

Lección 6. Análisis de datos geográficos

Los modelos de abstracción tipo campo son en realidad matrices de datos. Esto nos permite realizar con ellos todo tipo de operaciones matriciales:

- Superposición de matrices
- Operadores escalares
- Reclasificación

Operaciones matriciales

Los modelos de abstracción tipo campo y objeto son susceptibles de ser analizados mediante:

Técnicas Estadísticas Descriptivas

- Medidas de tendencia central y dispersión.
- Distribución de frecuencias.

Inferencias Estadísticas

- Teoría de muestras y estimación de parámetros.
- Contraste de hipótesis y diseño experimental.
- Inferencia bayesiana y métodos no paramétricos.

Técnicas Estadísticas

Análisis espectral

Los análisis espectrales se realizan sobre imágenes remotas tomadas mediante sensores digitales sensibles en diferentes rangos del espectro electromagnético. La combinación de bandas espectrales permite identificar y cuantificar procesos naturales relacionados con la atmósfera, los océanos y la biosfera.

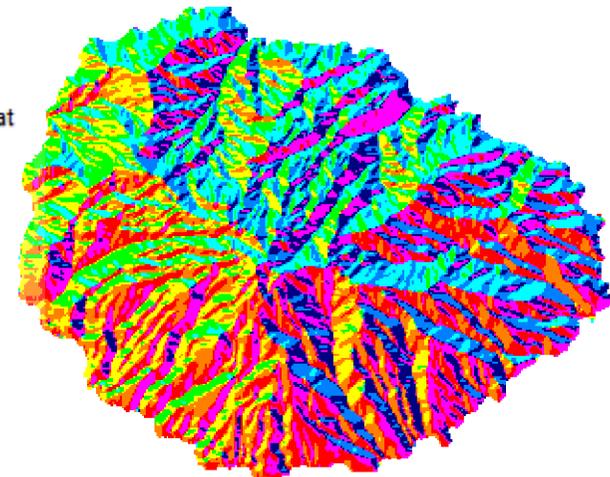
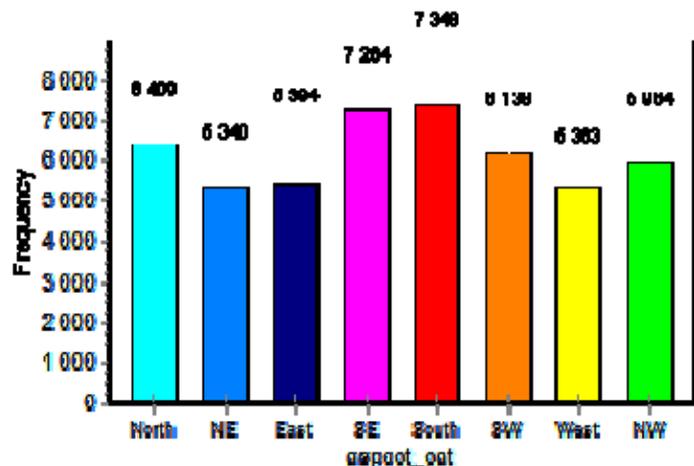
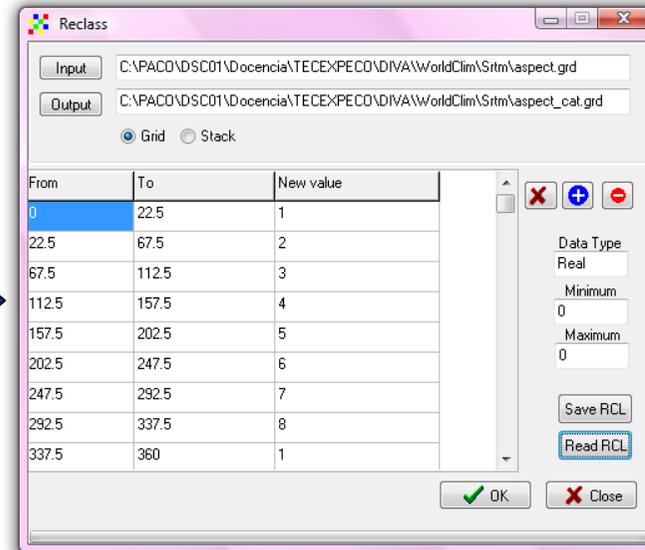
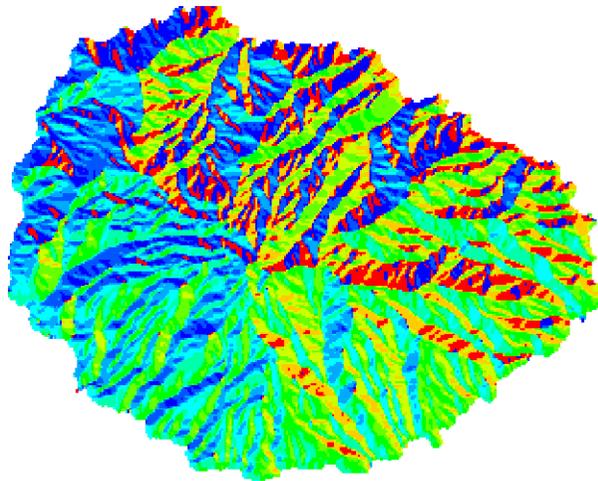
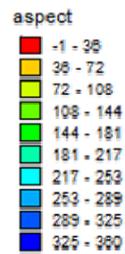
Análisis de continuidad espacial

Se trata de análisis en los que la posición geográfica de los valores del fenómeno representado mediante un modelo de abstracción, juega un papel principal. En este caso se habla de análisis de datos espacialmente explícitos:

- Modelos del terreno
- Patrones ecológicos
- Interpolación espacial y Geoestadística
- Autocorrelación

OPERACIONES MATRICIALES Y TÉCNICAS ESTADÍSTICAS

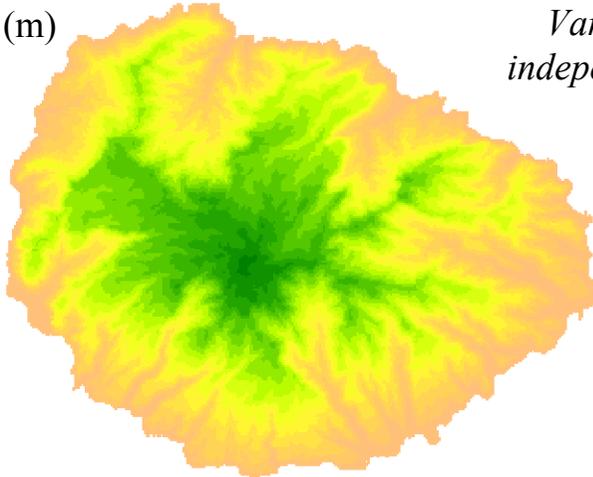
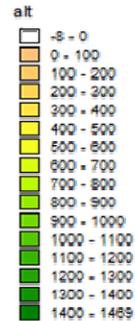
Ejemplo 1. Categorización de un variable continua (Orientación) y distribución de frecuencias de las categorías.



OPERACIONES MATRICIALES Y TÉCNICAS ESTADÍSTICAS

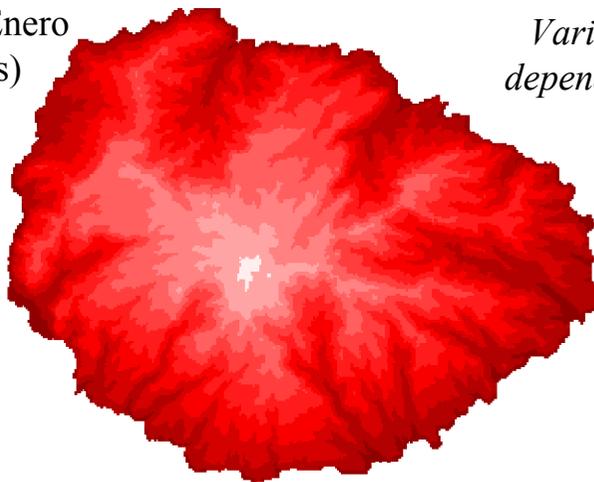
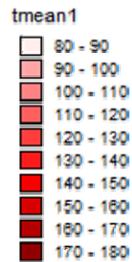
Ejemplo 2. Regresión lineal: Temperatura media de enero versus elevación del terreno.

Elevación (m)

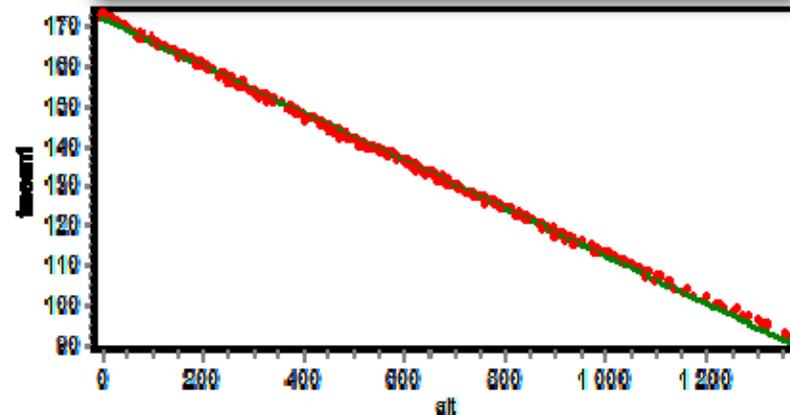
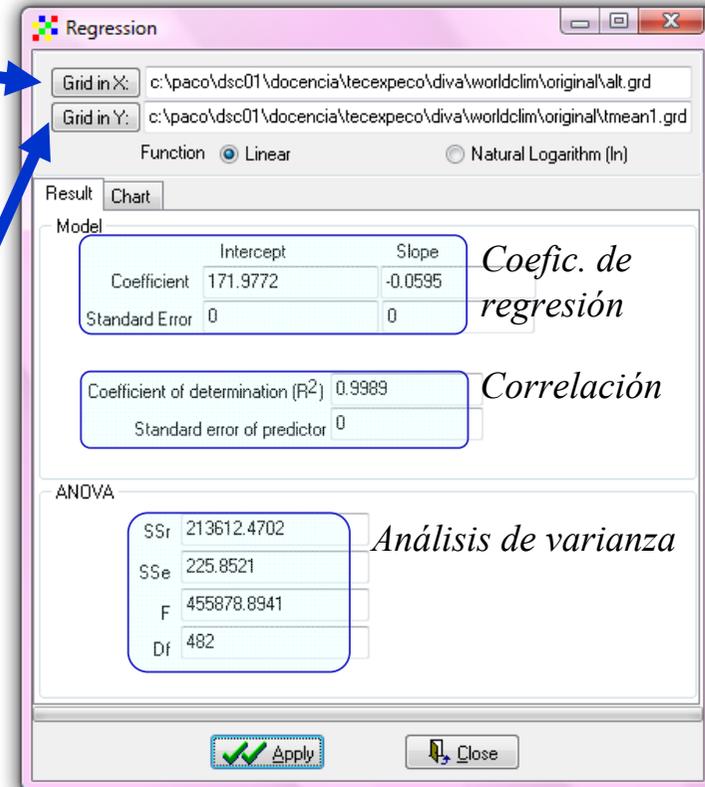


Variable independiente

Temp. Media Enero
(déc. grados)



