



Tema 5



Mediciones y cálculo de estadísticas en entorno raster

Asignatura obligatoria G-180 Los Sistemas de Información Geográfica I (Raster). Grado en Geografía y Ordenación del Territorio. Curso 3º. Universidad de Cantabria.

Profesora: Olga DE COS GUERRA.

Profesora titular de la Universidad de Cantabria.

Área: Geografía Humana.

Licencia: Creative Commons BY-NC-SA 4.0 Internacional



Abordaremos por primera vez las operaciones de análisis espacial que permiten realizar mediciones y comprobar estadísticamente pautas de distribución. Presta atención a este tema porque la explicación de las pautas de distribución de un hecho en el territorio no se puede hacer sólo mostrando mapas; veremos cómo cuantificar objetivamente los patrones espaciales.

Índice de contenidos

	Página
Bloque 1. Mediciones espaciales: cálculo de áreas y perímetros de base raster a partir de un SIG	3
1.1. Cálculo del área en modelizaciones SIG raster	
1.2. Cálculo del perímetro en modelizaciones SIG raster	
Bloque 2. Cálculo de estadísticos espaciales a partir de un SIG raster	5
2.1. Estadísticos centrográficos	
2.2. Autocorrelación espacial	
2.3. Correlación y regresión espacial	
Bloque 3. Operaciones de cálculo condicionado a una variable o capa de referencia mediante SIG	10
3.1. Extracción de valores	
3.2. Muestreo de capas raster	
Bibliografía	12

Bloque 1. MEDICIONES ESPACIALES: CÁLCULO DE ÁREAS Y PERÍMETROS DE BASE RASTER A PARTIR DE UN SIG

El análisis de superficies mediante herramientas SIG es una operación que se resuelve con rapidez y que resulta muy funcional. No obstante, existen importantes diferencias entre las maneras de operar de un SIG raster y un vectorial; además, el análisis de superficies o áreas en entorno vectorial es más preciso.



En este caso, nos centraremos especialmente en las **mediciones raster**, si bien, es importante incidir en que estos cálculos en los vectoriales cuentan con más precisión y quedan almacenados en la base de datos [asociados a cada elemento de referencia], con lo que pueden ser usados y consultados en diferentes momentos con facilidad.

1.1. Cálculo del área en modelizaciones SIG raster



El cálculo del área mediante SIG en niveles raster es una **operación zonal de capa única**. Con ello, se obtendrá un resultado para cada uno de los valores de la capa de entrada, que forman zonas y clases; así, los programas SIG para analizar el área recurren a un concepto de “clase”, de modo que ya no está condicionado el resultado por la ubicación de las celdas (vecindad), sino por su pertenencia a una determinada categoría de la leyenda; se trata, por tanto, de una operación del tipo zonal.



El AREA permite calcular la superficie que ocupa cada valor de la variable de una capa raster que contenga categorías.

Para el cálculo del área de una determinada variable cartografiada en modelo raster, la resolución es la clave de cálculo. De hecho, la resolución (tamaño de pixel) permite conocer la superficie que representa cada celda en el mundo real y conociendo ese valor se estima el área de forma automática a partir del número total de celdas que cuentan con el mismo valor en el mapa de entrada.

