

Tratamiento numérico de la información hidrogeológica: fases de estudio y ejemplos de aplicación

R. Jiménez Espinosa

Departamento de Geología, Universidad de Jaén, Campus Las Lagunillas, Edificio B3, 23071, Jaén.
E-mail: respino@ujaen.es

RESUMEN

En este trabajo se ha realizado una recopilación de algunos de los métodos estadísticos y geoestadísticos de estimación usados en hidrogeología. En primer lugar, se lleva a cabo una introducción a las características de los datos hidrogeológicos, para pasar a comentar los pasos a seguir en el tratamiento numérico de los mismos, comenzando por las técnicas estadísticas univariantes, y posteriormente se realiza una breve descripción de los fundamentos teóricos de uno de los métodos estadísticos multivariantes de más amplia aplicación en hidrogeología, como es el análisis de componentes principales. Dentro de las técnicas geoestadísticas, comenzaremos por el análisis variográfico o estructural, para continuar con el desarrollo de algunos de los métodos de estimación geoestadística univariante, como son el krigeaje ordinario, el de indicatrices, el residual o el espacio-temporal, así como la estimación de los factores del análisis estadístico multivariante.

De forma conjunta, se van presentando una serie de casos de aplicación sobre datos hidrogeológicos reales de algunas de las técnicas analizadas. Veremos el tratamiento de datos de contaminación en varios acuíferos, así como, la deducción de los procesos hidrogeoquímicos actuantes en un acuífero aluvial. Con estas aplicaciones se mostrarán las diversas casuísticas surgidas al realizar una aplicación práctica y la elección de los métodos más adecuados para llevarla a cabo.

Palabras clave: estadística, geoestadística, hidrogeología, hidrogeoquímica, krigeaje

Numerical treatment of hydrogeological information: steps and case studies

ABSTRACT

A recompilation of several statistical and estimation geostatistical methods has been made in this work. First, an introduction of hydrogeological data characteristics has carried out. Later, the common steps followed in a numerical treatment was commented, univariate and multivariate techniques, like principal component analysis. Related to geostatistical methods, the starting point of study is the variographic analysis and then we developed several univariate techniques of estimation, such as, ordinary, indicator, residual kriging, and space-temporal kriging. Also, a revision of multivariate approach to treatment of hydrogeological data was made through cartography of multivariate indices.

In some chapters, we present a series of study cases of application of several commented techniques to hydrogeochemical data. Thus, we present the treatment of pollution in different aquifers and the deduction of hydrogeochemical processes controlling the behaviour in an alluvial aquifer. With these applications, a show of different points of view in the study of hydrogeological data has been presented, indicating the most suitable method in each study case.

Key words: geostatistics, hydrogeochemistry, hydrogeology, kriging, statistics

Introducción

La utilización de cualquier método numérico en el campo de los recursos naturales, ya sean hidrogeológicos, u otros, implica la caracterización de los mis-

mos por medio del conocimiento de la cantidad y calidad del propio recurso. Desde el punto de vista estadístico clásico, podemos aproximarnos a la distribución de la variable, pero sin tener en cuenta la localización y distribución espacial de la misma. Estos

métodos resultan de gran utilidad para conocer la población muestral de la que se dispone, describir cada una de las variables, detectar valores aberrantes, establecer asociaciones entre variables y/o muestras, etc. No obstante, cuando se trabaja con fenómenos naturales interesa, además, conocer la influencia de la posición de la muestra en el espacio y/o en el tiempo. Así, la *Geoestadística* tiene por objeto la caracterización de la dispersión espacial y/o temporal de los fenómenos naturales, modelizándolos para poder explotarlos adecuada y racionalmente.

Al trabajar con datos hidrogeológicos, surge la primera de las dificultades, inherente a la propia naturaleza de los datos, como es el elevado coste de adquisición de la información experimental y la no duplicidad de la misma. Nuestro conocimiento del problema hidrogeológico es un conjunto de datos discreto y limitado de muestras. Pero por otro lado, se dispone de un gran volumen de datos que hay que tratar, lo que implica necesariamente utilizar métodos numéricos para su proceso. Además, esta información es de carácter multivariante, añadiendo una dificultad adicional al manejo de la misma, y es habitual disponer de datos georeferenciados y, en ocasiones, temporales.

Los objetivos de cualquier aplicación numérica serán:

- Reducción de la variabilidad: obtener la máxima información con un menor número de variables puestas en juego;
- Obtención de modelos numéricos: estos modelos nos permitirán describir la realidad del fenómeno natural para la casuística concreta de nuestro estudio;
- Localización de zonas potencialmente interesantes: en el caso de los problemas habituales en Hidrogeología, esto se puede referir a la posible localización de áreas contaminadas del acuífero, establecimiento de flujos de agua, zonas de recarga, etc.
- Relación *geología* - modelo numérico: es muy importante que cualquier modelo que generemos sea coherente y consistente con la realidad geológica, por lo que un conocimiento del medio en el que desarrollamos la investigación es imprescindible.

Análisis exploratorio de datos

El análisis de los datos típicamente comienza con la representación de los mismos y el cálculo de los estadísticos básicos. De esta forma nos familiarizaremos con los datos y podremos detectar patrones de regu-

laridad, pero estos se analizan de forma más clara cuando están organizados y presentados de forma que se pueda sacar la mayor cantidad de información con el mínimo de esfuerzo. De ahí que, la realización de tablas con los estadísticos más usuales y de gráficos de diversos tipos, sea de gran ayuda para llevar a cabo este cometido.

Aparte de los clásicos estadísticos univariantes de tendencia central o de dispersión, cuando se trabaja con datos hidrogeológicos resulta práctico determinar una serie de estadísticos robustos y resistentes:

- *histogramas y estadísticos de orden*: mín, Q1, Q2, Q3, RIC y máx (Fig. 1a)
- *gráficos de cajas o "box-plots"* (Fig. 1b)

Uno de los objetivos de este análisis exploratorio será determinar y definir la existencia de valores extremos, los denominados "outliers" estadísticos. Estos pueden ser de varios tipos: *"outliers verdaderos"*: forman parte del conjunto de datos, aunque son observaciones discordantes con la tendencia general, pero tienen una explicación geológica. Por ejemplo, una la presencia de un contaminante puntual en un acuífero, que nos daría un valor medido en el pozo próximo muy elevado respecto a los de alrededor. *b."outliers contaminantes"*: son datos explicables geológicamente, pero no pertenecen completamente al conjunto de datos, sino que forman parte de una unidad geológica diferente no identificada, y que nos puede hacer sospechar de la existencia de algo "raro". Un caso de este tipo sería en el de estar trabajando con dos acuíferos confinados superpuestos, de los cuales desconocemos la existencia de uno de ellos. *c."falsos outliers"*: no son datos reproducibles y suelen estar debidos a errores computacionales, de transcripción, mal muestreo, etc.

Curva de distribución de frecuencias

Con este gráfico (Fig. 1c) es posible detectar valores anómalos de dos categorías (Royer, 1984): BAG *"begining anomalous grade"*: aquellos puntos que caen en el tramo a partir del cual la curva de distribución sufre un punto de inflexión. SAG *"significant anomalous grade"*: obtención gráfica de la curva de frecuencias o se puede escoger arbitrariamente como el doble de BAG.