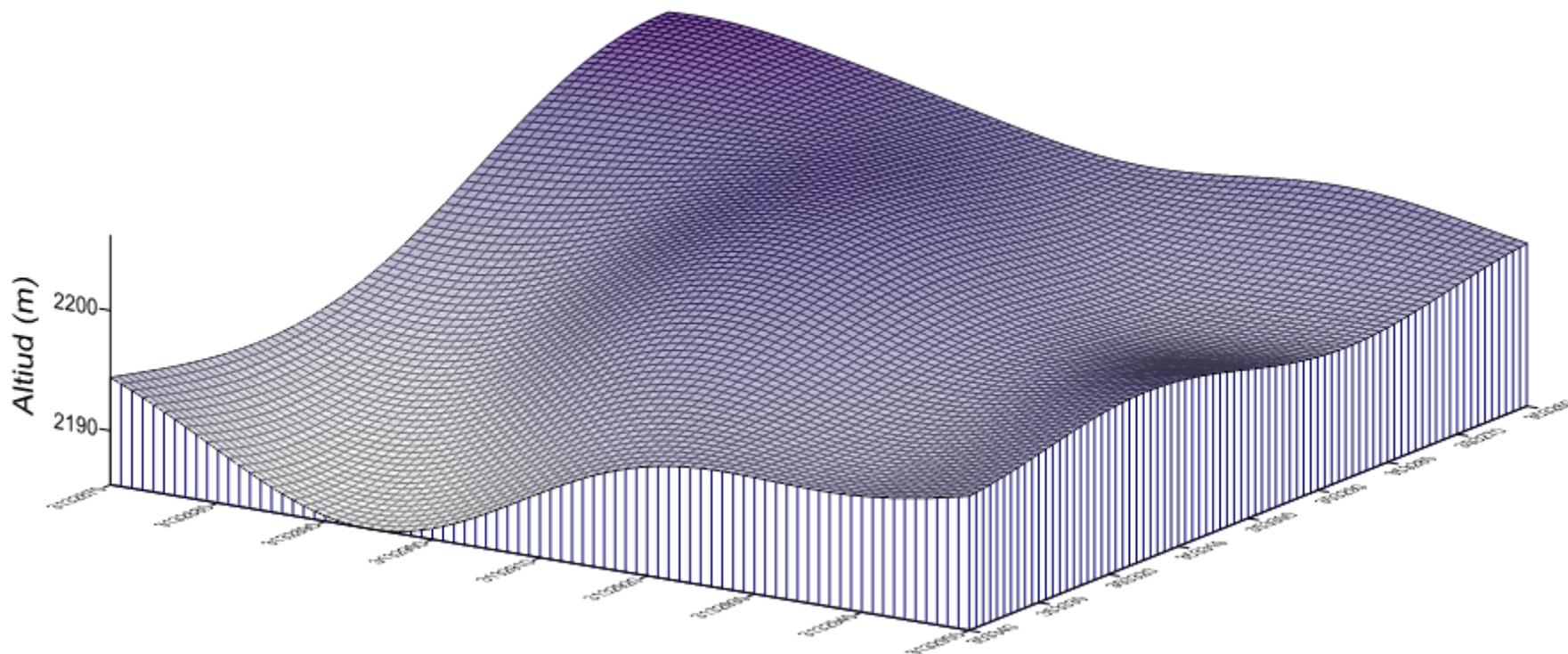
Variable dependiente



ANÁLISIS DE CONTINUIDAD ESPACIAL

Modelos del Terreno

MODELO DE ELEVACIÓN DEL TERRENO REPRESENTADO EN 3D



ANÁLISIS DE CONTINUIDAD ESPACIAL

Modelos del Terreno

PENDIENTE

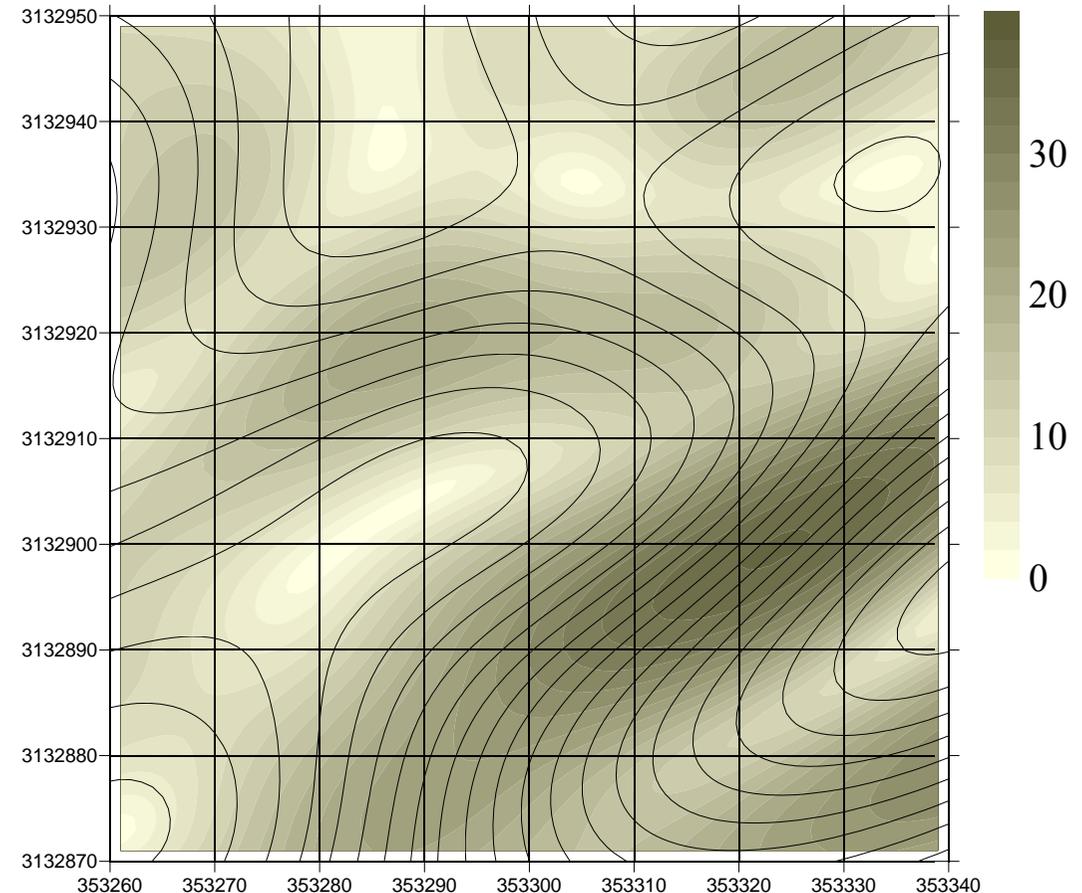
La pendiente de cada celdilla se determina como el vector resultante de la pendiente en dirección del eje X y la pendiente en dirección del eje Y.

La pendiente (S_T) se calcula de la siguiente forma:

$$S_T \approx \frac{360}{2\pi} \times \arctan \left[\sqrt{\left(\frac{Z_E - Z_W}{2\Delta x} \right)^2 + \left(\frac{Z_N - Z_S}{2\Delta y} \right)^2} \right]$$

siendo Z_N , Z_S , Z_E y Z_W la altura de las celdillas circundantes de la celdilla para la que se está calculado el valor de pendiente. Δx e Δy representa la resolución de la celdilla en dirección E-W y N-S respectivamente.

Los valores de pendiente del terreno se expresan en grados.

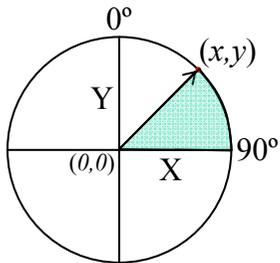


ANÁLISIS DE CONTINUIDAD ESPACIAL

Modelos del Terreno

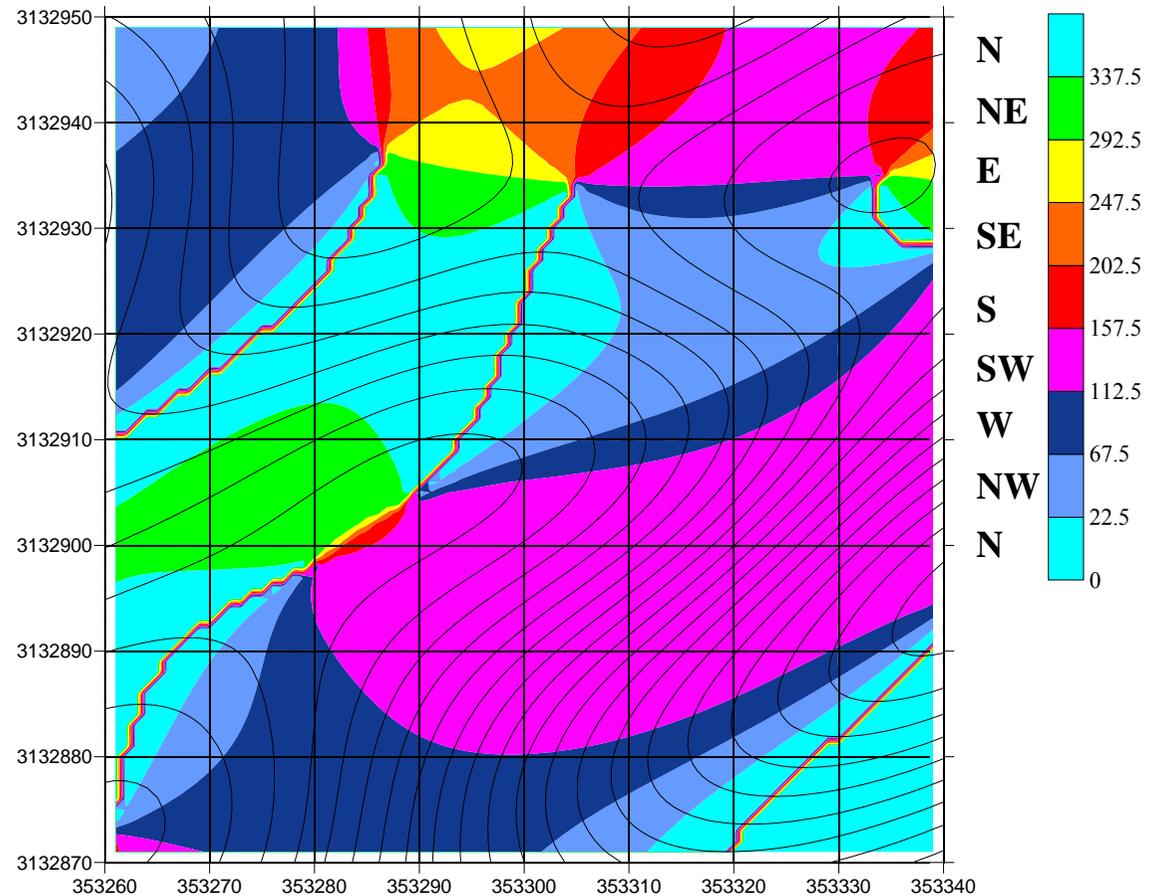
ORIENTACIÓN

La orientación (A_T) se calcula como el arco tangente de un plano definido por unas coordenadas X e Y especificadas. El arco tangente es el ángulo medido desde el eje X hasta la línea que contiene el origen (0,0) y el punto con las coordenadas (coord_x, coord_y):



$$A_T \approx 270 - \frac{360}{2\pi} \times a \tan 2 \left[\frac{Z_N - Z_S}{2\Delta y}, \frac{Z_E - Z_W}{2\Delta x} \right]$$

siendo Z_N , Z_S , Z_E y Z_W la altura de las celdillas circundantes de la celdilla para la que se está calculado el valor de orientación. Δx e Δy representa la resolución de la celdilla en dirección $E-W$ y $N-S$ respectivamente.



Los valores de orientación del terreno se expresan en grados considerando el Norte a 0° y el Este a 90° .