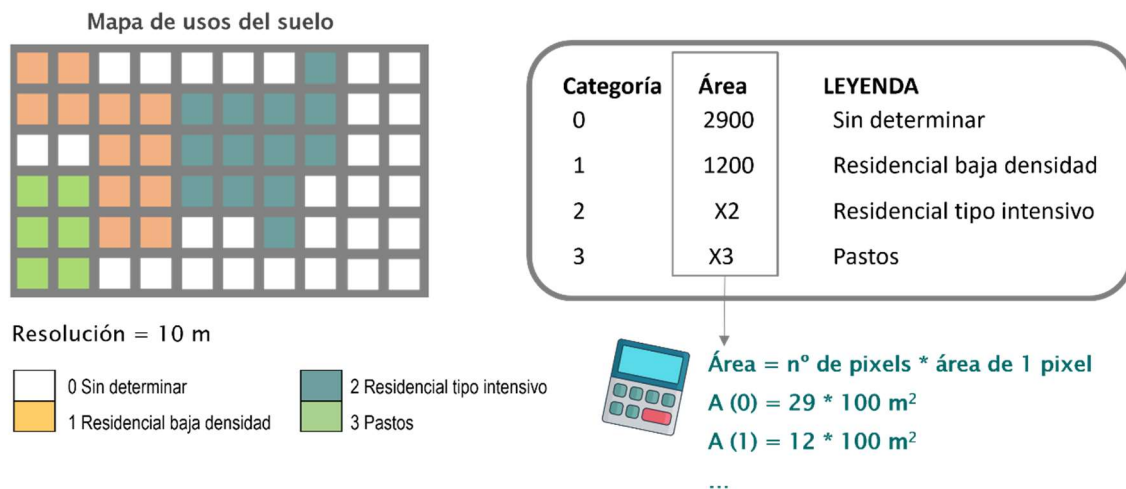
La **fórmula** de cálculo de base raster para el análisis de áreas es: el número de píxeles de cada valor de la variable multiplicado por el área de un píxel.

$$\text{Área} = \text{nº de pixels} * \text{área de 1 pixel}$$

De este modo, se mide la superficie asociada a cada valor de la variable del mapa (clase). Véase por ejemplo en la imagen, un mapa raster que tiene cuatro códigos de usos del suelo [0, 1, 2, 3] organizados en seis

manchas compactas (o zonas: grupos contiguos de celdas con el mismo atributo/valor).

Figura 1. Esquema de cálculo del área de una capa raster



En el ejemplo planteado en la Fig. 1, el cálculo del área permite obtener sólo cuatro valores de área: uno, para las tres manchas del tipo 0, uno para las celdas de código 1, uno para la mancha de tipo 2 y, finalmente, uno para la mancha de tipo 3.

1.2. Cálculo del perímetro en modelizaciones SIG raster

Al igual que el cálculo del área, el **perímetro es una operación zonal de capa única**. Los perímetros mediante un SIG raster se calculan en función del **número de lados de píxel** (celdas) que forman la **frontera exterior** de la zona.

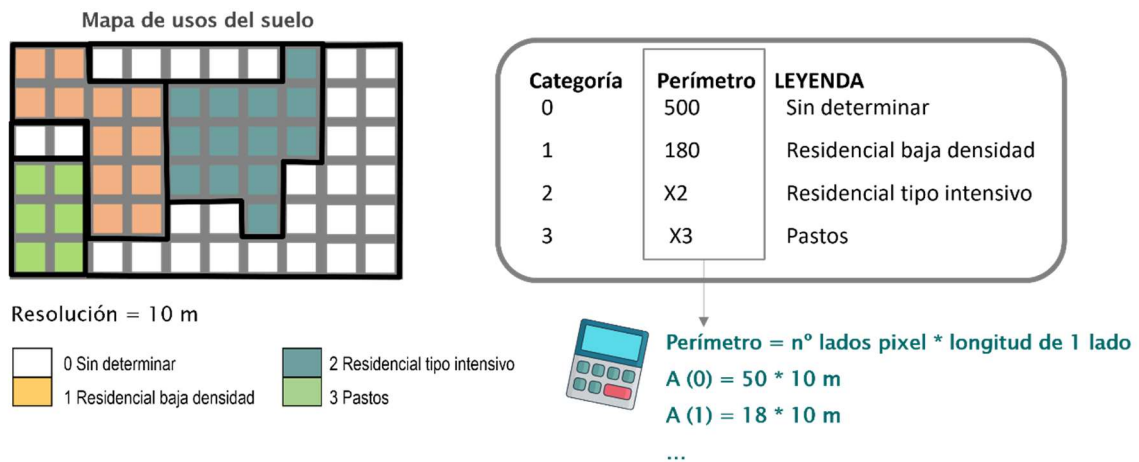
El procedimiento es semejante al del área –también interviene la resolución–; si bien, en el caso del perímetro se consideran el número de lados que definen un conjunto de celdas con cierto valor y ese número de lados se multiplica por la medida de longitud de un lado.

$$\text{Perímetro} = \text{n}^\circ \text{ lados pixel} * \text{longitud de 1 lado}$$

Tomando como referencia el mismo raster de base que en el área, en la Fig. 2 se muestra en línea gruesa de color negro el exterior envolvente de cada categoría de la leyenda raster (es decir, de cada clase).

