICDXI.XG. (1997). Estudio sobre a actualización do Inventario de Solos Contaminados, Xeraquización e Desenvolvemento dunha Lexislación para prtección de Solo en Galicia. Consellería de Industria e Comercio.. Dirección Weral de Industria. Xunta de Galixia.

Cocker, M.D. 1999. Geochemical mapping in Georgia, USA: a tool for environmental studies, geologic mapping and mineral exploration. Journal of Geochemical Exploration 67.345–360.

Consejo Europeo (2003). Medio Ambiente. Sesión nº 2517 (presse 165). Bruselas, 13 de junio de 2003. 18pp.

Covelo García, A. and Prego, R.. 2003. Heavy metal sedimentary record in a Galician Ria (NW Spain): background values and recent contamination. Marine Pollution Bulletin. 46: 1235-1262.

Cowell, F.A. 2003. Theil, Inequality and the structure of income distribution

Crock, JG.; Severson, RC.; Gough, LP. 1992. Determining base-lines and variability of elements in plants and soils near the kenai national wildlife refuge, Alaska. Water Air and Soil Pollution. 63 (3-4): 253-271.

Darnley, A.G., 1997. A global geochemical reference network: the foundation for geochemical baselines. J. of Geochemical Exploration, (60): 1-5.

Darnley, A.G.; Björklund, A.; Bølviken, B.; Gustavsson, N.; Koval, P.V.: Plant, J.A.; Steenfelt, A.; Tauchid, M.; Xie Xuejing; Garrett, R.G.; Hall, G.E.M., 1995. A Global Geochemical Database for Environmental and Resource Management: Recommendations for International Geochemical Mapping. Final report of IGCP Project 259. 2nd revised edition. Paris: UNESCO. 122 pp.

De Vivo, B.; Boni, M.; Marcello, A.; Di Bonito, M.; Russo, A., 1998. Baseline geochemical mapping of Sardinia (Italy). J. Geochem. Explor. 60, 77–90.

Deer, P.J. and Eklund, P..2003. A study of parameter values for a Mahalanobis distance fuzzy classifier. Fuzzy Sets an Systems. 137: 191-213.

Denneman, C. 1999. The Netherlands. En: Risk Assessment for Contaminated Sites in Europe. Vol 2. Policy Frameworks. Ed. Ferguson,C. y Kasamas, H. LQM Press. Nottingham NG7 2RD. UK. 107-121.

Diario Oficial de las Comunidades Europeas, 2002. Decisión Nº 1600/2002/CE del Parlamento Europeo y del Consejo, por la que se establece el Sexto Programa de Acción Comunitario en Materia de Medio Ambiente. 19.9.2002. L 242. 15pp.

Dirección General de Medio Ambiente, 2002. Guía Informativa. Dirección General de Medio Ambiente. Comisón Europea. Rue de la Loi, 200. B-1049 Bruxelles.

Edelgaard, I.; Dahlstrom, K. 1999. Denmark. En: Risk Assessment for Contaminated Sites in Europe. Vol 2. Policy Frameworks. Ed. Ferguson,C. y Kasamas, H. LQM Press. Nottingham NG7 2RD. UK. 29-39.

EEA, 2003. Europe's Environment: the thirth assessment. Environmental assessment Report nº 10.

Elberling, B.; Knudsen, K.L.; Kristensen, P.H. and Asmund, G.. 2003. Applying foraminiferal stratigraphy as a biomarker for heavy metal contamination and mining impact in a fiord in West Greenland. Marine Environmental research. 55: 235-256.

Environment Agency. 2002a. Soil Guideline Values for Arsenic Contamination. Environment Agency, Rio House, Waterside Drive, Aztec West, Almondsbury, BRISTOL, BS32 4UD. Website: www.environment-agency.gov.uk

Environment Agency. 2002b. Soil Guideline Values for Cadmiun Contamination. Environment Agency, Rio House, Waterside Drive, Aztec West, Almondsbury, BRISTOL, BS32 4UD. Website: www.environment-agency.gov.uk

Environment Agency. 2002c. Soil Guideline Values for Chromiun Contamination. Environment Agency, Rio House, Waterside Drive, Aztec West, Almondsbury, BRISTOL, BS32 4UD. Website: www.environment-agency.gov.uk

Environment Agency. 2002d. Soil Guideline Values for Mercury Contamination. Environment Agency, Rio House, Waterside Drive, Aztec West, Almondsbury, BRISTOL, BS32 4UD. Website: www.environment-agency.gov.uk

Environment Agency. 2002e. Soil Guideline Values for Nickel Contamination. Environment

Agency, Rio House, Waterside Drive, Aztec West, Almondsbury, BRISTOL, BS32 4UD. Website: www.environment-agency.gov.uk

Environment Agency. 2002f. Soil Guideline Values for Lead Contamination. Environment Agency, Rio House, Waterside Drive, Aztec West, Almondsbury, BRISTOL, BS32 4UD. Website: www.environment-agency.gov.uk

Environment Agency. 2002g. Soil Guideline Values for Selenium Contamination. Environment Agency, Rio House, Waterside Drive, Aztec West, Almondsbury, BRISTOL, BS32 4UD. Website: www.environment-agency.gov.uk

EPA. 2003. Soil Screening Guidance: Technical Background Document Table of Contents. Environmental Protection Agency. USA. Website: http://www.epa.gov/superfund/resources/soil/toc.htm

ETC/ACC. European Topic Centre on Air and Climate Change. (2001). Good practice Guidance for CLRTAP Emisión Inventories. UNECE Corinair Guidebook on Emisión Inventories. 42pp.

European Environment Agency. (2000). Questions to be answered by a state of the environment report. The first list. Technical Report nº 47. 116pp.

European Environment Agency. (2002). Assesment of data needs and data availability for development of indicators on soil contamination. Technical report nº 81. 79pp.

Fachinelli, A.; Sacchi, E. and Mallen, L.. 2001. Multivariate statistical and GIS-based approach to identify heavy metal sources in soil. Environmental pollution. 114: 313-324.

Farber, O. and Kadmon, R.. 2003. Assesment of alternative approaches for bioclimatic modeling with especial emphasis on the Mahalanobis distance. Ecological Modeling. 160: 115-130.

Ferguson, C.C. 1999. Assessing Risk from Contaminated Sites: Policy and Practice in 16 European Countries. Land Contaminated and Reclamation. 7(2). 33-54.

Fernandez Pierna, Wahl, F., de Noord, O.E. and Massart, D.L. 2002. Methods for outlayer detection in prediction. Chemometrics and Inteligent Laboratory Systems 63. 27-39.

Ferreira, A.; Inácio, M.M.; Morgado, P.; Batista, M.J.; Ferreira, L.; Pereira, V. and Pinto, M.S., 2001. Low density geochemical mapping in Portugal. Applied Geochemistry (16): 1323-1333.

Forstner, U. 1995. Land contamination by metals: global scope and magnitude of problem. En: Allen, H.E.; Huang, C.P.; Bailey, G.W.; Bowers, A.R. (eds). Metal speciation and contamination of soil. Lewis Pub. Londres. UK. 1-33.

Galant, J.C. and Hutchinson, M.F.. 1997. Scale dependence in terrain analysis. Mathematics and Computers in Simulation. 43: 313-321.

Gallego, J.L.R.; Ordoñez, A. and Loredo, J.. 2002. Investigation of trace element sources from an industrializated area (Avilés, northern Spain) using multivariate statistical methods. Environment International. 27: 589-596.

Giles, E.A. 2002. Calculating a standard error for the Gini coefficient: some further results. Working Paper EWP0202. University of Victoria.

Goovaerts, P.. 2000. Estimation or simulation of soil properties?. An optimization problem with conflicting criteria. Geoderma. 97: 165-186.

Gordeev, V.V.; Rachold, V. and Vlasova, I.E., 2004. Geochemical behaviour of major and trace elements in suspended particulate material of the Irtysh river, the main tributary of the Ob river, Siberia. Applied Geochemistry. *in press*.