ANÁLISIS DE CONTINUIDAD ESPACIAL

Modelos del Terreno

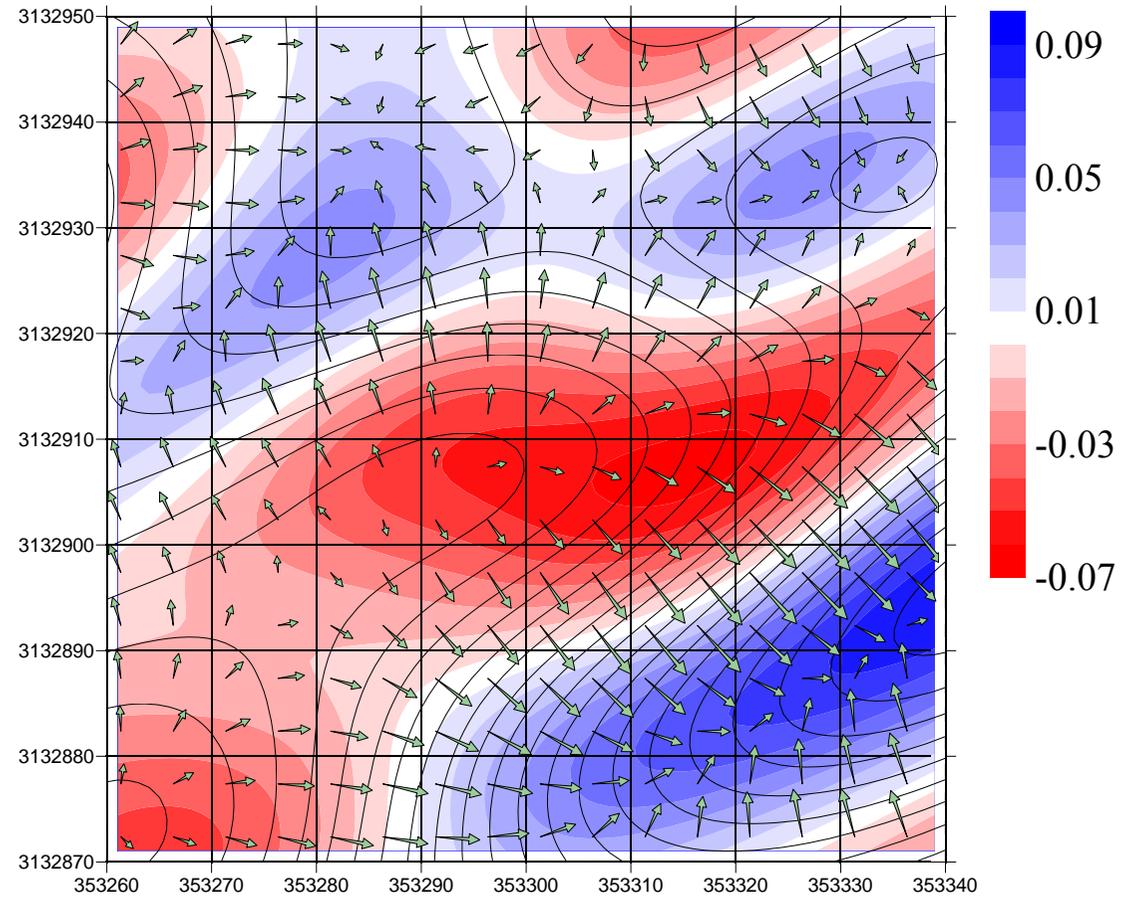
OPERADOR DE LAPLACE

Permite estimar las zonas de recarga y descarga en una superficie. Como resultado se obtiene una superficie con valores positivos en las zonas de recarga y negativos en las zonas de descarga. Se utiliza con variables físicas conservativas, como el agua superficial del suelo, cuya tasa local de flujo es proporcional al gradiente local.

El operador Laplaciano ($\nabla^2 Z_{(x,y)}$) se calcula de la siguiente forma:

$$\nabla^2 Z_{(x,y)} = \left(\frac{Z_E - 2 \cdot Z + Z_W}{\Delta x^2} \right) + \left(\frac{Z_N - 2 \cdot Z + Z_S}{\Delta y^2} \right)$$

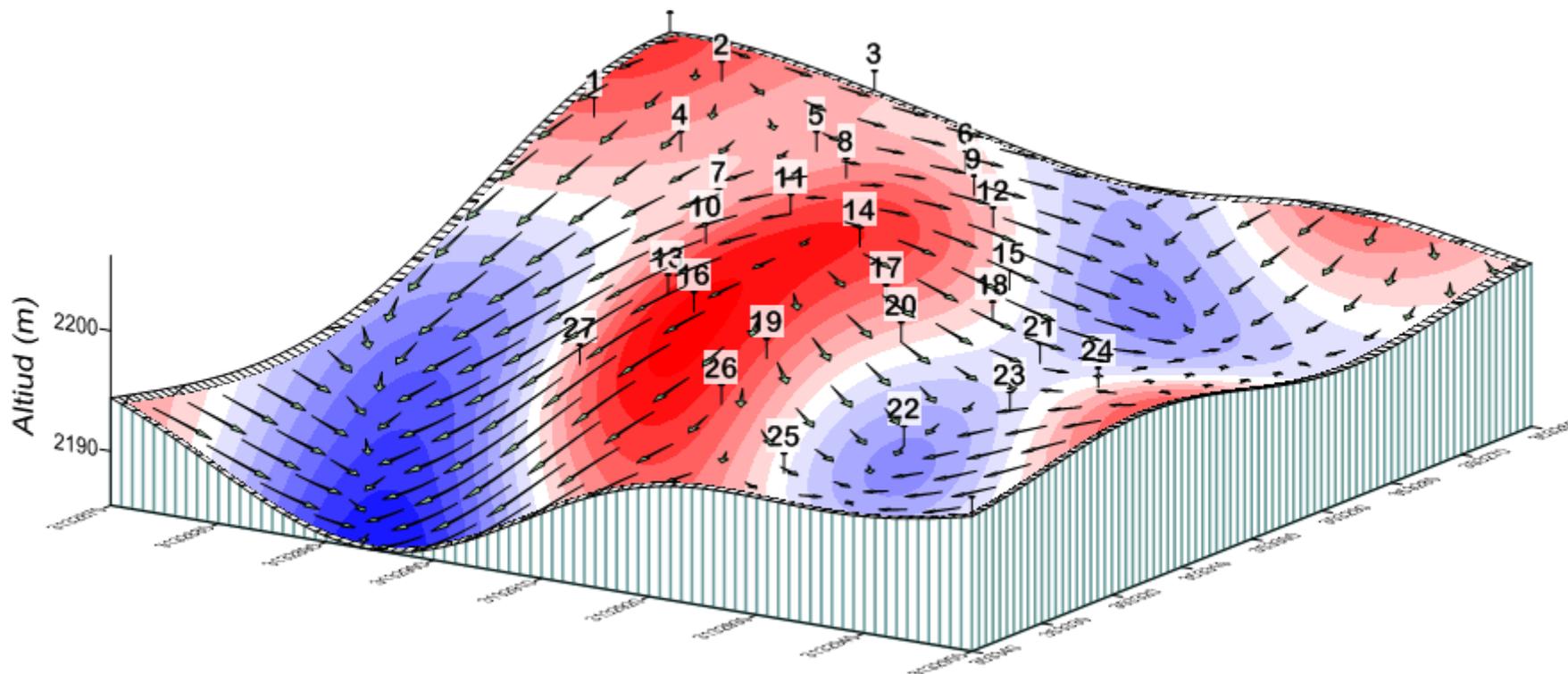
siendo Z la altura de la celdilla para la que se está calculado el valor del operador en la posición x, y . Z_E , Z_W , Z_N y Z_S hacen referencia a la altura de las celdillas circundantes de Z . Δx^2 y Δy^2 representa el cuadrado de la resolución de la celdilla en dirección E-W y N-S respectivamente.



ANÁLISIS DE CONTINUIDAD ESPACIAL

Modelos del Terreno

Operador de Laplace superpuesto al modelo de elevación del terreno en 3D.



ANÁLISIS DE CONTINUIDAD ESPACIAL

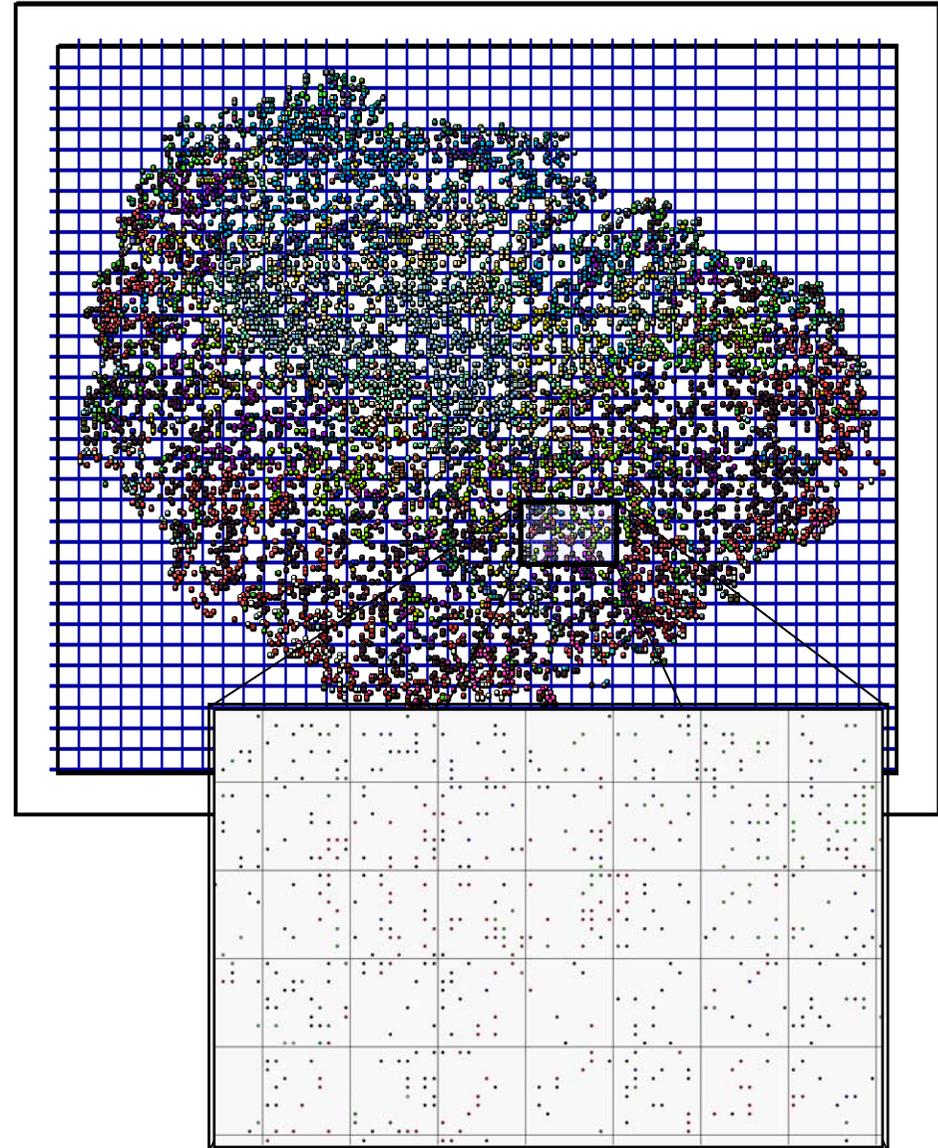
Patrones ecológicos

Estos análisis se basan en la localización y en un atributo adicional de un conjunto de puntos con dato (coord_x, coord_y, especie_i). Los puntos representan lugares donde un espécimen ha sido recolectado o se ha observado su presencia.