Descripción estadística multivariante

Los métodos estadísticos multivariantes aplicados a datos hidrogeológicos, además de proporcionar una forma rápida y cómoda de visualizar e interpretar la

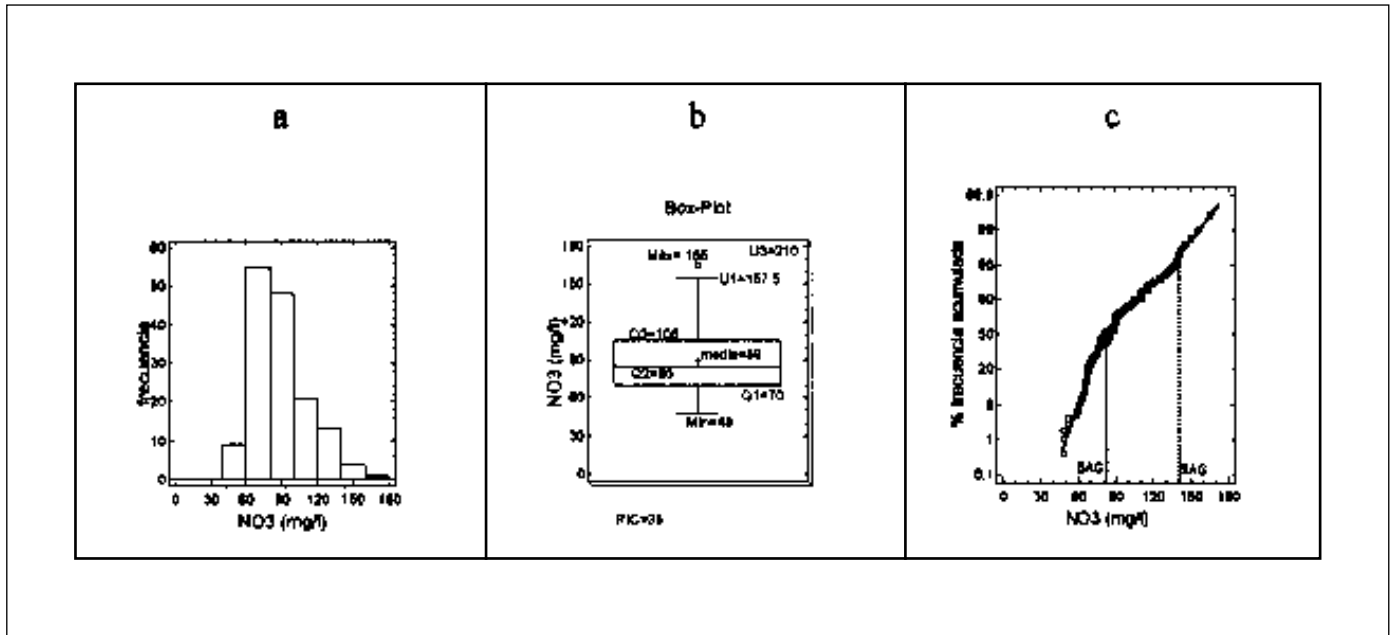


Fig. 1. Análisis exploratorio de datos del contenido en nitratos en el Acuífero Detrítico del Bajo Andarax. (a) histograma; (b) gráfico de cajas o "box-plot"; (c) curva de frecuencia acumulada
 Fig. 1. Exploratory data analysis of nitrate contents in Detrital Aquifer of Bajo Andarax. (a) histogram; (b) box-plot; (c) cumulative frequency curve

información multielemental para un gran número de muestras, implica la posibilidad de definición de patrones de carácter multivariable. Por ejemplo, las características de calidad de un agua subterránea en un área de estudio se manifiestan por la variación en la concentración, no de un elemento de forma individual, sino de una serie de elementos asociados entre sí. Por lo tanto, es necesario definir umbrales multielementales que separen anomalías y fondos de variación de carácter multivariable. Existen numerosos métodos de análisis estadístico multivariante, y no es el objeto de este trabajo hacer un resumen de todos ellos, sino que se lleva a cabo un planteamiento de trabajo usando un número limitado de técnicas aplicadas a casos reales de estudios hidrogeológicos. Por las características de los conjuntos de datos de los que se suelen disponer en este tipo de trabajos, se ha optado por considerar el análisis de componentes principales y el de correspondencias como unos de los métodos más apropiados para el tratamiento de datos hidrogeológicos.

Análisis de componentes principales

El *análisis de componentes principales* (ACP) es un método descriptivo de análisis de datos ampliamente usado en estudios geoquímicos (Esbensen *et al.*,

1987; Pereira y Sousa, 1988; Jiménez Espinosa *et al.*, 1993; 2001; entre otros muchos) y cuyo principal objetivo es separar las asociaciones de elementos inherentes en la matriz de correlación en una serie de grupos de elementos que sean capaces de explicar parte de la variabilidad observada en los datos originales, cuando se realiza el estudio en el llamado modo *R*. El propósito final es representar un gran número de elementos de los datos originales a través de un pequeño número de ejes factoriales denominados *componentes principales*, cada uno de los cuales es una función lineal (transformación) de los datos de partida, dando así una gran eficiencia en términos de síntesis de la información inicial, además de ser una herramienta útil desde el punto de vista de la interpretación. Otra posibilidad de estudio es ver las relaciones existentes entre las muestras, por medio del estudio en modo *Q*. En definitiva, la idea básica del ACP es proyectar las muestras y las variables en gráficos planos (2D) definidos por un pequeño número de ejes, minimizando la pérdida de información.

En el ACP, la matriz de partida *Z* puede ser tomada como una nube de *n* puntos en R^p o como una nube de *p* puntos en R^n . Como cualquier método descriptivo de análisis de datos, el ACP tiene como objetivo reducir la dimensionalidad del espacio en el que las muestras y variables están proyectados, perdiendo el mínimo de información. Para reducir la dimensión del

espacio se transforma el sistema de referencia de partida, de forma tal que la proyección de la nube en esos nuevos ejes (llamados componentes principales o ejes factoriales) va siendo sucesivamente menor, a medida que el orden de extracción de los ejes factoriales aumenta. Así, se obtiene el primer eje factorial, como aquel vector en el que la suma de los cuadrados de las proyecciones de la nube en ese eje es máxima. La segunda componente principal está contenida en el hiperplano perpendicular a la primera CP, y es en la dirección de ese hiperplano donde la suma de los cuadrados de las proyecciones de la nube en ese eje es máxima. El algoritmo continúa desarrollándose del mismo modo hasta extraer todos los ejes de la nube (ya sea p en R^p o n en R^n).

A través de este algoritmo, la nube de las n muestras queda así referida a un nuevo sistema de referencia de p ejes, tal que los ejes de orden elevado (en general superior a 3) contienen un porcentaje poco importante de la información sobre la localización de las muestras en el espacio. Si se acepta despreciar la información contenida en los ejes de orden superior a $p_1 < p$, se retienen sólo los p_1 primeros ejes, por lo que se puede hacer una descripción de las muestras en el nuevo sistema de referencia, en el cual la interpretación se facilita, ya que se analiza un pequeño número de gráficos planos cruzando las proyecciones en los ejes retenidos. Por otro lado, al escoger p_1 se conoce la pérdida de información inherente al método, que es la suma de los cuadrados de las proyecciones en los ejes de orden superior a p_1 . Se intenta establecer un equilibrio entre una pérdida de información mínima (lo que exige un máximo de ejes) y una facilidad de interpretación máxima (lo que exige la retención de un pequeño número de ejes).