El perímetro tiene en cuenta sólo los lados que quedan en cada polígono envolvente de categoría, bien sean limítrofes a otras categorías, o bien correspondan a límite de capa.

Figura 2. Esquema de cálculo del perímetro de una capa raster



Bloque 2. CÁLCULO DE ESTADÍSTICOS ESPACIALES A PARTIR DE UN SIG RASTER

2.1. Estadísticos centrográficos

Se trata de los estadísticos de centralidad y dispersión llevados al ámbito espacial. Entre los estadísticos posibles destacan dos medidas de centralidad: el centro medio y el centro mediano, cuyas características principales se presentan a continuación.

A) Centro medio

El centro medio se define como la media aritmética de las coordenadas X e Y de todos los píxeles que integran una determinada zona o categoría de la leyenda.

Figura 3. Esquema de cálculo del centro de gravedad





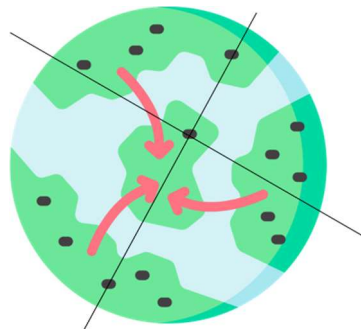
Su valor resultante implica, por tanto, la obtención de una **localización puntual** a partir de un par de coordenadas que se traducen en un archivo vectorial.

Con frecuencia **lleva asociadas medidas complementarias de dispersión**, como la desviación de las distancias o el radio estándar (análogo espacial directo de la desviación estándar) que informan del grado de dispersión de un hecho geográfico respecto a su centro medio.

B) Centro mediano

Es la traducción espacial de la mediana [valor que divide una serie en dos partes iguales y equivale al percentil 50]. Al igual que el centro medio, el centro mediano es también un punto. En este caso sus coordenadas son las medianas de las de los puntos analizados en cada eje (Olaya, 2020). Es menos sensible a valores extremos que el centroide (o media espacial).

Figura 4. Esquema gráfico del centro mediano



En el centro mediano se distinguen dos variantes de cálculo; según el enfoque inglés, es el punto de **intersección de dos ejes ortogonales** de modo que cada uno de ellos dividen las ubicaciones de partida en dos partes iguales (no es un punto único, sino que puede adoptar múltiples ubicaciones cumpliendo esta definición). Por otro lado, el enfoque norteamericano considera el centro mediano como el **punto de desplazamiento mínimo** a las ubicaciones de partida.

C) Autocorrelación espacial



La autocorrelación permite analizar en qué grado una variable medida en diferentes puntos del territorio se relaciona con ella misma (Bosque, 1992: 191). Esta definición enlaza, como vimos en temas anteriores, con la llamada **Primera ley de la Geografía** (Tobler¹, 1959) por

¹ TOBLER, W.R. (1959): "Automation and Cartography". *Geographical Review*, pp. 526–534.