Los puntos vienen representados por un modelo de abstracción tipo objeto, asociado a una base de datos que nos indica la especie observada en cada punto. Los principales parámetros que se calculan son la riqueza, estimaciones de la riqueza máxima, diversidad y diversidad β .

Para realizar los cálculos previamente se define una retícula que nos permite contabilizar los puntos presentes en cada cuadrícula en función de sus correspondientes atributos, que en nuestro caso se refieren a especies de vegetación.

Como resultado del análisis, se obtiene un modelo de abstracción tipo campo que representa el patrón de variación del macroparámetro analizado.

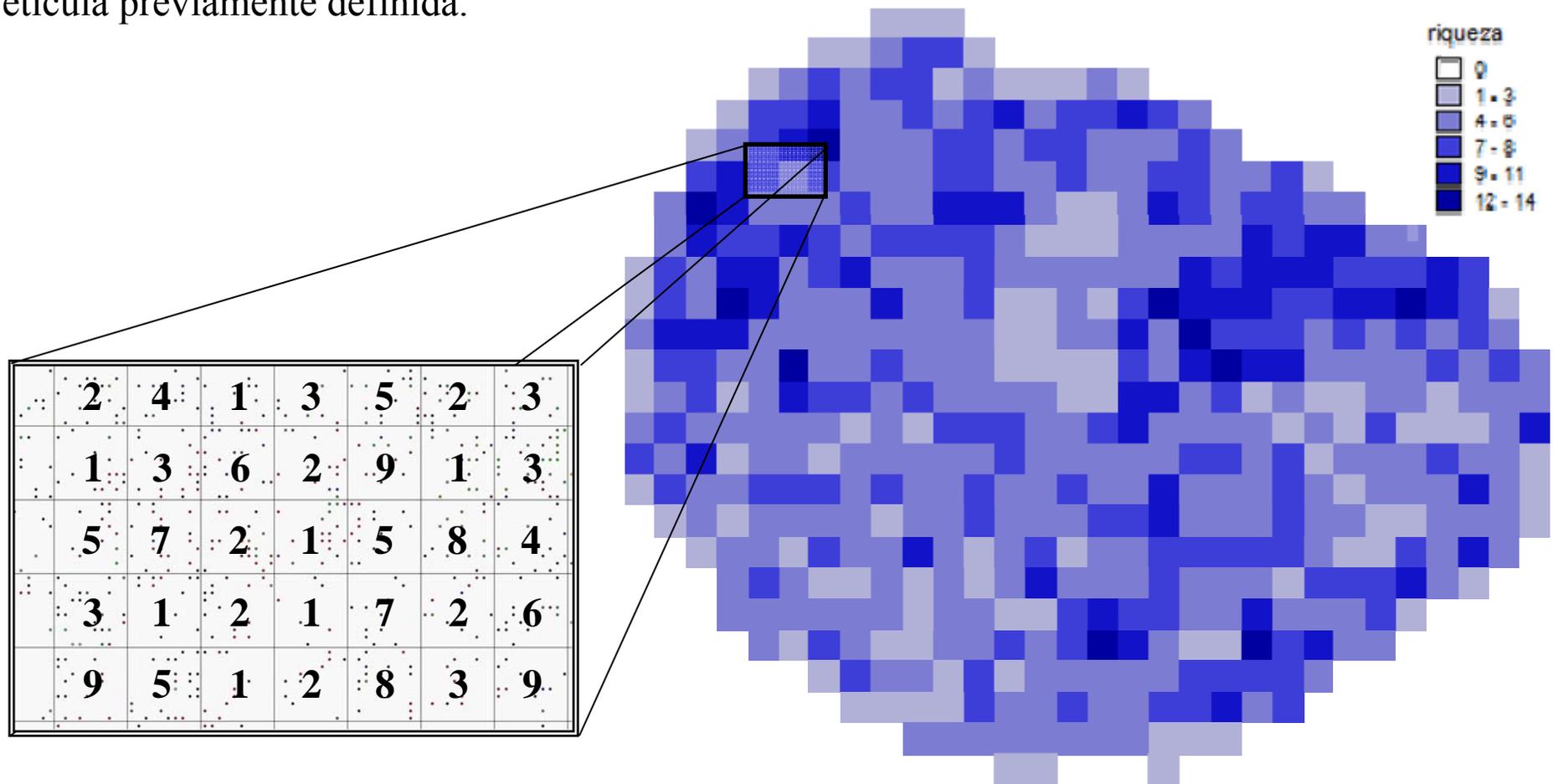


ANÁLISIS DE CONTINUIDAD ESPACIAL

Patrones ecológicos

RIQUEZA ESPECÍFICA

Se calcula como el número de especies diferentes presentes en cada celdilla de la retícula previamente definida.



ANÁLISIS DE CONTINUIDAD ESPACIAL

Patrones ecológicos

ESTIMADORES DE LA RIQUEZA MÁXIMA

El número de especies observado en una localidad depende en gran medida del esfuerzo invertido en el muestreo. De tal forma que si continuamos muestreando siempre encontraremos alguna especie nueva.

Realizar un censo completo de todas las especies presentes en una localidad resulta inviable, pero sí se pueden realizar estimaciones de la riqueza máxima en función de la riqueza observada.

Este cálculo aproximado permite comparar la abundancia de especies de diferentes localidades disminuyendo el sesgo impuesto por el esfuerzo de muestreo.

Además, calcular el número máximo de especies resulta útil para valorar si la información adicional que se gana continuando el proceso de muestreo justifica su coste.

ANÁLISIS DE CONTINUIDAD ESPACIAL

Patrones ecológicos

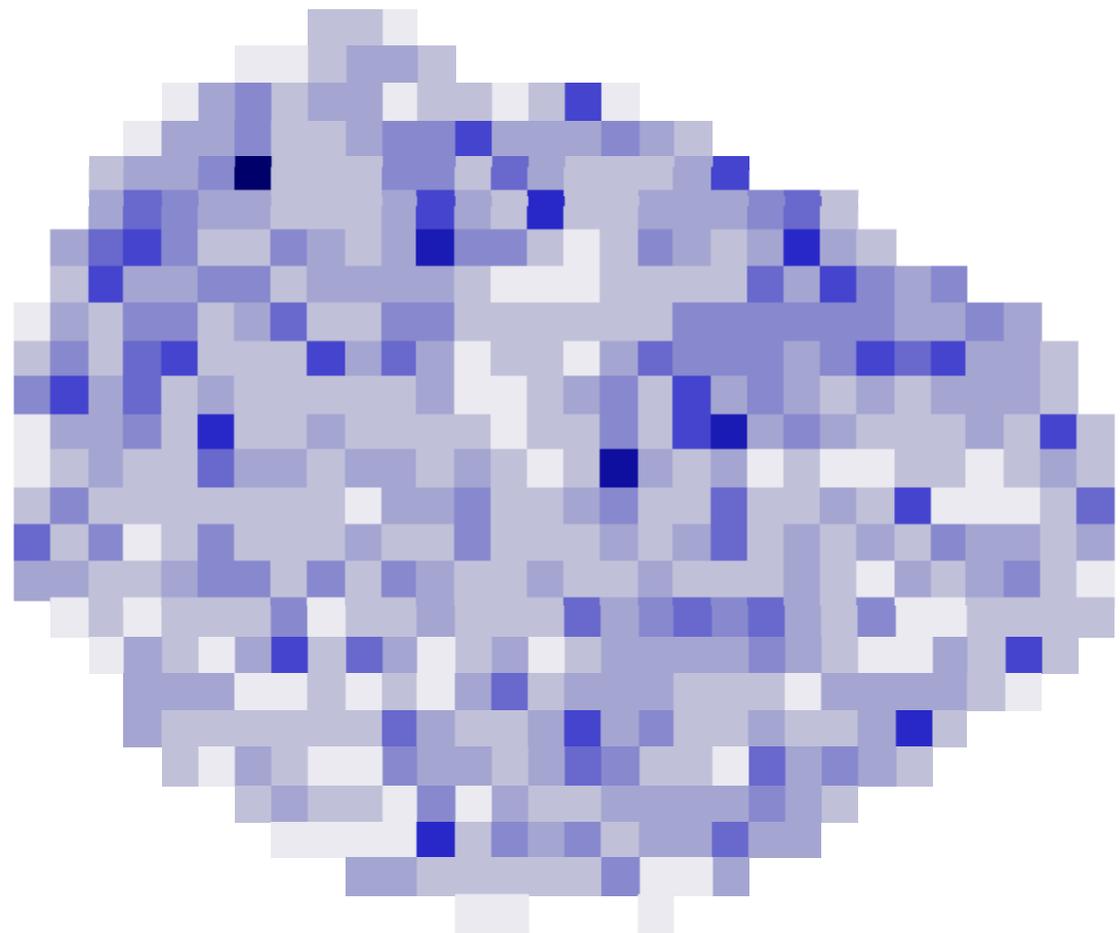
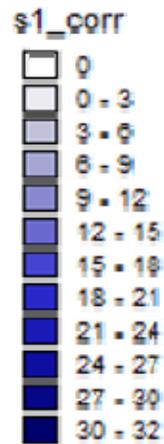
ESTIMADORES DE RIQUEZA MÁXIMA:

Chao 1 Corregido

Se trata de la versión corregida del estimador *Cho* original que ofrece estimaciones menos sesgadas por el esfuerzo de muestreo:

$$S_{Chao1} = S_{obs} + \frac{F_1^2}{2 \times (F_2 + 1)} - \frac{F_1 \times F_2}{2 \times (F_2 + 1)^2}$$

donde S_{obs} es el número total de especies observadas en una celdilla, F_1 el número de especies que tienen tan sólo un individuo y F_2 el número de especies que tienen 2 individuos exactamente.



