

Análisis e integración de datos espaciales en investigación de recursos geológicos mediante Sistemas de Información Geográfica

Juan P. Rigol-Sánchez^{1,*}, Mario Chica-Olmo², Eulogio Pardo-Igúzquiza², Víctor Rodríguez-Galiano² y Mario Chica-Rivas³

¹ Depto. de Geología, G.I. RSGIS-CEAMA, Universidad de Jaén, Campus Universitario Las Lagunillas, B-3, 23071, Jaén, España

² Depto. de Geodinámica, G.I. RSGIS-CEAMA, Universidad de Granada, Avenida Fuente Nueva s/n, 18071, Granada, España

³ Depto. de Análisis Matemático, Universidad de Granada, Av. Fuentenueva s/n, 18071, Granada, España

* jprigol@ujaen.es

Resumen

En la investigación geológica, como es el caso de la exploración de recursos minerales, es frecuente disponer de grandes volúmenes de datos espaciales, referentes a información temática diversa, costosa de obtener, que son adecuadamente tratados y analizados con ayuda de los sistemas de información geográfica (SIG). Se puede decir, que el objetivo final de este análisis de datos es la elaboración de mapas que indiquen las zonas o puntos donde es posible la presencia de un recurso (por ejemplo, mapas de favorabilidad minera). Para ello es necesario elaborar modelos espaciales predictivos que permitan la incorporación y combinación de las variables relevantes relacionadas con el fenómeno estudiado. Estos modelos, elaborados normalmente en entorno SIG, poseen naturaleza numérica diversa, que van desde modelos basados en la combinación de mapas mediante reglas lógicas, aritméticas, estadísticas o probabilísticas, hasta modelos más complejos basados en algoritmos de inteligencia artificial y minería de datos. En este artículo se presenta la aplicación de dos métodos de integración de datos espaciales, el método de suma ponderada multiclase y el de regresión logística múltiple, en un contexto aplicado de investigación de depósitos metálicos en el SE de España. Los modelos han sido implementados mediante SIG y han permitido generar mapas predictivos, por medio del cálculo de un índice de favorabilidad minera (IFM), que ha facilitado la selección de las zonas con mayor potencialidad para albergar depósitos minerales. Los resultados obtenidos indican que el rendimiento de los modelos es similar en muchos de los experimentos, con porcentajes de acierto de depósitos conocidos sólo un poco superiores en los modelos basados en regresión logística. No obstante, el método de suma ponderada multiclase puede ser aceptable en la mayoría de los casos.

Palabras clave: recursos geológicos, integración de datos espaciales, SIG, suma ponderada multiclase, regresión logística múltiple

Abstract

In modern investigation of geological resources, including mineral exploration, large amounts of spatial data are usually collected. These data correspond to diverse and costly thematic information and are adequately organized, visualized and analyzed using a geographical information system (GIS). The main objective of data analysis in this field is the creation of maps showing areas or points where a geological resource may be located (e.g., mineral favorability maps). To achieve this, predictive spatial models capable of incorporating and combining all relevant variables related to the resources have to be generated. These models are usually implemented using GIS and are typically based on numerical rules, ranging from combination of maps using logical, arithmetical, statistical or probabilistic rules, to complex models based on artificial intelligence and data mining algorithms. In this paper, an application of two spatial integration methods based on multiclass weighted sum and multiple logistic regression for exploration of metallic deposits

in SE Spain is presented. Spatial models have been developed using a GIS and have allowed to generate predictive maps showing a mineral favorability index (MFI), helping the selection of the areas having the highest potential for mineral deposits. Results indicate that both models achieve similar performance in most experiments, being percent of identification of known deposits slightly better for logistic regression. Nevertheless, model based on multiclass weighted sum perform well in most cases.

Keywords: geological resources, spatial data integration, GIS, multiclass weighted sum, multiple logistic regression.

1. Introducción

Desde épocas remotas, la búsqueda de recursos geológicos ha sido un deseo constante en las sociedades humanas, con un máximo impulso en nuestra historia reciente como consecuencia del desarrollo industrial. En la actualidad, el descubrimiento de nuevos depósitos minerales, de hidrocarburos o de aguas subterráneas depende de la aplicación de un conjunto variado de tecnologías más que de una sola técnica *per se*, al haber sido ya encontrados buena parte de los recursos fácilmente localizables (Evans, 1995). Como consecuencia, la investigación moderna de recursos geológicos se caracteriza por el uso conjunto de diversos y sofisticados métodos como geoquímica, geofísica, teledetección, etc. Por otro lado, los avances científicos en los métodos de análisis y de integración de datos o información representan un progreso de indudable interés hoy día en este ámbito de las geociencias. Este aspecto ha encontrado en los sistemas de información geográfica (SIG) un apoyo tecnológico que está facilitando enormemente el desarrollo, cada vez más extendido, de este tipo de aplicaciones, en las que en esencia se analiza información multitemática georreferenciada (Chica-Olmo et al., 1994, 1996, 1997, 2002; Cassard et al., 2008; Carranza, 2009).

El análisis espacial de la información generada en los estudios geocientíficos ha consistido tradicionalmente en la superposición de los datos analógicos recopilados en forma de mapas en una mesa para realizar un posterior análisis visual (Bonham-Carter, 1994). Con el uso de los sistemas informáticos, esta aproximación ha sido reemplazada por un proceso digital semiautomático, mucho más eficaz, realizado con la ayuda de programas informáticos SIG o CAD. Sin embargo, con cierta frecuencia, aún hoy día, el nuevo proceso es análogo al tradicional, en el sentido que el análisis se limita a la superposición de los datos digitales para generar nuevos documentos cartográficos de interés en la aplicación en curso. En los últimos años, no obstante, la etapa en que los datos básicamente se almacenan en bases de datos espaciales SIG para confeccionar cartografías o mapas puramente descriptivos, está evolucionando a una nueva fase caracterizada por un mayor peso del análisis numérico intensivo de los datos disponibles (Rigol-Sánchez, 2000).

Un caso típico de los estudios multidisciplinarios indicados es precisamente la exploración de recursos geológicos.