Estudio geoestadístico univariante

Todo el análisis realizado hasta el momento se ha basado en considerar los datos como valores numéricos independientes de su posición en el espacio. De cada muestra en una campaña de hidrogeología se conocen sus coordenadas geográficas, de ahí que sea importante tener en cuenta esta información espacial como una variable añadida al estudio. Desde esta nueva perspectiva, la geoestadística se presenta como un conjunto de técnicas apropiadas para abordar la problemática del estudio de datos hidrogeológicos. Obviamente, todo el tratamiento estadístico previo es de gran utilidad y no deben considerarse los métodos geoestadísticos como sustitutos de los estadísticos clásicos, sino como complementarios, ya que el correcto conocimiento del conjunto de datos

es un paso importante de cara a una fiel aplicación de los métodos geoestadísticos más convenientes. Una posible opción en el tratamiento geoestadístico univariante de datos hidrogeológicos puede implicar seguir una serie de pasos elementales (Fig. 2).

- *Análisis estructural o variográfico*: es el punto de partida de un estudio geoestadístico y supone una etapa fundamental, pues en ella se calculan los variogramas experimentales y teóricos.
- *Estimación geoestadística o krigeaje*: este es uno de los objetivos básicos de la geoestadística, y resulta de gran utilidad para llevar a cabo la identificación del problema hidrogeológico. Existen numerosos métodos de estimación y todos ellos reciben el nombre genérico de krigeaje o krigeado, aunque los más comunes suelen ser métodos basados en interpolación por medio de combinaciones lineales ponderadas de los datos experimentales, para lo cual el krigeaje utiliza también la información procedente del análisis estructural.
- *Interpretación de resultados*: es la etapa que lleva a la identificación del problema que se está tratando, utilizándose toda la información disponible y los resultados del tratamiento realizado sobre las diferentes variables.

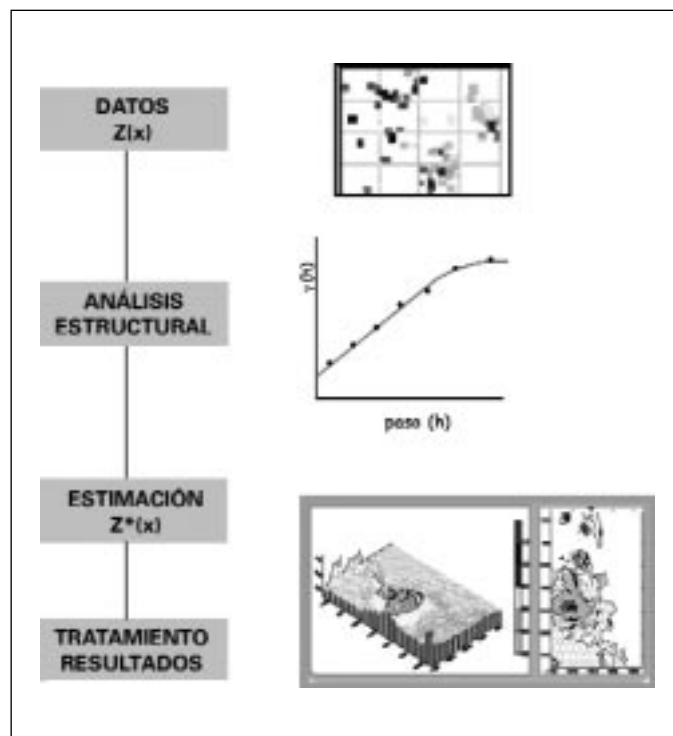


Fig. 2. Pasos esquemáticos en la realización de un estudio geoestadístico de datos hidrogeológicos
 Fig. 2. Schematic steps in the performance of a geostatistical study of hydrogeologic data

Análisis estructural o variográfico

El análisis estructural es el proceso de definición del modelo geoestadístico en el marco de los conceptos de la teoría de funciones aleatorias. Por lo tanto, supone tener en cuenta el tipo de hipótesis que se va a hacer sobre la variabilidad del fenómeno en estudio, es decir, implica determinar si la variable se puede considerar estacionaria o si se requiere la definición de una tendencia y, en caso de requerirla, la forma que tendrá esta deriva. Todo este proceso, recogido bajo la denominación de análisis estructural, conlleva una serie de pasos: (i) *inspección de los datos y verificación*: es necesario tener un conocimiento sobre los datos disponibles de la variable que va a ser objeto de estudio, esto implica reexaminar aquellas observaciones que puedan ser cuestionables, debido a errores de muestreo, de transcripción o cualquier otro tipo de sesgo introducido a los datos; (ii) *estimación de los valores del variograma*: basándose en las observaciones existentes, se estiman los valores de la función variograma, además, el variograma experimental se suele calcular en varias direcciones del espacio para identificar posibles anisotropías; (iii) *modelización del variograma*: teniendo en cuenta las características de la variable y del variograma experimental, se establece un modelo de variograma, cuyos parámetros se estiman a partir de los valores del variograma experimental; (iv) *validación del modelo impuesto*: la función variograma puede ser validada por medio de técnicas aplicadas directamente sobre las observaciones, por ejemplo por validación cruzada.

La función variograma se define como una función aleatoria intrínseca que representa la mitad de los incrementos cuadráticos medios de la variable para puntos distantes h :

$$\gamma(h) = \frac{1}{2} \text{Var}\{Z(x+h) - Z(x)\} \quad (1)$$

Para funciones estacionarias $E\{Z(x+h) - Z(x)\} = 0$, $\gamma(h)$ vale entonces:

$$\gamma(h) = \frac{1}{2} E\{[Z(x+h) - Z(x)]^2\} \quad (2)$$

con $\gamma(h) = 0$ cuando $h = 0$.

La importancia de esta etapa radica en dos cuestiones fundamentales, (i) el variograma es una herramienta básica para informar sobre las características espaciales del fenómeno, y esta información gráfica

está estrechamente relacionada con las características hidrogeológicas e hidrogeoquímicas de la variable, debiendo estar de acuerdo los resultados procedentes del estudio variográfico con la realidad física; (ii) para realizar cualquier aplicación geoestadística es necesario usar un modelo teórico (ver Chica Olmo, 1988).

Krigeaje ordinario

La particularidad de este método es que se utilizan funciones aleatorias estacionarias de esperanza desconocida, en donde se emplea un modelo probabilístico en el que los pesos asignados a las muestras de la vecindad del punto a estimar se escogen de tal forma que se asegura que el error medio cometido sea cero, y que la varianza del error sea minimizada, y es precisamente, la aproximación al problema por medio de las funciones aleatorias la que permite expresar el error, su media y su varianza (Matheron, 1970). La información disponible consta, por una parte, de una información experimental, compuesta por una serie de valores numéricos $\{Z(x_\alpha), \alpha = 1 \dots n\}$, medidos en distintos puntos del dominio; y por otro lado, una información estructural, relativa a la FA Z e integrada por el variograma. El sistema de krigeaje ordinario queda expresado en los siguientes términos:

$$\sum_{\beta=1}^n \lambda_\beta \bar{\gamma}(v_\alpha, v_\beta) + \mu = \bar{\gamma}(v_\alpha, x_0), \quad \forall \alpha, \beta = 1, n \quad (4)$$

$$\sum_{\beta=1}^n \lambda_\beta = 1$$

y la correspondiente varianza de estimación es:

$$\sigma_k^2 = \sum_{\alpha=1}^n \lambda_\alpha \bar{\gamma}(v_\alpha, x_0) + \mu - \bar{\gamma}(x_0, x_0) \quad (5)$$

Ejemplo de aplicación del krigeaje ordinario: Acuíferos del Campo de Dalías

El Campo de Dalías se sitúa en la provincia de Almería, entre las cuencas vertientes de los ríos Adra y Andarax, ocupando una superficie de unos 330 km². El límite septentrional está constituido por las estribaciones de la Sierra de Gádor, mientras que los bordes restantes los ocupa el mar. En el Campo existen más de 20.000 hectáreas de cultivos bajo plástico y unas 3.500 al aire libre, lo que supone unas impor-

tantes necesidades de agua para el regadío de estos terrenos, que procede de los acuíferos de la zona.