ANÁLISIS DE CONTINUIDAD ESPACIAL

Patrones ecológicos

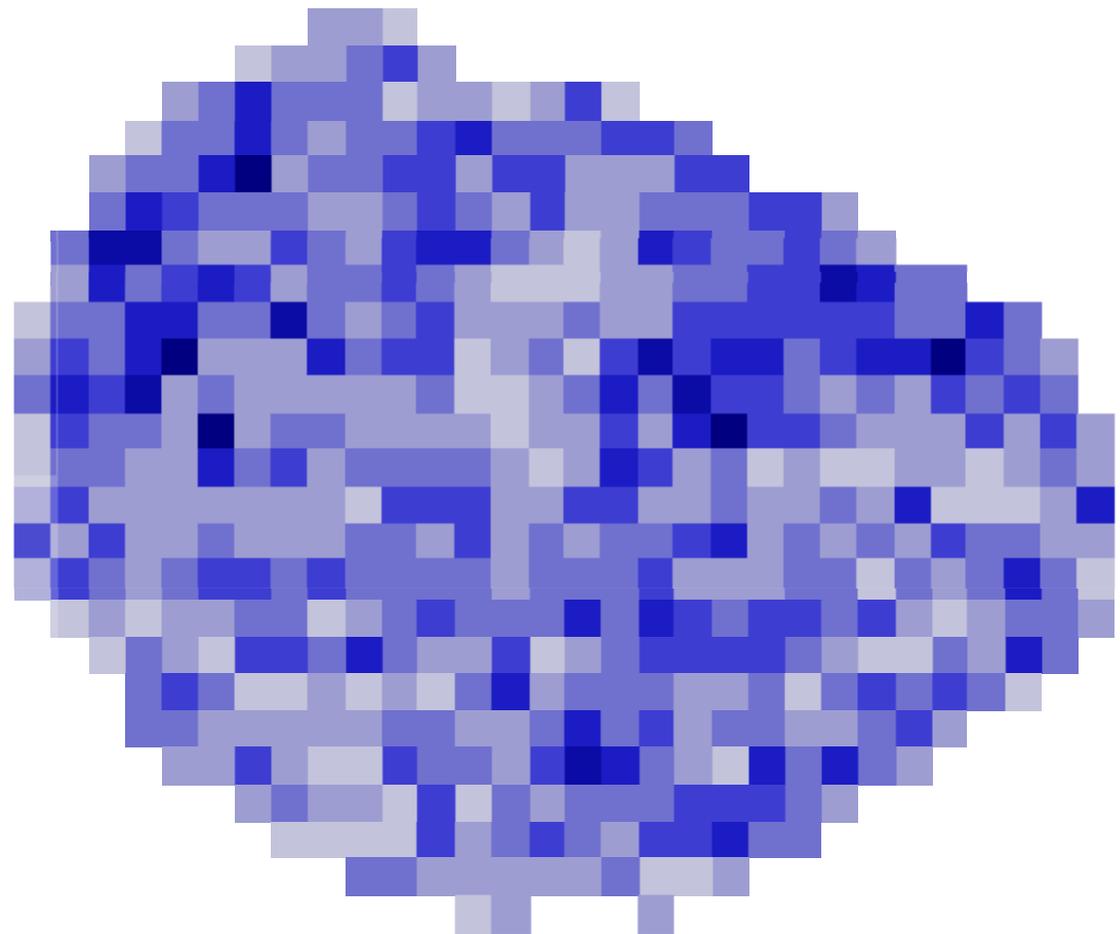
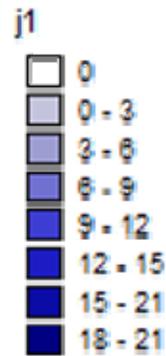
ESTIMADORES DE RIQUEZA MÁXIMA:

Jacknife de primer orden

Se trata de un estimador basado en el número de muestras para un área. Para crear las muestras se divide cada celdilla de la retícula en sub-áreas (de 4 a 9).

$$S_{jack1} = S_{obs} + Q_1 \left(\frac{m-1}{m} \right)$$

donde S_{obs} es el número total de especies observadas en todas las muestras tomadas en una celdilla, Q_1 el número de especies que se han encontrado, exactamente, en una sola muestra y m que representa el número total de muestras tomadas en la celdilla.



ANÁLISIS DE CONTINUIDAD ESPACIAL

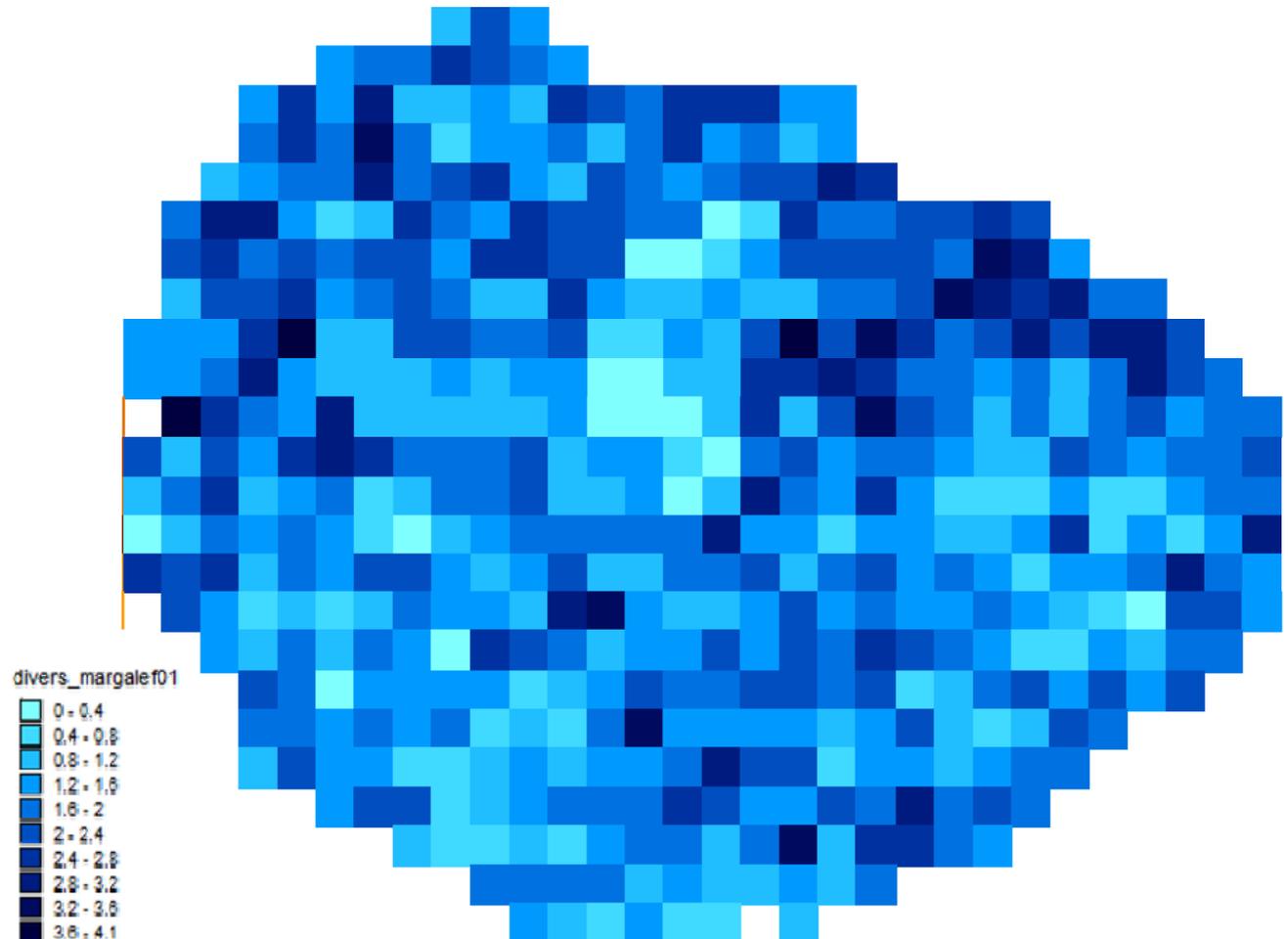
Patrones ecológicos

INDICES DE DIVERSIDAD:

Índice de Margalef

Se estima en función del número total de especies por celda (S) y el número total de observación por celda (N):

$$D_{Mg} = (S - 1) / \ln(N)$$



ANÁLISIS DE CONTINUIDAD ESPACIAL

Patrones ecológicos

INDICES DE DIVERSIDAD:

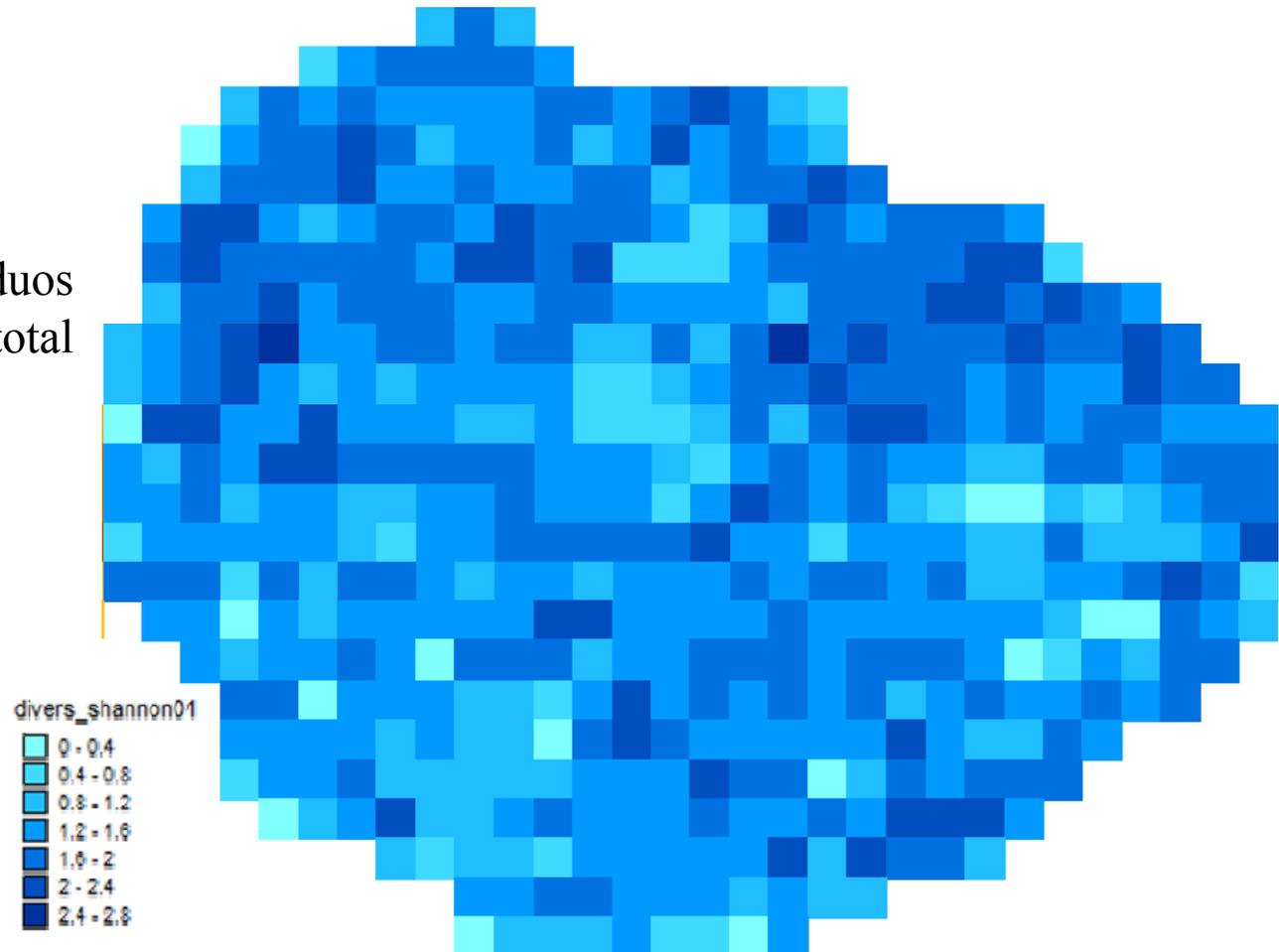
Índice de Shannon

Se estima en función de la abundancia proporcional de especies (p_i) por celda:

$$H' = -\sum p_i \cdot \ln p_i$$

$$p_i = n_i / N$$

siendo n_i el número de individuos de la especie i , y N el número total de observaciones por celda.