En la misma se recolecta normalmente un gran volumen de datos que no pueden ser analizados de forma efectiva sin un sistema de gestión de datos espaciales adecuado y eficiente (Burrough, 1986); es decir, sin un Sistema de Información Geocientífica. Por otro lado, el estudio de fenómenos naturales complejos como el señalado requiere de la manipulación, no sólo de las capas individuales de información, sino también del análisis de las relaciones, especialmente las espaciales, que existen entre ellas, para poder llegar así a obtener los resultados deseados (Bonham-Carter, 1994). Esos resultados son sintetizados, en el caso de la exploración geológica y minera, en la cartografía predictiva de zonas con alto interés prospectivo. Al respecto, Moon y Whateley (1995) destacan que la habilidad clave en la delimitación de un sector prioritario u objetivo para perforación minera es la integración de la información obtenida en las distintas campañas de exploración. Consecuentemente, el procesamiento de los datos debe ser considerado una fase fundamental en el proceso de investigación de recursos geológicos, "*datos más análisis igual a geoinformación*" (Rigol-Sánchez, 2000; Chica-Olmo et al., 2002). Para ello es necesario elaborar modelos espaciales predictivos que permitan la incorporación y combinación de las variables relevantes relacionadas con el fenómeno estudiado. Tales modelos, elaborados en entorno SIG, son variados según su base metodológica, desde aquellos que combinan los mapas con sencillas reglas lógicas, aritméticas, estadísticas o probabilísticas, hasta aquellos otros algo más complejos que emplean reglas no lineales.

Por último, no debe olvidarse que la investigación de los recursos geológicos tiene siempre su fundamento científico en un modelo conceptual geológico, que sirva de guía para la determinación y comprensión de las principales características que éstos presentan, especialmente las que pueden proporcionar pistas sobre su localización (Bustillo-Revuelta y López-Jimeno, 1996). Sobre este pilar científico se construyen y elaboran prácticamente todas las investigaciones posteriores, en las que debemos incluir, igualmente, a la modelización espacial con SIG.

En este trabajo se presenta la aplicación de dos modelos distintos de integración de datos geocientíficos basados en SIG para la exploración de depósitos metálicos en un distrito minero del SE de España. El objetivo final de este estudio es obtener mapas de favorabilidad o potencialidad minera, que faciliten la localización de los sectores más favorables para contener un depósito mineral.

2. Materiales y métodos

2.1. Área de estudio

El área de estudio corresponde a los distritos mineros de Rodalquilar y San José, situados en el extremo SE de la Península Ibérica, dentro de la provincia de Almería, España (Figura 1). El área coincide en su mayor parte con el campo volcánico de edad miocena de Cabo de Gata, que forma la sierra del mismo nombre. El área se extiende a lo largo de la costa, y en ella se encuentran muchas de las mineralizaciones filonianas de metales básicos y preciosos del SE de España (Arribas-Rosado y Arribas-Moreno, 1995). El campo volcánico está situado en la parte oriental de la Cordillera Bética, un complejo cinturón orogénico de edad alpina que se formó como consecuencia de la colisión de las placas tectónicas Africana y Europea.

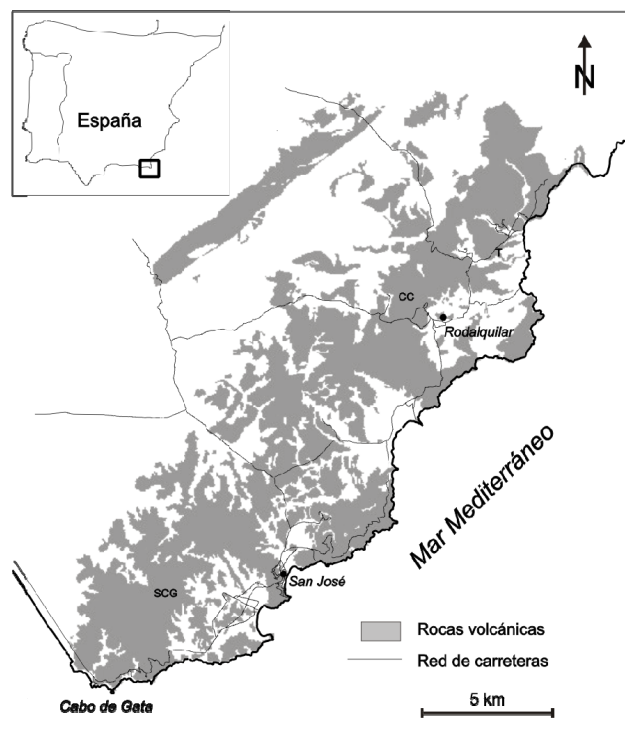


Figura 1. Localización geográfica del área de estudio de Rodalquilar-San José, Almería, España (SCG: Sierra de Cabo de Gata, CC: Cerro del Cinto, T: Los Tollos).

El conjunto volcánico calco alcalino de Cabo de Gata es el más importante, volumétricamente, de los materiales volcánicos emergidos, y corresponde a una de las cuatro series volcánicas descritas en el orógeno bético, concretamente a la serie calco alcalina *s.s.* (Cabo de Gata y Mar de Alborán) de López-Ruiz y Rodríguez-Badiola (1980). Se caracteriza por presentar una variedad de litologías desde andesitas basálticas a riolitas, pertenecientes a una serie calco alcalina de contenido medio en potasio, con predominio de los términos intermedios (andesitas y dacitas)

(Fernández-Soler, 1996). Este magmatismo actuó desde el Langhiense al Tortoniense Superior (15-7 Ma). Los depósitos messinienses que sellan la actividad volcánica están representados por sedimentos carbonatados y margosos (complejos arrecifales en torno a los grandes relieves). Posteriormente, se formaron niveles de evaporitas en las zonas más deprimidas y subsidentes, y carbonatos estromatolíticos en las zonas costeras. Los depósitos pliocenos y cuaternarios cubren también parcialmente las rocas volcánicas.

Como consecuencia de la actividad volcánica y sub-volcánica, se desarrollaron durante el Mioceno una serie de sistemas hidrotermales simultáneos a la emisión de las rocas volcánicas que produjeron alteraciones de rango menor, sin interés económico, ligadas a la circulación de fluidos marinos a través de poros y fracturas de las volcanitas emplazadas en medios marinos (Fernández-Soler, 1996). Asimismo, se produjeron grandes áreas de alteraciones-mineralizaciones correspondientes a sistemas de circulación hidrotermal (epitermal), acompañados comúnmente con disposiciones zonales del tipo de alteración. Estos campos hidrotermales se vieron favorecidos en muchos casos por los sistemas de fracturación creados por la formación de calderas volcánicas (Arribas *et al.*, 1989), aunque en otros casos se sitúan en relación con los sistemas de fracturación regionales (Hernández *et al.*, 1987). El distrito de San José se caracteriza principalmente por la asociación epitermal Pb-Zn-(Ag-Cu-Au). En general, se trata de depósitos relacionados con cuerpos de brechas hidrotermales, orientadas según patrones de fracturación regionales NE-SO y NO-SE, que encajan en zonas de alteración de tipo adularia-sericita (Castroviejo, 1992). Los yacimientos epitermales del distrito de Rodalquilar están constituidos por venas de Pb-Zn-(Ag-Cu-Au) de baja sulfuración y, principalmente, por depósitos de Au-alunita-(Cu-Te-Sn) de alta sulfuración, también llamados de tipo ácido-sulfatado, que incluyen venas, brechas hidrotermales y depósitos diseminados (Arribas *et al.*, 1988). Los sistemas hidrotermales también originaron la formación de numerosos yacimientos de arcillas industriales como la bentonita.