Gotoh, S.; Udoguchi, A. 1993. Japan´s policies on soil enviroment protection – History and presents status. En: Contaminated Soil´93. Ed. Arendt, F.; Annokkée, G.J.; Bosman, R; Van der. Brink, W.J.; Kluwer Academic Pub. 3-10.

Gregorauskiene V.; Kadunas, V., 1997. Experience and goals of geochemical mapping for environmental protection in Lithuania. J. Geochem. Explor. 60, 67-76.

Gzyl, J. 1999. Soil protection in Central Eastern Europe. J. Geoch. Expl. 66. 333-337.

Herbert, S. 1999. United Kingdom. En: Risk Assessment for Contaminated Sites in Europe. Vol 2. Policy Frameworks. Ed. Ferguson,C. y Kasamas, H. LQM Press. Nottingham NG7 2RD. UK. 165-176.

Huisman, D.J.; Vermeulen, F.J.H.; Baker, J.; Veldkamp, A.; Kroonemberg, S.B. and Klaver, G. Th., 1997. A geological interpretation of heavy metal concentrations in soils and sediments in the southern Netherlands. Journal of Geochemical Exploration (59): 163-174.

IHOBE, 1993. Investigación de la Contaminación del Suelo. Vol 8. Calidad del suelo. Valores indicativos de evaluación. Departamento de Urbanismo, Vivienda y Medio Ambiente. Gobierno Vasco.

Isaakidis, A.; Boura, F.; Liakopoulus, A. 1999. Greece. En: Risk Assessment for Contaminated Sites in Europe. Vol 2. Policy Frameworks. Ed. Ferguson,C. y Kasamas, H. LQM Press. Nottingham NG7 2RD. UK. 77-84.

Ji, S.M.; Zang, L.B.; Yuan, J.L.; Wan, Y.H.; Zhang, X.; Zhang, L. and Bao, G.J. 2002. Method of monitoring wearing and breakage states of cutting tools based on Mahalanobis distance features. J. of Materials Procesing Technology. 129. 114-117.

Juang, K.W.; Chen, Y.S. and Lee, D.Y.. 2004. Using secuential indicator simulation to assess the uncertainly of delineating heavy-metal contaminated soils. Environmental Pollution. 127: 229-238.

Kabata Pendias, A. 1995. Agricultural Problems related to excesive trace metals contens of soils. En: Salomons, W.; Forstner, U; Mader, P. 1995. Haevy metals. Problems and solutions. Springer. New York. 3-18.

Kabata Pendias, A. 2001. Trace elements in soils and plants. CRC Press. New York. USA.

Kawamura, K.; Gotoh, S. 1995. Japan´s national soil quality standards as expanded in 1994. Contaminated Soil´95. Kluwer Academic Pub.735-738.

Kemsley, E.K.. 2001. A hibrid classification method: discrete canonical variate analysis using a genetic algorithm. Chemometrics and Inteligent Laboratory Systems. 55: 39-51.

Kooistra, L.; Salas, E.A.L.; Clevers, J.G.P.W.; Wehrens, R.; Leuven, R.S.E.W.; Nienhuis, P.H. and Buydens, L.M.C.. 2004. Exploring field vegetation reflectance as an indicator of soil contamination in river food plains. Environmental Pollution. 127: 281-290.

Lasaga, A.C.. 1984. Chemical kinetics of water-rocks interactions. Journal of Geophisical Research, 89: 4009-4025.

Leon (de), A.R. and Carriere, K.C.. (2004). A generalized Mahalanobis distance for mixed data. Multivariate Analysis. *in press*.

Lintinen, P.; Savolainen, H.; Jarva, J. 2003. Suggested new guideline values for Cu, Cr, Ni an Zn and comparison with concentrations in soil parent material in Finland. CONSOIL´03. Edicción en CD.

Lis, J.; Pasieczna, A.; Strzelecki, R.; Wolwkowicz, S.; Lewandowski, P. 1997. Geochemical and radiactivity mapping in Poland. J. Geoch. Explor. 60, 39-53.

Matschullat, J.; Ottenstein, R. and Reimann, C., 2000. Geochemical background – can we calculate it?. Environmental Geology 39(9) 990-1000.

McGrath, S.P.; Loveland,P.J. 1992. The soil Geochemical Atlas of England and Wales. Blackie Academic and Professional. Gloasgow.

McNeill, J.R., 2001. Something New Under the Sun. An Environmental History of the Twentieth-Century World. W. W. Norton.

Miguel, de E.; Callaba, A.; Arranz, J.C.; Cala, V.; Chacón, E.; Gallego, E.; Alberruche, E.; Alonso, C.; Fernández-Cantelli, P.; Iribarren, I.; Palacios, H. 2002. Determinación de niveles de fondo y niveles de referencia de metales pesados y otros elementos traza en suelos de la Comunidad de Madrid. Ministerio de Ciencia y Tecnología. Consejería de Medio Ambiente. Instituto Geológico y Minero de España. Madrid.

Miko, S.; Halamiæ, J.; Peh, Z.; Galoviæ, L. 2001. Geochemical Baseline Mapping of Soils Developed on Diverse Bedrock from Two Regions in Croatia Geologia Croatica. ZAGREB. 54/1. 53-118.

Milani, A.; Carella, F.; Petruzelli, G.; Jean, P.; Unzo, N. 1995. Soil quality criteria and remediation goals for regione Lombardia´s legislation a on soil quality protection and contaminated sites reclamation. En: Contaminated Soil´95. Ed. Van der. Brink, W.J.; Bosman, R; Arendt, F. Kluwer Academic Pub. 681-690.

Muller, D.; Schadmann, M. 1999. Austria. En: Risk Assessment for Contaminated Sites in Europe. Vol 2. Policy Frameworks. Ed. Ferguson,C. y Kasamas, H. LQM Press. Nottingham NG7 2RD. UK. 5-10.

Navas, A.; Machín, J. 2002. Spatial distribution of heavy metals and arsenic in soils of Aragón (norheast Spain): controlling factors and environmental implications. Appied Geoch. 17, 961-973.

Norman, F. 1999. Sweden. En: Risk Assessment for Contaminated Sites in Europe. Vol 2. Policy Frameworks. Ed. Ferguson,C. y Kasamas, H. LQM Press. Nottingham NG7 2RD. UK. 151-159.

Odor, L.; Horváth, I.; Fúgedi, U. 1997. Low-density gechemical mapping in Hungary. J. Geoch.