ANÁLISIS DE CONTINUIDAD ESPACIAL

Patrones ecológicos

DIVERSIDAD β (*Turnover*)

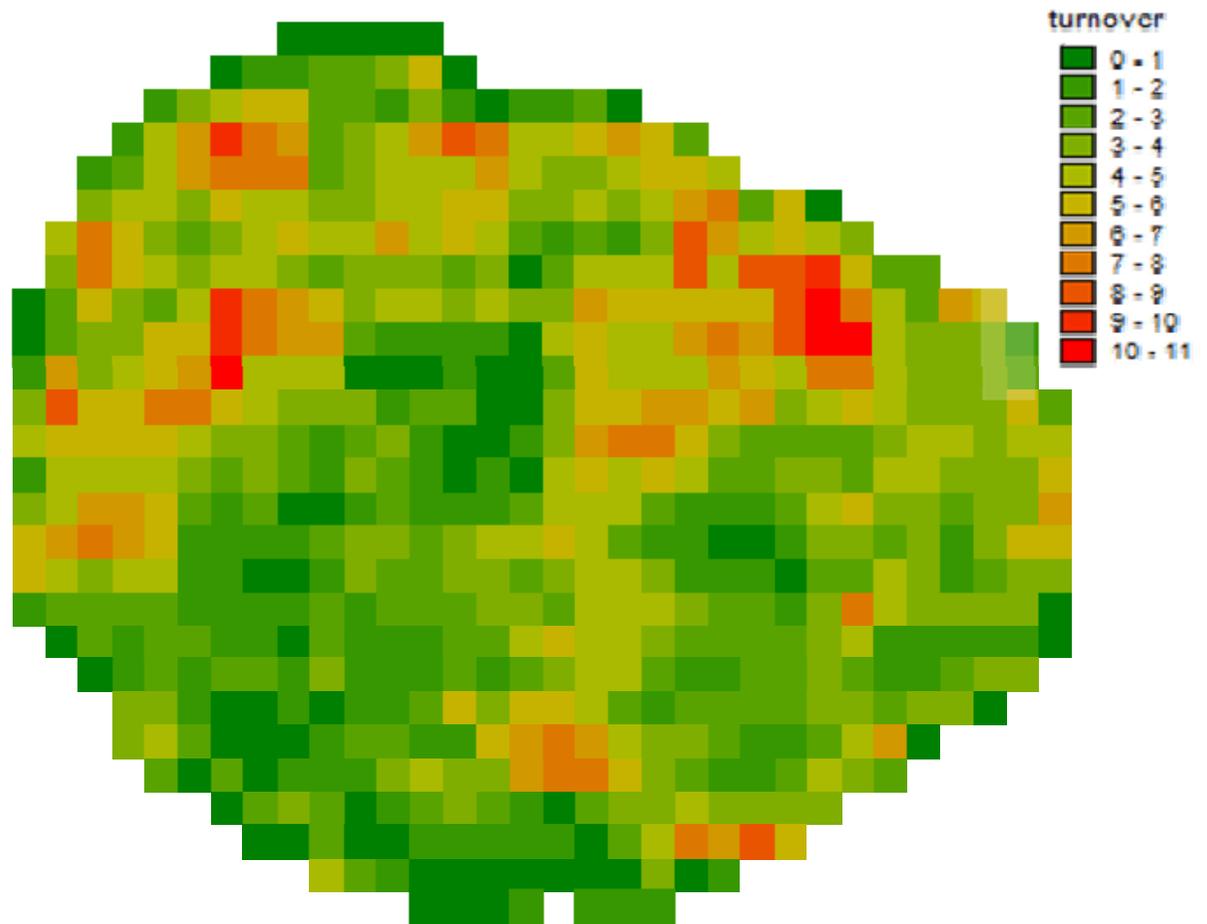
El turnover es una medida de la tasa de cambio del ensamblaje de especies en el espacio. Indica en que medida son diferentes localidades cercanas.

Imaginemos dos áreas grandes con un número similar de especies, pero una con especies diferentes en todas las celdas de la retícula y otra con la misma especie en todas las celdas de la retícula. La primera localidad tendría una tasa de cambio alta, mientras que la segunda localidad tendría una tasa de cambio baja.

La diversidad β de un celdilla se calcula teniendo en cuenta el número de especies presentes en las 8 celdillas que la rodean.

$$\beta_w = (S/\alpha) - 1$$

siendo S el número total y α el número medio de especies encontradas en todas las celdillas consideradas.



ANÁLISIS DE CONTINUIDAD ESPACIAL

Interpolación espacial y Geoestadística

La geoestadística se basa en la ley Tobler: “... *Todas los objetos están relacionados entre sí, pero los objetos más próximos en el espacio tienen una relación mayor que los distantes*”.

Se asume que existe un cierto grado de dependencia espacial entre los diferentes puntos que representan un fenómeno tipo objeto y, en base a esta dependencia, es posible modelizar, mediante técnicas de interpolación espacial, fenómenos de tipo campo.

Si tan sólo se quiere saber cual es la mejor representación gráfica de la continuidad del fenómeno estudiado, basta con aplicar técnicas simples de interpolación espacial:

- Inverso de la Distancia al Cuadrado
- Triangulación
- Función de Base Radial

Si además, queremos conocer los parámetros que definen el patrón de variación espacial y el grado de incertidumbre de estas estimas se utilizarían técnicas más complejas:

- Análisis de texturas mediante semivariogramas
- Interpolación espacial por el método Kriging

ANÁLISIS DE CONTINUIDAD ESPACIAL

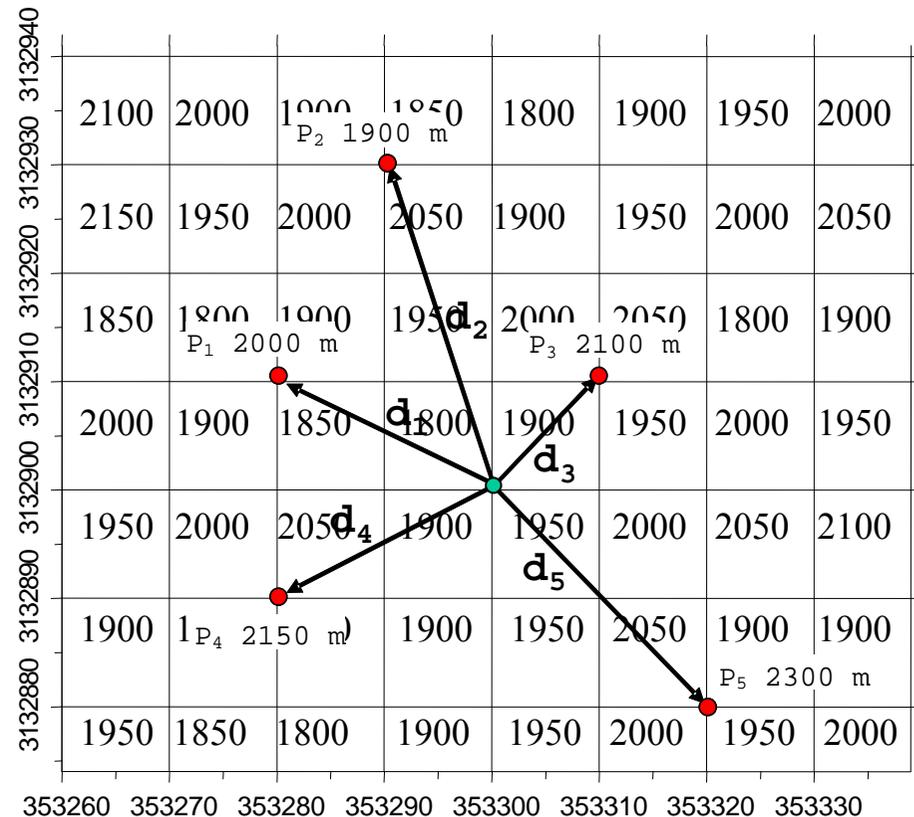
Inverso de la distancia al cuadrado

Primero se define una rejilla o grid sobre la cual se proyectan los puntos con los atributos de un fenómeno tipo objeto. Si dicho fenómeno es la altitud, los atributos de los puntos será la elevación sobre el nivel del mar.

El valor de cada celdilla que no contiene un punto con el atributo del fenómeno analizado se calcula como la media ponderada de los valores de los puntos con dato que la rodean.

La ponderación se realiza en función del inverso de la distancia que separa a la celdilla de cada punto con dato. De tal forma que los puntos cercanos se ponderan más que los puntos lejanos.

El resultado final, en nuestro ejemplo, sería un Modelo Digital de Elevación del terreno (DEM) en formato raster.

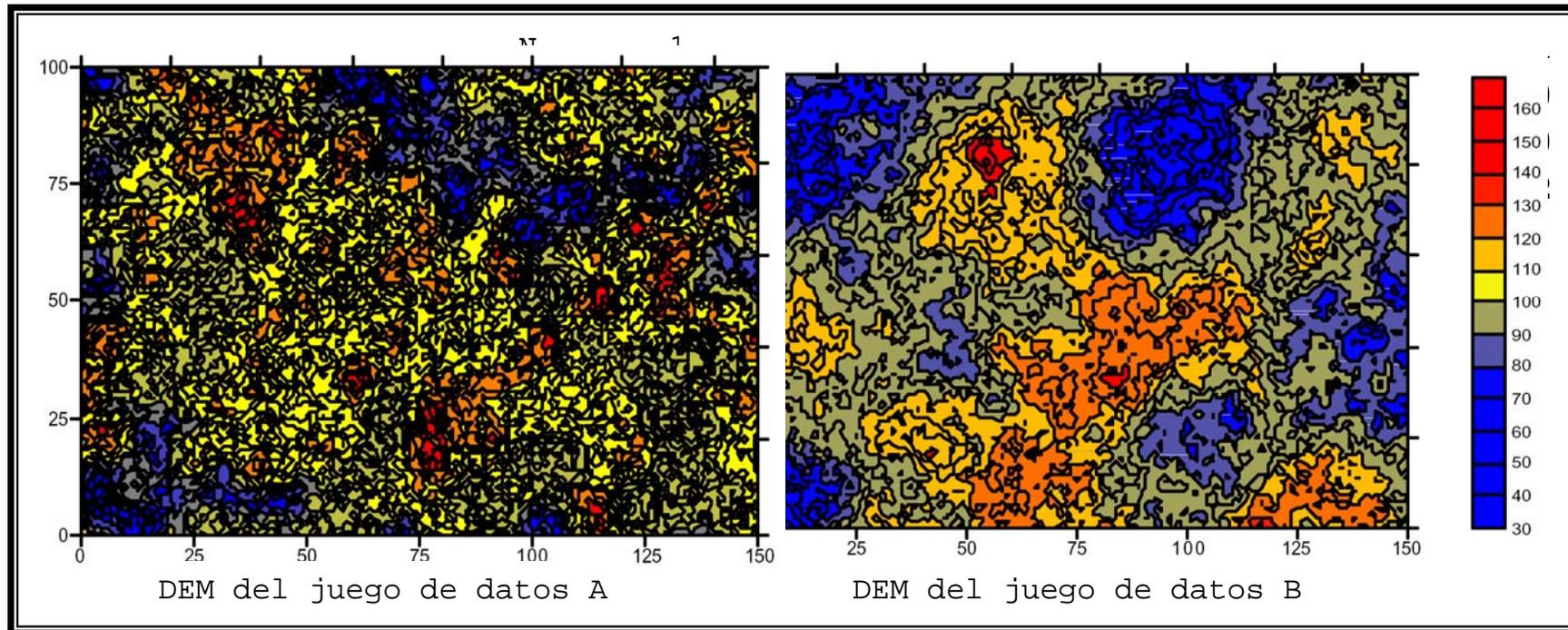


$$P = \frac{\sum_{i=1}^N P_i \frac{1/d_i^2}{\sum_{i=1}^N 1/d_i^2}}$$

ANÁLISIS DE CONTINUIDAD ESPACIAL

Análisis de texturas mediante semivariogramas

Supongamos que queremos estimar las diferencias entre dos modelos de abstracción tipo campo. Para ello, se podría representar la distribución de frecuencias de los valores de cada uno de ellos y calcular el valor medio y la desviación típica.



El hecho de que la media y la desviación típica de ambos DEM sean iguales no es suficiente motivo para pensar que ambos DEM son iguales, ya que puede variar la continuidad espacial de los datos.