Para realizar este estudio se ha dispuesto de un gran volumen de datos geológicos y mineros, geoquímicos, geofísicos, imágenes de satélite, etc. En la Tabla 1 se indican los conjuntos de datos primarios utilizados. A partir de éstos se han obtenido un gran número de conjuntos de datos secundarios mediante distintas operaciones de análisis espacial, con la finalidad de facilitar el proceso de integración de datos.

2.2. Métodos de integración de datos espaciales

Los modelos de integración de datos espaciales mediante SIG son categorizados en distintos grupos en función de la metodología empleada o el objetivo perseguido. En términos generales, un modelo SIG puede ser conside-

Tabla 1. Datos primarios disponibles para el estudio junto con algunas de sus características.

Nombre	Escala / Proyección	Adquisición	Tipo espacial	Atributos	Tipo de atributo
Geología	1:25,000 / UTM	Digitalización manual	Polígono	Unidad; símbolo	Nominal; ordinal
Fracturas	1:25,000 / UTM	Digitalización manual	Línea	Tipo; orientación; símbolo	Nominal; razón; ordinal
Filones	1:25,000 / UTM	Digitalización manual	Línea	Tipo; orientación; símbolo	Nominal; razón; ordinal
Lineamientos	1:25,000 / UTM	Digitalización manual	Línea	Tipo; orientación; símbolo	Nominal; razón; ordinal
Indicios mineros	- / UTM	Digitalización teclado	Punto	Elemento; importancia; símbolo	Nominal; ordinal; ordinal
Estructuras volcánicas	1:25,000 / UTM	Digitalización manual	Línea	Tipo; símbolo	Nominal; ordinal
Magnetometría campo	- / UTM	Base de datos digital	Punto	C. magnético; anomalía residual	Razón; razón
Gravimetría campo	- / UTM	Base de datos digital	Punto	C. gravitatorio; anomalía resi.	Razón; razón
Magnetometría aerotransportada	1:25,000 / UTM	Digitalización manual	Línea, polígono	C. magnético; lineamiento m.	Razón; nominal
Geoquímica de rocas y suelos	- / UTM	Base de datos digital	Punto	Concentración 32 elementos	Razón
Geoquímica de rocas y suelos 1995	- / UTM	Base de datos digital	Punto	Concentración 32 elementos	Razón
Mineralogía	- / UTM	Digitalización teclado	Punto	Concentración 12 fases	Razón
Radiometría de campo	- / UTM	Digitalización teclado	Punto	Radiancia (6 bandas)	Intervalo
Imagen TM	- / - (píxel 30m)	Cinta magnética	Raster	Radiancia (7 bandas)	Intervalo
Imagen HRV pancromática	- / - (píxel 10m)	Cinta magnética	Raster	Radiancia	Intervalo
Imagen HRV multispectral	- / - (píxel 20m)	Cinta magnética	Raster	Radiancia (3 bandas)	Intervalo
Imagen ATM	- / - (píxel 5m)	Cinta magnética	Raster	Radiancia (11 bandas)	Intervalo
Elevación	1:25,000 / UTM	Digitalización manual	Línea, punto	Elevación	Razón (superf)
Red hidrográfica	1:25,000 / UTM	Digitalización manual	Línea	Nombre; símbolo	Nominal; ordinal
Uso del suelo	1:50,000 / UTM	Digitalización manual	Polígono	Tipo; símbolo	Nominal; ordinal
Red viaria	1:25,000 / UTM	Digitalización manual	Línea	Tipo; símbolo	Nominal; ordinal
Poblaciones	1:25,000 / UTM	Digitalización manual	Polígono	Nombre; símbolo	Nominal; ordinal

rado como el proceso de combinación de un conjunto de mapas o capas de entrada para producir un mapa de salida (Burrough, 1986; Aronoff, 1989; Berry, 1993):

$$\text{Mapa}_{\text{Salida}} = f(\text{Mapa}_1, \text{Mapa}_2, \text{Mapa}_3, \dots, \text{Mapa}_n)$$

La función f presenta formas diferentes y puede ser categorizada en tres tipos dependiendo de la naturaleza de la relación expresada (Bonham-Carter, 1994): (a) basada en teorías y principios de la física y la química, (b) empírica, basada en observaciones de los datos (estadística o heurística), o (c) algún tipo de mezcla entre teoría y empirismo, lo que da origen a otra clasificación de los modelos en teóricos, empíricos e híbridos, respectivamente. Luo (1990) y, posteriormente, Fabbri y Chung (1996) proponen una aproximación unificada a la integración de datos espaciales para predicción mediante funciones de favorabilidad, en la que cada capa o mapa relevante en el modelo es clasificada en función de su favorabilidad a la presencia de un objeto concreto de interés (por ejemplo, un depósito mineral). Las capas clasificadas, o funciones de favorabilidad, son combinadas siguiendo reglas que pueden ser lógicas, aritméticas, probabilísticas, etc. (Figura 2).

Las funciones de favorabilidad son deducibles mediante la aplicación de métodos diversos de análisis espacial, que pueden ir desde una simple reclasificación de las clases temáticas de un mapa hasta métodos estadísticos avan-

zados (por ejemplo, estimación geoestadística). Se pretende con estos métodos caracterizar la distribución espacial de las variables relacionadas con los recursos, para así detectar las zonas anómalas de interés prospectivo; es decir, la delimitación de objetivos parciales de exploración. Cada uno de estos objetivos parciales, en forma de capa de información SIG, indica la favorabilidad a la presencia potencial de depósitos, y es obtenido a partir del tratamiento de los datos generados por una técnica de exploración (por ejemplo, geoquímica). En general, los objetivos parciales no permiten individualmente indicar de forma excluyente la presencia de un depósito mineral, por lo que es necesario combinar todos ellos para generar los objetivos finales con la mayor posibilidad de éxito (Bonham-Carter, 1994).