Explor. 60, 55-66.
Pais, I.; Benton, J.J. 1977. The Handbook of Trace Elements. St. Lucie Press. Boca Raton. Florida.
Palacios-Orueta, A.; Pinzón, J.E.; Ustin, S.L. and Roberts, D.A..1999. Remote sensing of soils in Santa Monica Mountains: II Hierachical foreground and background analysis. Remote Sensing Environment. 68: 138-151.
Passaro, D.; Lima, A.; Jorge, C.; Ferreira da Silva, E. 1999. Portugal. En: Risk Assessment for Contaminated Sites in Europe. Vol 2. Policy Frameworks. Ed. Ferguson,C. y Kasamas, H. LQM Press. Nottingham NG7 2RD. UK.129-135.
Petzold, E.; Luering, E.M.; Morariu, A. 2003. Groundwater and soil contaminations in Romania. Monitoring and remediation feasibilities. CONSOIL´03. Edicción en CD.
Plant, J.; Smith, D., Smith, B. and Williams, L., 2001. Environmental Geochemistry at the global scale. Applied Geochemistry (16): 1291-1308.
Podlesakova, E.; Nemecek, J. 1995. Contamination and pollution of soils in Czech Republic. Artículo 101. Edicción en CD.
Quercia, F. 1999. Italy. En: Risk Assessment for Contaminated Sites in Europe. Vol 2. Policy Frameworks. Ed. Ferguson,C. y Kasamas, H. LQM Press. Nottingham NG7 2RD. UK. 99-106.
Rapant, S.; Raposova, M.; Bodis, D.; Marsina, K.; Slaninka, I. 1999. Environmental – geochemical mapping program in the Slovak Republic. Journal of Geochemical Exploration 66. 151–158.
Reimann, C.; Caritat, P. 1998. Chemicals elements in the environments. Factsheets fir the geochemist and environmental scientist. Springer. New York.
Reimann, C.; Caritat, P.; Halleraker, J.H.; Volden, T.; Ayras, M.; Niskavaara, H.; Chekushin, V.A. and Pavlov, V.A.. 1997. Rainwater composition in eight artic catchements in Northern Europe (Finland, Norway and Russia). Atmospheric Environment. 31: 159-170.
Reimann, C.; Koller, F.; Kashulina, G.; Niskavaara, H. and Englmaier, P.. 2001. Influence of extreme pollution on the inorganic chemical composition of some plants. Environmental Pollution. 115: 239-252.
Reimann, C.;Kashulina, G.; Caritat, P.; Niskavaara, H. 2001. Multi-element, multi-medium regional geochemistry in the European Arctic: element concentration, variation and correlation. Applied Geochemistry 16. 759-780.
Ross SM (ed). 1994. Toxic metals in Soil-Plant Systems. John Wiley and Sons Ltd.
Rühling, A., 2002. A European survey of atmospheric heavy metal deposition in 2000-2001. Environment Pollution, (120): 23-25.
Rühling, A., 2002. A European survey of atmospheric heavy metal deposition in 2000-2001. Environment Pollution, (120): 23-25.
Safizadeh, M.H. 2002. Minimizing the bias and variance of the gradient estimate in RMS simulation studies. European J. of Operational Res. 136. 121-135.
Salminen, R.; Gregorauskiene, G. 2000. Considerations regarding the defnition of a geochemical baseline of elements in the surfcial materials in areas differing in basic geology. Applied Geochemistry 15. 647-653.
Salminen, R.; Tarvainen, T., 1997. The problem of defining geochemical baselines: a case study of selected elements and geological materials in Finland. J. Geochem. Explor. 60, 91-98.
Salminen, R.; Tarvainen, T.; Demetriades, A.; Duris, M.; Fordyce, F.M.; Gregorauskiene V.; Kahelin, H.; Kivisilla, J.; Klaver, G.; Klein, P.; Larson, J.O.; Lis, J.; Locutura, J.; Marsina, K.; Mjartanova, H.; Mouvet, C.; O'Connor, P.; Odor, L.; Ottonello, G.; Paukola, T.; Plant, J.A.; Reimann, C.; Schermann, O.; Siewers, U.; Steenfelt, A.; Van der Sluys, J.; de Vivo, B.; Williams, L., 1998. FOREGS geochemical mapping field manual. Geological Survey of Finland, Guide 47.
Salomons, W.; Forstner, U; Mader, P. 1995. Heavy metals. Problems and solutions. Springer. New York.
Sauvalle, B. y Darmendrail, D. 1999. France. En: Risk Assessment for Contaminated Sites in Europe. Vol 2. Policy Frameworks. Ed. Ferguson,C. y Kasamas, H. LQM Press. Nottingham NG7 2RD. UK. 49-60
Shacklette, H.T.; Boerngen, J.G. 1984. Elemental concentrations in soils and other surficial materials of the conterminous United States .In USGS Prof. Paper 1270. US Gov. Printing Office, Washington, DC.

Sheppard, S.C.; Gaudet. C.; Sheppard, M.L.; Cureton, P.M.; Wong, M.P. 1992. The development of assessment and remediation guidelines for contaminated soils - a review of the science. Can. J .Soil SCi. 72: 359-394.

Singha, M.; Mullerb, G.; Singha, I.B. 2003. Geogenic distribution and baseline concentration of heavy metals in sediments of the Ganges River, India. Journal of Geochemical Exploration 3976 (2003) 1 –17. Article in press

Stuckens, J.; Coppin, P.R. and Bauer, M.E.. 2000. Integrating contextual information with per-pixel classification for improved land cover classification. Remote Sensing Environment. 71: 282-296.

Szabó, P. 1993. Agro-environmental assessment and remediation. Case study in Hungary. En: Contaminated Soil´93. Ed. Arendt, F.; Annokkée, G.J.; Bosman, R; Van der. Brink, W.J.; Kluwer Academic Pub. 1593-1601.

Töpfer, K. (2002). Change and Challenge. A state of the environment briefing for the global environment facility. UNEP. 26pp.

Van Dick, E.; Cornelis, C. 1999. Belgium. En: Risk Assessment for Contaminated Sites in Europe. Vol 2. Policy Frameworks. Ed. Ferguson,C. y Kasamas, H. LQM Press. Nottingham NG7 2RD. UK. 11-28.

Vegter, J.J. 1995. Soil Protection in The Netherlands. En: Salomons, W.; Forstner, U; Mader, P. 1995. Haevy metals. Problems and solutions. Springer. New York. 79-100.

Vollmer, M.K.; Gupta, S.K. and Krebs, R.. 1997. New standards on contaminated soil in Switzerland comparison with Dutch and German quality criteria. 3erd International conference on Biogeochemistry of the Trace Elements. Paris (France), May 15-19. R. Proust ed. Ed. INRA. pp. 445-457.

Vrana, K.; Rapant, S.; Marsina, K.B. ; Mankovska, B.; Curlık, J.; Sefcık, P.; Daniel, J.; Lucivjansky´ , L.; Lexa, J.; Pramuka, S. 1997. Geochemical atlas of the Slovak republic at scale of 1:1,000,000. J. Geochem. Explor. 60, 7–37.

Wenger, C.; Ziegler, U. 1999. En: Risk Assessment for Contaminated Sites in Europe. Vol 2. Policy Frameworks. Ed. Ferguson,C. y Kasamas, H. LQM Press. Nottingham NG7 2RD. UK. 161-164.

Williamsa,T.M.; Dunkleya, P.N.; Cruzb, E.; Acitimbayb, V.; Gaiborb, A.; Lopezb, E.; Baezb, N.; Aspdena, J.A. 2000. Regional geochemical reconnaissance of the Cordillera Occidental of Ecuador: economic and environmental applications Applied Geochemistry 15. 531-550.

Woitke, P.; Wellmitz, J.; Helm, D.; Kube, P.; Lepom, P. and Litheraty, P. 2003. Analysis and assessment of heavy metal pollution in suspended solids and sediments of the river Danube. Chemosphere. 51: 633-642.

Xie, X.; Hangxin, C. 2001. Global geochemical mapping and its implementation in the Asia–Pacific region Applied Geochemistry 16. 1309–1321

Xie, X.J.; Mu, X.Z.;Ren, T.X., 1997. Geochemical mapping in China. J. Geochem. Explor. 60, 99–113.

Xuejing, X. and Hangxin, C., 2001. Global geochemical mapping and its implementation in the Asia-Pacific region. Applied Geochemistry (16): 1309-1321.

Yohn, S.S.; Long, D.T.; Fett, J.D.; Patino, L.; Giesy, J.P. and Kannan, K.. 2002. Assesing environmental change through chemical-sediment chronologies from inland lakes. Lakes&Reservoirs, Research and Management 7:217-230.

Zhang, C.; Selinusb, O.; Schedinb, U.J. 1998. Statistical analyses for heavy metal contents in till and root samples in an area of southeastern Sweden. The Science of the Total Environment. 212. 217-232

Resultados y discusión.

En primer lugar, se ha procedido a realizar una serie de cálculos preliminares siguiendo los criterios tradicionales, con el fin de demostrar de manera fehaciente que los valores obtenidos son muy poco indicativos del estado y evolución de los procesos de contaminación, a la vez que se justifica la necesidad de abordar estos problemas desde otro punto de vista y por tanto de desarrollar otras metodologías que permitan avanzar en su estudio e interpretación

Se han realizado los siguientes cálculos y estadísticos:

1º.- Estadísticos descriptivos univariantes de todas las variables y contrastes de bondad de ajuste a las distribuciones normal y lognormal.

