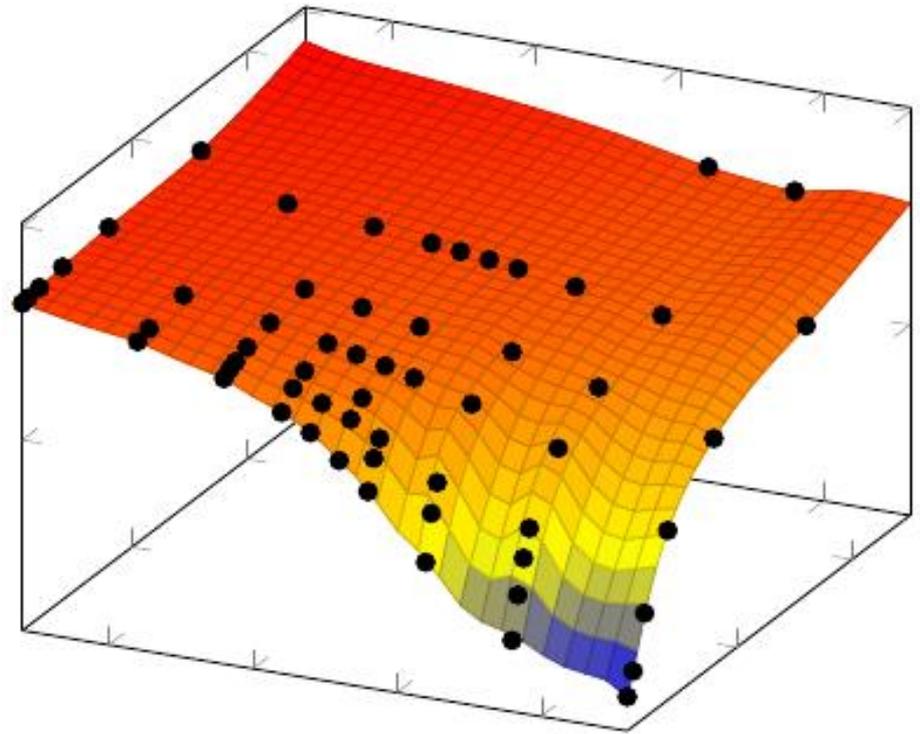


# **Sistema de Información Geográfica Avanzado**

**Ingeniero Agrimensor  
Tecnólogo en Cartografía**

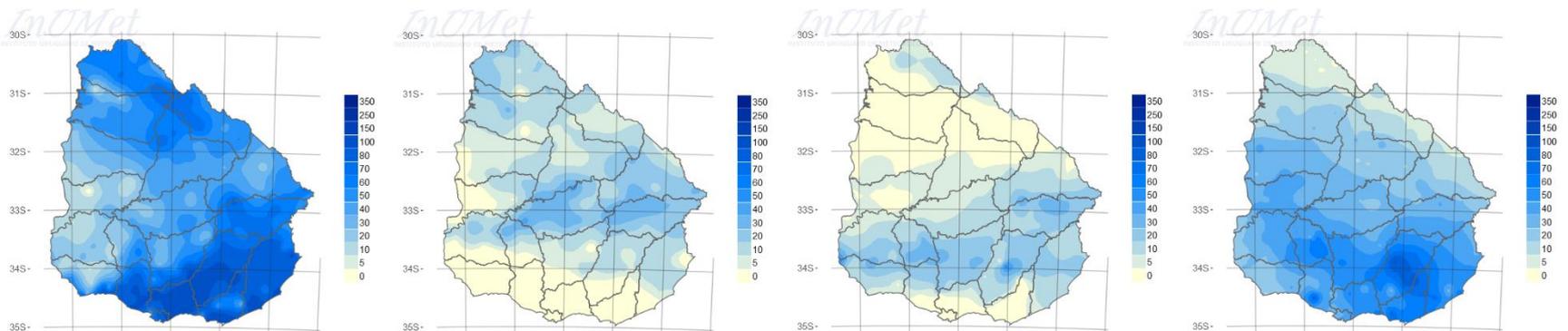
# Introducción (I)

- En este módulo se formula el problema de interpolación espacial de **datos dispersos** como un método para la predicción y la representación de los campos de variables múltiples.



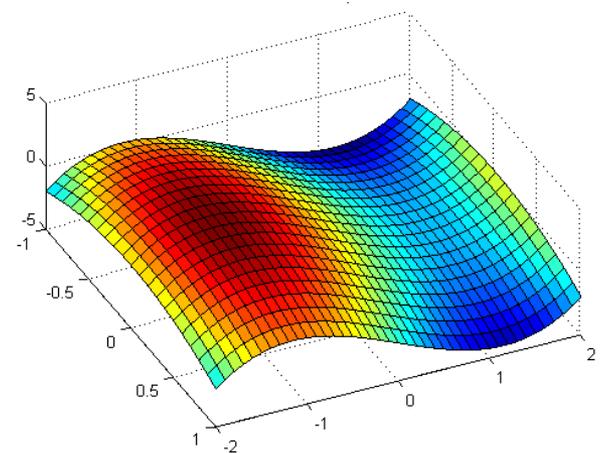
## Introducción (II)

- Veremos el papel y las cuestiones específicas de interpolación para aplicaciones SIG y veremos algunos métodos basados en la localización, geoestadística y el concepto de variación.
- Se ilustrarán algunas propiedades de los métodos de interpolación mediante ejemplos de 2D y 3D de interpolaciones de elevación, precipitación, y los datos de las concentraciones químicas.



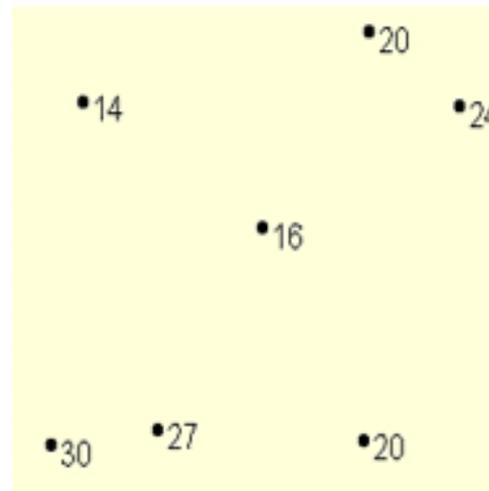
## Introducción (III)

- Las distribuciones **espaciales** y **espacio-temporales** de los fenómenos físicos y socioeconómicos pueden aproximarse mediante funciones dependiendo de la ubicación en un espacio multidimensional; como escalar multivariante, vector o campos tensoriales.
- Ejemplos: **elevaciones, fenómenos climáticos, las propiedades del suelo, la densidad de población, los flujos de materia, etc.**



## Introducción (IV)

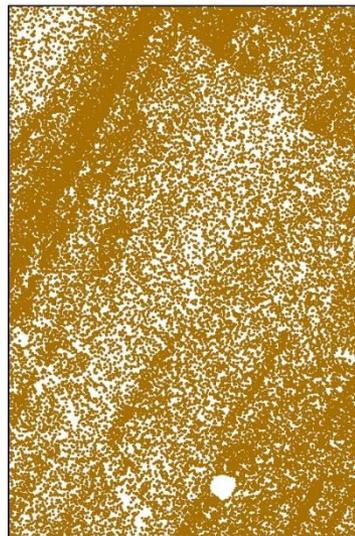
- Mientras que la mayoría de estos fenómenos se caracterizan por datos de **puntos medidos o digitalizados** (a menudo distribuidos irregularmente en el espacio y el tiempo) la visualización, análisis y modelado dentro de una SIG se basan generalmente en una representación en un **modelo ráster**.



13	14	16	20	23
14	14	16	19	24
18	16	16	18	22
24	22	19	19	21
30	27	23	20	20

# Introducción (V)

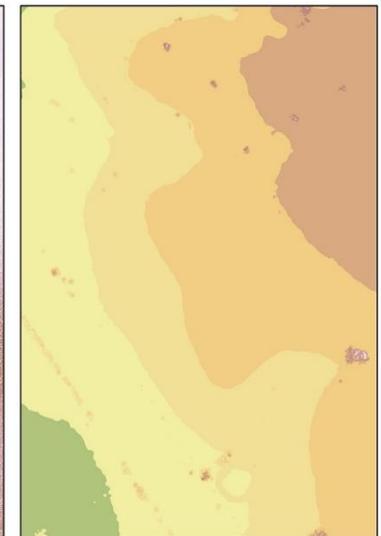
- **Los fenómenos se pueden medir utilizando diversos métodos** (teledetección, muestreo sitio, etc.) y conducen a los conjuntos de datos heterogéneos con diferentes representaciones digitales y resoluciones que deben ser combinados para crear un único modelo espacial del fenómeno en estudio.



Raw LAS points (ground)



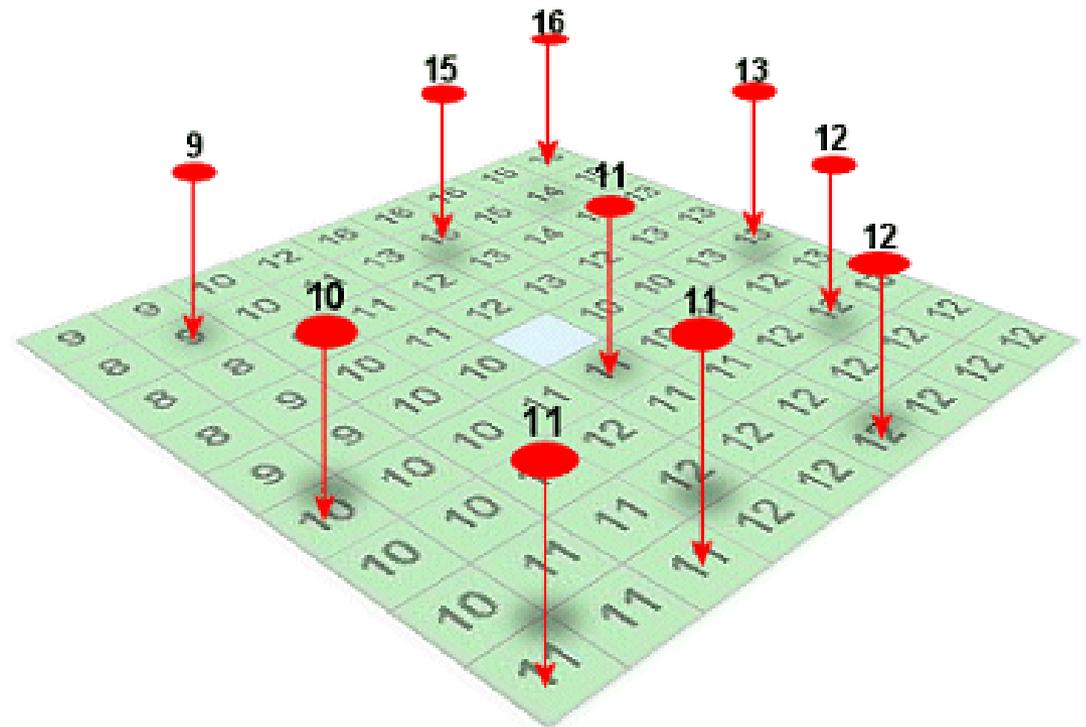
Point to Raster (Conversion)



IDW Interpolation

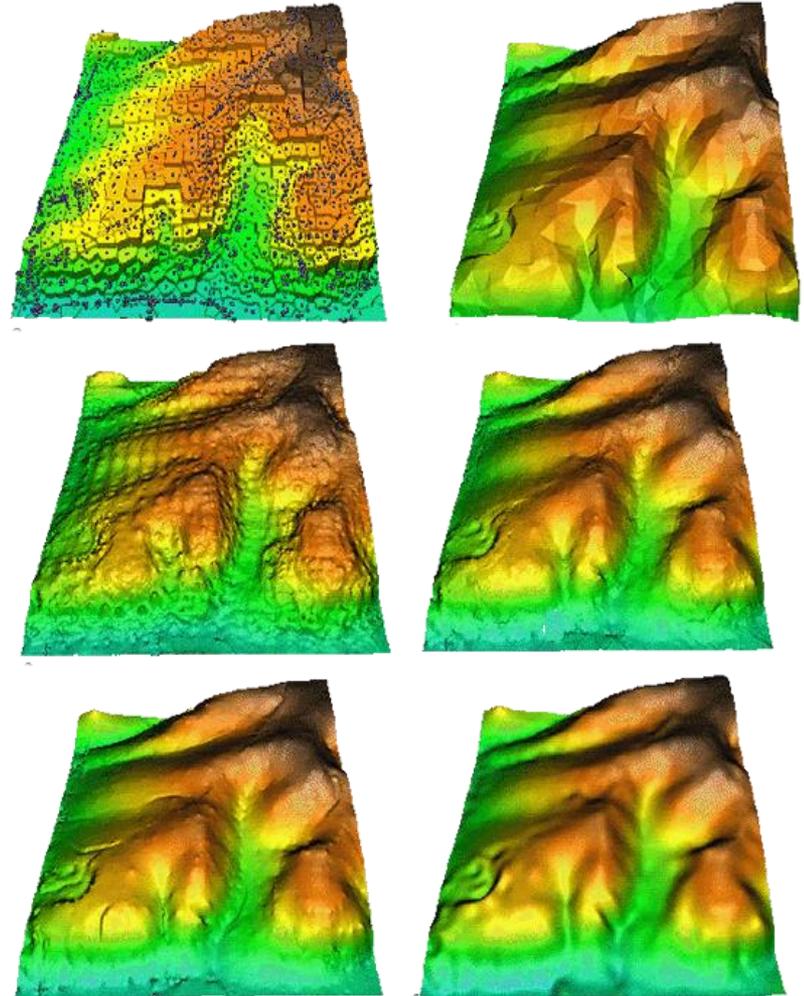
## Introducción (VI)

- Muchos métodos de interpolación y aproximación se han desarrollado para predecir los valores de los fenómenos espaciales en ubicaciones no muestreadas.