En nuestro ejemplo se puede observar que la textura del DEM de ambos juegos de datos es muy diferente. La continuidad espacial de los datos se analiza mediante:

VARIOGRAMAS

ANÁLISIS DE CONTINUIDAD ESPACIAL

Análisis de texturas mediante semivariogramas

El semivariograma es una función básica que describe la variabilidad espacial de un determinado fenómeno.

Para cualquier variable analizada, podemos calcular la varianza encontrada entre todos los pares de puntos separados por un distancia d_1 . A continuación, podríamos calcular la varianza encontrada entre todos los pares de puntos separados por una distancia mayor d_2 , y así sucesivamente podríamos continuar aumentando la distancia entre pares hasta calcular la varianza encontrada entre todos los pares separados por la máxima distancia que separa a nuestro conjunto de datos.

La varianza entre pares separados por intervalos de distancia se conoce como semivarianza (γ) y se estima como:

$$\gamma(h) = \frac{1}{2} N_{(h)} \sum_{i=1}^N [Z_x - Z_{(x+h)}]^2$$

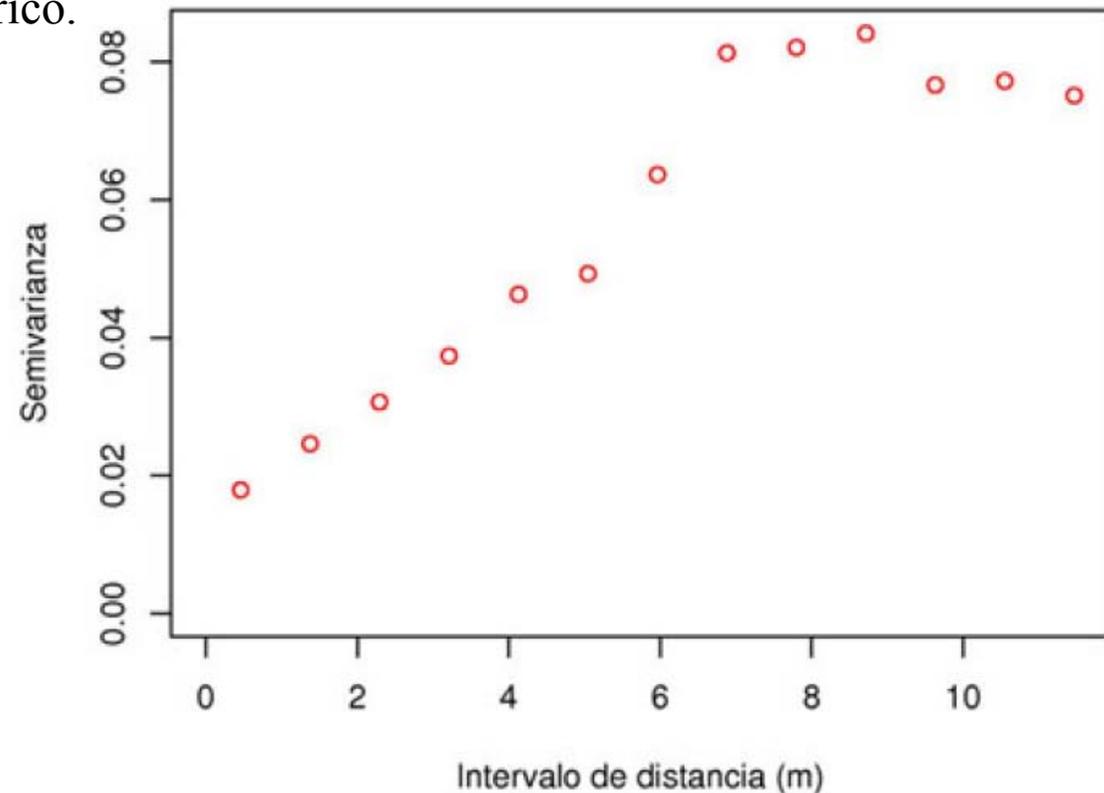
Donde, $\gamma(h)$ es la semivarianza para todas las muestras localizadas en el espacio separado por el intervalo de distancia h . $N_{(h)}$ es el número total de pares de muestras separados por un intervalo de distancia h . $Z_{(x)}$ es el valor de la muestra en una localización x y $Z_{(x+h)}$ es el valor de la muestra a la distancia de intervalo h desde x .

ANÁLISIS DE CONTINUIDAD ESPACIAL

Análisis de texturas mediante semivariogramas

La representación gráfica de todas estas semivarianza en función de la distancia que separa a las muestras es el semivariograma empírico.

Normalmente sólo se representan en el semivariograma distancias aproximadas a la mitad de la dimensión de la parcela pues conforme aumenta la distancia el número de pares descende y la semivarianza puede llegar a mostrar un comportamiento errático.

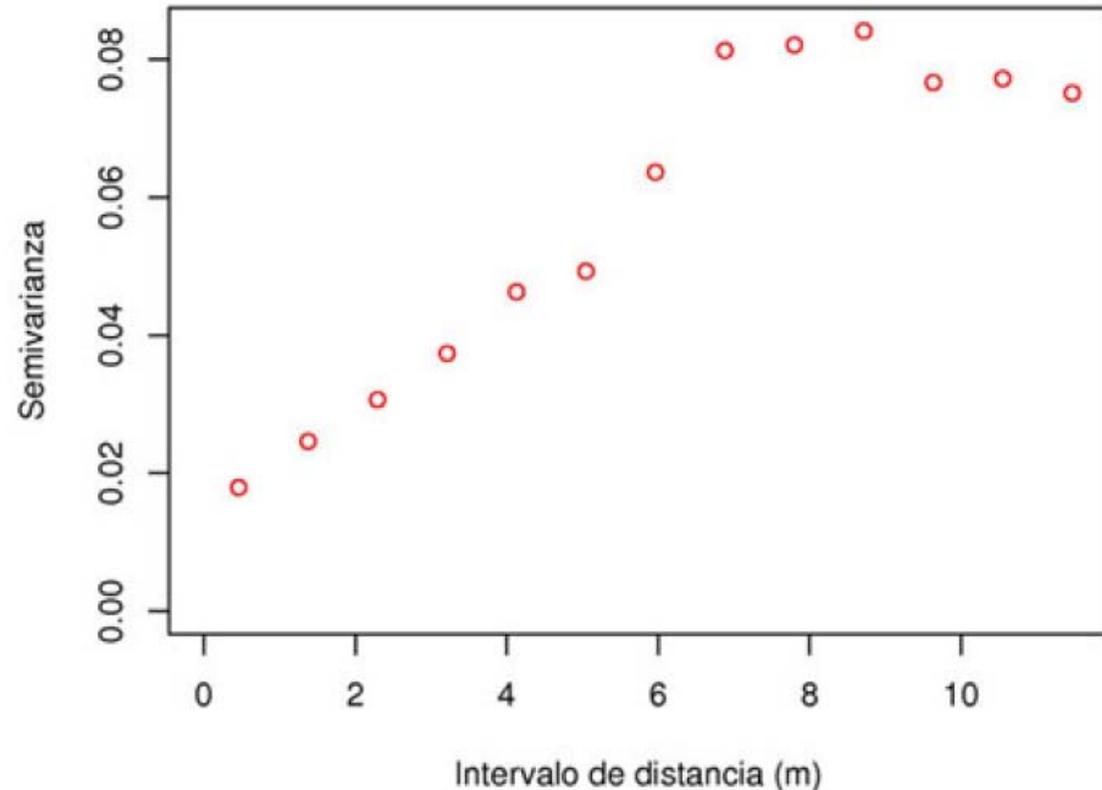


Como regla general, el número mínimo de pares para representar un punto en el semivariograma debería ser superior a 30 y el número de puntos en el espacio objeto de estudio no debería ser inferior a 50. Estas reglas generales sirven de guía, pero no deben ser tomadas como un dogma insalvable.

ANÁLISIS DE CONTINUIDAD ESPACIAL

Análisis de texturas mediante semivariogramas

La semivarianza aumenta a medida que aumenta la distancia que separa las muestras (o puntos con datos) hasta que se llega a un intervalo de distancia aproximada de 8 m, a partir del cual la semivarianza alcanza un máximo y no aumenta más. Este nivel máximo de semivarianza debe coincidir con la varianza del conjunto de puntos con dato.



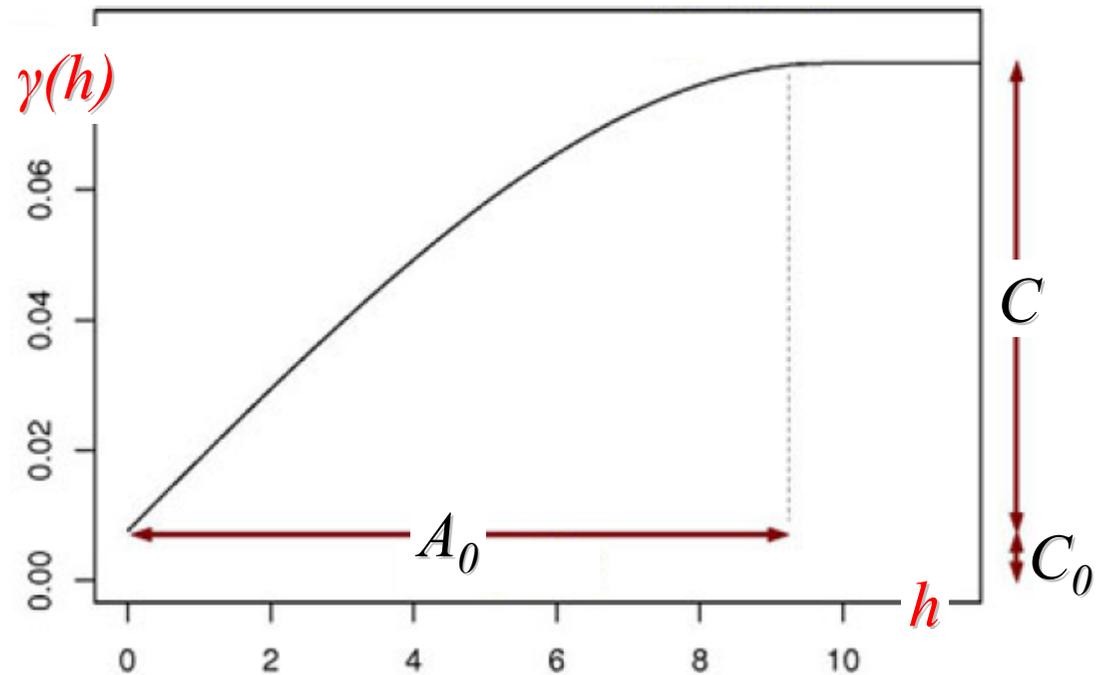
El semivariograma proporciona bastante información del comportamiento espacial de la variable. Sin embargo, es necesario ajustar una función para cuantificar el grado y escala de variación espacial.