2º.- Análisis de regresión y correlaciones entre variables dos a dos.

3º. Análisis de componentes principales.

4º.- Análisis de factores (correlación y covarianza).

5.-º Análisis de *cluster* por varios métodos.

Obteniendo los resultados que se muestran a continuación.

Análisis de correlación y regresión bivariante.

Tal y como se muestra en la matriz de correlación adjunta, las correlaciones entre los distintos elementos son altas sólo en Co-Cr-Ni, elementos que suelen estar relacionados desde el punto de vista mineralógico, ($r > 0.9$), pero no son significativas, la razón de que esto ocurra, se debe a la distribución de los datos, basta observar los gráficos de ajuste correspondientes, para notar unas agrupaciones de gran densidad de puntos en las partes inferiores, que dan como resultado un r elevado y una serie de puntos dispersos, que parecen ser los causantes de la falta de significación.

Otros elementos que suelen estar relacionados, como el Pb y el Zn, dan coeficientes de correlación significativos, pero sensiblemente inferiores a los habituales en mineralogía, que suelen encontrarse en torno a 0.9, en este caso el r calculado es de 0.56, que probablemente se deba a a dos causas, una primera relacionada con procesos de meteorización y remoción posterior de materiales y una segunda a efectos antrópicos.

El resto de los elementos entre sí y con los ya citados presenta coeficientes de correlación bajos y no significativos.

Además del coeficiente de correlación clásico (r de Pearson), se han calculado los coeficientes de correlación Rho de Spearman, recordemos que es similar al de Pearson, pero utilizando métodos no paramétricos, obteniéndose resultados similares, como se muestra en la tabla.

Análisis multivariante.

Los métodos multivariantes ensayados, han sido un análisis de factores/componentes principales según correlaciones y según covarianzas y un análisis de *clusters* por varios métodos.

Como puede verse en la tabla adjunta, los análisis de componentes principales realizados por los dos métodos arrojan unos buenos resultados, tanto por el método de correlaciones como por el de covarianzas, dando pocos factores y un alto porcentaje de varianza explicada.

En el caso de componentes principales por covarianza, resultan tres factores que explican el 98% de la varianza, dominados por los elementos Cr, Cd y Zn.

Otro tanto ocurre al utilizar las correlaciones, donde tres factores explican el 78.4%, y que en este caso están comandados por Co-Cr, Pb-Zn y Cd.

Pese a los, aparentemente, buenos resultados que proporcionan los análisis de factores/componentes principales, desde el punto de vista de este trabajo no son demasiado fiables, pues basta observar la matriz de correlación del apartado anterior para ver que las correlaciones entre elementos son muy bajas en general, y por tanto los resultados que puedan obtenerse tomándolas como base de partida son como poco dudosos. Éste parece ser uno de los casos a los que Davis (1978) se refiere como ajustes dudosos, debido a la sensación de *falsa seguridad* que proporciona un buen ajuste que no se ve apoyado por resultados preliminares.

En cuanto a los análisis de *clusters* jerárquicos, realizados por distintos métodos, basta ver la tabla correspondiente para notar que proporcionan resultados diferentes según el método utilizado, lo que en general los invalida, pues en el mejor de los casos se puede utilizar el método que más nos convenga, que no tiene por qué ser el más objetivo

Resumiendo, tal y como se refleja en los resultados mostrados, las metodologías tradicionales son muy poco precisas a la hora de ser aplicadas a estudios sobre contaminación, pues:

1º.- Los contrastes de bondad de ajuste de todas las variables muestran que no existe un buen ajuste a ningún tipo de distribución de las que se utilizan habitualmente, normal o lognormal, y por tanto los parámetros estadísticos son muy poco representativos del conjunto de los datos.

2º.- La correlación entre las distintas variables es muy baja, y en la mayoría de los casos no es siquiera significativa.

3º.- Los métodos multivariantes ensayados proporcionan, en el mejor de los casos, unos resultados muy vagos.

4º.- Los análisis de *cluster* realizados, muestran distintos resultados en función del método de cálculo/agrupación utilizado.

Obtención de niveles característicos.

Tal y como se ha explicado en el capítulo anterior, la metodología que se propone para la obtención de niveles característicos

Sistemas de agrupación.

Los sistemas de agrupación propuestos proporcionan unos buenos resultados como puede verse en las tablas resumen adjuntas (Tablas 1 a 8).

En ellas se observa que, independientemente del número de agrupaciones resultante, la pérdida de información, medida a través de la variación de entropía, es mínima, por lo que las agrupaciones resultantes parecen un buen sistema de agrupación de datos sin pérdida de información, por lo que puede decirse que los resultados obtenidos son mejores que los que proporcionan los métodos convencionales.

En cuanto a los resultados que proporcionan los distintos métodos de agrupación, como puede verse en la tabla resumen, se observa que no existen diferencias significativas en cuanto a variación de entropía se refiere, entre los distintos métodos ensayados, por lo que la elección de alguno de ellos, desde un punto de vista general, debe hacerse por el número de grupos que genere cada uno de ellos, si bien la elección del de máximos relativos 2 (MR2) y máximos relativos 2 con detección de inliers/outliers, son los que presentan una mejor relación nº grupos/*ganancia* de información, y por tanto una menor distorsión de la misma, lo que resulta crucial a la hora de interptetar los resultados.

En las figuras y mapas que se acompañan puede observarse la distribución de anomalías obtenidas a partir de cada uno de los métodos utilizados.

No obstante se hace una descripción somera de como se estima la variación de la cantidad de información en el apartado siguiente.

ESTIMACIÓN DE ENTROPÍAS (IGUALDAD DE DIVERSIDADES)

El concepto de diversidad aparece en numerosos campos de investigación asociado siempre a la idea de variabilidad de los elementos de una determinada población. Así, la entropía de Shannon permite cuantificar esta diversidad desde un punto de vista teórico, pero puede suceder que no se conozca su valor más que a nivel muestral, siendo necesario desarrollar expresiones para estimar de forma óptima la entropía de toda la población. En nuestra situación, dado que los tamaños muestrales son suficientemente grandes, el estadístico para el contraste

$$H_0: H(\Theta) = D_0$$
$$H_1: H(\Theta) \neq D_0$$

viene dado por:

$$Z = \frac{\sqrt{n}[H(\theta) - D_0]}{\hat{\sigma}}$$

siendo

$$\hat{\sigma}^2 = \sum_{i=1}^{n} p_i (\log p_i)^2 - H(\theta)^2$$

la estimación de la varianza. Su distribución asintótica se ajusta por el teorema de Slutsky (Ferguson, 1996) a un modelo Gausiano tipificado.

Igualmente, una vez calculadas las entropías para diferentes elementos, el objetivo siguiente consiste en averiguar si existen diferencias significativas entre ellas, es decir si la pérdida de información consecuencia de los distintos métodos de agrupación son dignas de tenerse en cuenta. Para ello contrastaremos la hipótesis nula

$$H_0: H(\Theta_1) = H(\Theta_2)$$

frente a la alternativa

H_1: $H(\Theta_1) \neq H(\Theta_2)$

Dado que el tamaño muestral es en todos los casos grande (n=278), se utilizará el siguiente estadístico de contraste (Pardo, 1997):

$$Z = \frac{\sqrt{n_1 n_2}\,[H(\theta_1) - H(\theta_2)]}{\sqrt{n_2 \hat{\sigma}_1^2 + n_1 \hat{\sigma}_2^2}}$$

donde n_i denota el tamaño muestral y s_i^2 es la cuasivarianza muestral asociada a la entropía $H(\Theta_i)$, que se distribuye asintóticamente según una normal N(0,1). Como $n_1 = n_2 = 278$, la expresión del estadístico queda reducida a:

$$Z = 16,67333 \frac{H(\theta_1) - H(\theta_2)}{\sqrt{\hat{\sigma}_1^2 + \hat{\sigma}_2^2}}$$

En la Tabla XX se muestran los distintos valores de entropía, así como los resultados de los tests de hipótesis resultantes de contrastar la entropía de cada método de agrupación frente a la variable que representa los elementos sin agrupar.