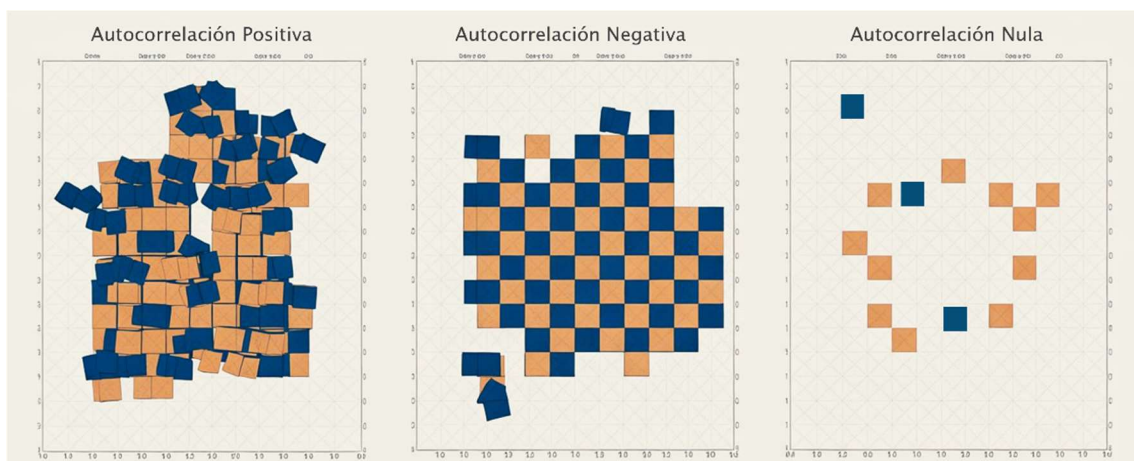
la que los elementos próximos en el espacio tienden a parecerse más que los que se encuentran alejados entre sí.

Su utilidad podría orientarse, por ejemplo, a analizar o comprobar si la aparición de un determinado valor de una variable en un lugar hace más (o menos) posible que ocurra ese mismo valor en las zonas limítrofes. Con frecuencia, la autocorrelación espacial se incorpora en estudios de procesos causales de “contagio espacial o difusión espacial” (Bosque, 1992), que permite comprender, razonar y explicar determinados procesos espaciales.

Existen **tres tipos** de resultados en el análisis de autocorrelación:

- **Autocorrelación positiva**, si existe tendencia a que el hecho registrado en un lugar también aparezca en su entorno próximo.
- **Autocorrelación negativa**, si existe tendencia a que el hecho registrado en un lugar hace poco probable que aparezca en su entorno próximo.
- **Autocorrelación nula**, si no existe relación entre la presencia de un hecho en una determinada posición y su entorno próximo.

Figura 5. Ejemplo teórico de comportamiento de una variable según su tipo de autocorrelación espacial



La autocorrelación espacial puede expresarse a partir de diferentes medidas o estadísticos, destacando especialmente las medidas de autocorrelación específicas que relacionan las distancias entre diferentes puntos del espacio analizado y las diferencias de valores temáticos de la variable estudiada.

En algunos programas SIG la autocorrelación se estima a partir del **índice de Morans**, que presenta dos variantes:

- **Rook** (semejante a la torre en el tablero de ajedrez) → analiza valores en línea recta vertical y horizontal respecto a cada celda.
- **King** (semejante al rey en el tablero de ajedrez) → analiza valores tanto en línea recta (igual que la torre) y en diagonal.

Se trata de un **índice acotado**, con valores comprendidos entre -1 (autocorrelación negativa, es decir, celdas contiguas con valores muy diferentes) y $+1$ (autocorrelación positiva, es decir celdas contiguas con valores muy parecidos).

Índice acotado entre -1 (máxima autocorrelación negativa) y 1 (máxima autocorrelación positiva) con nivel intermedio 0 para la ausencia de autocorrelación