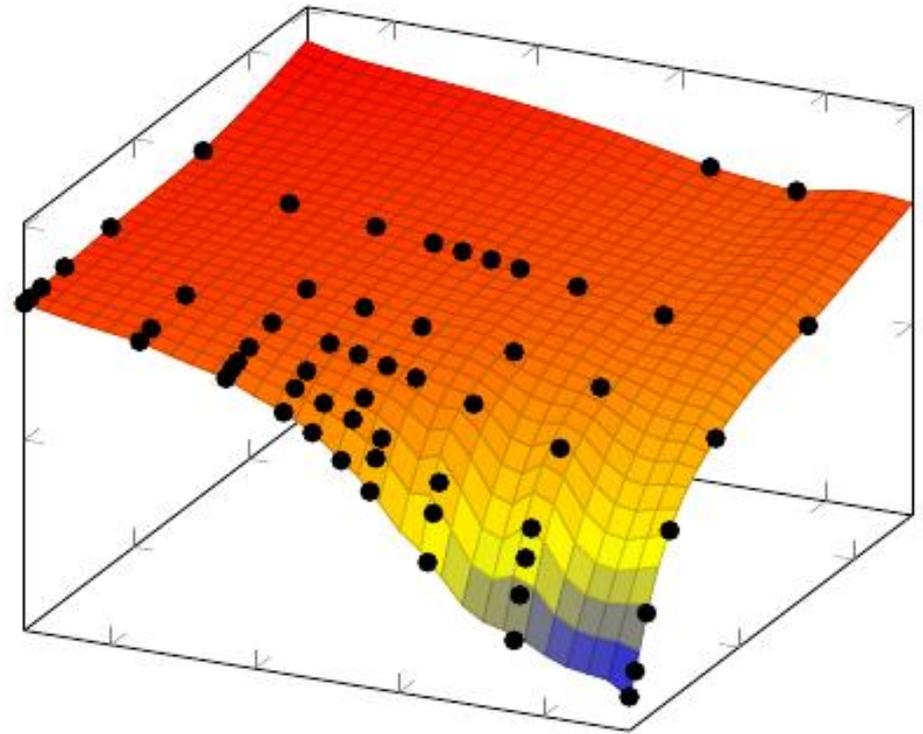
## Introducción (VII)

En las aplicaciones de SIG, estos métodos han sido diseñados para apoyar las transformaciones entre diferentes representaciones discretas y continuas de los campos espaciales y espacio-temporales, para transformar datos de punto o líneas irregulares a la representación del ráster, o volver a muestrear entre diferentes resoluciones de ráster.



# Métodos de interpolación espacial (I)

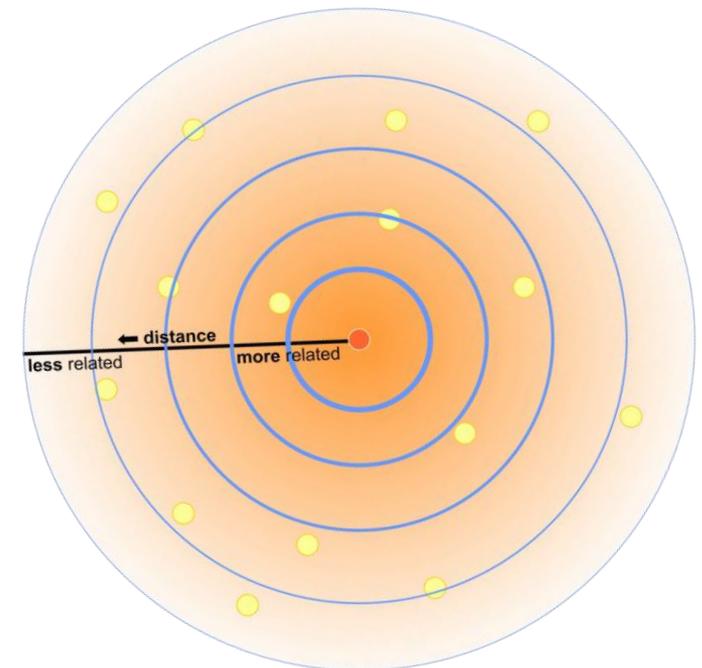
- **¿Por qué?** Muchas de las variables en el espacio se conocen con puntos de muestreo y se pretende **predecir valores** en zonas no muestreadas.



# Métodos de interpolación espacial (II)

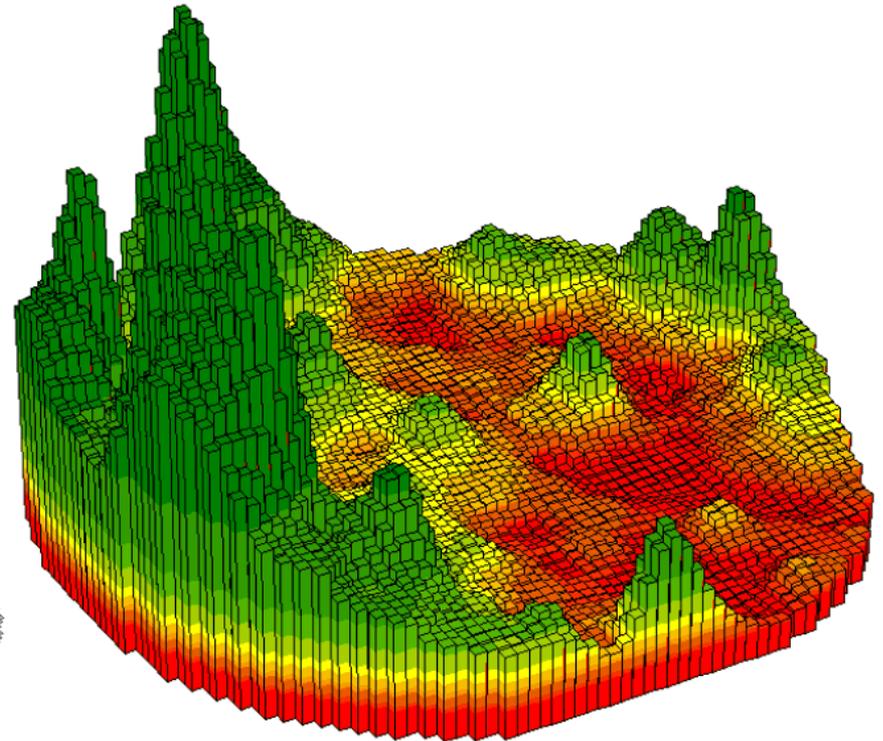
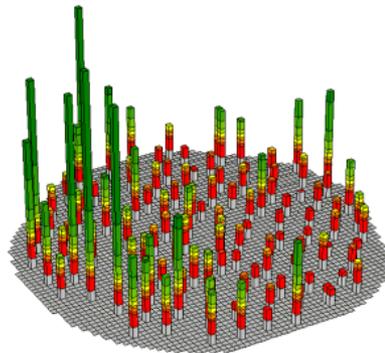
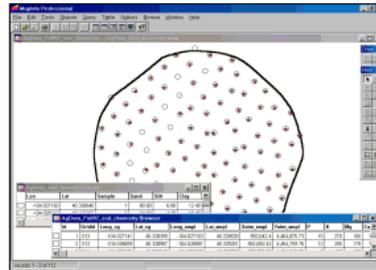
- **Ley de Tobler** (o **Principio de Autocorrelación espacial**): *“las observaciones que se encuentran próximas unas con otras presentan valores similares en contraste con las observaciones que se encuentran alejadas, las cuales muestran valores poco parecidos”.*

*El sonido es más fuerte cerca que lejos de la sirena.  
Cuando llueve, es más probable que llueva a un metro de distancia que a 500 metros.*



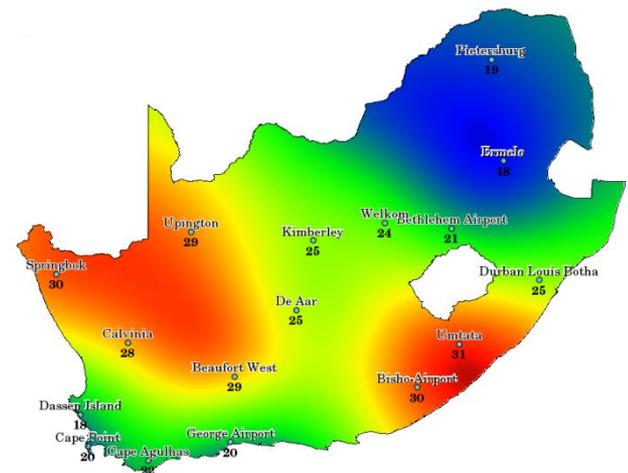
# Métodos de interpolación espacial (III)

- Por tanto, la interpolación espacial **se centra en el análisis y simulación de una muestra de datos, así como su comportamiento en el espacio e influencia en otros puntos.**



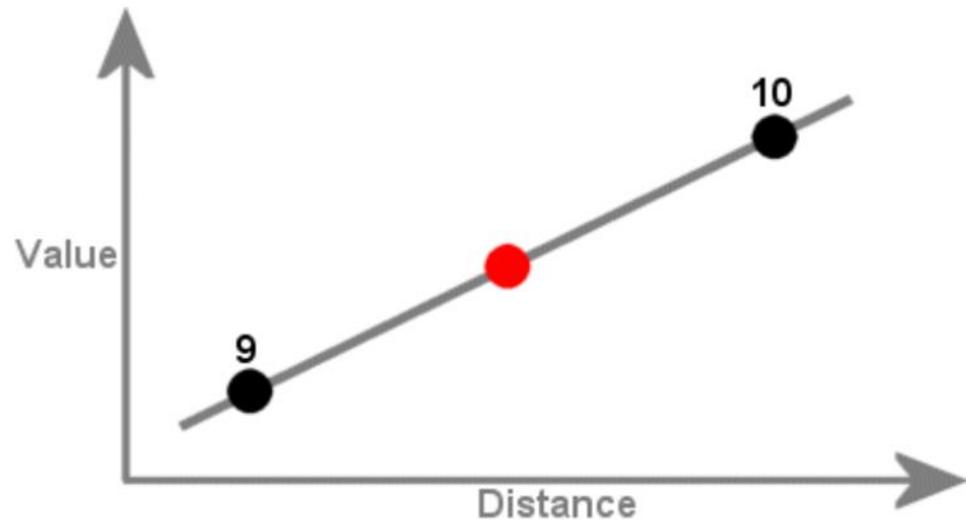
# Métodos de interpolación espacial (IV)

- La interpolación espacial es el proceso de utilizar puntos con valores conocidos para estimar valores desconocidos en otros puntos.
- Por ejemplo: para realizar un mapa de precipitación para el país no es posible colocar estaciones meteorológicas distribuidas uniformemente para hacer una cobertura total.



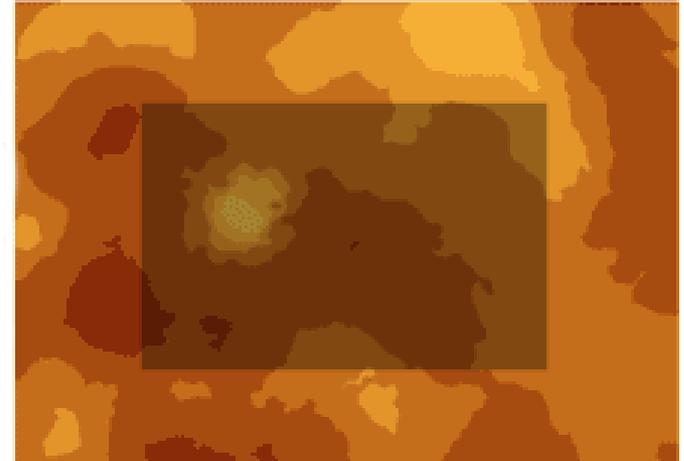
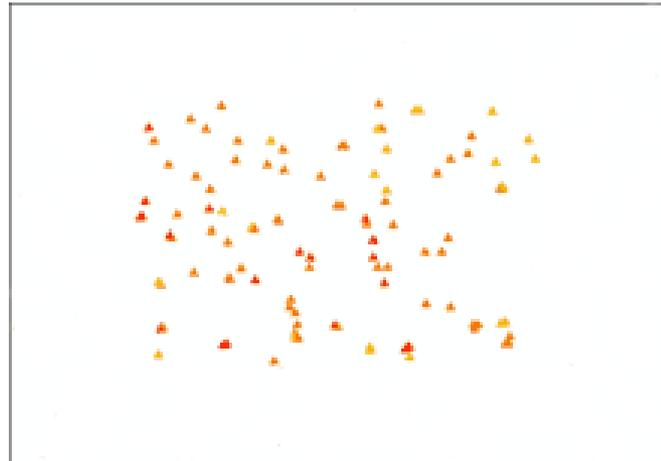
# Métodos de interpolación espacial (V)

- Llamamos interpolación a la **obtención de un conjunto nuevos de valores partiendo de un conjunto discreto de valores conocidos.**