Los modelos utilizados para la predicción de recursos geológicos, son típicamente modelos predictivos de tipo empírico, estadísticos o heurísticos, ya que en la génesis de la mayor parte de los recursos geológicos intervienen numerosos y complejos factores físico-químicos difícilmente “predecibles” mediante teorías expresadas matemáticamente (Bonham-Carter, 1994). Esto da lugar a una división adicional de los modelos espaciales empíricos en dos tipos: basados en el conocimiento y basados en los datos (Bonham-Carter, 1994; Pendock y Nedeljkovic, 1997). En el primer caso, los parámetros son estimados sobre la base de la opinión de un experto en el tema, y en el segundo son obtenidos del análisis de las relaciones espaciales entre las capas independientes y la capa dependiente. Los modelos

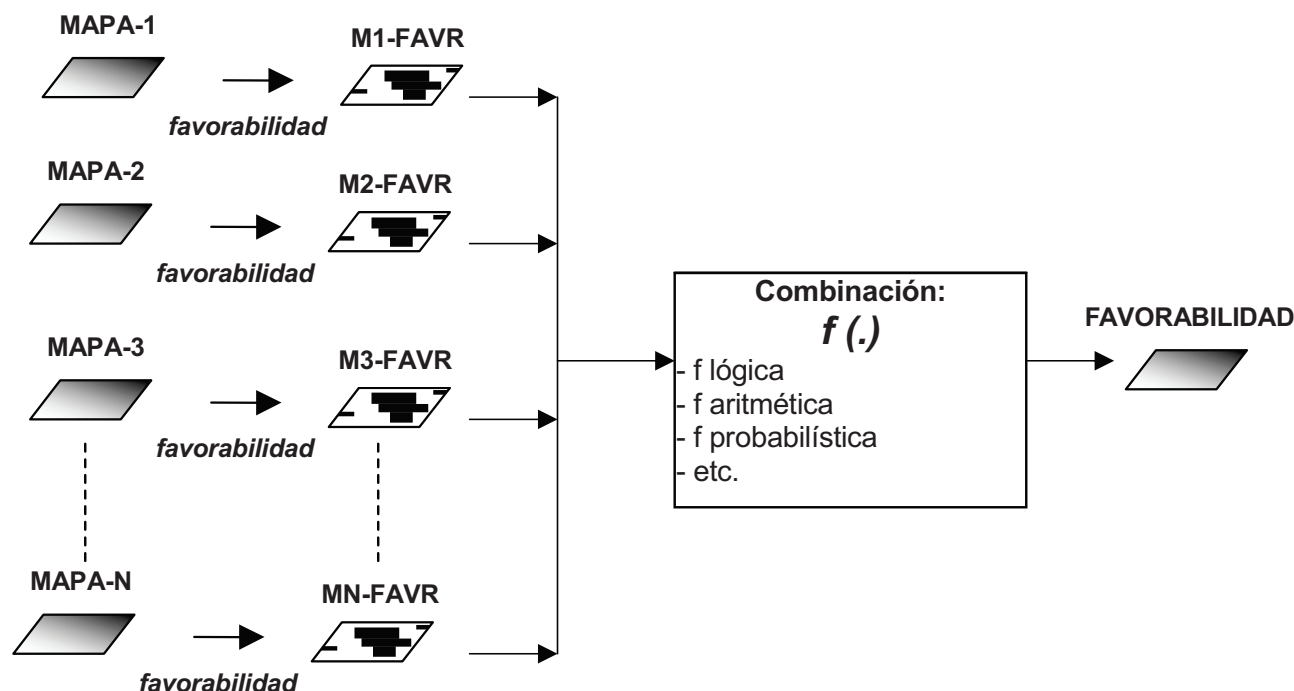


Figura 2. Modelo de integración de mapas (capas de información) de tipo predictivo (Rigol-Sanchez, 2000).

basados en el conocimiento hacen uso de funciones de integración tales como la lógica booleana, la suma ponderada o la lógica difusa, mientras que los modelos basados en los datos utilizan típicamente funciones como la regresión múltiple, el análisis discriminante, métodos probabilísticos bayesianos o incluso redes neuronales (Bonham-Carter *et al.*, 1989; Agterberg *et al.*, 1993; Bonham-Carter, 1994; Rigol-Sanchez *et al.*, 2003). Los modelos basados en la suma ponderada y en la regresión múltiple son, debido a sus características, dos de los métodos más utilizados.

2.2.1. Modelos basados en la suma ponderada

Este método es simple, está basado en la multiplicación de cada capa de entrada por un peso o factor de ponderación que indica su importancia en el modelo, y la posterior suma de las capas así modificadas. El resultado es una nueva capa con valores numéricos, cuya magnitud representa en una escala relativa la favorabilidad de cada zona. Esta aproximación posibilita una selección flexible de las zonas de interés, y permite, además, tener en cuenta la distribución espacial de las mismas en el proceso de selección (Bonham-Carter, 1994).

El método presenta la ventaja adicional de poder ser aplicado indistintamente a capas con dos o más clases. En el caso de capas de entrada binarias, éstas llevan asociado solamente el valor del peso asignado. En cambio, con las capas multiclase se introduce mucha más información en el modelo, puesto que se asigna un valor de favorabilidad diferente a cada una de las clases. En estas aproximaciones también se hace intervenir información derivada de los datos, en el sentido que los umbrales para la binarización

o definición de las clases pueden ser obtenidos mediante el análisis de las relaciones entre un mapa de indicios y los mapas de entrada (Rigol-Sánchez, 2000).

La aplicación del método de suma ponderada a mapas de entrada multiclase implica la asignación no sólo de un peso a cada capa de entrada al modelo, sino también el establecimiento de un peso diferente a cada una de las clases que compone cada mapa. La función de combinación utilizada en este caso se representa mediante la expresión simple:

$$F = \frac{\sum_i^n P_i * Pcl_{ij}}{\sum_i^n P_i}$$

en la que F es la puntuación que indica la favorabilidad estimada por el modelo; P_i es el peso para el mapa de entrada i , y Pcl_{ij} es la puntuación para la clase j del mapa i . El valor de j depende de la clase temática que aparezca en cada píxel o polígono del mapa i . En esta situación, la puntuación no está necesariamente en el intervalo $[0, 1]$, sino en un rango que varía en función de la magnitud de los pesos usados. La suma ponderada de capas multiclase es un método frecuentemente utilizado en la modelización espacial con SIG, debido fundamentalmente a su sencillez conceptual, fácil aplicación y notable flexibilidad a la hora de incorporar información relevante (en forma de pesos) en el proceso de combinación. No obstante, Bonham-Carter (1994) argumenta que la principal limitación del método radica en su naturaleza lineal aditiva.