2.3. Correlación y regresión espacial

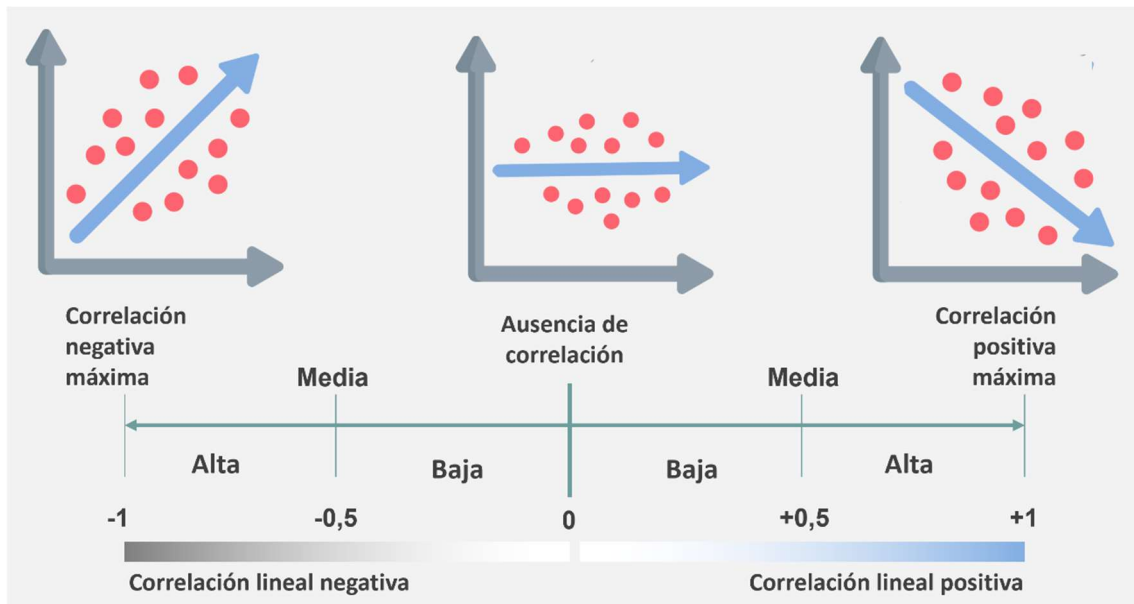
En estadística espacial mediante SIG cobran especial importancia los conceptos de **correlación y regresión**, que en términos generales (estadísticos) hacen referencia al conjunto de métodos que permiten comparar dos o más conjuntos de datos (en nuestro caso mapas, capas o variables espaciales) y sirven para determinar la fuerza de la asociación entre las variables de partida, indicando el tipo de relación existente.

Es bien sabido que uno de los estadísticos de correlación más extendido es el **coeficiente de correlación de Pearson**, que permite medir el grado y tipo de asociación lineal entre dos variables; así, sólo se genera un elevado coeficiente de correlación cuando la relación entre las variables es lineal.


Su valor se encuentra acotado entre -1 y $+1$, informando de intensidad de la relación y del signo o tipo de la misma. Tal como muestra el esquema explicativo adjunto:

- **Correlación positiva** (con máxima intensidad en $+1$) implica que a medida que aumenta el valor de la variable del mapa X también aumenta el valor de la variable del mapa Y.
- **Correlación negativa** (con máxima intensidad en -1) implica que a medida que aumenta el valor de la variable del mapa X disminuye el valor de la variable del mapa Y.
- **Correlación lineal nula** → valor de Pearson 0 .

Figura 6. Esquema de interpretación de la correlación lineal



Otro indicador ampliamente utilizado –derivado del coeficiente de correlación de Pearson– es el **coeficiente de determinación**, que consiste en r^2 y se interpreta como la explicación del grado de variación de una variable en función de la variación experimentada por la otra. Se interpreta en tantos por ciento (%).

 En los casos en los que la correlación es intensa cobran especial importancia los análisis de regresión espacial. Recordemos, que la regresión (como concepto estadístico) permite predecir los valores que asumirá una variable Y (dependiente), a partir de los de otra X (independiente), entre las que se plantea que existe cierta relación.

El planteamiento general se basa en determinar la ecuación de regresión mediante la cual se podrá estimar el valor de la variable dependiente (Y) cuando se tenga un valor conocido de la variable independiente (X).