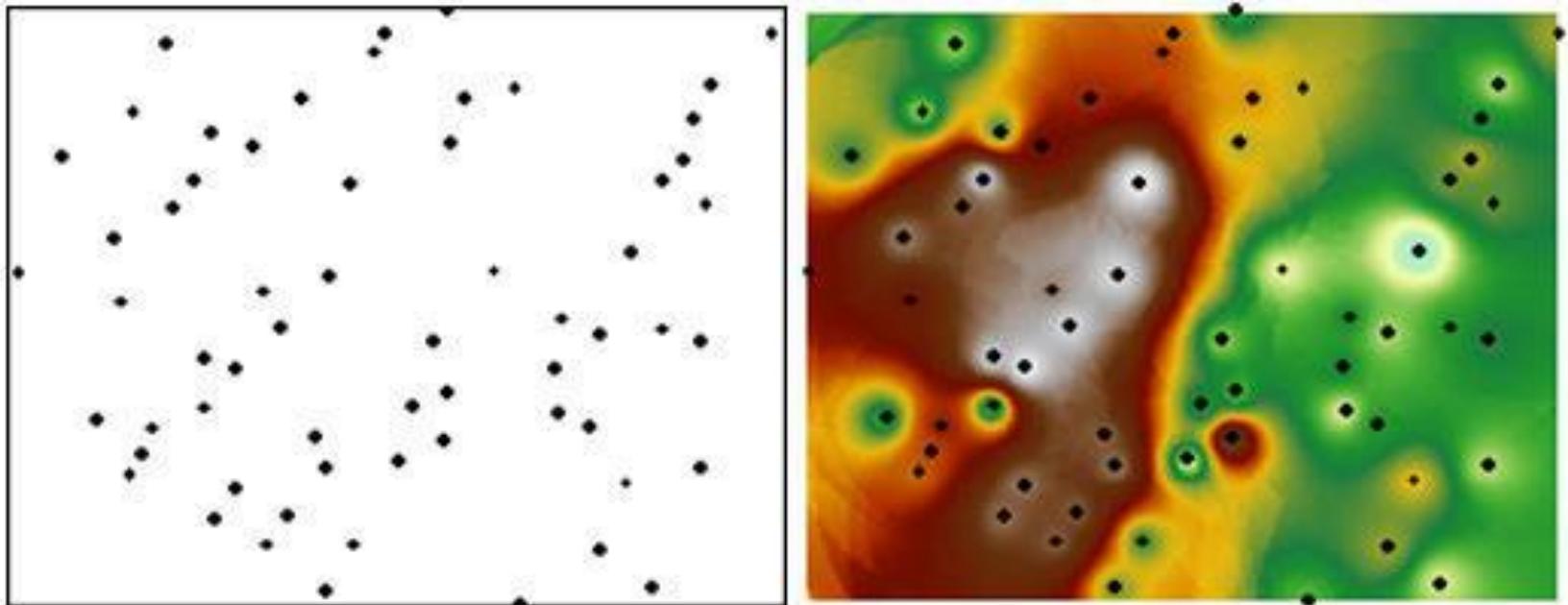
## Métodos de interpolación espacial (VI)

- El proceso de **interpolación espacial** consiste en la estimación de los valores que alcanza una variable  $Z$  en un conjunto de puntos definidos por un par de coordenadas  $(X, Y)$ , partiendo de los valores de  $Z$  medidos en una muestra de puntos situados en el mismo área de estudio, la estimación de valores fuera del área de estudio se denomina **extrapolación espacial**.



## Métodos de interpolación espacial (VII)

- Cuando se trabaja con un GIS la interpolación espacial suele utilizarse para obtener **capas ráster** que representan la variable a interpolar. **En esos casos cada pixel de la capa ráster constituye un punto en el que hay que realizar la interpolación.**



# Formulación de problemas (I)

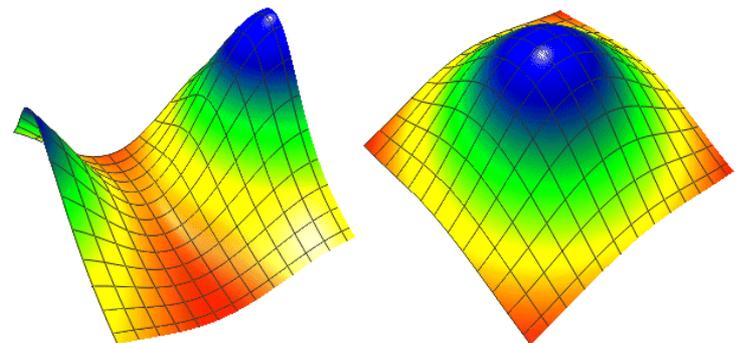
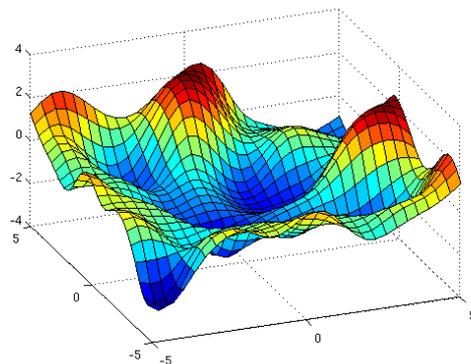
La **Formulación General de la Interpolación Espacial**

es: *Dado los  $N$  valores de un fenómeno estudiado  $z_j$*

*(con  $j = 1, \dots, N$ ) medidos en puntos discretos*

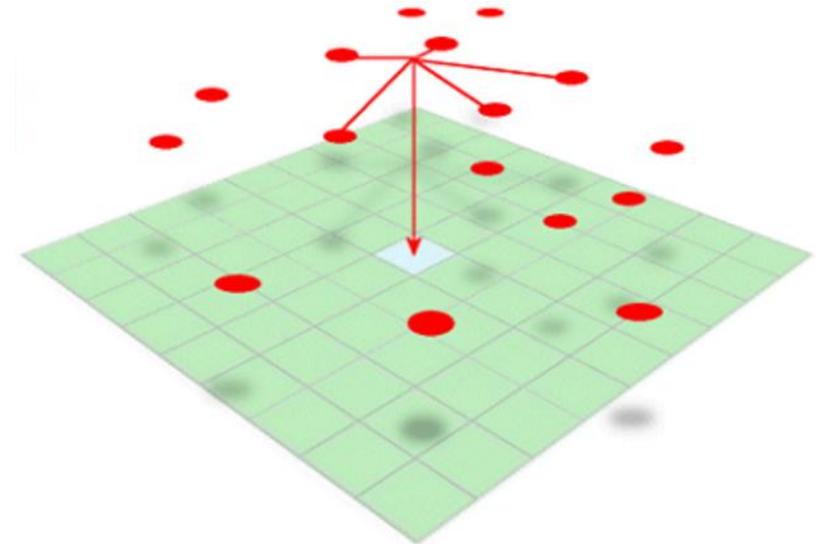
*$r_j = (x_j^1, x_j^2, \dots, x_j^d)$  (con  $j = 1, \dots, N$ )*

*dentro de una determinada región de un espacio  $d$ -dimensional, encontrar la función  $F(r)$  que cumpla la condición:  $F(r_j) = z_j$ .*



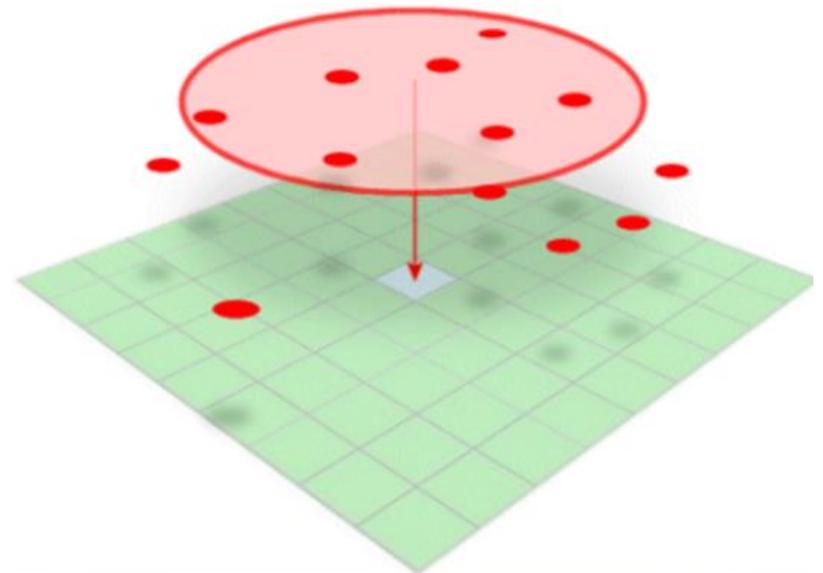
# Tamaños de la muestra (I)

- La mayoría de los métodos de interpolación permiten controlar el **número de muestras puntos utilizados para estimar los valores de celda.**
- Por ejemplo: *si se limita su muestra a cinco puntos, el interpolador utilizará los cinco más cercana puntos para estimar los valores de celda .*



## Tamaños de la muestra (II)

- **La distancia a cada punto de muestreo variará dependiendo de la distribución de los puntos.**
- Ante una gran cantidad de puntos de muestra, reducir el tamaño de la muestra a utilizar acelera el proceso de interpolación debido a que un conjunto más pequeño de números se utilizan para estimar cada valor de celda.

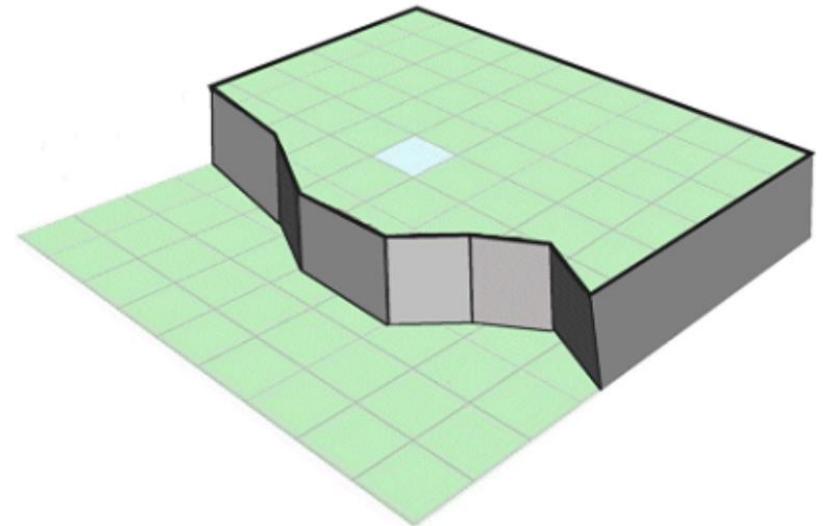


## Tamaños de la muestra (III)

- **Radio de búsqueda:** Es posible controlar el tamaño de la muestra mediante la definición de un **radio de búsqueda**; el número de la muestra de puntos que se encuentra dentro una radio de búsqueda puede variar dependiendo de cómo se distribuyen los puntos.
- Es posible optar por utilizar algunos o todas las muestras que caen dentro de este radio para calcular la célula valor.
- Una **radio de búsqueda variable** continuará expandiéndose hasta que el tamaño de la muestra se encuentra especificado. Un **radio de búsqueda fijo** sólo utilizará las muestras contenidas dentro de ella, independientemente de la cantidad encontrada.

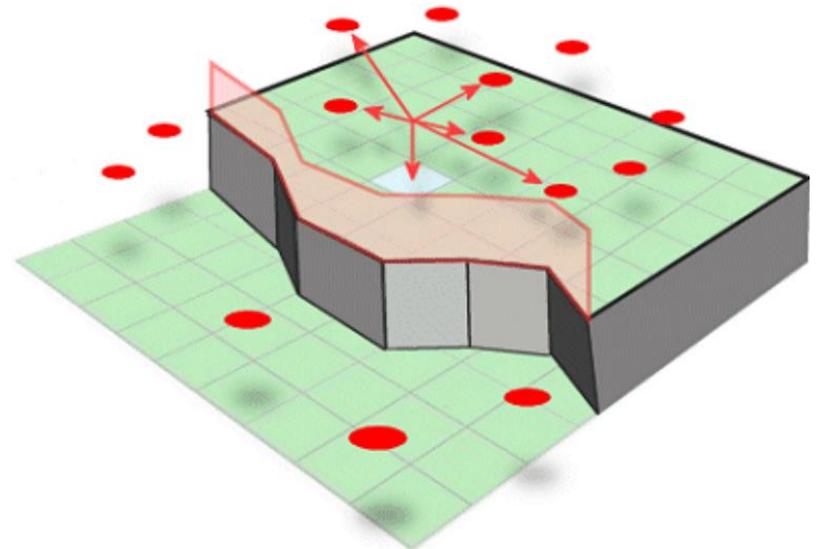
# Barreras (I)

- Las barreras físicas que existen en la geografía (acantilados o ríos) presentan un desafío particular al intentar modelar una superficie utilizando interpolación.
- **Los valores a cada lado de una barrera que representa una interrupción repentina en la geografía son drásticamente diferente.**



## Barreras (II)

- La mayoría de los interpoladores intentan suavizar estas diferencias mediante la incorporación de un promedio de los valores y en ambos lados de la barrera. El método de IDW permite incluir barreras en el análisis. **La barrera impide que el interpolador de la utilización de puntos de las muestras en un lado de ella.**

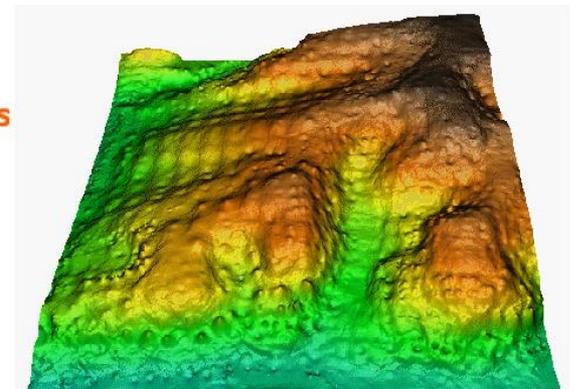
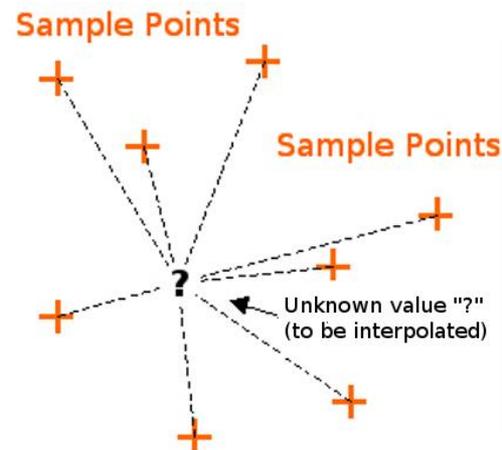


# Interpolación IDW (I)

- **Este método presupone que la variable que se representa cartográficamente disminuye su influencia a mayor distancia desde su ubicación de muestra.**
- **Por ejemplo:** *al interpolar una superficie de poder adquisitivo de los consumidores para analizar las ventas minoristas de un sitio, el poder adquisitivo de una ubicación más distante tendrá menos influencia porque es más probable que las personas compren cerca de sus casas.*
- **Es un método de interpolación local a partir de puntos.** Los resultados de la interpolación usualmente se muestran en un **modelo ráster 2D**.