2.5. Compatibilización de la metodología propuesta con los *baseline* .

Tal y como queda reflejado en los capítulos anteriores, los métodos que se están utilizando en la actualidad para la detección de puntos o zonas contaminadas se basan en el concepto de *baseline*, es decir, en la obtención de un valor o conjunto de valores, en el caso de que se considere más de un *baseline*, de concentración de *origen natural*, que se supone representativo de la zona de estudio, y que permite diferenciar las zonas contaminadas de las que no lo están, y éste valor se suele obtener a través de la asignación directa de un parámetro estadístico.

En esencia, a través del concepto de *baseline*, se trata de extender una serie de valores puntuales, correspondientes a los puntos muestrales, a una superficie acotada, de forma que el valor obtenido sea representativo de la misma, lo cual, desde un punto de vista práctico, lo convierte en un proceso de integración areal de datos puntuales, que como tal puede realizarse de varias formas, que van desde la asignación de su valor a partir de un parámetro estadístico (Gregorauskienè y Kardunas, 1997; Salminen y Tarvainen, 1997; Salminen y Gregorauskiene, 2000, Tarvainen y Kallio, 2002; De Miguel *et al.*, 2002; ...), al uso de criterios de tipo más probabilístico como el rango intercuartílico (Canadian Soil Survey).

Desde el punto de vista de este trabajo, parece más lógica la utilización de criterios de tipo probabilístico como el rango intercuartílico, pues asumiendo riesgo que supone el uso de distribuciones de frecuencias como distribuciones de probabilidad en un muestreo único, a la hora de determinar un valor de concentración como referencia, si se toma el rango intercuartílico, se estará considerando un intervalo de valores que representan al menos al 50% de los puntos muestreados, lo cual le confiere robustez, desde el punto de vista estadístico y representatividad desde el punto de vista conceptual. Sin embargo su precisión es menor que la de métodos paramétricos como aquéllos que consideran un intervalo de confianza para la media o la mediana, pues su recorrido, es decir la diferencia entre los valores que definen el intervalo (25-75%), es mayor.

A tenor de lo visto, el ideal sería alcanzar un equilibrio entre robustez y precisión, de manera que se pudiera obtener un intervalo de valores que fuese representativo de al menos el 50% de los datos y que a la vez tuviese un recorrido menor. Esta idea no es nueva, de hecho autores como Rousseeuw y Leroy (1988) y Rousseew y Croux (1993), proponen lo que llaman *shortest half*, literalmente la mitad más corta, que consiste en el intervalo de valores que contiene al 50% de los datos con recorrido menor, como estimador robusto de escala. Así, se puede disponer de un intervalo de valores tan representativo como el rango intercuartílico, pues al igual que éste, contiene al 50% de los datos, que simultáneamente es más preciso puesto que su recorrido es menor, y que presenta la ventaja adicional de estar menos influenciado por valores extremos.

Dado que para la obtención, tanto del rango intercuartílico como del *shortest half*, que podríamos traducir como el 50% más denso, se hace uso únicamente de la posición los valores en la distribución de frecuencias, en distribuciones correspondientes a variables como las que se tratan es este trabajo, en las que frecuentemente encontramos valores repetidos, resultaría bastante útil definir un parámetro de *cobertura*, tomándolo como el porcentaje de datos, incluido entre los valores que delimitan el intervalo correspondiente al 50% más denso, de forma que además de disponer de un intervalo más preciso que el rango intercuartílico, se mejoraría su representatividad, ya que puede incluir un porcentaje mayor de las muestras.

Además, el hecho de utilizar un intervalo frente a un sólo valor presenta una serie de ventajas:

1º

Las variables se caracterizan por una serie de parámetros (media, mediana, moda, recorrido, rango intercuartílico, etc.). No hay que olvidar que en este caso, las variables están referidas a coordenadas geográficas (x,y), lo que permite dividir el área de estudio en casillas según una cuadrícula, y

asignar valores a cada casilla, de forma que se obtiene la distribución areal del o los *baselines*.

Si dibujamos un mapa de isolíneas, veremos unas concentraciones que corresponden a máximos y mínimos (que salvo que estén cerradas, pueden sufrir variaciones con las dimensiones del área considerada), y que además, en el caso de los máximos corresponden a modas areales de la distribución de la variable, su número, situación, valor y forma, son los que definen las pautas de comportamiento de la variable en la zona de estudio.

Se considera que el valor o intervalo de valores más representativo (para una variable) de la zona estudiada es el que ocupa mayor superficie, que por extensión es el que se da con mayor frecuencia y trasladándolo a probabilidad es el que es más probable encontrar.

Dentro de los intervalos clásicos, el rango intercuartílico representa el 50% de los datos y además por debajo del valor que alcance la variable en su límite superior se encuentra el 75% de los valores, se trata de un método "seguro" pero poco preciso. Si ese 50% de los valores lo hacemos correr por la escala hasta reducir al máximo su amplitud, tenemos el 50% más denso, que supone el intervalo más pequeño con mayor probabilidad de aparición, por lo tanto es el más preciso, pues no existe otro intervalo de las mismas dimensiones que se pueda presentar con un grado de probabilidad mayor. Este intervalo es tan representativo como el rango intercuartílico (50% de la muestra) pero más preciso, pues su amplitud es menor. Si la distribución es de tipo normal coinciden. Además, tiene la ventaja de que está menos influenciado por valores extremos. Como valor más representativo de este intervalo, podríamos tomar su mediana (50% de los valores arriba y abajo) que sigue la línea del planteamiento, o la media que representa, desde el punto de vista geométrico el "centro de gravedad" del intervalo. La mediana presenta la ventaja de ser el centroide desde el punto de vista del rango de los datos una vez ordenados de menor a mayor, pero a la hora de tomar un valor característico que represente a todo el intervalo, parece que la media es más

adecuado, pues es sensible a los valores extremos que también cuentan, además de dejar la puerta abierta a lagunas en el muestreo.

2.3. Propuesta de un sistema de agrupación.

Tal y como se refleja en los objetivos de este trabajo, la definición de niveles característicos, pasa por conseguir un método de agrupación de datos tal, que los subgrupos de datos resultantes cumplan la propiedad de que cada elemento que pertenezca a un grupo se parezca más al resto de los componentes de ese grupo que a los de otro, por próximo que se encuentre, y sin hacer uso de condiciones *apriorísticas*, que supongan un comportamiento predeterminado de las variables y que por tanto puedan introducir un sesgo en la interpretación de los valores resultantes. Es decir, se busca un método que establezca una serie de agrupaciones de valores que por su similitud individual y su *independencia* con respecto a grupos adyacentes permita considerar *homogéneos* a los valores que lo constituyen.

Normalmente, cuando se dispone de un conjunto de datos en los que se desea establecer subgrupos o clasificarlos, se suele hacer uso de criterios o medidas de semejanza, que en el caso de variables cuantitativas se basan en un criterio de minimización de las diferencias o distancias existentes entre los valores entre sí o con respecto a un parámetro. Para la obtención de éstas medidas de distancias o semejanzas, hay multitud de sistemas propuestos: distancias de Cook, euclidianas, de Mahalanobis, de Matushita, etc., a partir de las cuales se pueden agrupar los datos en función de su parecido, desde el punto de vista de que existan menos diferencias entre ellos, siguiendo un método específico de cálculo para cada una de ellas. Las agrupaciones resultantes, se realizan mediante iteraciones, de manera que se van estableciendo agrupaciones a distintos niveles en cada iteración, quedando reflejado el *factor de semejanza* a través de la distancia que separa a un nuevo punto o agrupación de un subgrupo previamente establecido. En tanto que, cuando se trata de variables cualitativas, se utilizan las llamadas distancias de Hamming, que consisten en el contaje y comparación de de atributos diferentes para cada muestra.