- X → Variable independiente: predictora y explicativa
- Y → Variable dependiente: resultado de predicción y explicada

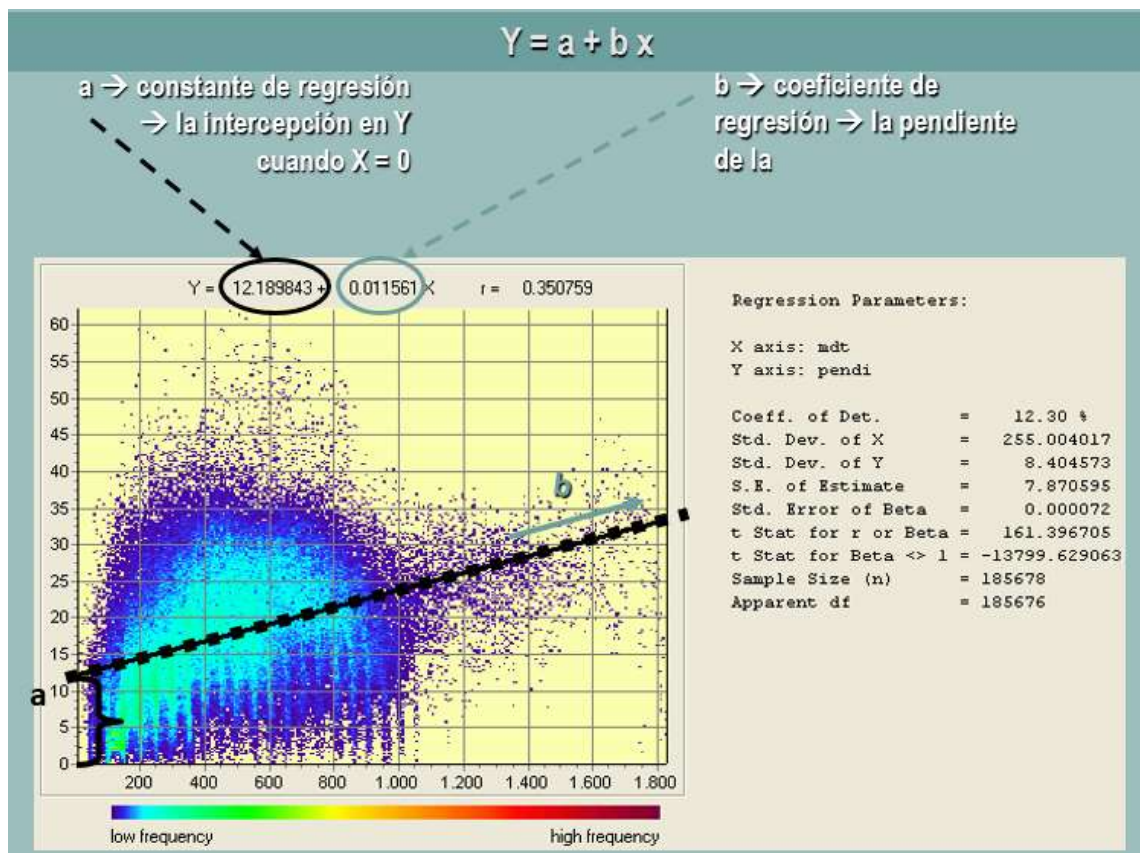


El método de regresión mediante un SIG **permite estimar** a partir del mapa de la variable X (independiente) y del mapa de la variable Y (dependiente) una **ecuación** que permitiría despejar cual sería el valor Y de un pixel para el que se conoce el valor X.

$$\text{Ecuación de regression lineal} \rightarrow Y = a + b x$$

Siendo: “Y” el valor promedio pronosticado de Y para cualquier valor de X; “a” la constante de regresión y “b” el coeficiente de regresión

Figura 7. Resultado de un análisis de regresión de las capas MDT (x) y Pendientes (y)



Tal como se observa en la Fig. 7, el análisis de regresión se basa en la selección de una serie de casos -celdas- (enumerados como pares de datos) para los que se dispone de ciertas observaciones de dos variables, que en el ejemplo son: mapa X [MDT] y mapa Y [pendientes]. A continuación, se dibuja un diagrama de dispersión (nube de puntos) que permite visualizar la relación existente entre los dos mapas considerados y a partir de ella determinar la ecuación de regresión.

Bloque 3. OPERACIONES DE CÁLCULO CONDICIONADO A UNA VARIABLE O CAPA DE REFERENCIA MEDIANTE SIG

En el presente apartado se incluyen algunas de las operaciones de cálculo principales en las que los resultados obtenidos están condicionados a un elemento de referencia representado cartográficamente en una capa de definición de elementos, que actúa como una variable discreta (una vía de comunicación, usos del suelo, cuenca visual, etc.).