El objetivo de este ejemplo (Jiménez Espinosa, *et al.*, 1996) es mostrar la evolución espacial y temporal de los nitratos en las aguas subterráneas del Campo de Dalías, definiendo dos sistemas acuíferos en la zona, uno profundo, S1, y otro superficial, S2, realizando su cartografía por krigeaje ordinario.

El análisis espacial del sistema S1 a través de los diferentes meses (Fig. 3) indica que el área correspondiente al acuífero de Aguadulce aparece como la de mayor concentración en nitratos. Por otra parte, en la parte O del Campo, en la zona de Balanegra (acuífero de Balanegra) es donde se localizan los valores más altos. En esos ámbitos se produce una evolución desde zonas elevadas más o menos puntuales hasta una tendencia de aumento generalizado por todo el área de estudio, apareciendo zonas más amplias con

valores elevados en noviembre de 1990. En relación con el sistema S2 (Fig. 4), los contenidos en nitratos son lógicamente mayores. En este sistema la evolución es diferente a S1, pues se pasa de tener unas concentraciones elevadas en 1988 y 1989, a reducir estos valores, siendo noviembre de 1990 el mes que menos áreas con valores elevados presenta. En este sistema aparecen otra vez las zonas oriental y occidental del Campo de Dalías (zona de Roquetas y Balanegra, respectivamente) con gran concentración de valores máximos.

La conclusión importante que se deduce de este estudio es que la evolución temporal de los contenidos en nitratos en estas aguas es creciente en los niveles acuíferos profundos. El origen de esta contaminación se encuentra principalmente en el uso de fertilizantes para la agricultura intensiva de la comarca.

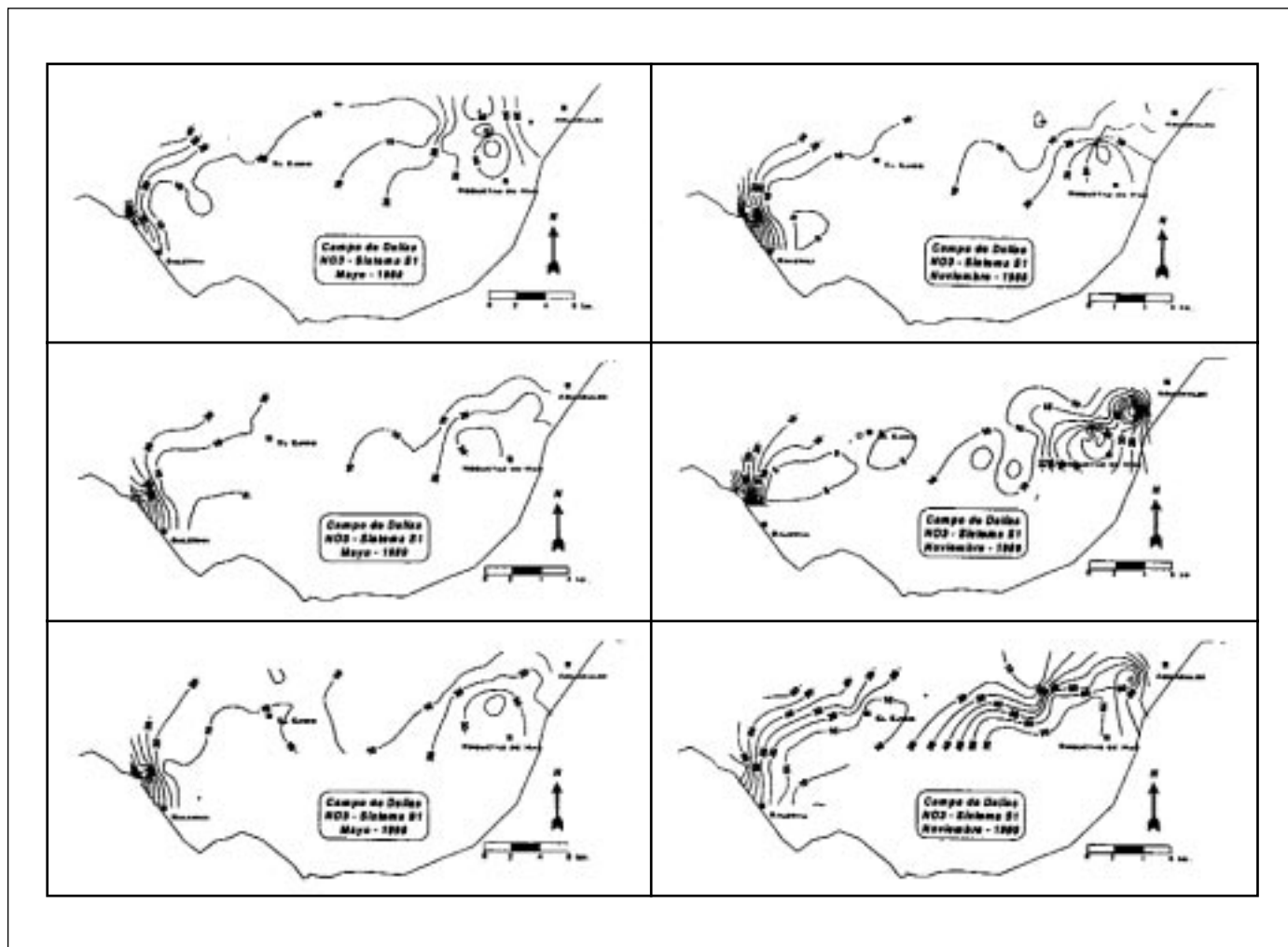


Fig. 3. Mapas de evolución temporal de nitratos en el sistema acuífero S1 del Campo de Dalías
 Fig. 3. Temporal evolution maps of nitrate in the Campo de Dalías aquifer system S1

Krigeaje de indicatrices

Como ya se ha comentado, uno de los problemas que conllevan los datos hidrogeológicos es que las variables suelen presentar distribuciones bastante sesgadas, con largas colas y rangos de variación importantes, normalmente debido a la existencia de algunos valores extraños. En estas ocasiones los variogramas experimentales son bastante sensibles a la presencia de este tipo de datos. Una solución interesante para evitar estos problemas puede ser considerar métodos de estimación no paramétricos libres de distribución, como el krigeaje de indicatrices (Journel, 1983). La función indicatriz se define como una función establecida a partir de una codificación binomial, definida para un determinado valor de corte z :

$$i(x; z) = 0, \text{ si } z(x) \leq z, \text{ si } z(x) > z \quad (6)$$

La proporción exacta de valores $z(x)$ que aparecen bajo el límite z , dentro de un área A determinada puede escribirse como:

$$\phi(A; z) = \int_A i(x; z) dx \in [0, 1] \quad (7)$$

La función variograma de estas funciones se puede definir como:

$$\gamma_I(h; z) = \frac{1}{2} E\{[I(x+h; z) - I(x; z)]^2\} \quad (8)$$

El krigeaje de las distribuciones espaciales implica la aplicación de los algoritmos clásicos a estas nuevas variables, siendo el más usado el del krigeaje ordinario:

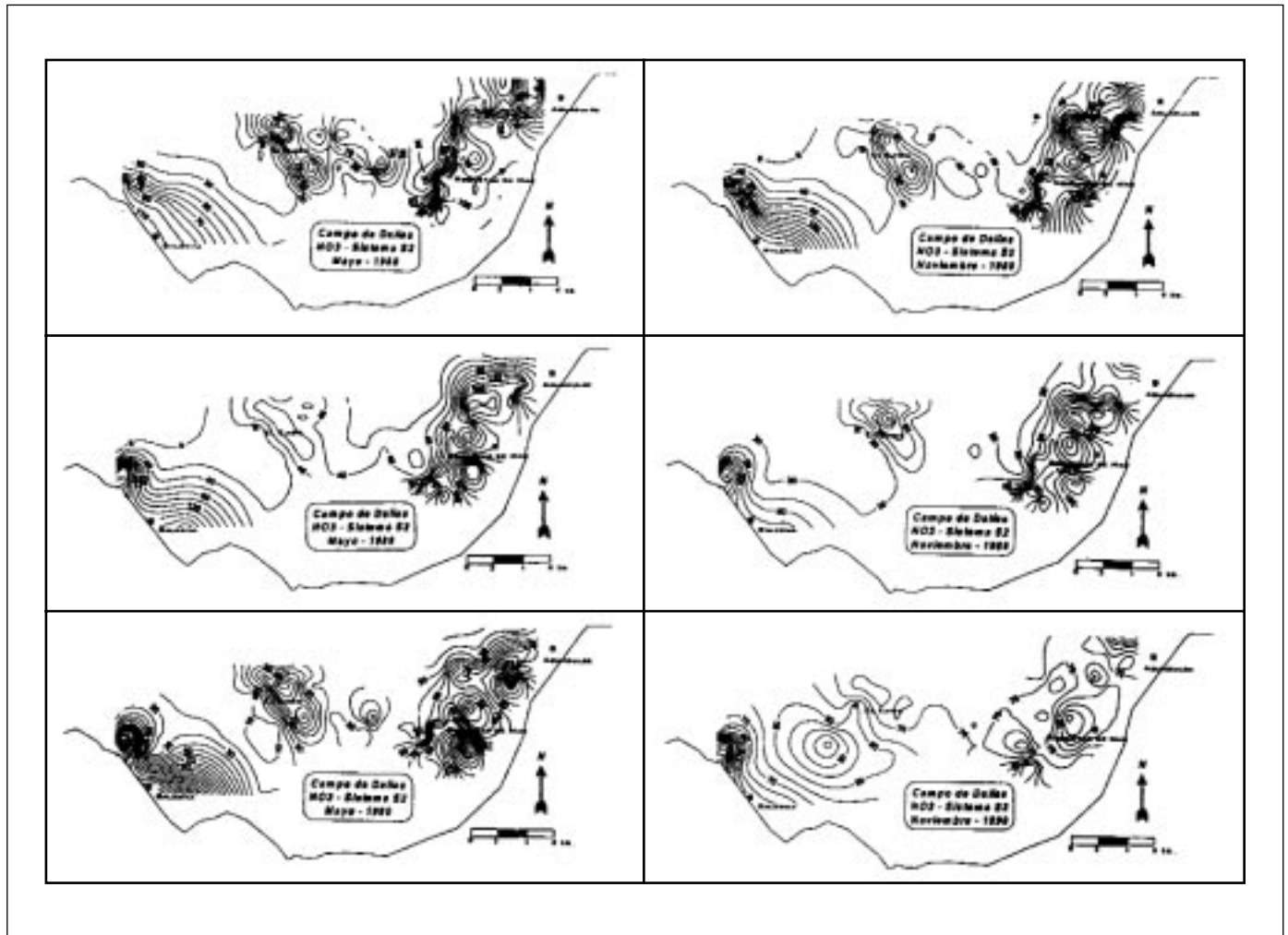


Fig. 4. Mapas de evolución temporal de nitratos en el sistema acuífero S2 del Campo de Dalías
 Fig. 4. Temporal evolution maps of nitrate in the Campo de Dalías aquifer system S2

$$\sum_{\beta=1}^n \lambda_{\beta}(x; z) C_I(v_{\alpha}, v_{\beta}; z) - \mu = C_I(v_{\alpha}, V; z), \quad \forall \alpha, \beta = 1, n$$

$$\sum_{\beta=1}^n \lambda_{\beta} = 1 \quad (9)$$

sistema de $n+1$ ecuaciones con $n+1$ incógnitas, que son los n ponderadores λ_{α} y el parámetro de Lagrange μ .

La varianza de estimación o varianza de krigeaje se podría expresar del siguiente modo:

$$\sigma_K^2 = C_I(V, V; z) + \mu - \sum_{\alpha=1}^n \lambda_{\alpha}(x; z) C_I(v_{\alpha}, V; z) \quad (10)$$

Krigeaje residual

Una función aleatoria $Z(u)$ es estrictamente estacionaria si su función de distribución es invariante respecto a cualquier translación de vector h . En el caso de una función aleatoria no estacionaria (su esperanza matemática depende de u) esta variable $Z(u)$ puede expresarse como la suma de una componente determinista $m(u)$ y de un residuo $R(u)$ estacionario con media nula y variograma:

$$\gamma_{\varepsilon}(h) : \gamma_{\varepsilon}(h) = \frac{1}{2N} \sum_{i=1}^N [\varepsilon(u) - \varepsilon(u+h)]^2 \quad (11)$$

Entre los diferentes métodos existentes para el tratamiento de funciones aleatorias no estacionarias se puede utilizar el krigeaje residual ordinario que se resume en los siguientes pasos (Samper y Carrera, 1990; Fernández-Cortés, *et al.*, 2002):

1. Suponer conocida la forma de la deriva basándose en razonamientos físicos.
2. Ajustar por mínimos cuadrados ordinarios la deriva $m(u)$ a los datos medidos.
3. Calcular los residuos $R(u)$ del ajuste anterior mediante la siguiente expresión: $R(u) = Z(u) - m(u)$.
4. Calcular el variograma experimental, $\gamma_{\varepsilon}^*(h)$, de los residuos ajustándole un variograma teórico $\gamma_{\varepsilon}(h)$.
5. Realizar el krigeaje ordinario de los residuos utilizando el anterior variograma teórico.
6. Obtener el valor estimado de variable $Z^*(u_0)$ en un punto no muestreado u_0 como la suma del valor de la deriva $m(u_0)$ calculada en el paso (2) más el valor estimado del residuo, $R(u_0)$.

Estudio geoestadístico multivariante de los factores del análisis de datos

La aproximación al estudio de las variables hidrogeo-

lógicas se ha llevado a cabo desde esta doble perspectiva, considerando de forma independiente ambos tipos de metodologías, realizando primero un estudio estadístico, y posteriormente, un estudio geoestadístico. En este apartado se va a abordar toda la problemática de forma conjunta, englobando los dos tipos principales de técnicas empleadas bajo una misma metodología que consiste en usar los resultados de las técnicas de análisis de datos para realizar el estudio geoestadístico. Para ello se usan los factores procedentes de los métodos multivariantes y se utilizan para realizar el estudio geoestadístico: análisis estructural y estimación por krigeaje.

Este tipo de tratamiento numérico de la información implica una serie de ventajas respecto al enfoque separado realizado hasta el momento. Permite la utilización conjunta de los dos tipos de técnicas que se han mostrado como las más apropiadas para tratar los datos de prospección geoquímica, pues las nuevas variables, resultado de combinaciones lineales de las variables originales, son las empleadas para la aplicación geoestadística. Esto supone una importante mejora respecto a la geoestadística multivariante, pues en una sola variable queda recogida toda la información multivariable correspondiente a una asociación de elementos; mientras que desde la óptica de la geoestadística multivariante es necesario realizar una serie de pasos intermedios para llegar a obtener los resultados del cokrigeaje.