# Interpolación IDW (II)

- La interpolación mediante **distancia inversa ponderada** determina los valores de pixel de un ráster a través de una combinación ponderada linealmente de un conjunto de puntos de muestra.
- **La superficie que se interpola debe ser la de una variable dependiente de la ubicación.**



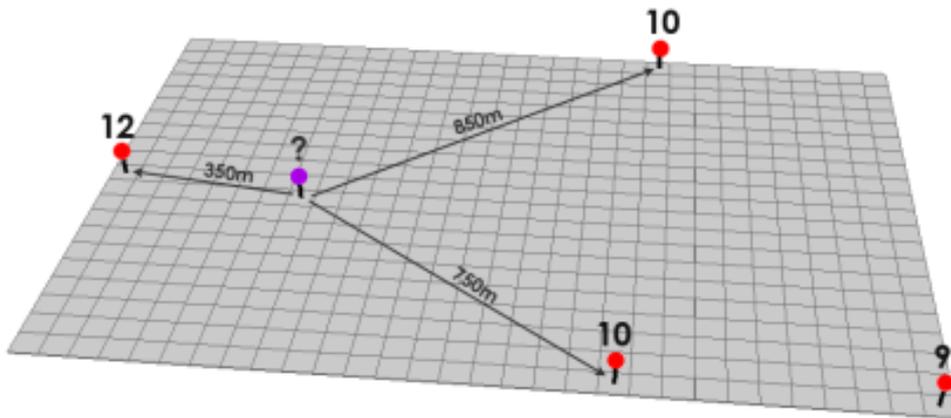
## Interpolación IDW (III)

- Se basa en el **Principio de Autocorrelación Espacial**.
- **La ponderación es por distancia:** *los valores medidos más cercanos a la ubicación de predicción tienen más influencia en el valor predicho que los más lejos. IDW asume que cada punto medido tiene una influencia local que disminuye con la distancia.*
- **Se da mayor peso a los puntos más cercanos a la ubicación de predicción**, y los pesos disminuye en función de la distancia, de ahí el nombre de la distancia inversa ponderada.

# Interpolación IDW (IV)

## ■ Formulación matemática:

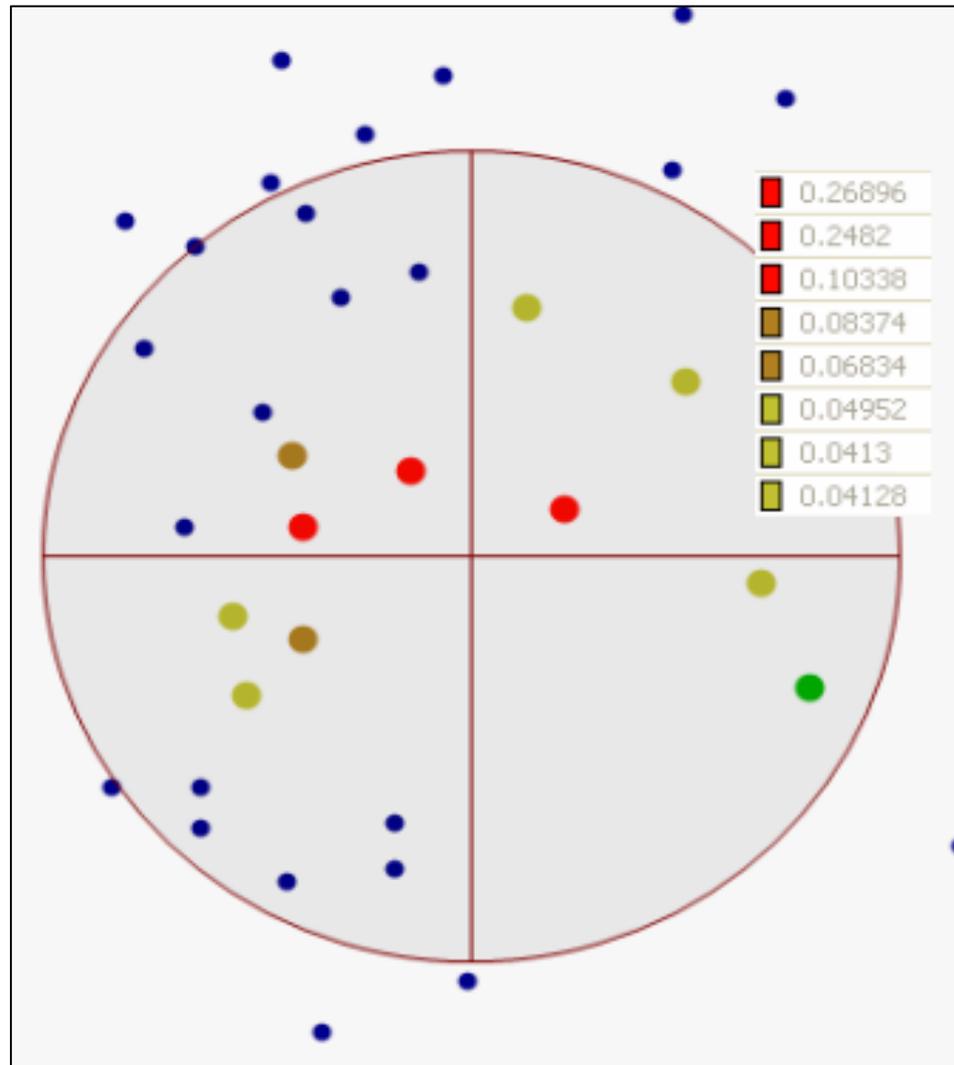
$$z_p = \frac{\sum_{i=1}^n \left( \frac{z_i}{d_i} \right)}{\sum_{i=1}^n \left( \frac{1}{d_i} \right)}$$



Para un **valor de potencia** de 1 el valor del pixel sería:  
 $((12/350) + (10/750) + (10/850)) / ((1/350) + (1/750) + (1/850)) = 11.1$

En cambio, para un **valor de potencia** de 2 el valor del pixel sería:  
 $= ((12/350^2) + (10/750^2) + (10/850^2)) / ((1/350^2) + (1/750^2) + (1/850^2)) = 11.4$

# Interpolación IDW (V)

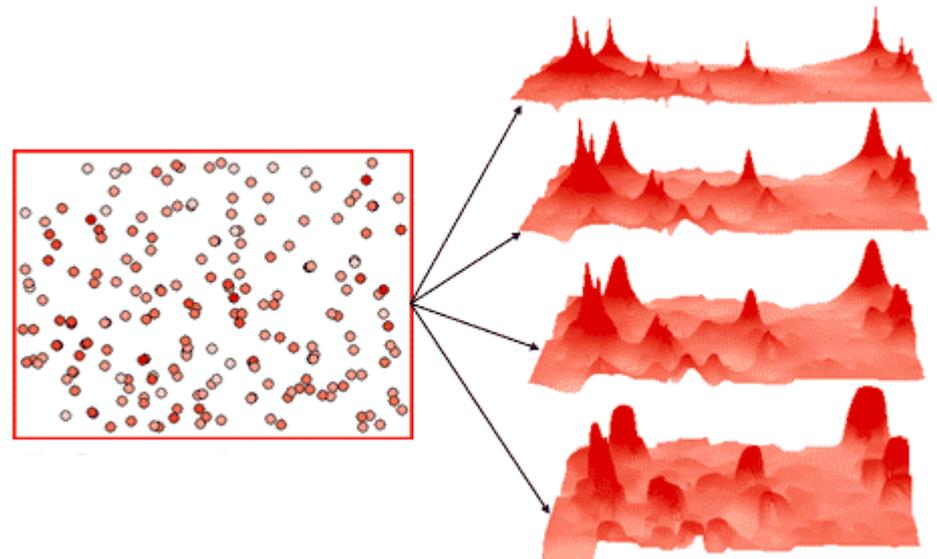
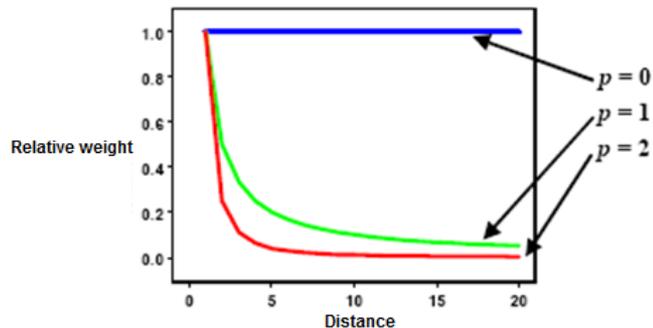


## Interpolación IDW (VI)

- Los pesos son proporcionales a la inversa de la distancia (entre el punto de datos y la ubicación de predicción) elevado al **valor de potencia** ( $p$ ). Como resultado, a medida que aumenta la distancia, los pesos disminuyen rápidamente.
- La velocidad a la que disminuyen los pesos es dependiente del valor de  $p$ .
- Si  **$p = 0$** : no hay una disminución con la distancia, y como cada  $\lambda_i$  de peso es el mismo, la predicción será la media de todos los valores de datos en la zona de búsqueda.

# Interpolación IDW (VII)

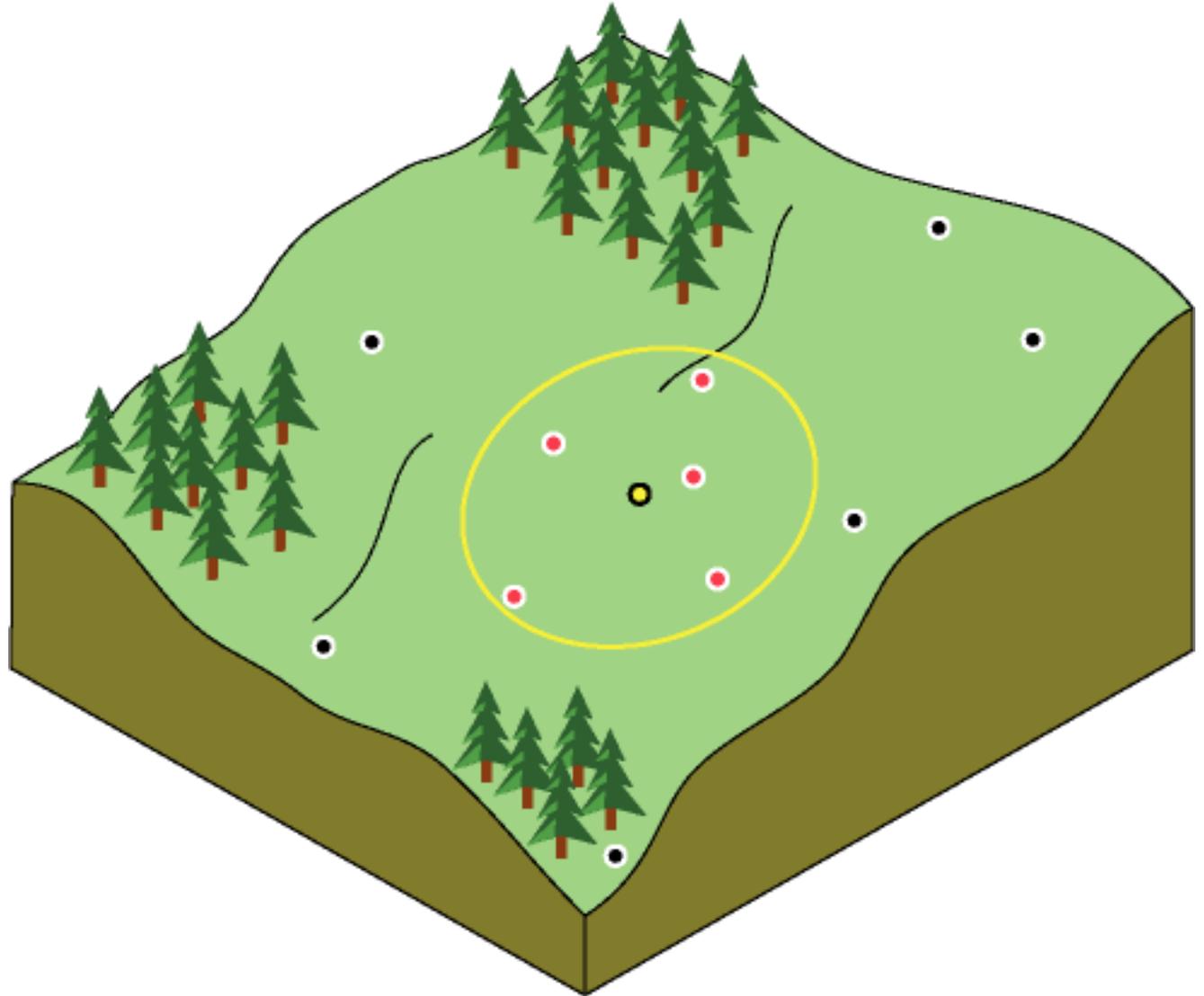
- A medida que  $p$  aumenta, las ponderaciones de puntos distantes disminuyen rápidamente. Si el valor  $p$  es muy alta, solamente los puntos circundantes inmediatas influirán en la predicción.



## Interpolación IDW (VIII)

- La **zona de búsqueda**: Por aplicar el Principio de Autocorrelación Espacial y, para acelerar los cálculos, se excluyen los puntos más distantes ya que tendrán poca influencia en la predicción.
- Como resultado, una práctica común para limitar el número de valores medidos se especifica la zona de búsqueda.
- La forma de la zona limita hasta dónde y dónde buscar los valores medidos para ser utilizado en la predicción.
- Otros parámetros de la vecindad restringen los lugares que se utilizarán dentro de esa forma.