Por ejemplo, si tenemos una capa vectorial de categorías de *Corine Land Cover* y deseamos obtener la altitud media de cada cobertura a partir de la información que contienen los píxeles del MDE.

3.1. Extracción de valores

La función extracción de valores permite –como su nombre indica– extraer u obtener información calculada para cierto mapa de definición de unidades espaciales en función de una capa raster que se procesa.



Se trata de una operación, por tanto, de **capa múltiple**.

Esta información procesada puede ser variada; así, es posible extraer un valor máximo, mínimo, medio... de variables como la altitud, la pendiente, la distancia a un elemento, etc. El resultado de esta operación suele ser una tabla con una serie de valores numéricos obtenidos para cada elemento de referencia.

3.2. Muestreo de capas raster

Una variante de la extracción de valores es el muestreo de capas raster a partir de una capa de referencia. Esta operación **cuenta con la particularidad de partir de un vectorial de puntos** –a modo de puntos muestrales–, aunque pueden responder a cualquier tipo de entidad puntual [puntos kilométricos, localizaciones singulares, paradas en un recorrido, emisores/receptores, antenas, etc.

Para ese vectorial se añaden una o varias capas raster de la misma zona que son las variables que se procesan para obtener los resultados de muestreo. Estas variables pueden ser diversas: altitudes, pendientes, orientaciones, distancias, etc. Como resultado de la operación se obtiene un valor numérico (valor de celda) para cada punto muestral y para cada capa raster incorporada en la operación.

Por ejemplo, si se hace un muestreo para 3 capas raster a partir de un vectorial de puntos que contiene 2 localizaciones puntuales, el resultado de la operación serán 6 valores calculados: el valor del punto 1 en las 3 capas raster y el valor que corresponda al punto 2 en las 3 capas raster.

BIBLIOGRAFÍA

- BOSQUE SENDRA, J. (1992): *Sistemas de Información Geográfica*. Madrid, Rialp.
- CEBRIÁN, J.A. (1992): *Información geográfica y Sistemas de Información Geográfica* Santander, Servicio de Publicaciones de la Universidad de Cantabria.
- COMAS, D. y RUIZ, E. (1993): *Fundamentos de los sistemas de información geográfica*. Barcelona: Ariel.
- EQUIPO SEXTANTE (2008): *Manual práctico de Sextante en GvSIG*. En: http://forge.osor.eu/plugins/scmsvn/viewcvs.php/*checkout*/docs/LaTeX/ManualPracticoSEXTANTE/ManualPracticoSEXTANTE.pdf?root=sextante&rev=54&pathrev=54
- FELICÍSIMO, A.M. *Glosario de términos usados en el trabajo con sistemas de información geográfica*. En: www.etsimo.uniovi.es/~feli/pdf/glosario.pdf
- GONZÁLEZ AGUAYO, R.: *Diccionario de términos SIG*.
- GUTIÉRREZ PUEBLA, J. y GOULD, M. (1994): *SIG: Sistemas de Información Geográfica*. Madrid, Síntesis.
- HAINING, R. (1994): *Designing spatial data analysis modules for GIS*. En: Fotheringham and Rogreson (eds.) *Spatial Analysis and GIS*. Taylor and Francis: London, 45–64. Adaptado por Giscampus.org.
- LONGLEY, P., GOODCHILD M.F., MAGUIRRE, D.J., RHIND, D.W. (2011): *Geographic Information Systems and Science (3rd Edition)*. Chichester: John Wiley & Sons, cop. 2011.
- MORENO JIMÉNEZ, A. (1999): *Población y espacio en la Comunidad de Madrid. Análisis y aplicaciones a nivel microgeográfico*. Consejería de Hacienda. Madrid
- OLAYA, V. (2020): *Sistemas de Información Geográfica*. En: <http://volaya.github.io/libro-sig/>
- TOBLER, W.R. (1959): "Automation and Cartography". *Geographical Review*, págs. 526–534.