Estos métodos que, aparentemente suponen la solución al problema que se plantea, de hecho su uso es cada vez más común tanto en el campo en el

que nos movemos como en otros similares (Bandyopadhadyay, 2004; Buurman *et al.*, 2004; Farber y Kadmon, 2003; Globocanin *et al.*, 2004; Lacassie *et al.*, 2004; Meng y Maynard, 2001; Meyer *et al.*, 2004; Motelay-Massei *et al.*, 2004; Tayfur *et al.*, 2003; Therfeld *et al.*, 2003; Yücer y Demil, 2004; Zhang, Y.G. *et al.*, 2004; Zhang, B. *et al.*, 2004;...) se incluyen en los llamados métodos de análisis de *clusters*., que como ya se ha visto anteriormente presentan varios inconvenientes para su uso generalizado.

1º.- El tipo de medida de similaridad a utilizar, ya que según se tome una u otra, ej. euclidianas o de Mahalanobis, los resultados pueden ser diferentes, debido a que cada una de ellas sigue un método de cálculo diferente, asumiendo por tanto un comportamiento predeterminado de las variables, lo cual está en contra del planteamiento inicial.

2º.- Independientemente del sistema de obtención de similaridades (distancias), todas ellas parten del supuesto de que se está trabajando con una distribución continua y en la mayoría de los casos multivariante, lo que presupone un cierto grado de relación *a priori* de las variables (colinearidad o covariación), de manera que si no existe tal relación obtenemos unos resultados sesgados de muy difícil interpretación.

3º.- La conveniencia de estandarizar o normalizar las variables, para que en el cálculo de las similitudes no se les conceda más importancia a unas que a otras, pues si se realiza la estandarización se pierde precisión.

4º.- El orden de introducción de las variables puede influir en el resultado, ya que el cálculo de las similaridades al añadir una nueva variable, está condicionado por los cálculos previos.

5º.- Sea cual sea el método de *clustering* utilizado, k-medias o jerárquico, obliga a decidir de forma subjetiva los *grupos homogéneos* que se persiguen, el primero por definir desde el inicio el número de grupos resultantes y el segundo por obligar a decidir subjetivamente en qué iteración se considera que están formados los *grupos homogéneos*.

Pese a los inconvenientes citados, los métodos que produzcan *agrupaciones homogéneas*, conceptualmente parecen los más adecuados para trabajos como el presente, hay que recordar que se parte de la hipótesis de que *fenómenos iguales que actúan de idéntica manera, con la misma intensidad y en el mismo orden dan lugar a resultados idénticos*, y por tanto este comportamiento debe quedar reflejado en una cierta similitud en los datos obtenidos, lo que obliga a hacer una serie de correcciones en los planteamientos clásicos de los análisis de *clusters*:

1º.- No presuponer la existencia de una relación estadística o funcional entre las distintas variables.

2º.- No considerar *a priori* que las variables se ajustan a un tipo de distribución definida ni que son continuas.

Estas correcciones obligan a:

1º.- Estudiar las variables una a una y por separado, dejando las consideraciones sobre relaciones entre ellas para más tarde.

2º.- Desarrollar un sistema nuevo de cálculo de similitudes o distancias, que no se vea afectado por supuestos de comportamiento y continuidad de los datos.

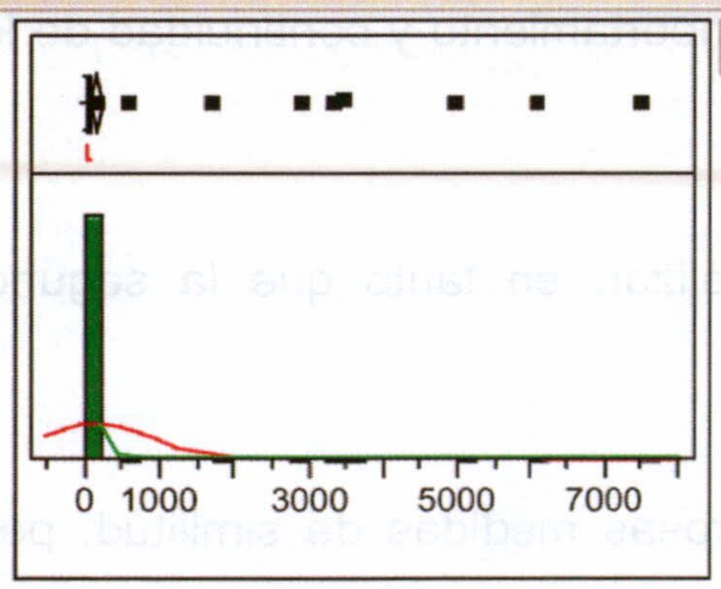
La primera corrección es fácil de realizar, en tanto que la segunda requiere un esfuerzo un poco mayor.

Como ya se ha citado, existen numerosas medidas de similitud, pero todas ellas presuponen un comportamiento continuo de la variable o un cierto grado de ajuste a una distribución conocida. En el caso de éste tipo de trabajos, en el que podemos encontrar una variabilidad de datos tal que sugiere en muchos casos la existencia de varias poblaciones mezcladas, -hay que recordar que algunos autores como Salminen y Tarvainen (1997), Salminen y

Gregorauskienè (2000), etc., proponen el uso de varios niveles de referencia-, la *medida de similitud* que se proponga debe seguir otros derroteros, de forma que el criterio a seguir no suponga necesariamente un comportamiento continuo y más o menos lineal de las variables en cuestión.

Partiendo de la base de que la información de cualquier tipo contenida en una variable reside en la distribución de la misma, se deduce fácilmente que los parámetros clásicos, como la media, la desviación típica, otros momentos, la mediana o la moda, aportan muy poca información si los datos no se ajustan a una distribución conocida, normal, lognormal, etc., pues pierden tanto sus propiedades de inferencia, como su significado conceptual, pasando a ser meros indicadores del comportamiento general de los datos que en el mejor de los casos solo aportan información cualitativa, como puede verse en la Figura 2, en la que puede verse cuales serían las distribuciones normales y lognormales correspondientes a las medias y desviaciones obtenidas, así como comparar los valores correspondientes a la media y a la mediana, resultando que el 50% de los datos tiene valores iguales o inferiores a 66.00 y el valor medio asignado es de 176.56.

Cr



Quantiles			Moments	
100.0%	maximum	7530,0	Mean	176,56162
99.5%		6970,0	Std Dev	730,0359
97.5%		1750,0	Std Err Mean	43,706106
90.0%		115,0	upper 95% Mean	262,59857
75.0%	quartile	89,0	lower 95% Mean	90,524665
50.0%	median	66,0	N	279
25.0%	quartile	46,0	Sum Wgt	279
10.0%		31,0	Sum	49260,692
2.5%		21,0	Variance	532952,42
0.5%		5,6	Skewness	7,602706
0.0%	minimum	0,0	Kurtosis	61,911104
			CV	413,47372
			N Missing	0

Fig. 2. Estadística descriptiva del Cr.

En otras disciplinas como la econometría frecuentemente se añaden a las tablas de la estadística descriptiva otros parámetros e índices de concentración, como:

- Mediala (MI), valor de la variable tal que la suma de las observaciones mayores que él es igual a la suma de las observaciones menores que él.

- Índice de Gini, alcanza valores entre 0 y 1, el valor 0 indica concentración mínima, por tanto la muestra está uniformemente repartida a lo largo de todo el rango, y el 1 concentración máxima, de forma que un sólo valor acumula el 100% de los datos. Se calcula a partir de la Curva de Lorentz.

- Índice de Theil, que se basa en la entropía (medida de desorden) de la distribución de frecuencias, obteniéndose a partir de la fórmula:

$$T = 1 + (\sum pi*\log(pi)/\log(n))$$

Donde:

pi es el procentaje que representa el valor i frente a la muestra total.