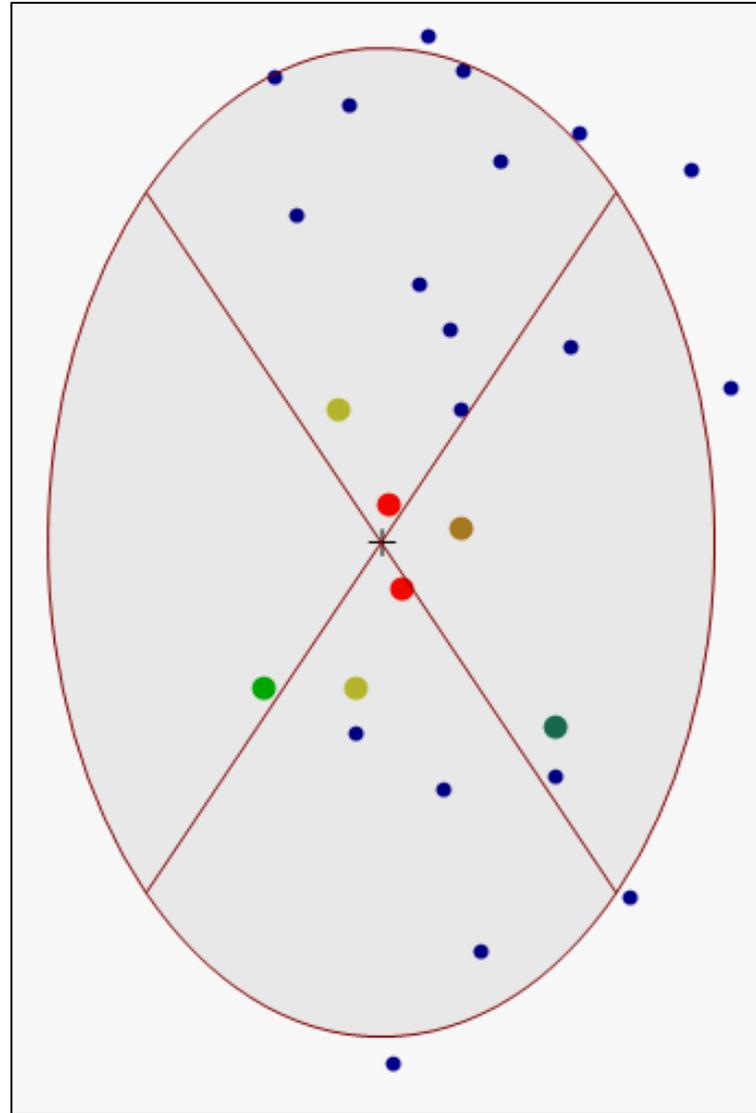
# Interpolación IDW (IX)



# Interpolación IDW (X)

- La forma de la zona está influenciada por los datos de entrada y la superficie que está intentando crear.
- Si no hay influencias direccionales en los datos, se deberán tener en cuenta puntos por igual en todas las direcciones. Esto hace que la zona de búsqueda sea circular.
- Sin embargo, si existe una influencia direccional en sus datos (por ej: viento predominante), es posible cambiar la forma de la zona de búsqueda a una elipse con el eje mayor paralelo con el viento.

# Interpolación IDW (XI)



# Interpolación IDW (XII)

## Desventajas del IDW:

- **La calidad del resultado de interpolación puede disminuir si la distribución de los puntos de datos de la muestra es desigual.**
- **Los valores máximos y mínimos en la superficie interpolada pueden ocurrir únicamente en los puntos de datos de la muestra.** Esto a menudo resulta en pequeños picos y pozos alrededor de los puntos de datos de la muestra.

## Interpolación IDW (XIII)

### ¿Cuándo usar IDW?

Por lo que comentado la salida depende del valor de la potencia ( $p$ ) y de la estrategia al definir la zonas de búsqueda. **Es un método exacto de interpolación, donde los valores máximos y mínimos en la superficie interpolada solamente ocurre en las proximidades de los puntos muestreados.**

## Interpolación IDW (XIV)

- La salida es sensible a las **agrupaciones (clusters)** y a los **valores atípicos (outliers)**.
- IDW asume que el fenómeno que está siendo modelado es impulsado por la variación local, que puede ser modelada mediante la definición de una adecuada zona de búsqueda.

## Interpolación IDW (XV)

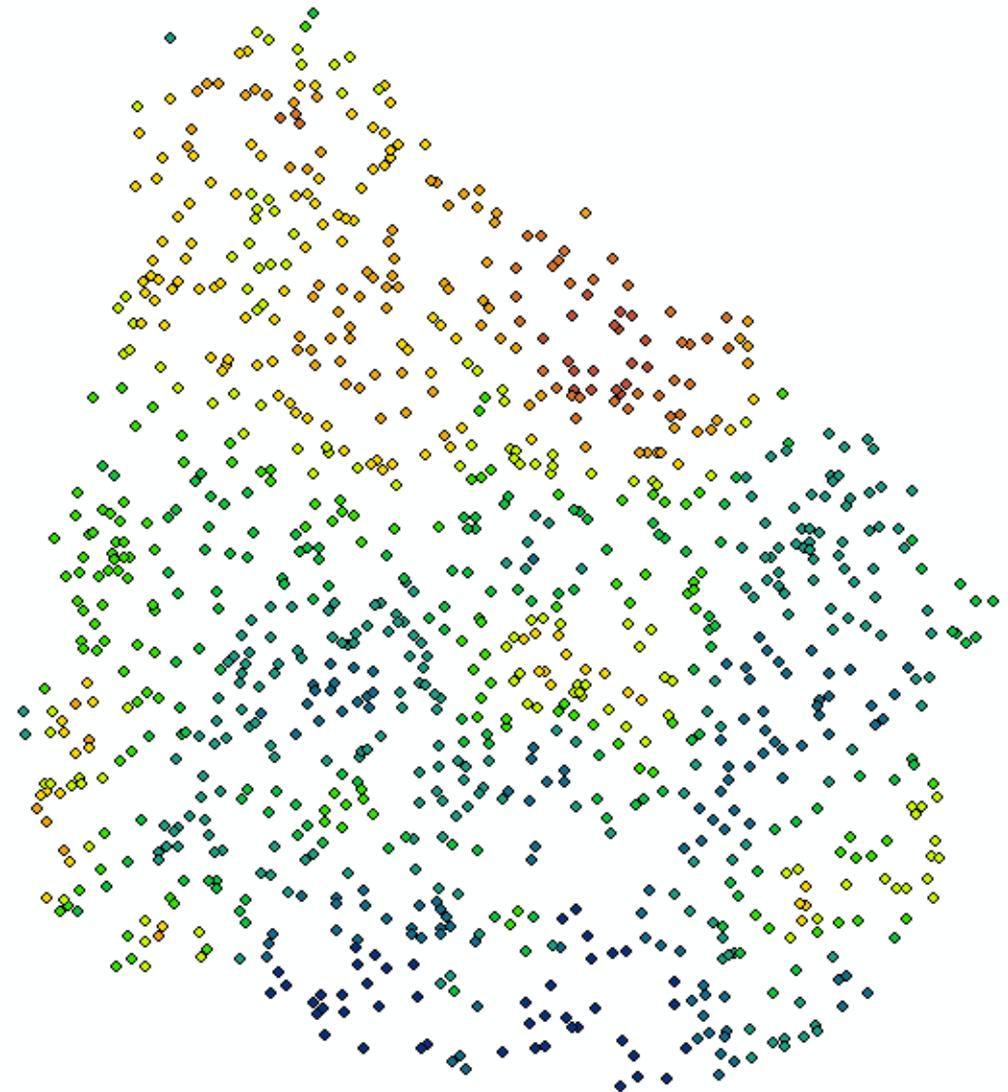
- La salida es sensible a las **agrupaciones (clusters)** y a los **valores atípicos (outliers)**.
- IDW asume que el fenómeno que está siendo modelado es impulsado por la variación local, que puede ser modelada mediante la definición de una adecuada zona de búsqueda.

# Ejemplo de Interpolación IDW (I)

Precipitación

Acumulado

- ◆ 201.000001 - 242.000000
- ◆ 172.000001 - 201.000000
- ◆ 148.000001 - 172.000000
- ◆ 130.000001 - 148.000000
- ◆ 113.000001 - 130.000000
- ◆ 97.000001 - 113.000000
- ◆ 83.000001 - 97.000000
- ◆ 68.000001 - 83.000000
- ◆ 47.000001 - 68.000000
- ◆ 20.000000 - 47.000000