Este método se utilizó para generar mapas de favorabilidad usando distintas combinaciones de los mapas relevantes multiclase que estaban disponibles en la base de datos de exploración. En este estudio se llevaron a cabo numerosos experimentos de los que seguidamente se presenta un ejemplo ilustrativo; para ello se utilizaron los modelos de depósito y las guías de exploración para elaborar los criterios a aplicar en la evaluación de la hipótesis de potencialidad minera. El estudio se orientó a la selección de zonas propicias para la presencia de depósitos de sulfuros metálicos (Pb-Zn-Cu-Ag-Au). Para este análisis se empleó una reclasificación en cuatro niveles de favorabilidad para todos los mapas de entrada (funciones de favorabilidad): (1) muy poco favorable, (2) poco favorable, (3) favorable y (4) muy favorable. Para la asignación de puntuación a las cuatro clases de cada mapa (Pcl_{ij}), se optó por seguir un esquema común con valores 2, 4, 8 y 16. Para el peso asignado a cada mapa (P_i) se definió un esquema análogo a porcentajes, con pesos individuales acotados en el rango [0, 100] y cuya suma total es 100. En la Tabla 2 se detallan las capas utilizadas para los depósitos de sulfuros polimetálicos y los pesos asignados a cada una. El resultado de la aplicación de este modelo es una nueva capa ráster en la que cada celda toma un valor en el rango [2, 16]. Este valor se ha interpretado como un índice de favorabilidad minera (IFM), que indicaría las zonas con más potencial para contener un depósito mineral del tipo en cuestión.

Tabla 2. Ponderación de las capas utilizadas en el modelo de suma ponderada multiclase para predicción de depósitos de sulfuros polimetálicos.

CAPA/MAPA	PESO
Fotointerpretación (TM+SPOT)	15
Clasificación TM	10
Radiometría de campo	2
Subtotal teledetección	27
Geología de detalle	15
Fracturas y lineamientos	12
Estructuras volcánicas	8
Indicios mineros	--
Subtotal geología	35
Mineralogía	2
CP1 Geoquímica detalle	13
CP2 Geoquímica detalle	1
CP3 Geoquímica detalle	19
Subtotal geoquímica	35
Aeromagnetometría	3
Subtotal geofísica	3
TOTAL	100

2.2.2. Modelos basados en regresión logística múltiple

El análisis de regresión es un método estadístico bien conocido y extendido en numerosas disciplinas científicas, usado para explorar las relaciones entre distintas variables sobre la base de observaciones de esas variables. El método implica la derivación de una relación matemática entre

un conjunto de variables predictivas o explicativas independientes y una condición dependiente específica (Davis, 1986). El objetivo es, por tanto, explicar las variaciones en el conjunto de observaciones de la variable dependiente en términos de las variables independientes.

Además de la identificación de un buen modelo, un objetivo adicional es la obtención de una buena estimación de los coeficientes de regresión. Asimismo, en algunas situaciones, como es el caso práctico planteado, hay un objetivo último que es proporcionar una ecuación predictiva. No obstante, como se mencionaba anteriormente, el conocimiento del fenómeno estudiado es un aspecto importante a la hora de especificar un modelo de regresión, y en general, cualquier modelo basado en los datos. Uno de los principales problemas que se encuentran en regresión múltiple con datos espaciales se debe al incumplimiento de los supuestos del método de mínimos cuadrados, que es el proceso utilizado para ajustar y hacer inferencia, o bien debido a la naturaleza de los datos (Haining, 1990). Si la variable dependiente a modelizar es categórica, como es el caso de la presencia o ausencia de un recurso geológico, el modelo de regresión lineal normal no es adecuado y se debe emplear un modelo lineal generalizado como la regresión logística múltiple (Petrucci et al., 1999).

La regresión logística múltiple permite identificar las variables importantes en la predicción de la probabilidad de presencia de un depósito, en la que la presencia o ausencia del depósito se define como una variable dicotómica dependiente. Los coeficientes de regresión para cada variable derivados de las observaciones experimentales en el área de estudio, son utilizados como pesos en un algoritmo aplicable a la base de datos SIG, por ejemplo, mediante álgebra de mapas (Tomlin, 1991). De esta forma, se obtiene un mapa que muestra en términos de probabilidad (intervalo [0,1]) la presencia de un depósito mineral, equivalente al IFM. Cuantitativamente, la relación de dependencia entre la ocurrencia y el conjunto de variables predictivas se expresa como:

$$P_x = p(D) = 1 / (1 + \{\exp[-(B_0 + B_1 X_1 + \dots + B_p X_p)]\})$$

donde D es presencia/ausencia de depósitos, $X_1 \dots X_p$ son el conjunto de variables o capas de información de exploración (por ejemplo, alteración hidrotermal, distancia a fracturas, anomalía geoquímica, etc.), y $B_0 \dots B_p$ son los coeficientes derivados de la regresión logística. Es decir, D es la variable dependiente y $X_1 \dots X_p$ son las variables independientes.

Para aplicar la regresión logística múltiple a los datos de exploración se obtuvieron, en primer lugar, los datos adecuados de entrada para construir el modelo en un paquete estadístico externo al SIG. Todas las variables predictoras se transformaron linealmente al intervalo [0,1] y la variable dependiente (presencia de depósitos) en una capa binaria. Para ilustrar el proceso descrito se presenta un experimento en el que se construyó y aplicó

un modelo para la predicción de depósitos de oro en el distrito de Rodalquilar. En este experimento se utilizó un subconjunto de siete capas de información: (1) distancia a estructuras de fracturación; (2) componente principal 1 de la geoquímica, relacionada esencialmente con la litología, con valores positivos para elementos metálicos asociados a rocas básicas y negativos para elementos metálicos asociados a rocas ácidas; (3) componente principal 2 de la geoquímica, relacionada con elementos metálicos típicos de las mineralizaciones de oro; (4) anomalía gravimétrica residual; (5) anomalía magnética residual; (6) cociente TM 5/7, y (7) cociente TM 3/1; estas dos últimas variables se refieren a cocientes de datos radiométricos de las bandas 5, 7, 3 y 1 de una imagen Landsat TM de la estación seca de verano (7/7/2001). Se interpretan como indicadores de la abundancia de arcillas de alteración hidrotermal y de óxidos de hierro, respectivamente (Sabins, 1996; Vincent, 1997). Para los análisis se seleccionaron 49 indicios mineros, depósitos de oro conocidos, situados en el distrito. Este conjunto (unos) se completó con una muestra de 56 celdas, *a priori* estériles (ceros), mediante muestreo aleatorio estratificado (selección de un punto aleatorio en el interior de la celda).