Ejemplo de aplicación: Acuífero detrítico del Bajo Andarax

El estudio de la evolución hidrogeoquímica del Acuífero detrítico del Bajo Andarax (Sánchez Martos, *et al.*, 2001) presenta una serie de dificultades relacionadas con la diversidad de sus aguas, puesto que existen procesos termales, intrusión marina, altos contenidos en sulfato y boro. A esto se une la complejidad geológica del área. A partir de la información multivariante y compleja, mediante un análisis de componentes principales, se han establecido una serie de variables factoriales que sintetizan la información hidrogeoquímica. Sobre estas nuevas variables se ha realizado un estudio geoestadístico que permite trabajar en un espacio multivariable más reducido y pretende establecer la distribución espacial de las variables a lo largo del acuífero mediante el cálculo de sus variogramas y el diseño de mapas de calidad de las aguas subterráneas usando krigeaje ordinario. De esta manera es posible estudiar la evolución hidrogeoquímica del Acuífero detrítico considerando todos los parámetros hidrogeoquímicos y señalando espacialmente el desarrollo de los principales procesos.

Además, el establecimiento de estas nuevas variables permite localizar aquellos sectores donde pudieran superponerse varios procesos físico-químicos.

F1: la cartografía de F1 (Fig. 5) permite diferenciar nítidamente dos regiones: (1) la zona de la rambla de Gador-Santa Fe (zona NO de la cuenca), con bajos contenidos salinos; y (2) la zona de la rambla de Tabernas y su prolongación desde dicha rambla hacia el propio cauce del río Andarax, hasta llegar a la desembocadura del río en el área del delta, con mayores concentraciones iónicas. Esta distribución espacial puede relacionarse claramente con el tipo de material drenado en cada caso. En primer lugar, en el área de Gergal-Santa Fe, la distribución espacial de la variable F1 se corresponde con el tramo del acuífero cuya litología corresponde con depósitos fluviodeltáicos pliocuaternarios y la alimentación está directamente relacionada con las aguas superficiales del río Andarax, lo que condiciona la baja salinidad de sus aguas. Por otro lado, las zonas con valores mayores de las variables F1 corresponden con los afloramientos miocenos y pliocenos margoso-limosos con evaporitas de la rambla de Tabernas. También aquí se alcanzan los contenidos más elevados en boro y está también relacionado con altos contenidos en sulfatos (Sánchez Martos *et al.*, 1999). Además, el ión nitrato también está correlacionado con la variable F1, por lo que se observan pequeños focos puntuales de valores elevados en la zona de Huerca-Benahadux, debido a posible contaminación agrícola. Respecto a la variación temporal de F1 en los dos muestreos dispo-

nibles, la tendencia general se mantiene constante, no obstante, parecen acentuarse las concentraciones en septiembre, debido principalmente a que el muestreo se realizó al comienzo del año hidrológico y antes del período de lluvias, por lo que las concentraciones en elementos salinos son mayores que en mayo, algo más diluidas por la entrada de agua limpia en el acuífero.

F2: analizando la cartografía de esta variable F2 (Fig. 6), las mayores concentraciones de este factor aparecen principalmente en la margen derecha del río Andarax, al sur de la rambla de Tabernas. Justamente esta zona coincide con la de la mayor concentración de fracturas en el entorno de la cuenca del Bajo Andarax. Esta distribución espacial para esta variable puede relacionarse con la circulación de aguas más profundas al resto de las del Acuífero Detrítico, favorecida por las existencias de estas numerosas fallas de dirección NO-SE que condicionan este flujo de calor que aumenta la temperatura de las aguas y permite la movilización del litio, elemento que presenta una elevada movilidad geotérmica. Con relación a una posible variación temporal en el acuífero, básicamente la distribución espacial es bastante aproximada en ambos meses, aunque valores algo más elevados en septiembre, debido posiblemente a la necesidad de sacar aguas de niveles piezométricos más profundos en una época con menor disponibilidad de agua en el acuífero.

F3: en el mapa de esta variable (Fig. 7) aparecen dos regiones localizadas principalmente al sur de la

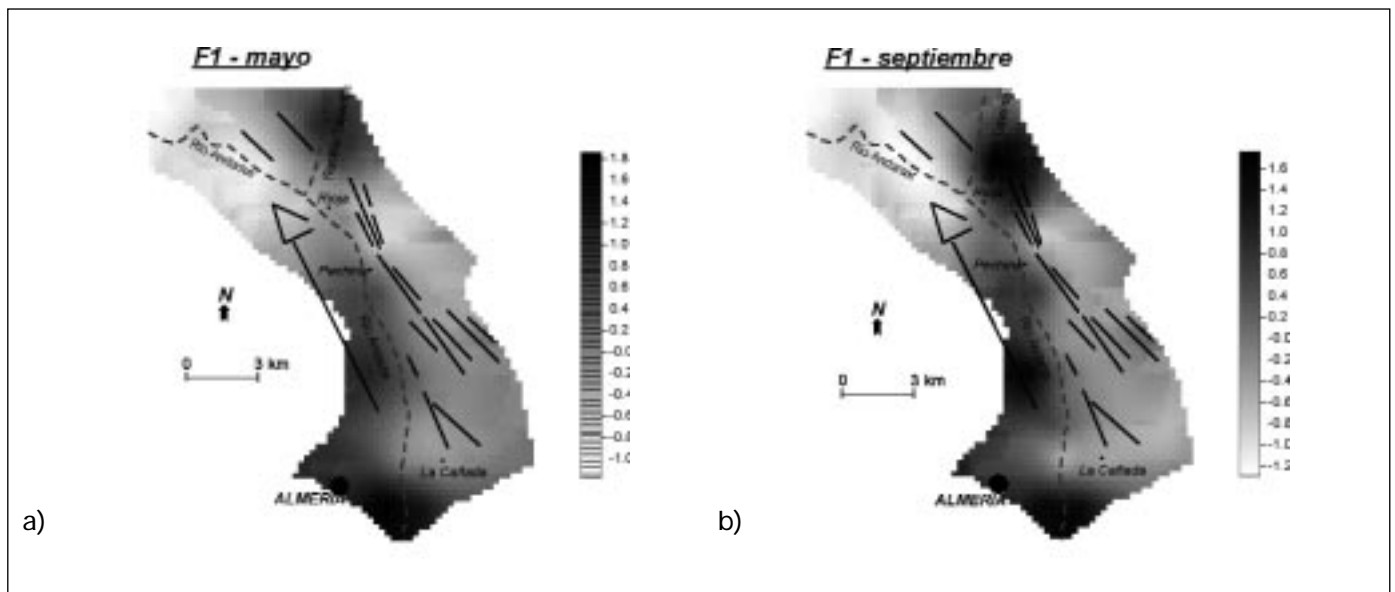


Fig. 5. Cartografía de la componente F1: a) mayo y b) septiembre
 Fig. 5. Plot of the F1 component: a) may and b) september