# Ejemplo de Interpolación IDW (II)

rasIDWp2  
Value  
High : 240.987  
Low : 20.175

IDW

Input point features  
Precipitacion

Z value field  
Acumulado

Output raster  
C:\Temp\New File Geodatabase.gdb\rasIDWp2

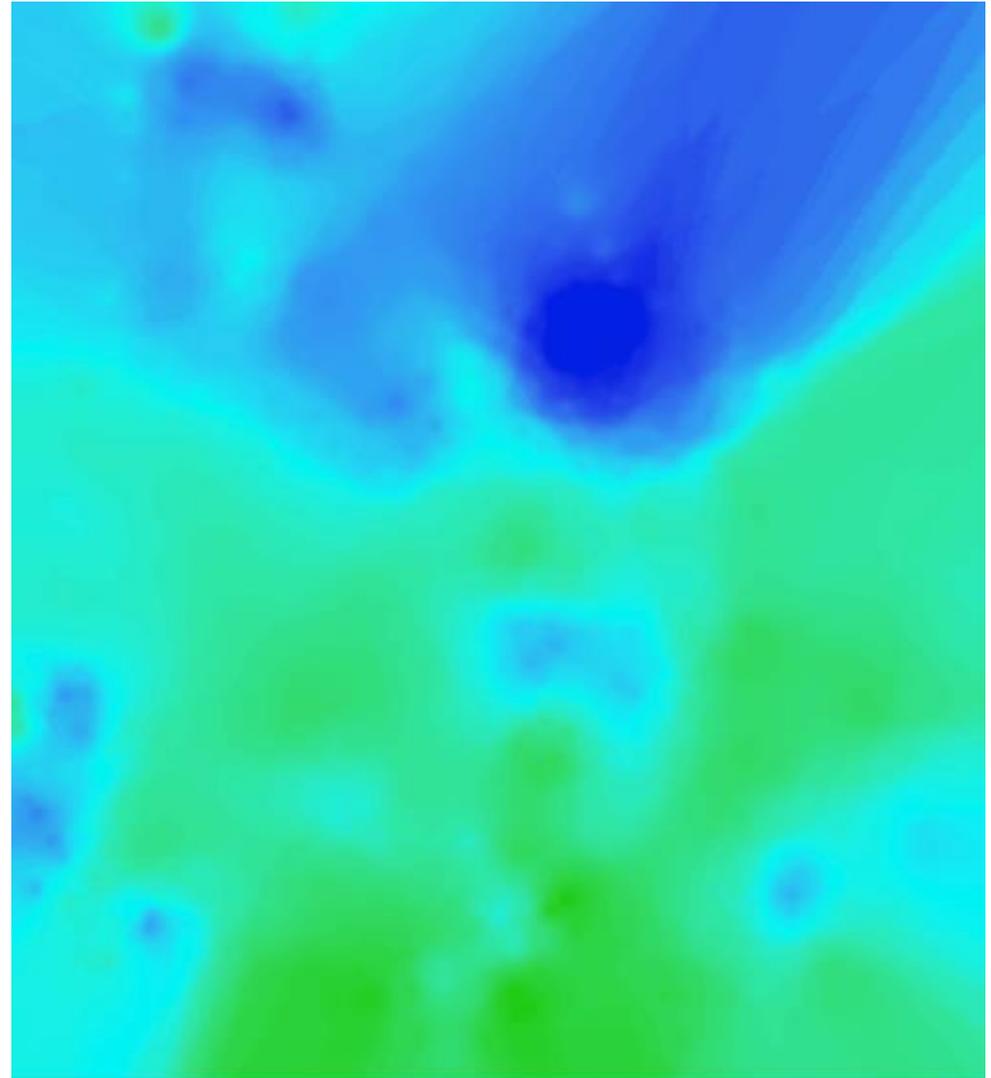
Output cell size (optional)  
2500

Power (optional)  
2

Search radius (optional)  
Variable

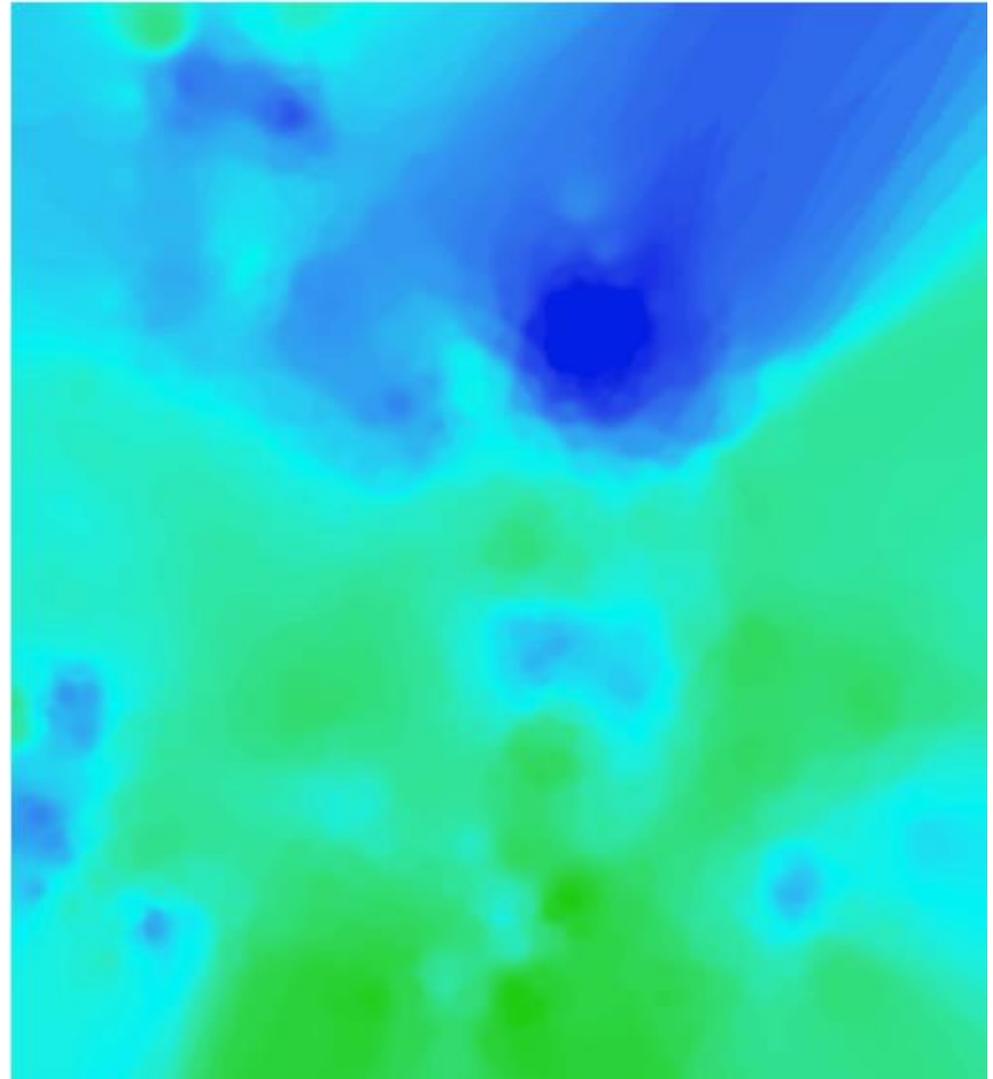
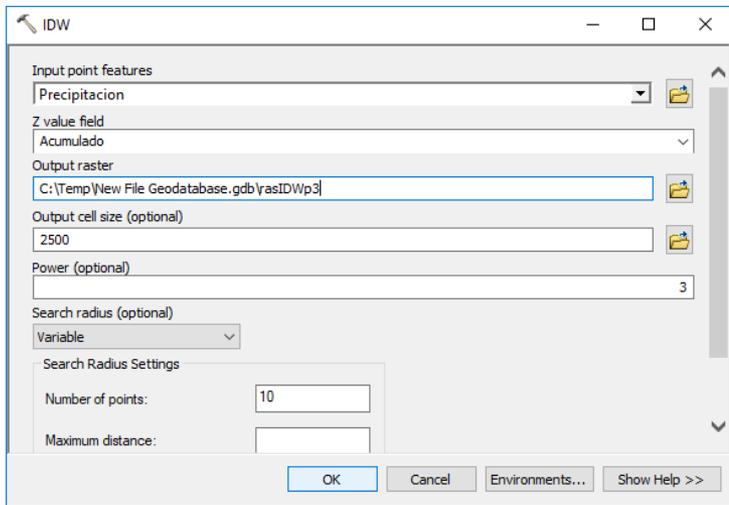
Search Radius Settings  
Number of points: 10  
Maximum distance:

OK Cancel Environments... Show Help >>



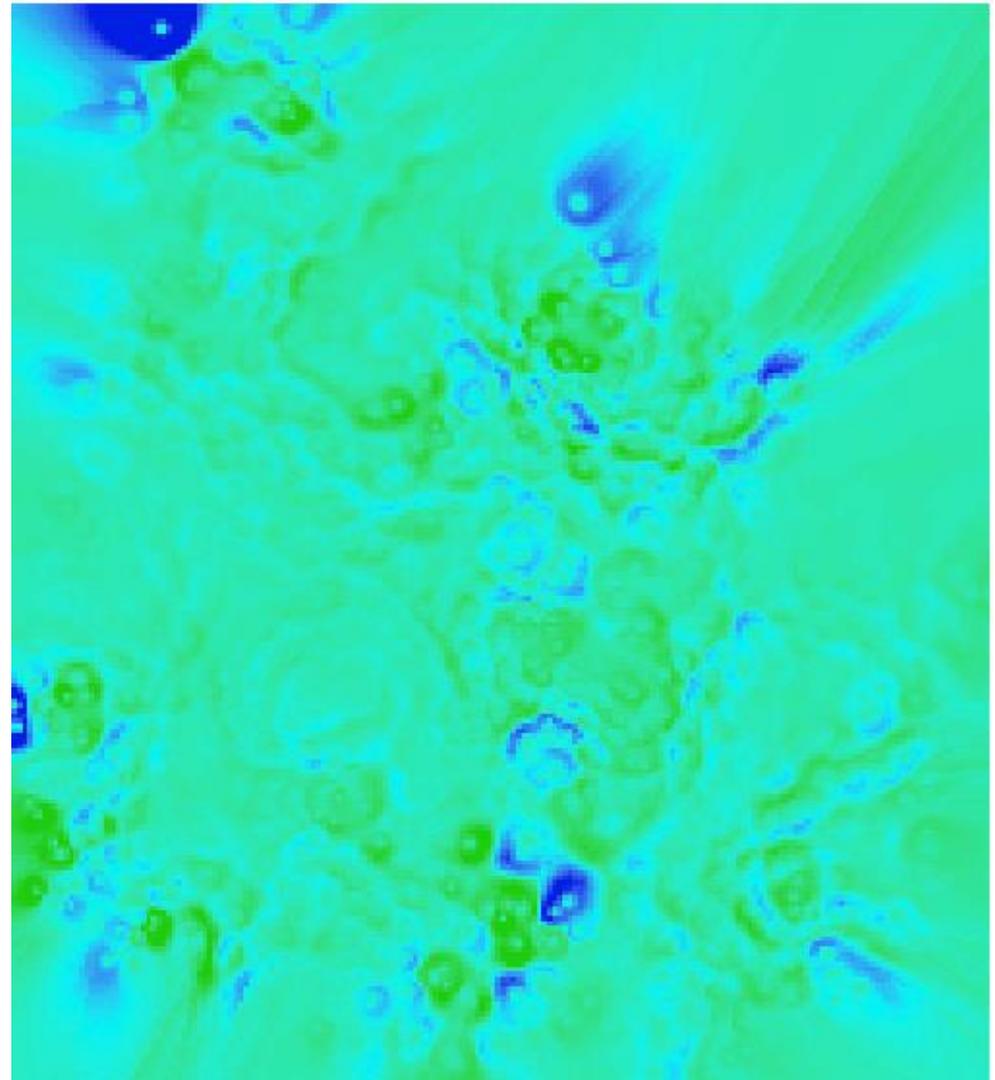
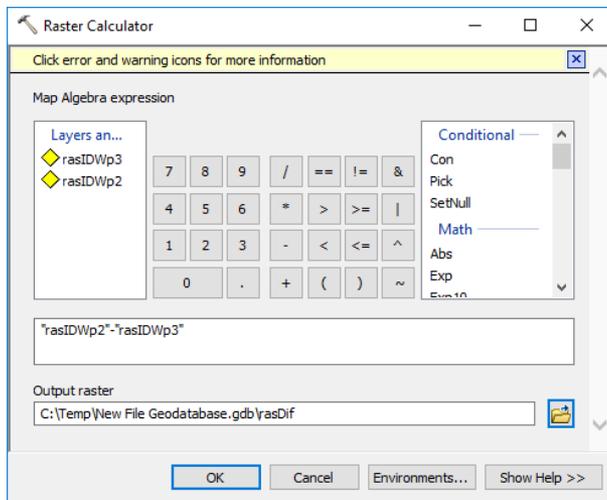
# Ejemplo de Interpolación IDW (III)

rasIDWp3  
Value  
High : 241.873  
Low : 20.006



# Ejemplo de Interpolación IDW (IV)

rasDif  
Value  
High : 19.8091  
Low : -7.42789



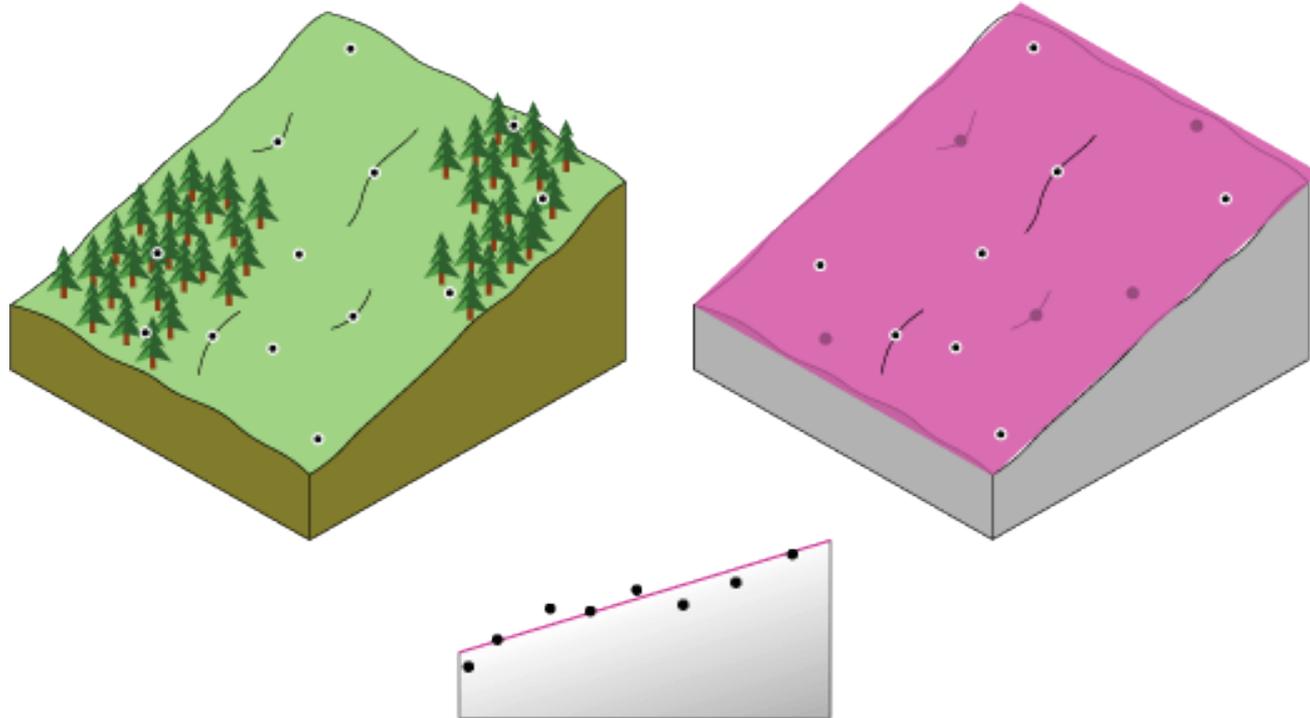


# DEMO de Interpolación IDW (I)

# Interpolación Polinómica Global (I)

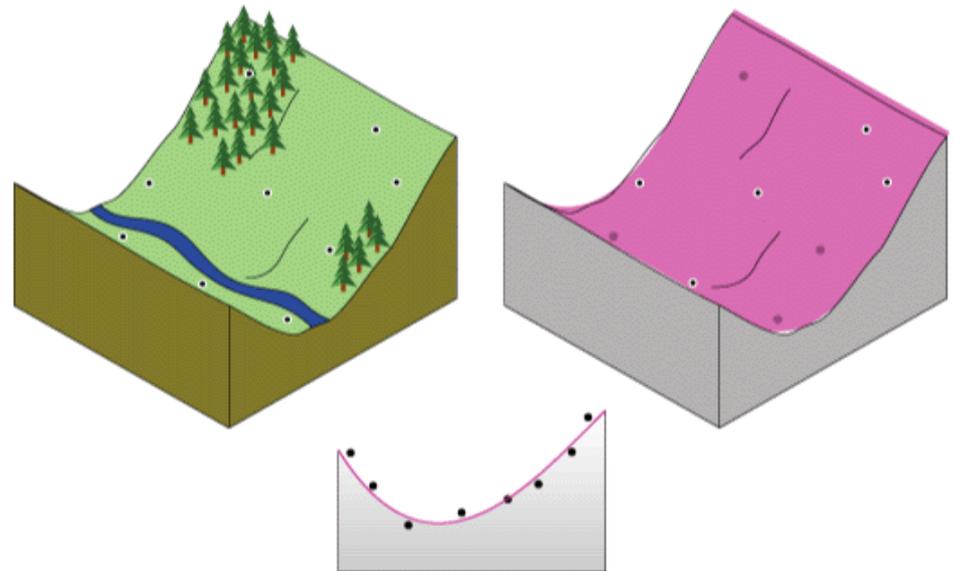
- Este método de interpolación **ajusta a una superficie suave definida por una función matemática (polinomio) para los puntos de la muestra de entrada.**
- La superficie polinómica global cambia gradualmente y captura el modelo a gran escala.
- Conceptualmente, la interpolación polinómica global mas sencilla (orden 1) es como tomar un pedazo de papel y ajustarlo a los puntos de muestreo.

# Interpolación Polinómica Global (II)



# Interpolación Polinómica Global (III)

- Un *papel plano* no es preciso para representar una geografía que contiene un valle. Doblando el papel una sola vez, se obtendrá un mejor ajuste.
- Agregando un término a la fórmula matemática se producirá un resultado más parecido: un doblez en el plano.



## Interpolación Polinómica Global (IV)

- El *papel plano* corresponde a una polinomio de orden 1 (un plano); el *papel con un doblez* corresponde a una polinomio de orden 2 (cuadrático); un *papel con dos dobleces* corresponde a un polinomio de orden 3 (cúbico).
- No es esperable que el resultado de la interpolación pase a través de los puntos de muestreo, con lo que la interpolación polinómica global de un **método inexacto**. Algunos puntos estarán por encima, y otros estarán por debajo.

## Interpolación Polinómica Global (V)

- **La cantidad de puntos de la muestra por encima del resultado de la interpolación es similar a la cantidad de puntos de la muestra por debajo del resultado de la interpolación.**
- La superficie es obtenida utilizando una regresión de ajuste de mínimos cuadrados; minimiza las diferencias al cuadrado entre los valores elevados y la hoja de papel.

# Interpolación Polinómica Global (VI)

## ¿Cuándo utilizar este método de interpolación?

- El resultado de este método arroja una **superficie suavizada que representa las tendencias graduales de la superficie en el área de interés.**
- **Se utiliza para ajustar una superficie a un conjunto de puntos de muestreo cuando la superficie tiene muy poca variación en el área de estudio.** Por ejemplo: variación de la polución en un área industrial.
- El uso de polinomios de orden inferior describen algunos procesos físicos (tales como la contaminación y la dirección del viento).