3. Resultados y discusión

Los resultados de la aplicación de los modelos SIG descritos para integración de datos espaciales en investigación de recursos geológicos son mapas de un índice

de potencialidad o favorabilidad minera (IFM). El mapa que se muestra en la Figura 3 indica la favorabilidad a la presencia de depósitos de sulfuros polimetálicos, obtenida con el método de suma ponderada multiclasa. En el mismo se observa que las zonas con mayor potencialidad se sitúan en torno al distrito de San José, concretamente en la parte central y noroccidental de la Sierra de Cabo de Gata. En este caso, el IFM alcanza valores máximos, superiores a 13 (tonos rojos), solamente en un pequeño sector. En el distrito de Rodalquilar los valores de IFM son medios, principalmente al sur y este del sector conocido como Cerro del Cinto, junto a la localidad de Rodalquilar. El análisis del porcentaje de indicios situados en cada clase del mapa de favorabilidad generado, así como el área porcentual que ocupa cada una de las clases, indica que las clases con valor IFM ≥ 6 , que representan una extensión del 15 % del área total, contienen el 84 % de los indicios. Es evidente el interés de estos resultados pues reducen considerablemente el *área objetivo* de futuros reconocimientos mediante sondeos de investigación. En términos generales, el modelo de suma ponderada multiclasa ha resultado una técnica de gran utilidad y muy potente para la creación de mapas de favorabilidad minera. El número de combinaciones de capas y esquemas de ponderación es enorme, lo que proporciona a este método de modelización de una gran flexibilidad.

En el caso de la aplicación del modelo de regresión logística múltiple en el distrito de Rodalquilar, dio como resultado la siguiente expresión en la que se muestran los coeficientes del modelo para cada variable independiente:

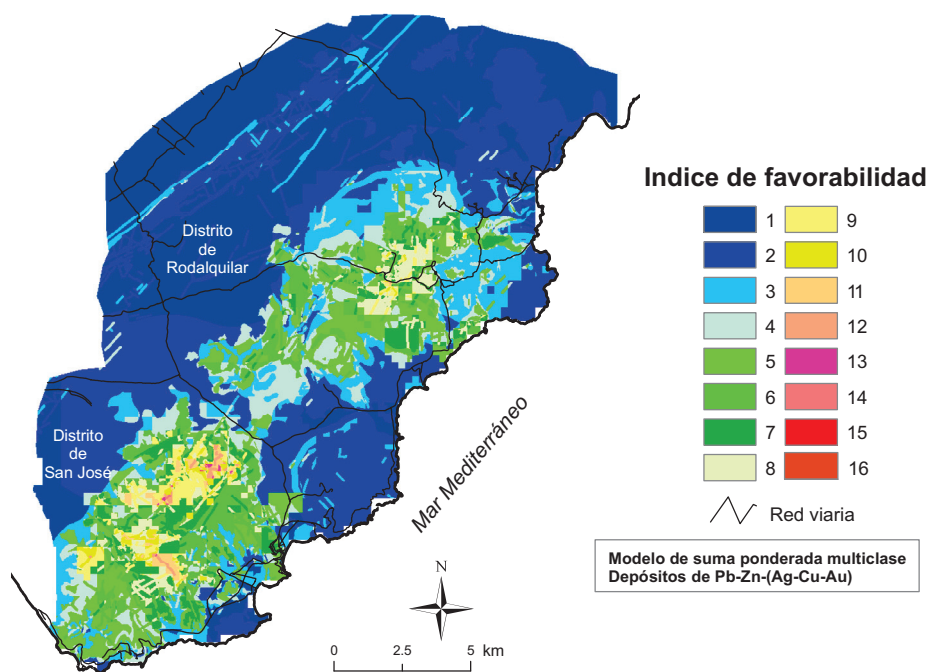


Figura 3. Mapa de favorabilidad a la presencia de depósitos de sulfuros polimetálicos en los distritos de Rodalquilar y San José generado con el modelo de suma ponderada multiclasa. Nota: la clase 1 corresponde a las zonas para las cuales no se ha dispuesto de todas las cubiertas de información.

$$P_x = 1 / (1 + (\exp(-(-19.4 - 2.5 \text{ gqpc1} + 4.9 \text{ gqpc2} - 1.7 \text{ agr} + 15.3 \text{ gravr} + 39.5 \text{ tm57} + 4.6 \text{ tm31} - 23.8 \text{ distlff}))))))$$

Las capas correspondientes a la distancia a lineamientos, fracturas y filones (*distlff*), el componente principal 1 de la geoquímica (*gqpc1*) y la anomalía magnética residual (*magr*) presentan coeficientes negativos, lo que señala que al aumentar su valor disminuye la probabilidad de presencia de depósitos. El resto de variables presentan coeficientes positivos indicando la relación contraria. El coeficiente mayor en valor absoluto corresponde al cociente de las bandas radiométricas Landsat TM 5/7 (*tm57*), relacionado con las alteraciones hidrotermales, seguido de la distancia a lineamientos, fracturas y filones (*distlff*) y la anomalía gravimétrica residual (*gravr*). El componente principal 1 y la anomalía magnética residual presentan una contribución relativamente pequeña al modelo. El análisis de los coeficientes y el valor del estadístico de Wald asociado sugieren que la distancia a estructuras de fracturación (*distlff*) y el cociente de las bandas Landsat TM 5/7 (*tm57*) son los más interesantes en la determinación de la presencia o ausencia de depósitos minerales. La bondad del ajuste se estimó sobre un subconjunto independiente de datos, observando que el modelo es capaz de estimar de forma moderadamente aceptable la probabilidad de presencia de depósitos en puntos (celdas) desconocidas ($R^2=0.47$).

Los coeficientes estimados se utilizaron para aplicar el modelo en el SIG y generar la capa de favorabilidad de

presencia de depósitos. El mapa resultante se muestra en la Figura 4 en una escala de 0 a 10. El examen visual del mismo muestra su parecido a la capa de entrada TM 5/7, circunstancia lógica dado que esta capa presenta el coeficiente de regresión logística más elevado. Las zonas de mayor interés se sitúan en torno al sector del Cinto y al sur del área de los Tollos, las cuales ocupan una extensión relativamente grande. También aparece una zona de valores altos al sur de la carretera que parte de Rodalquilar hacia el oeste, pero bastante discontinua. Se observa además una zona en el borde izquierdo del área analizada que presenta valores medios y altos. En el borde norte se observa una pequeña zona también de interés potencial alto. La comparación de la distribución espacial de los indicios mineros conocidos respecto a las distintas clases de favorabilidad de los mapas obtenidos indica que más del 85 % de los indicios se sitúa en las clases con valor IFM ≥ 5 . Si se consideran las clases con valor IFM ≥ 8 , el porcentaje de indicios situados en ellas es del 75 %, mientras que el área supone el 25 % del área total.