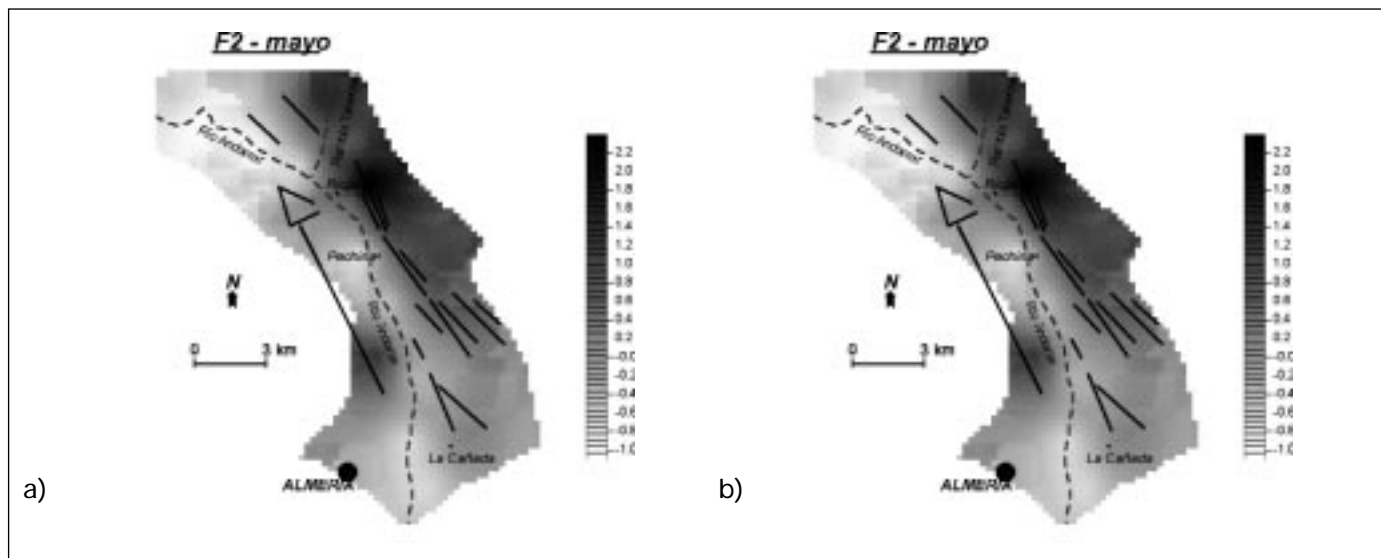


Fig. 6. Cartografía de la componente F2: a) mayo y b) septiembre
 Fig. 6. Plot of the F2 component: a) may and b) september

cuenca, coincidiendo con el delta y las áreas costeras adyacentes: (1) un área correspondiente al trazado del río hasta la desembocadura que aparece con valores bajos; (2) las zonas próximas al delta que presentan mayores concentraciones en F3. Los valores bajos del trazado del río corresponden con las aguas de menor salinidad, en relación con el flujo subterráneo que se produce a lo largo del acuífero detrítico. Este flujo subterráneo tiene menor salinidad que las aguas del sector costero y sus aguas son diferentes puesto que tienen una componente sulfatada, mientras que en el delta su naturaleza es clorurada.

Esta distribución espacial de la variable refleja la complejidad hidroquímica de las aguas de la zona sur del acuífero. La zona del delta puede recoger una mezcla de aguas:

- (i) por un lado, a la zona de la desembocadura propiamente dicha llegarían las aguas procedentes del enriquecimiento en sulfatos de la rambla de Tabernas;
- (ii) por otra parte, la demanda de agua en la zona es elevada debido a la intensa actividad agrícola y a la existencia de importantes núcleos de población, entre otros Almería capital. Esto provoca que en ciertas épocas del año se alcancen valores piezométricos negativos. Si además tenemos en cuenta la asociación de elementos en la variable F2: cloruros, sodio y conductividad, se puede considerar que un factor determinante en el quimismo de estas aguas haya sido la intrusión marina mezclado con el lavado de depósitos marinos salinos cuaternarios.

Esta mezcla de procesos ha podido aclararse por medio del ACP que ha permitido separar dos factores diferentes, uno que engloba a los procesos de enriquecimiento por yesos (variable F1) y otro factor que asocia a los procesos propios de intrusión salina y/o lavado de las sales cloruradas de depósitos marinos Cuaternarios.

La variabilidad temporal de F3 es prácticamente inapreciable, manteniéndose los valores aproximadamente constantes a lo largo del período de estudio considerado.

Modelos geoestadísticos espacio-tiempo

La conjugación de ambas dimensiones, temporal y espacial, dará como resultado un modelo espacio-temporal de la variable hidrogeológica (Fernández Cortés, *et al.*, 2002).

Al conjugar el espacio y el tiempo en el estudio de un fenómeno es necesario definir un nuevo dominio denominado espacio-temporal (Kyriakidis y Journel, 1999). Se considera un dominio finito D en el espacio y un dominio finito en el tiempo T , donde $D \subseteq R^d$ y $T \subseteq R^t$, y generalizando se considera un espacio bidimensional ($d=2$). Una variable aleatoria espacio-temporal $Z(u,t)$ es una variable que puede tomar una serie de realizaciones en cualquier punto del espacio $u \in D$ y en cualquier instante de tiempo $t \in T$, de acuerdo con una función de probabilidad. Esta variable $Z(u,t)$ está perfectamente caracterizada por una función de distribución acumulada ($F(u,t;Z)$) que da la

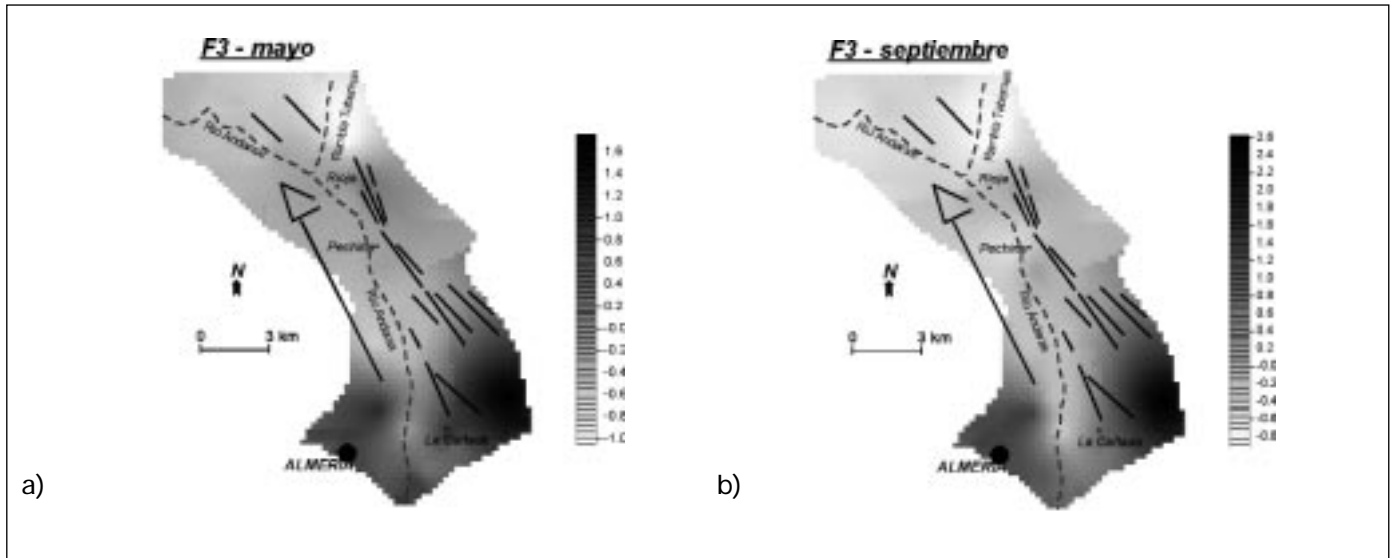


Fig. 7. Cartografía de la componente F3: a) mayo y b) septiembre
 Fig. 7. Plot of the F3 component: a) may and b) september

probabilidad de que la variable Z que en la localización u en el espacio y en el instante t en el tiempo no sea mayor a un límite z : $F(u,t;z) = \Pr \{Z(u,t) \leq z\}, \forall z, (u,t) \in DxT$.