## Interpolación Polinómica Global (VII)

### ¿Cuándo utilizar este método de interpolación?

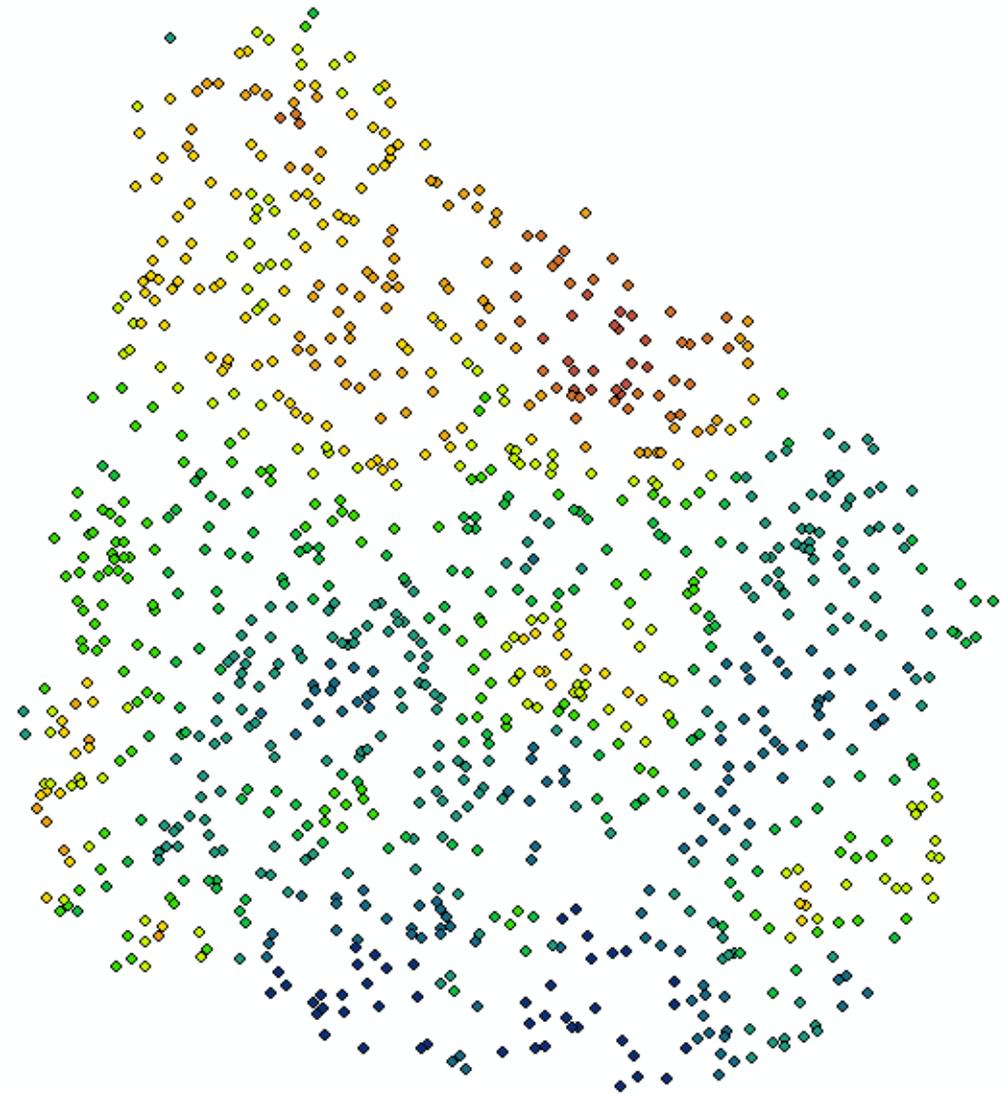
- Sin embargo, debe tenerse en cuenta que cuanto más complejo el polinomio, más difícil es atribuir significado físico a la misma.
- Las superficies calculadas son altamente susceptibles a los **valores atípicos** (extremadamente altos y bajos valores), especialmente en los bordes.

# Ejemplo de Interpolación Polinómica (I)

Precipitación

Acumulado

- ◆ 201.000001 - 242.000000
- ◆ 172.000001 - 201.000000
- ◆ 148.000001 - 172.000000
- ◆ 130.000001 - 148.000000
- ◆ 113.000001 - 130.000000
- ◆ 97.000001 - 113.000000
- ◆ 83.000001 - 97.000000
- ◆ 68.000001 - 83.000000
- ◆ 47.000001 - 68.000000
- ◆ 20.000000 - 47.000000

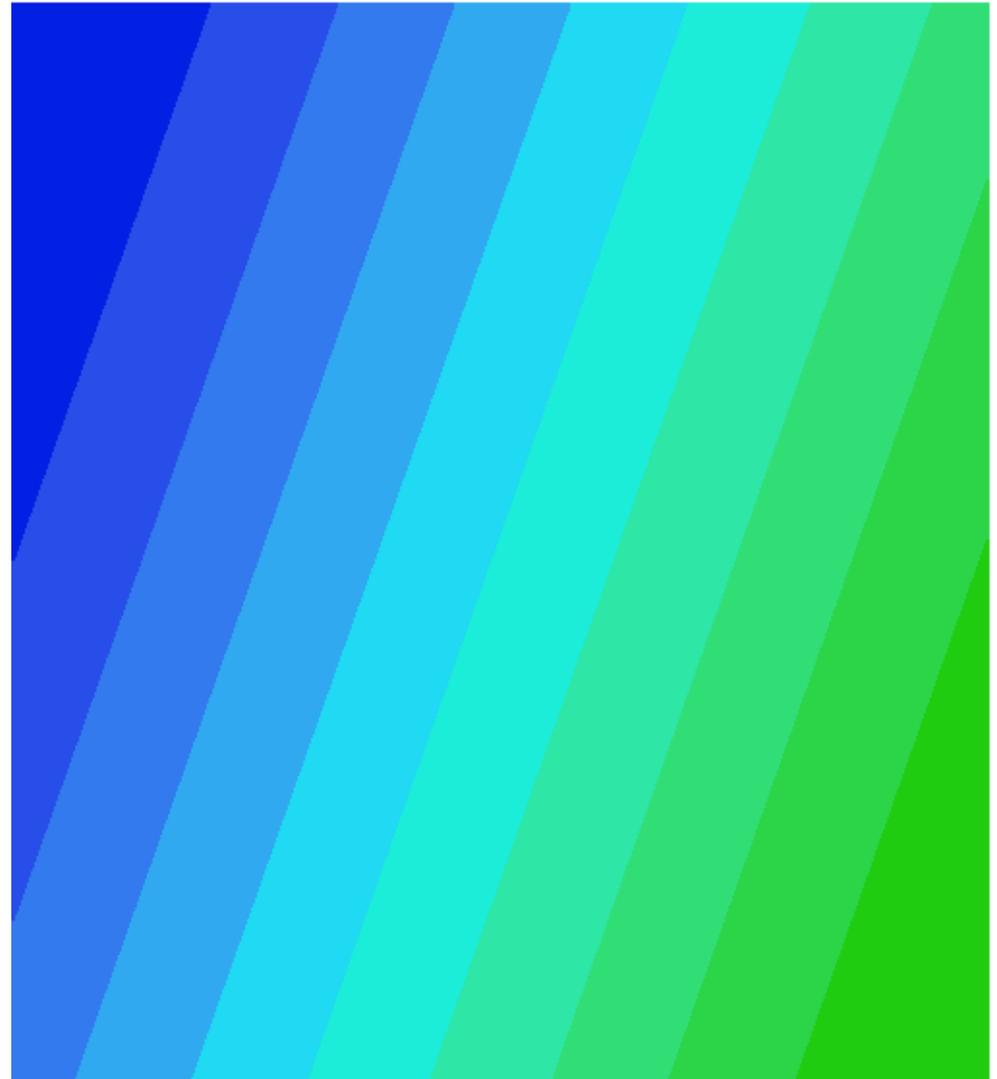


# Ejemplo de Interpolación Polinómica (II)

rasPoli1

<VALUE>

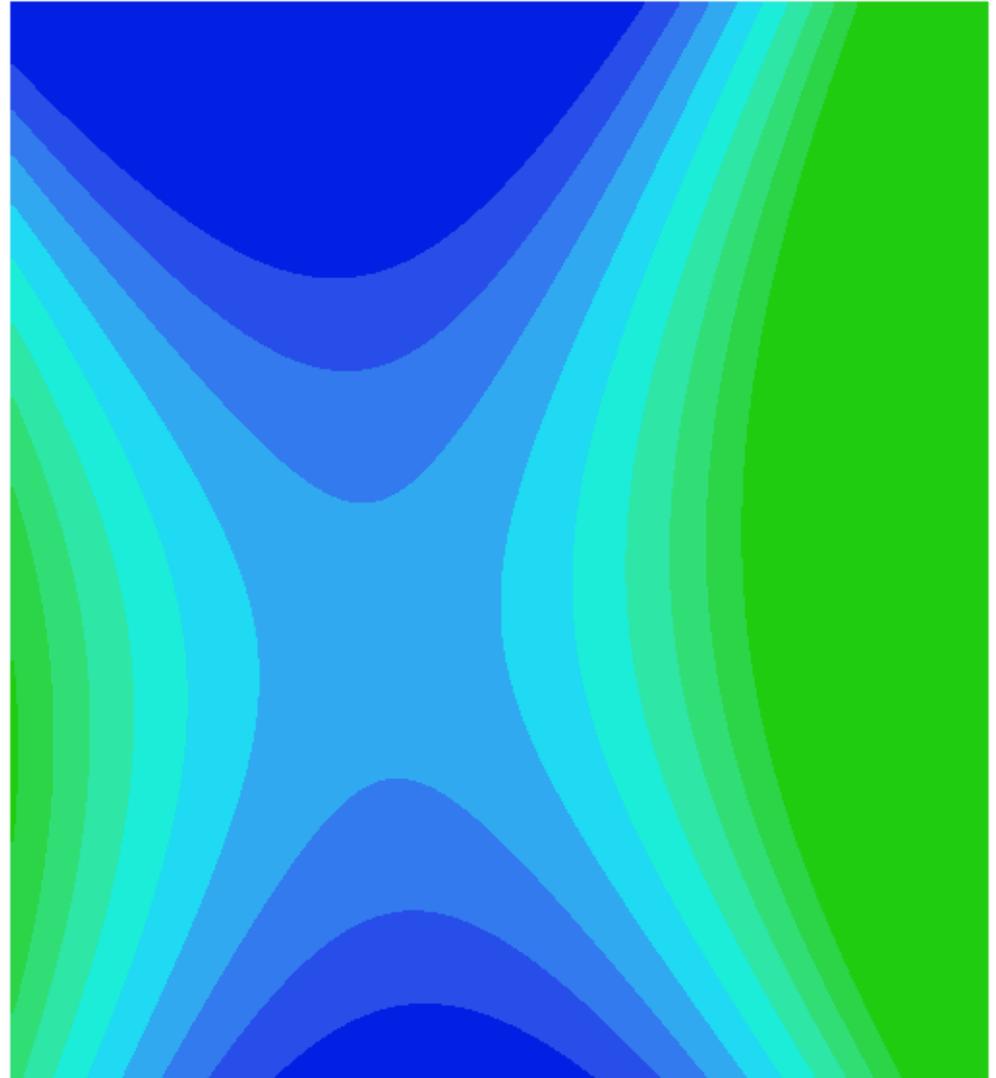
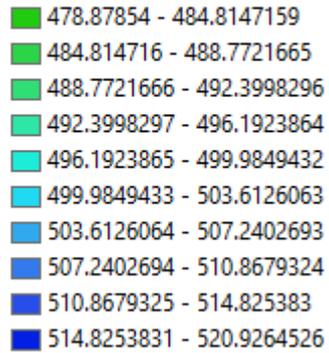
-  478.87854 - 484.8147159
-  484.814716 - 488.7721665
-  488.7721666 - 492.3998296
-  492.3998297 - 496.1923864
-  496.1923865 - 499.9849432
-  499.9849433 - 503.6126063
-  503.6126064 - 507.2402693
-  507.2402694 - 510.8679324
-  510.8679325 - 514.825383
-  514.8253831 - 520.9264526