4. Conclusiones

En este trabajo se han mostrado dos ejemplos de modelos de integración de datos geoespaciales para la investigación de recursos geológicos. Los modelos han sido implementados mediante SIG y han permitido generar mapas predictivos indicando las zonas con mayor potencialidad para contener depósitos minerales.

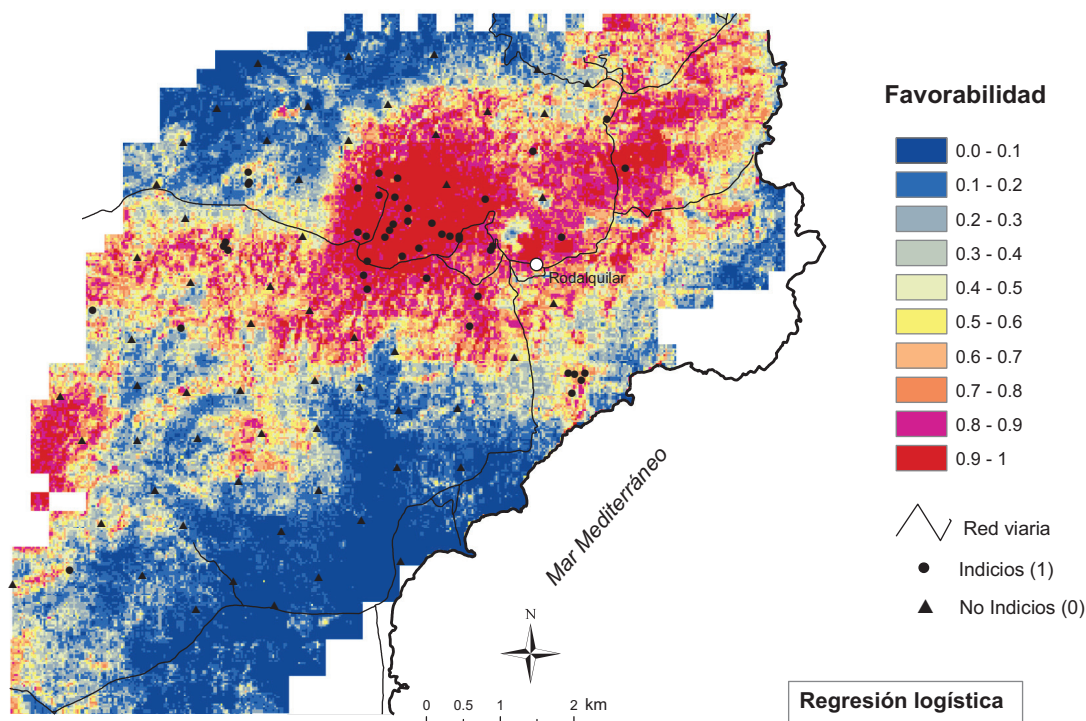


Figura 4. Superficie de probabilidad indicando favorabilidad a la presencia de depósitos minerales en el distrito de Rodalquilar generada mediante un modelo de regresión logística.

Los resultados obtenidos indican que los mapas de favorabilidad minera (IFM) obtenidos muestran cierta dependencia del método utilizado. No obstante, de los distintos experimentos realizados se desprende que el rendimiento de los modelos es similar en muchos de los casos, con porcentajes de acierto sólo un poco superiores en los modelos basados en regresión logística. Esto parece por otra parte lógico, puesto que estos métodos utilizan directamente las observaciones para ajustar los parámetros del modelo, mientras que en los basados en el conocimiento pasan por el tamiz de la interpretación subjetiva del investigador. Esto puede interpretarse como una ventaja o una desventaja dependiendo de la aplicación. El método más habitual de suma ponderada multiclasa puede ser aceptable en muchos casos. Este método permite la incorporación en el modelo de una gran cantidad de información, a pesar de su sencillez y facilidad de aplicación, utilizando las funciones básicas proporcionadas por todos los paquetes SIG. En este método los parámetros de la función de integración son elegidos en base al juicio subjetivo (experiencia) del investigador sobre la importancia relativa de las distintas capas y clases.

La naturaleza computacional de todas estas metodologías, junto con la estructura ráster de las capas de información, es ideal para la producción de mapas finales integrados de objetivos de exploración. Esta estrategia integrada parece ser el camino a seguir en la tarea cada vez más compleja y difícil de la investigación de recursos geológicos.

Agradecimientos

El trabajo se ha beneficiado de la financiación del proyecto GEOSDA CGL2010-17629 del Ministerio de Ciencia e Innovación de España y del Grupo de Investigación *Geoestadística, Teledetección y SIG* (RNM122) de la Junta de Andalucía, España.