Al igual que el índice de Gini, alcanza valores entre 0 y 1, el 0 indica concentración mínima y el 1 concentración máxima.

Éstas medidas de concentración o agrupamiento, proporcionan más información sobre la distribución de los datos, indicando en el caso de la mediala dónde se encuentran situados los valores más altos y más bajos, y en el caso de los índices de Gini y Theil el grado de concentración o dispersión general, pero no dan más que una información vaga de cuales son las zonas de mayor densidad de puntos, aunque algunos autores proponen métodos de segmentación o agrupación de los datos para delimitar las agrupaciones de concentraciones similares, Akita (2003) utiliza criterios apriorísticos similares a

los clásicos de análisis de varianza, Dikhanov (1996) utiliza los cuantiles, Kapur y Kesawan (1992) y Reesor y McLeish (2002) proponen métodos multivariantes y transformaciones de las variables, etc.. Pero en definitiva, estas propuestas, en lo que a este trabajo se refiere, se diferencian muy poco de las técnicas ya comentadas que parten de supuestos apriorísticos.

Visto lo anterior, se confirma que no existe ningún método que permita crear agrupaciones de datos de manera objetiva, sin hacer ningún tipo de suposición *a prioi*, por lo que para su desarrollo hay que partir de cero.

Para ello, se considera:

a) Las agrupaciones homogéneas, están refrendadas matemáticamente por las teorías de puntos autoconsistentes y puntos principales de Flury (1993 y 1995) y Tarpey (1998 y 2000), que definen puntos principales como el conjunto de k puntos que representa óptimamente una distribución en términos de error cuadrático medio y puntos autoconsistentes al conjunto de k puntos que coinciden con su media condicionada a los dominios de atracción que generan de acuerdo con la distancia minimal (González Caballero y Peralta Sáez, 2003). Es decir, que los conjuntos de puntos resultantes de una eventual agrupación, contienen un punto principal y que sus límites están fijados por los dominios de atracción que éste genera o Dominios de Voronoi, lo que en definitiva supone una serie de agrupaciones de puntos o valores que cumplen la condición de parecerse más entre sí que a cualquier otro punto o valor que pertenezca a un grupo distinto.

b) Las agrupaciones que se obtengan no deben sufrir pérdida de información o distorsión de la misma, lo que equivale a que la entropía correspondiente a la distribución completa calculada de forma discreta, sea igual a la de la suma de las entropías de

los grupos resultantes, pudiendo medir la entropía a través de los índices de Theil.

De esta manera se dispone, desde el punto de vista matemático, de los dos criterios necesarios para establecer las agrupaciones, uno derivado de los índices de concentración de Theil, pues las agrupaciones que se obtengan deben cumplir que la entropía de la distribución no varía al realizar su cálculo en forma discreta (punto a punto) o en intervalos, con lo cual se demuestra que no hay pérdida de información en los grupos y un segundo, basado en las teorías de puntos principales, que obliga a que los puntos o valores incluidos en un mismo grupo cumplan la condición de parecerse más entre sí que a otro punto perteneciente a un intervalo diferente.

Tanto Flury como Tarpey, destacan en sus trabajos la enorme dificultad que supone la obtención de los puntos principales de una distribución de tamaño medio, pues se basan en métodos de cálculo iterativos del tipo k-medias, que necesitan mucho tiempo de cálculo y un uso intensivo de ordenador.

Pero si se enfoca el problema de otra manera, haciendo uso de la filosofía de la teoría de la información (Shannon, 1948), se puede simplificar el problema del cálculo sensiblemente.

ENTROPÍA DE SHANON

El concepto de *entropía* de un sistema físico fue introducido por primera vez por Clausius en 1864 como una medida del desorden del sistema, que puede expresarse en términos de otras macrocoordenadas que sí pueden medirse de forma directa. No obstante, este concepto clásico de entropía de Clausius no era de naturaleza probabilística, y fue Bolzmann en 1896 quien definió la entropía en termodinámica estadística como una medida de las posiciones y velocidades de todas las partículas incluidas en el sistema físico, enfatizando así su significado probabilística.

En base al desarrollo de la teoría de la probabilidad durante la primera mitad del siglo XX, Shannon introdujo en 1948 la entropía en abstracto, por analogía con la expresión de Bolzmann, como una medida de la incertidumbre de experimentos probabilísticas arbitrarios, es decir una medida cuantitativa sobre la cantidad de información proporcionada por el experimento. Esto permitió a Kullback y Leibler (1951) definir la *divergencia* como una medida de la distancia entre dos poblaciones.

Consideremos un experimento aleatorio cuyos posibles resultados son a_1, a_2, …, a_n con probabilidades respectivas p_1, p_2, …, p_n ($pi \geq 0$, $1=1,2,\ldots,n$, $p_1+p_2+\cdots+p_n=1$). Es evidente que dicho experimento puede representarse mediante una variable aleatoria X con función de masa $P(X= a_i)=p_i$. Se denomina entropía de la variable X a la expresión:

$$H(X) = -\sum_{i=1}^{N} p_i \log p_i$$

Los logaritmos se pueden tomar respecto a cualquier base que sea mayor que la unidad. De hecho, si se toma en base dos, a la unidad correspondiente se le denomina *BIT* (Binary digit) y puede definirse como la entropía correspondiente a una variable aleatoria con dos resultados equiprobables:

$$H(X) = -\frac{1}{2}\log_2 \frac{1}{2} - \frac{1}{2}\log_2 \frac{1}{2} = \log_2 2 = 1\ BIT$$

Si se toman en base 10, como es usual, la unidad correspondiente se denomina *DIT* (Decimal digit o unidad de Hartley) y se define como la entropía de una variable aleatoria con diez resultados equiprobables:

$$H(X) = -\frac{1}{10}\log_{10} \frac{1}{10} - \cdots - \frac{1}{10}\log_{10} \frac{1}{10} = \log_{10} 10 = 1\ DIT$$

Si se toman logaritmos neperianos, la unidad correspondiente se denomina *NAT* y su interpretación encaja entro de la entropía como medida de la incertidumbre asociada a variables aleatorias continuas.

ESTIMACIÓN DE ENTROPÍAS (IGUALDAD DE DIVERSIDADES)

El concepto de diversidad aparece en numerosos campos de investigación asociado siempre a la idea de variabilidad de los elementos de una determinada población. Así, la entropía de Shannon permite cuantificar esta diversidad desde un punto de vista teórico, pero puede suceder que no se conozca su valor más que a nivel muestral, siendo necesario desarrollar expresiones para estimar de forma óptima la entropía de toda la población. En nuestra situación, dado que los tamaños muestrales son suficientemente grandes, el estadístico para el contraste

H_0: $H(\Theta) = D_0$

H_1: $H(\Theta) \neq D_0$

viene dado por:

$$Z = \frac{\sqrt{n}[H(\theta) - D_0]}{\hat{\sigma}}$$

siendo

$$\hat{\sigma}^2 = \sum_{i=1}^{n} p_i (\log p_i)^2 - H(\theta)^2$$

la estimación de la varianza. Su distribución asintótica se ajusta por el teorema de Slutsky (Ferguson, 1996) a un modelo Gausiano tipificado.