La variable aleatoria en el espacio y en el tiempo se descompone en una componente media modelizada de la tendencia $m(u,t)$ y en una componente residual $R(u,t)$ que modeliza las fluctuaciones entorno a la tendencia tanto en el tiempo como en el espacio. Dependiendo de si la estimación de la tendencia se aborda de una forma determinista o estocástica, tendremos dos clases de modelos de función espacio-temporal individual: determinista o estocástico, respectivamente. La componente determinista de la tendencia $m(u,t)$ es a menudo expresada como;

$$m(u,t) = \sum_{l=0}^L \sum_{k=0}^K b_{kl} f_{kl}(u,t), \forall (u,t) \in DxT \quad (12)$$

donde $m(u,t)$ está comprendido en el dominio KL, $f_{kl}(u,t)$ elegidas como aceptables para el ajuste de la variación media de los datos de la variable aleatoria $Z(u)$ y $b_{kl}, k = 0, 1, \dots, K, l = 0, 1, \dots, L$. son los coeficientes que deben de ser ajustados.

Los modelos de tendencia:

$$m(u,t) = \sum_{l=0}^L \sum_{k=0}^K b_{kl} f_{kl}(u,t), \forall (u,t) \in DxT \quad (13)$$

son considerados probabilísticos si los coeficientes b_{kl}

que multiplican las funciones deterministas $f_{kl}(u,t)$ son estimados como variables aleatorias;

$$M(u,t) = \sum_{l=0}^L \sum_{k=0}^K B_{kl} f_{kl}(u,t), \forall (u,t) \in DxT$$

con

$$E\{B_{kl}\} = b_{kl}, k = 0, 1, \dots, K, l = 0, 1, \dots, L \quad (14)$$

Por lo tanto los valores esperados de la tendencia pueden ser estimados directamente de los valores medidos de la variable $Z(u,t)$ mediante regresión. De forma general los modelos con una componente media estocástica pueden ser expresados como:

$$Z(u,t) = M(u,t) + R(u,t) = M_u(t) + R(u,t), \forall (u,t) \in DxT \quad (15)$$

La tendencia temporal determinista $m_u(t)$ en cada localización u_u es en primer lugar ajustada por mínimos cuadrados, independientemente de otras localizaciones, usando funciones polinómicas de tiempo. Los coeficientes calculados en estos polinomios son regionalizados en el espacio mediante krigeaje, consiguiendo de esta forma caracterizar la tendencia espaciotemporal $M(u,t)$ para localizaciones u en el espacio e instantes t en el tiempo no muestreados.

La otra componente necesaria para definir la variable aleatoria en el espacio y en el tiempo es el residuo $R(u,t)$ o variación de la variable respecto a la tendencia anteriormente modelizada. Bajo un modelo espaciotemporal de media estacionaria m , la covarianza espaciotemporal $C_R(h,\tau)$ de la función de probabilidad de los residuos $R(u,t)$ puede ser asimilada a la covarianza estacionaria $C(h,\tau)$ de la función de probabilidad de la variable aleatoria $Z(u,t)$;

$$C_R(h,\tau) = E\{R(u,t) * R(u+h,t+\tau)\} = E\{[Z(u,t) - m] * [Z(u+h,t+\tau) - m]\} = C(h,\tau) \quad (16)$$

La función de covarianza es una medida de la correlación de los datos. En este caso un cálculo de la continuidad espaciotemporal de los residuos mediante la función covarianza permite la caracterización de la componente residual $R(u,t)$ del modelo de $Z(u,t)$ con componente estocástica

$$[Z(u,t) = M(u,t) + R(u,t) = M_u(t) + R(u,t)] \quad (17)$$

Conclusiones

En este trabajo se presenta una metodología de tratamiento de la información hidrogeológica, que va desde el momento en que se dispone de la información experimental hasta que se presenta un modelo coherente con la realidad geológica de la zona de trabajo. En este sentido, se han planteado diferentes métodos de estudio para abordar las diferentes causas presentes en un estudio hidrogeológico, que abarca un amplio espectro de posibilidades simultáneas o no, como son información espacial, multivariante, temporal, etc.

Se han comentado las técnicas de tratamiento estadístico clásicas, y diferentes métodos de estimación geoestadística, ofreciendo varios ejemplos de aplicación que presentan las ventajas de las técnicas empleadas y que pueden servir de guía para usuarios no habituales de estos tipos de tratamientos.

Referencias

Chica Olmo, M. 1988. *Análisis geoestadístico en el estudio*

- de la exploración de los recursos minerales*. Tesis Doctoral. Universidad de Granada, 387 pp.
- Esbensen, K., Lindquist, L., Lundholm, L., Nisca, D. y Wold, S. 1987. Multivariate modelling of geochemical and geophysical exploration data. *Chernometrics and Intelligent Laboratory Systems*, 2, 161-175.
- Fernández Cortés, A., Calaforra, J.M., Jiménez Espinosa, R., Sánchez Martos, F. y Ruiz Potero, C. 2002. Caracterización espacio-temporal de la temperatura del aire de una cavidad mediante técnicas geoestadísticas (Sistema Covadura, Karst en yesos de Sorbas). En: Carrasco, F., Durán, J.J., Andreo, B. (Eds.). *Karst and Environment*, 555-565 pp.
- Jiménez Espinosa, R. 1993. *Análisis geoestadístico de la distribución espacial de anomalías geoquímicas en exploración minera*. Tesis Doctoral. Universidad de Granada. 442 pp.
- Jiménez Espinosa, R., Molina Sánchez, L., Pulido Bosch, A. y Navarrete, F. 1996. Influencia de la agricultura intensiva en el contenido de nitratos de las aguas del Campo de Dalías: evolución espacial y temporal. *Geogaceta*, 20(6), 1281-1284.
- Jiménez Espinosa, R., Sánchez Martos, F. y Pulido Bosch, A. 2001. Caracterización hidrogeoquímica del acuífero detrítico del Bajo Andarax (Almería) por medio de técnicas estadísticas y geoestadísticas. *V Simposio sobre Agua en Andalucía*. Universidad de Almería.
- Journel, A.G. 1983. Nonparametric estimation of spatial distributions. *Mathematical Geology*, 15, 445-468.
- Journel, A.G. y Huijbregts, Ch.J. 1978. *Mining Geostatistics*. Academic Press. Londres. 600 pp.
- Kyriakidis, P.C. y Journel, A.G. 1999. Geostatistical Space-Time Models: A review. *Mathematical Geology*, 31(6), 651-684.
- Matheron, G. 1970. *La theorie des variables regionalisées et ses applications*. Centre Géostatistique et de Morphologie Mathématique, Ecole National Supérieur des Mines de Paris, 5/ 212 pp.
- Pereira, H.G. y Sousa, A.J. 1988. *Análise de dados para o tratamento de quadros multidimensionais*. CVRM - Instituto Superior Técnico, Lisboa. 115 pp.
- Samper, F.J. y Carrera, J. 1990. *Geoestadística: Aplicaciones a la hidrología subterránea*. Centro Internacional de Métodos Numéricos en Ingeniería, Universidad Politécnica de Cataluña, Barcelona, 484 p.
- Sánchez Martos, F. y Pulido-Bosch, A. 1999. Boron and the origin of salinization in an aquifer in Southeast Spain. *C.R. Acad. Sci. Paris, Earth & Planetary Sciences*, 328, 751-757
- Sánchez Martos, F., Jiménez Espinosa, R. y Pulido Bosch, A. 2001. Mapping groundwater quality variables using PCA and geostatistics: a case study of Bajo Andarax, SE Spain. *Hydrogeological Science Journal*.

Recibido: Mayo 2003

Aceptado: Septiembre 2003