Bibliografía

- Agterberg, F.P., Bonham-Carter, G.F., Cheng, Q., Wright, D.F., 1993, Weights of evidence modeling and weighted logistic regression for mineral potential mapping, *en* Davis, J.C., Herzfeld, U.C., (eds.), *Computers in Geology – 25 Years of Progress*: New York, Oxford University Press, 13-32.
- Aronoff, S., 1989, *Geographic information systems: a management perspective*: Ottawa, Canada, WDL Publications, 294 p.
- Arribas-Rosado, A., Arribas-Moreno, A., 1995, Caracteres metalogénicos y geoquímica isotópica del azufre y del plomo de los yacimientos de minerales metálicos del sureste de España: *Boletín geológico y minero*, 106, 23-62.
- Arribas Jr., A., Rytuba, J.J., Cunningham, C.G., Kelly, W.C., Rye, R.O., Castroviejo, R., 1988, Rodalquilar deposits, Spain, first example of caldera-related epithermal mineralization in Europe, Part 2: Ore genesis (abstract): *Geological Society of America Abstracts with Programs*, 20, A351.
- Arribas Jr., A., Rytuba, J.J., Rye, R.O., Cunningham, C.G., Podwysocki, M.H., Kelly, W.C., Arribas, A., McKee, E.H., Smith, J.G., 1989, Preliminary study of the ore deposits and hydrothermal alteration in the Rodalquilar caldera complex, southeastern Spain: *U.S. Geological Survey Open-File Report* 89-327, 44 p.
- Berry, J.K., 1993, *Beyond Mapping: concepts, algorithms, and issues in GIS*: Fort Collins, Colorado, USA, GIS World Books, 266 p.
- Bonham-Carter, G.F., 1994, *Geographic information systems for geoscientists: modelling with GIS*: Ottawa, Canada, Pergamon, 398 p.
- Bonham-Carter, G.F., Agterberg, F.P., Wright, D.F., 1989, Weights of evidence modelling: a new approach to mapping mineral potential, *en* Agterberg, F.P., Bonham-Carter, G.F. (eds.), *Statistical applications in the earth sciences: Geological Survey of Canada Paper* 89-9, 171-183.
- Burrough, P.A., 1986, *Principles of geographical information systems for land resources assessment*: Oxford, Reino Unido, Clarendon Press, 194 p.
- Bustillo-Revuelta, M., López-Jimeno, C., 1996, Recursos minerales: tipología, prospección, evaluación, mineralurgia e impacto Ambiental: Madrid, Gráficas Arias Montano, 372 p.
- Carranza, E.J.M., 2009, Controls on mineral deposit occurrence inferred from analysis of their spatial pattern and spatial association with geological features: *Ore Geology Reviews*, 35, 383-400.
- Cassard, D., Billa, M., Lambert, A., Picot, J.C., Husson, Y., Lasserre, J.L., Delor, C., 2008, Gold predictivity mapping in French Guiana using an expert-guided data-driven approach based on a regional-scale GIS: *Ore Geology Reviews*, 34, 471-500.
- Castroviejo, R., 1992, Ore Geology of the Santa Barbara Pb-Ag and Au District (SE Spain) in *Mineral Deposits Studies Group*: Aberdeen, Scotland, The Geological Society, 34, 471-500.
- Chica-Olmo, M., Delgado-García, J., Abarca-Hernández, F., Rigol-Sánchez, J.P., 1994, Elementos básicos para la construcción de un SIG para exploración minera (resumen), *en* *Comunicación III*: Madrid, España, Asociación Española de Sistemas de Información Geográfica.
- Chica-Olmo, M., Abarca-Hernández, F., Rigol-Sánchez, J.P., Delgado-García, J., 1996, Etapas del diseño de un sistema de información para exploración de recursos minerales (resumen), *en* *Congreso Nacional y Conferencia Internacional de Geología Ambiental y Ordenación del Territorio y Medio Ambiente*: Granada, España, 2, 23-26.
- Chica-Olmo, M., Rigol-Sánchez, J.P., Abarca-Hernández, F., Delgado-García, J., 1997, Desarrollo de un sistema de información para exploración de recursos minerales, *en* Casanova, J.L., Sanz Justo, J. (eds.), *Teledetección: Usos y Aplicaciones*: Valladolid, España, Secretariado de publicaciones e intercambio científico, Universidad de Valladolid, 317-322.
- Chica-Olmo, M., Abarca, F., Rigol, J.P., 2002, Development of a Decision Support System based on remote sensing and GIS techniques for gold-rich area identification in SE Spain: *International Journal of Remote Sensing*, 23, 4801-4814.
- Davis, J.C., 1986, *Statistics and data analysis in geology*: New York, John Wiley & Sons, 656 p.
- Evans, A.M., 1995, *Ore, Mineral Economics and Mineral Exploration*, in Evans, A.M. (ed.), *Introduction to Mineral Exploration*: Oxford, UK, Blackwell Science Ltd., 3-15.
- Fabbri, A.G., Chung, C.F., 1996, Predictive spatial data analysis in the geosciences, in Fisher, M., Scholten, H.J., Unwin, D. (eds.), *Spatial analytical perspectives on GIS*: London, Taylor and Francis, 147-159.
- Fernández-Soler, J.M., 1996, El volcanismo calco-alcalino del Parque Natural Cabo de Gata-Níjar (Almería): Estudio volcanológico y petrológico: Almería, Andalucía, España, Sociedad Almeriense de Historia Natural - Consejería de Medio Ambiente, Junta de Andalucía, 295 p.
- Haining, R., 1990, *Spatial data analysis in the social and environmental sciences*: Cambridge, UK, Cambridge University Press, 410 p.
- Hernández, J., De Larouzière, F.D., Bolze, J., Bordet, P., 1987, Le magmatisme néogène bético-rifain et le couloir de décrochement "trans-Alboran": *Bulletin de la Société Géologique de France*, 8, 257-267.
- López-Ruiz, J., Rodríguez-Badiola, E., 1980, La región volcánica neógena del sureste de España: *Estudios Geológicos*, 36, 5-64.

- Luo, J., 1990, Statistical mineral prediction without defining a training area: *Mathematical Geosciences*, 22, 253-260.
- Moon, C.J., Whateley, M.K.G., 1995, From Prospect to Predevelopment, *en* Evans, A. (ed.), *Introduction to mineral exploration*: Oxford, Reino Unido, Blackwell Science Ltd., 63-91.
- Pendock, N., Nedeljkovic, V., 1997, Integrating geophysical data sets using probabilistic methods: *International Journal of Remote Sensing*, 18, 1627-1635.
- Petrucelli, J.D., Nandram, B., Chen, M., 1999, *Applied statistics for engineers and scientists*: Nueva Jersey, USA, Prentice Hall, 944 p.
- Rigol-Sánchez, J.P., 2000, *Aplicación de Sistemas de Información Geográfica y Teledetección en exploración minera*: Granada, España, Universidad de Granada, tesis doctoral, 415 p.
- Rigol-Sánchez, J.P., Chica-Olmo, M., Abarca, F., 2003, Artificial neural networks as a tool for mineral potential mapping with GIS: *International Journal of Remote Sensing*, 24, 1151-1156.
- Sabins, F.F., 1996, *Remote Sensing: principles and interpretation*, Nueva York, W.H. Freeman Company, 432 p.
- Tomlin, C.D., 1991, Cartographic modeling, *en* Maguire, D.J., Goodchild, M.F., Rhind, D.W. (eds.), *Geographical information systems: principles and applications*: Essex, Reino Unido, Longman Scientific & Technical, 361-374.
- Vincent, R.K., 1997, *Fundamentals of geological and environmental remote sensing*: Nueva Jersey, EUA, Prentice Hall, 370 p.

Manuscrito recibido: Octubre 15, 2009

Manuscrito corregido recibido: Diciembre 17, 2009

Manuscrito aceptado: Enero 02, 2010