Igualmente, una vez calculadas las entropías para diferentes elementos, el objetivo siguiente consiste en averiguar si existen diferencias significativas entre ellas, es decir si la pérdida de información consecuencia de los distintos métodos de agrupación son dignas de tenerse en cuenta. Para ello contrastaremos la hipótesis nula

H_0: $H(\Theta_1) = H(\Theta_2)$

frente a la alternativa

H_1: $H(\Theta_1) \neq H(\Theta_2)$

Dado que el tamaño muestral es en todos los casos grande (n=278), se utilizará el siguiente estadístico de contraste (Pardo, 1997):

$$Z = \frac{\sqrt{n_1 n_2}[H(\theta_1) - H(\theta_2)]}{\sqrt{n_2 \hat{\sigma}_1^2 + n_1 \hat{\sigma}_2^2}}$$

donde n_i denota el tamaño muestral y s_i^2 es la cuasivarianza muestral asociada a la entropía $H(\Theta_i)$, que se distribuye asintóticamente según una normal N(0,1). Como $n_1 = n_2 = 278$, la expresión del estadístico queda reducida a:

$$Z = 16,67333 \frac{H(\theta_1) - H(\theta_2)}{\sqrt{\hat{\sigma}_1^2 + \hat{\sigma}_2^2}}$$

Así:

1º.- Si se ordenan los datos en forma creciente, obtenemos una sucesión de valores en la que los datos quedan situados de forma que los más parecidos, es decir los que guardan mayor similitud, se encuentran más próximos.

2º.- Si en vez de utilizar el criterio tradicional de calcular las diferencias/similitudes entre puntos y estudiar la distribución de los dominios obtenidos sobre los valores, se estudia el comportamiento de las distancias entre puntos consecutivos, se observa que los valores quedan agrupados automáticamente entre máximos relativos de las distancias calculadas.

Siguiendo este planteamiento, se obtienen agrupaciones que cumplen las condiciones impuestas, en primer lugar no hay pérdida de información, lo que queda demostrado a los efectos de este trabajo en que no hay variación en los índices de Theil y por tanto en la entropía de la distribución, y además los valores quedan agrupados siguiendo un criterio de máxima similitud puesto que el valor medio de cada una de estas agrupaciones constituye un punto principal de la distribución de los mismos, resultando *agrupaciones autoconsistentes* univariantes, lo que en definitiva resulta ser un método de obtención de *clusters* univariante, que además presenta una serie de ventajas sobre los métodos tradicionales, que esencialmente consisten en:

a) Obtener directamente las agrupaciones finales sin necesidad de cálculos iterativos.
b) Conocer de manera cuantitativa la pérdida de información que se produce en el proceso de clasificación comparando el índice de Theil de la distribución en forma discreta (original) y el obtenido al calcularlo con los intervalos resultantes.
c) En el caso de resultar un número de grupos demasiado elevado para interpretarlo con facilidad, se pueden recalcular las agrupaciones sabiendo la información que se pierde en cada iteración.

2.2. Condiciones a cumplir por una nueva metodología.

Tal y como se ha planteado el problema en el capítulo anterior, la metodología a desarrollar debe de cumplir una serie de condiciones:

a) Estar libre de supuestos de comportamiento previos, pues no se puede afirmar *a priori* que las variables sigan una distribución determinada, ni que existan relaciones funcionales entre ellas.

b) El establecimiento de grupos de *datos homogéneos* debe hacerse con el mínimo de pérdida de información, siguiendo criterios de similaridad, que permitan su explicación a partir de la distribución de los factores condicionantes o patrones explicativos. Además debe asumir una cierta heterogeneidad tanto en el factor explicativo como en el muestreo.

c) Ha ser sensible tanto a los *outliers* (valores extremos) como a los *inliers* (valores anómalos situados en el interior de la distribución), no detectables por métodos convencionales.

d) Debe de ser *compatible*, en la medida de lo posible, con los métodos que se han venido utilizando hasta ahora.

Por lo tanto, el cumplimiento de estas condiciones, implica:

1º) Estudiar las variables por métodos libres de cualquier supuesto de comportamiento *a priori*.

2º) Desarrollar un método de agrupación que siga simultáneamente criterios de similaridad y continuidad.

3º) Elaborar un sistema de detección de *outliers* e *inliers*.

4º) Proponer un método de integración de datos que haga *compatibles* los resultados con otras metodologías.

En cuanto al punto 1º, se puede afirmar que el estudio de las variables una a una y por separado mediante el uso de técnicas no paramétricas, parece ser la solución más idónea, pues está libre de cualquier supuesto de ajuste a un tipo de distribución predeterminada así como de cualquier supuesto de relación funcional entre ellas.

La propuesta de estudiar las variables por separado, puede parecer equivocada desde el punto de vista de los trabajos tradicionales, pues se trata de técnicas univariantes que no aportan nada al conocimiento sobre cuales son las relaciones existentes entre los distintos elementos contaminantes y de su *comportamiento en conjunto*, pero hay que tener en cuenta que el uso de las técnicas de análisis multivariante, tanto las tradicionales (De Vivo *et al.*, 1997; Fachinelli *et al.*, 2001; Filzmoser, 1999; Gallego *et al.* 2002; Gordeev *et al.*, 2004; Huisman *et al.* 1997; Reimann *et al.* 2001; Preda y Cox, 2002; ...) como las modificaciones y desarrollos recientes (Buurman *et al.* 2004; Farber y Kadmon, 2003; Globocanin *et al*, 2004; Lacassie *et al.* 2004; Meng y Maynard, 2001; Meyer *et al.*, 2004; Motelay-Massei *et al.*, 2004; Tayfur *et al.*, 2003; Therfeld *et al.*, 2003; Yücer y Demil, 2004; ...), suponen en definitiva, el establecimiento de una *estructura general de comportamiento* de las variables en base a las relaciones existentes entre ellas, lo que además de producir una pérdida de información, que puede ser crucial, debida a la reducción de dimensionalidad de datos y variables, a los efectos de este trabajo se traducen en una serie de supuestos que están en contra de las observaciones y el conocimiento previo disponible:

a) Continuidad de las variables, no considerando de esta manera la existencia de umbrales, geoquímicos, como los que se pueden producir en contactos entre materiales de composición muy diferente como pueden ser una roca ígnea básica, una metamórfica o una sedimentaria, o de rupturas y cambios de pendiente entre otros.

b) Existencia de relaciones funcionales en el conjunto de los datos, ya que la estructura obtenida, es la resultante de la combinación de todos los

valores disponibles, considerando así, de manera implícita, que las relaciones entre las concentraciones de los distintos elementos contaminantes es parecida en todos los tipos de sustrato o roca, que la alteración de distintos minerales, y por tanto la liberación de elementos contaminantes, sigue una dinámica similar, etc.

Aunque, en aquellos casos en los que se den las condiciones adecuadas de homogeneidad de sustrato, como de condiciones del medio, las técnicas de análisis multivariante se puedan aplicar con buenos resultados.

Los tres puntos siguientes: 2º) Desarrollo de un método de agrupación que siga simultáneamente criterios de similitud y continuidad. 3º) Elaboración de un sistema de detección de *outliers* e *inliers*. y 4º) Propuesta de un método de integración de datos que haga *compatibles* los resultados con otras metodologías. No presentan una solución tan fácil como el primero pues exigen el desarrollo de nuevas metodologías de cálculo, que se abordan el los capítulos siguientes.

1.4. Objetivos concretos.

Una vez identificados los problemas que se presentan a la hora de identificar suelos contaminados, estimar el grado de contaminación y generalizar los valores de referencia. En este trabajo se van a proponer tres objetivos que se supone mejoren el conocimiento en las tareas citadas:

1º Definir el concepto de *niveles característicos* como *los intervalos de valores de las concentraciones de metales pesados que guardan una mayor similitud en agrupaciones de puntos muestrales realizadas en base a variables explicativas* (ej. Litología, topografía, clima, etc.).

2º Desarrollar la metodología de análisis de datos necesaria para poder agrupar de forma objetiva las muestras y eventualmente realizar una clasificación.

3º Una vez caracterizadas las *concentraciones normales* de los distintos elementos, *niveles característicos*, en base a ellos localizar zonas de concentración anómala, que permita aseverar si se trata de puntos contaminados o no y en caso de que resulten positivos inferir en que medida lo están.

4º Ajustar un sistema de umbralización que permita representar y extender los niveles de referencia (*baselines*) a otras escalas de trabajo, siguiendo criterios de maximización de su probabilidad de aparición.