ANÁLISIS DE CONTINUIDAD ESPACIAL

Análisis de texturas mediante semivariogramas

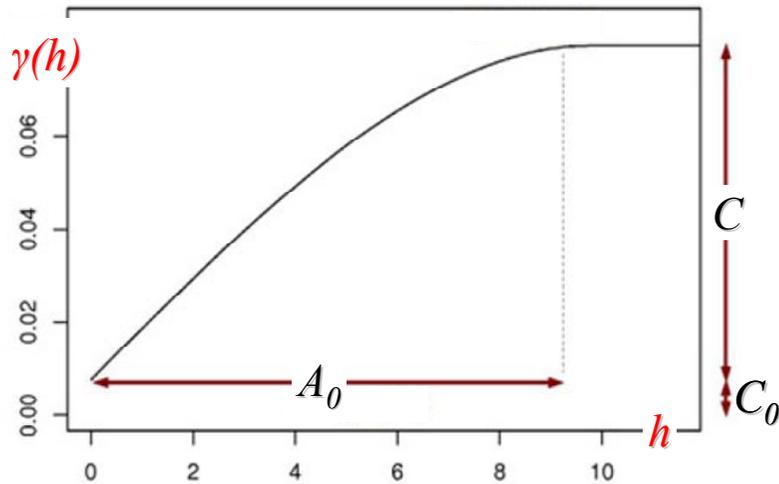
El ajuste a una función permite extraer una serie de parámetros que son los que van a ser usados para la interpolación geostatística (kriging) y que definen el grado y escala de variación espacial. Estos parámetros son:

- A_0 : Rango
- C : Varianza estructural
- C_0 : Nugget o varianza no explicada
- C_0+C : Sill o varianza total
- C/C_0+C : Proporción de la varianza explicada por el espacio



ANÁLISIS DE CONTINUIDAD ESPACIAL

Análisis de texturas mediante semivariogramas



- A_0 : Rango
- C : Varianza estructural
- C_0 : Nugget
- C_0+C : Sill
- C/C_0+C : % varianza explicada

El rango (A_0) es la distancia a la que la semivarianza deja de aumentar. El rango, por tanto, indica la distancia a partir de la cual los puntos son espacialmente independientes unos de otros, y representa el tamaño de grano o mancha que representa la variable.

La varianza estructural (C) es la varianza explicada de los puntos para el rango (A_0).

El nugget (C_0) es la varianza no explicada por el modelo, y se calcula como la intercepción con el eje Y. Se conoce también como varianza error puesto que la varianza de dos puntos separados por 0 metros (la intercepción con el eje Y) debería ser cero. Es por ello que esta varianza está normalmente indicando variabilidad a una escala inferior a la muestreada.

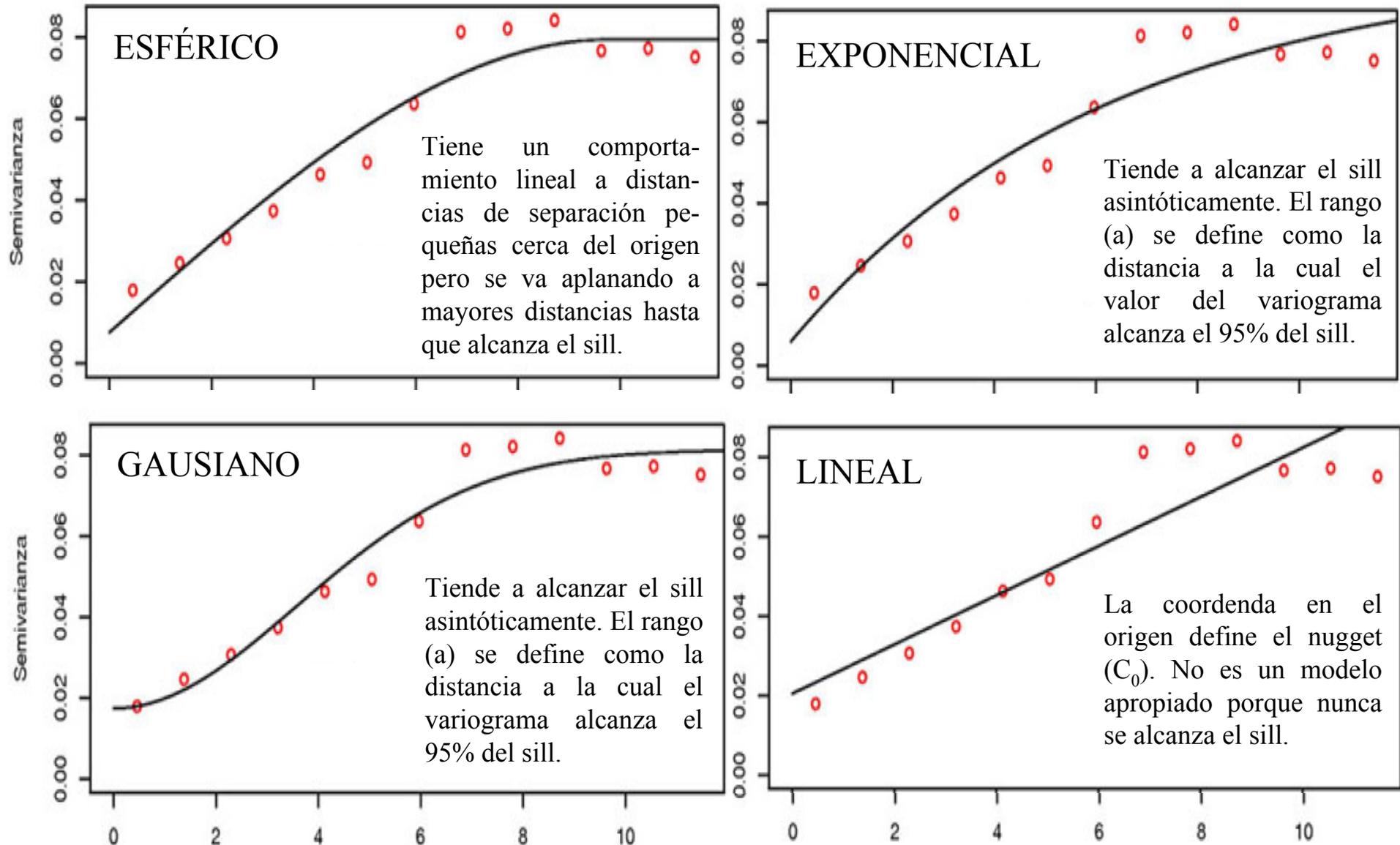
La máxima semivarianza encontrada entre pares de puntos (C_0+C) debe coincidir con la varianza de la población.

La proporción de la varianza explicada respecto a la varianza total $C/(C_0 + C)$ nos da el grado de variación espacial, y por tanto el grado de incertidumbre a la hora de interpolar puntos en el espacio. Un alto cociente nos indica una variable espacialmente muy predecible.

ANÁLISIS DE CONTINUIDAD ESPACIAL

Análisis de texturas mediante semivariogramas

MODELOS DE SEMIVARIOGRAMA



ANÁLISIS DE CONTINUIDAD ESPACIAL

Análisis de texturas mediante semivariogramas

ANISOTROPÍA

Hasta ahora hemos supuesto que la variación del valor de nuestra variable con el espacio es igual en todas las direcciones de éste. Si esto ocurre decimos que la variable tiene un **comportamiento isotrópico**.

Pero no siempre es así, y puede ser que la variación espacial sea diferente en las distintas direcciones del espacio. Entonces decimos que tiene un **comportamiento anisotrópico**.

Si tras una inspección visual sospechamos que puede ocurrir este fenómeno es interesante realizar semivariogramas considerando por separado varias direcciones del espacio (**semivariogramas direccionales**).

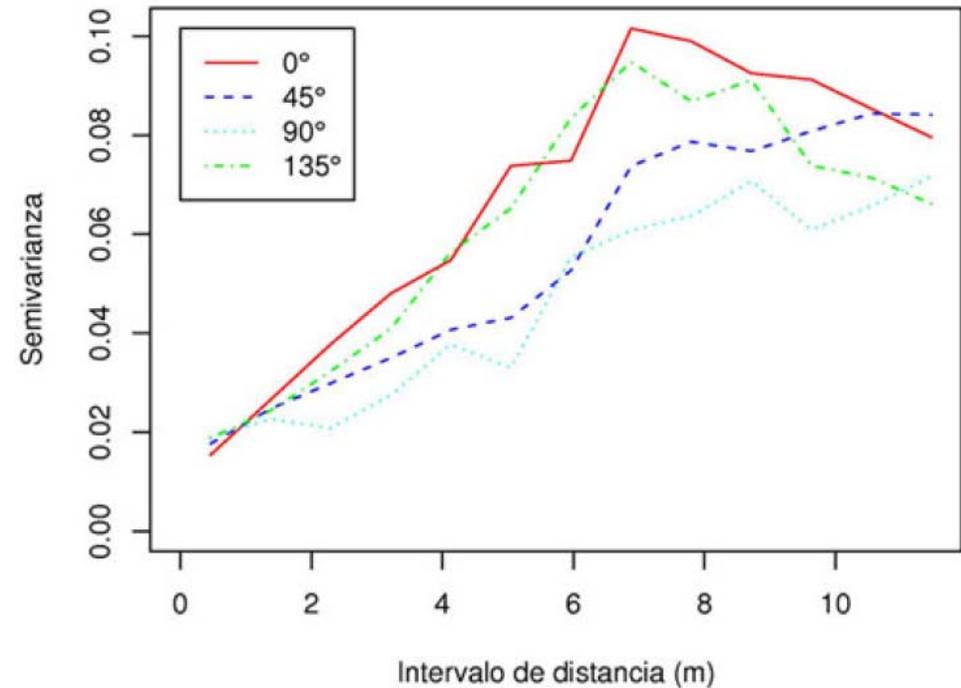
ANÁLISIS DE CONTINUIDAD ESPACIAL

Análisis de texturas mediante semivariogramas

ANISOTROPIA

En este ejemplo, no existe un claro comportamiento anisotrópico en la distribución de la variable, ya que los rangos y *nuggets* son aproximadamente similares.

A pesar de ello, en las direcciones NE-SO (45°) y E-O (90°) el *sill* alcanza valores menores que en las otras dos direcciones, siendo también su comportamiento en las distancias mayores diferente.



Si el *sill* disminuye, pero los *nugget* son similares el porcentaje de varianza explicada por el espacio ($C/C0+C$) también disminuye en las direcciones 45° y 90°.

Siendo así, los puntos situados en esas direcciones deberían tener un peso menor a la hora de la interpolación que puntos situados en las otras dos direcciones.

ANÁLISIS DE CONTINUIDAD ESPACIAL

Análisis de texturas mediante semivariogramas

ESTACIONARIEDAD

El empleo de técnicas geostatísticas requiere la asunción de estacionaridad de segundo orden, es decir, al menos la varianza debe ser igual en las diferentes zonas del área de estudio.

La falta de estacionariedad puede deberse bien a la existencia de anomalías en el espacio, o bien a la existencia de una tendencia o gradiente espacial cuya dimensión es mayor que el área de estudio.

La estacionariedad puede ser un problema a la hora de la interpolación de puntos en el espacio mediante técnicas geostatísticas basadas en el análisis de la continuidad espacial.

Cuando los datos muestran una tendencia a seguir gradientes espaciales (p.ej. gradiente altitudinal) puede resultar más conveniente aplicar técnicas de Análisis de Regresión lineal para ajustar modelos externos que expliquen la variación espacial del fenómeno en función de estos gradientes y realizar posteriormente una interpolación espacial.

ANÁLISIS DE CONTINUIDAD ESPACIAL

Interpolación espacial mediante Kriging

Su fortaleza estriba en el uso del semivariograma que nos indica la capacidad predictora que tiene cada punto en función de la distancia que lo separa con otro punto. A la hora de realizar un kriging hay que suministrar los parámetros del semivariograma isotrópico o los anisotrópicos.

A diferencia de otras técnicas, el método Kriging permite que cada interpolación lleve asociado un grado de incertidumbre que puede también ser representado en el espacio en forma de varianza o desviación estándar.

Por tanto a cada punto del espacio interpolado se le puede asociar una distribución teórica, lo que además permite la posibilidad de realizar simulaciones probabilísticas, representando el resultado del kriging como la probabilidad de que la variable alcance un determinado valor.