# Ejemplo de Interpolación Polinómica (III)

rasPoli2

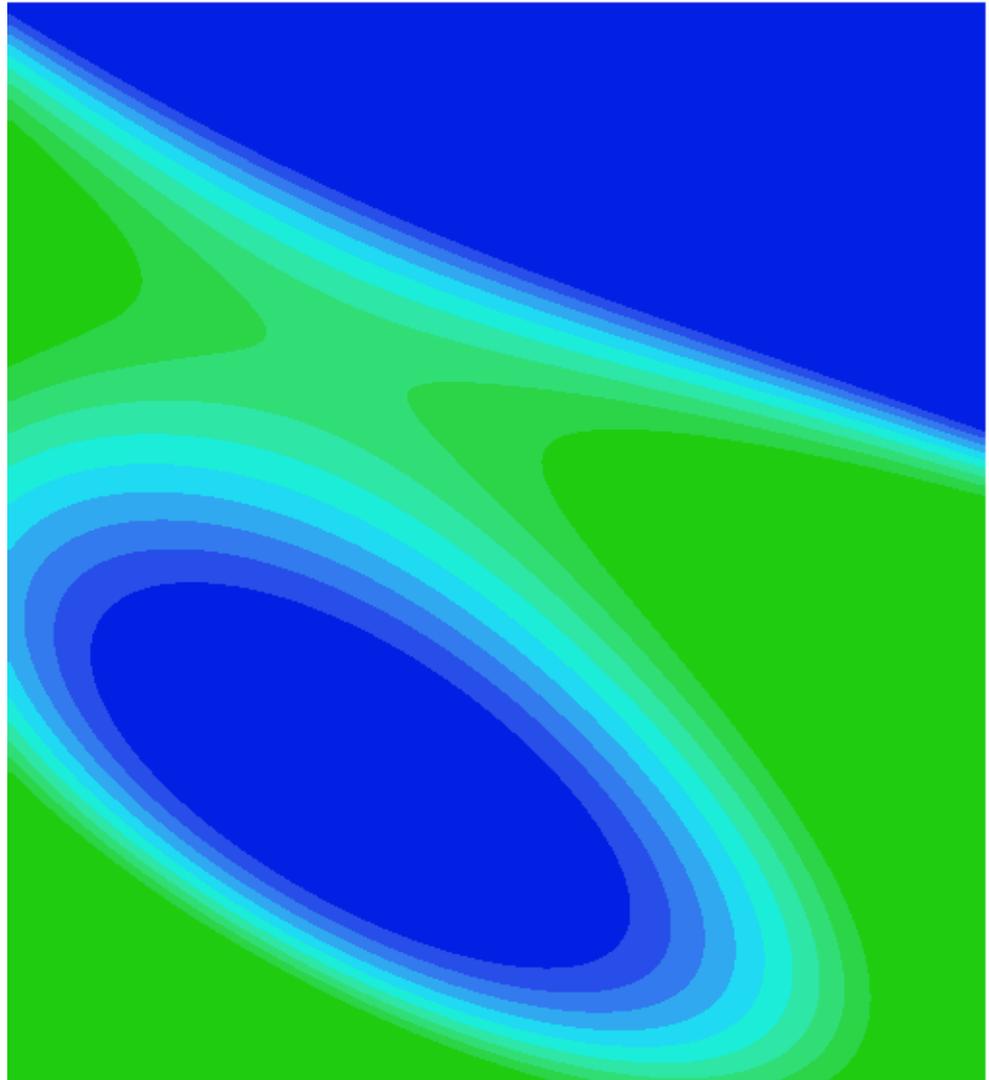
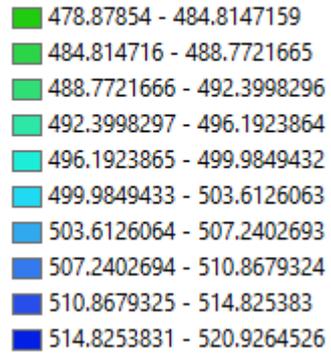
<VALUE>



# Ejemplo de Interpolación Polinómica (IV)

rasPoli3

<VALUE>

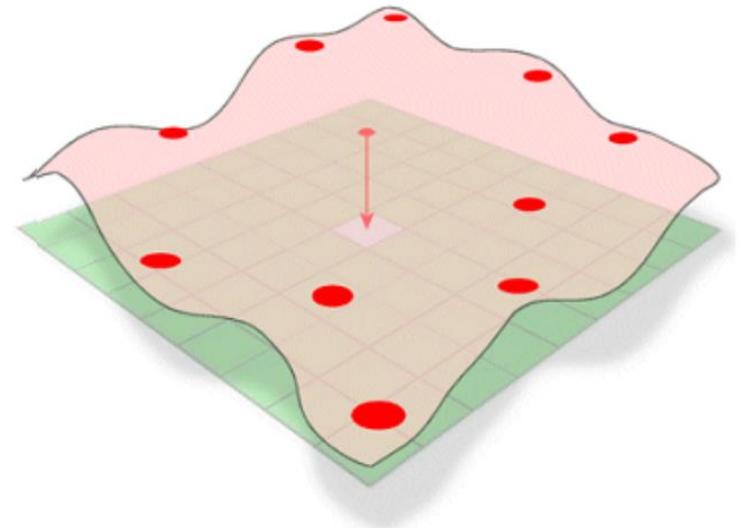




# Demo de Interpolación Polinómica (I)

# Spline (I)

- El método de **Spline** curva una hoja de goma que pasa a través de los puntos de entrada y minimiza a la vez la curvatura total de la superficie.
- Ajusta una función matemática a una cantidad especificada de puntos de entrada más cercanos mientras pasa a través de los puntos de muestra.



## Spline (II)

- Es útil para interpolaciones que varían levemente, como la elevación, la altura del agua, contaminación, etc.
- Impone las siguientes dos condiciones en el resultado: la **superficie debe pasar exactamente por los puntos de datos** y la **superficie debe tener una curvatura mínima**.
- Lo último se logra haciendo que la acumulación cuadrática de los términos de la derivada segunda de la superficie tomada sobre cada punto de la superficie debe ser mínimo.

## Spline (III)

- Esto asegura una superficie suave (continua y diferenciable), junto con superficies continuas de derivada primera.
- En las inmediaciones de los puntos de muestreo pueden ocurrir cambios rápidos en la pendiente (derivada primera), por lo que este modelo no es adecuado para estimar la derivada segunda (curvatura).

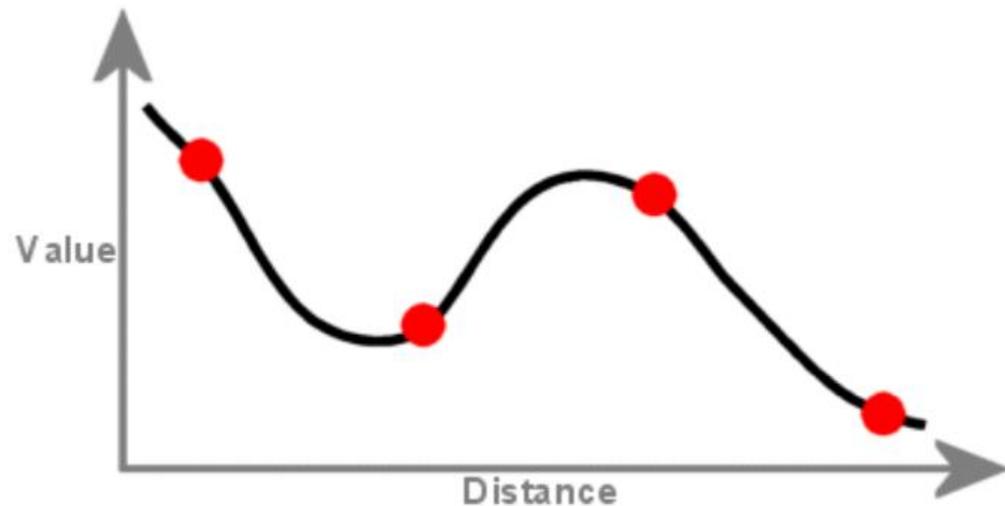
# Spline (IV)

## Tipos de Spline:

- **Regularizado:** crea una superficie suave que cambia gradualmente con valores que pueden estar fuera del rango de datos de muestra. Modifica el criterio de minimización para que los términos de la derivada tercera se incorporen a los criterios de minimización. Esta técnica es útil si es necesario calcular la derivada segunda de la superficie interpolada.
- **De Tensión:** controla la rigidez de la superficie en función del carácter del fenómeno modelado. Crea una superficie menos suave con valores que están más restringidos por el rango de datos de la muestra.

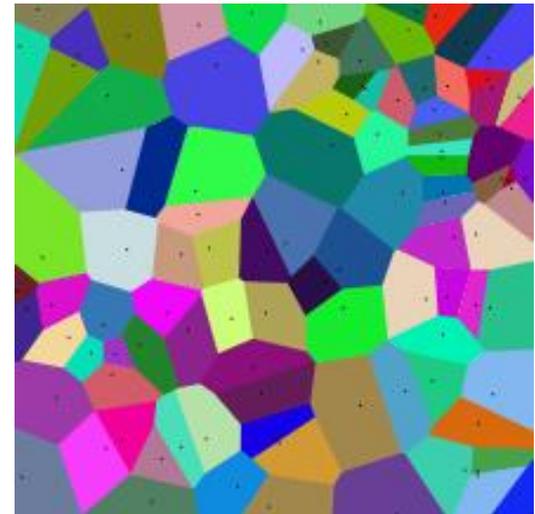
## Spline (V)

- Este método es útil si se quiere estimar valores que están por encima o debajo de los valores máximos de los valores de muestreo.



## Vecino más cercano (I)

- También llamada **interpolación por proximidad**.
- Es el método mas simple de interpolación.
- Selecciona el valor del punto más cercano y no tiene en cuenta los valores de los puntos vecinos.
- Produce una interpolación por tramos constante.
- Se basa en los **Diagramas de Voronoi** o de **Thiessen**.



# Vecino natural (I)

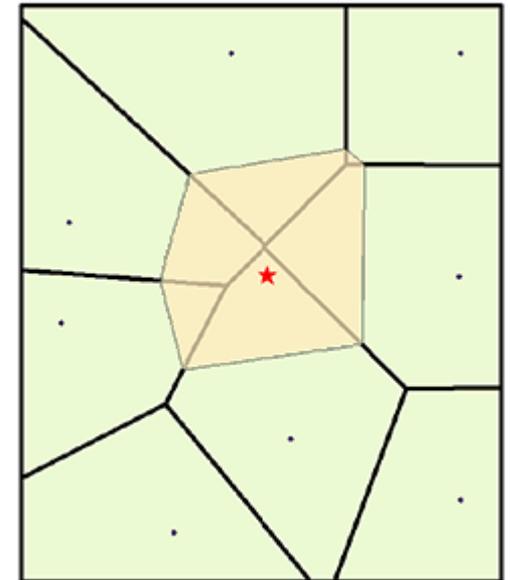
- Se basa en los **Diagramas de Voronoi**.
- Este método **encuentra el subconjunto de muestras de entrada más cercano a un punto de consulta y aplica ponderaciones sobre éstas basándose en áreas proporcionales para interpolar un valor**.
- Conocido también como **Interpolación de Sibson** o de **Robo de área**.
- Es un **método local** que utiliza sólo un **subconjunto de muestras que circundan a un punto de consulta y asegura que los valores interpolados estén dentro del rango de las muestras utilizadas**.

## Vecino natural (II)

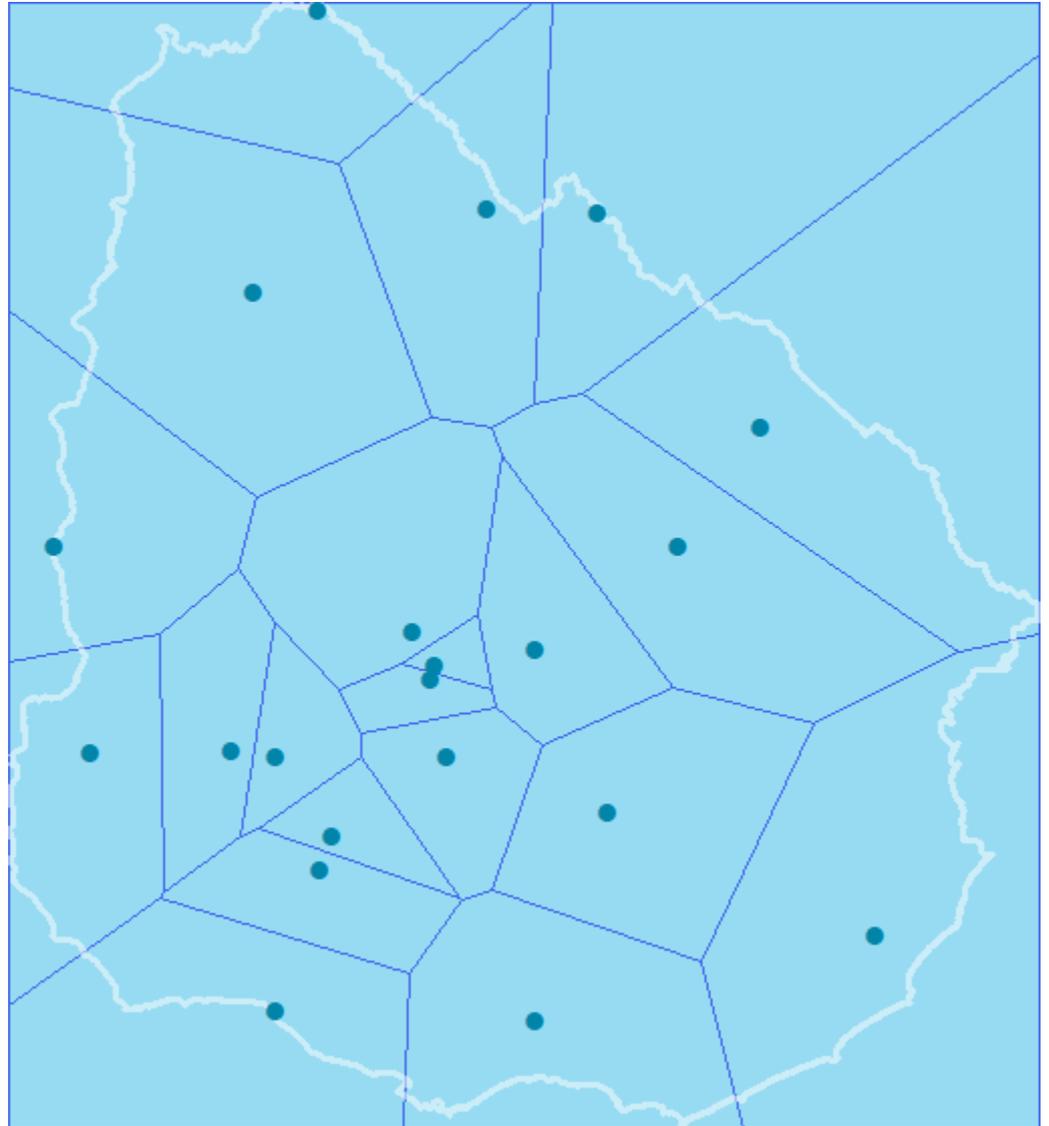
- **No produce tendencias ni produce picos, depresiones, crestas o valles que no estén ya representados por las muestras de entrada.**
- La superficie **pasa por las muestras de entrada** y es suave en todas partes, excepto en las ubicaciones de las muestras de entrada.
- **Los vecinos naturales de un punto son aquellos que están asociados con los polígonos de Voronoi adyacentes.**

## Vecino natural (III)

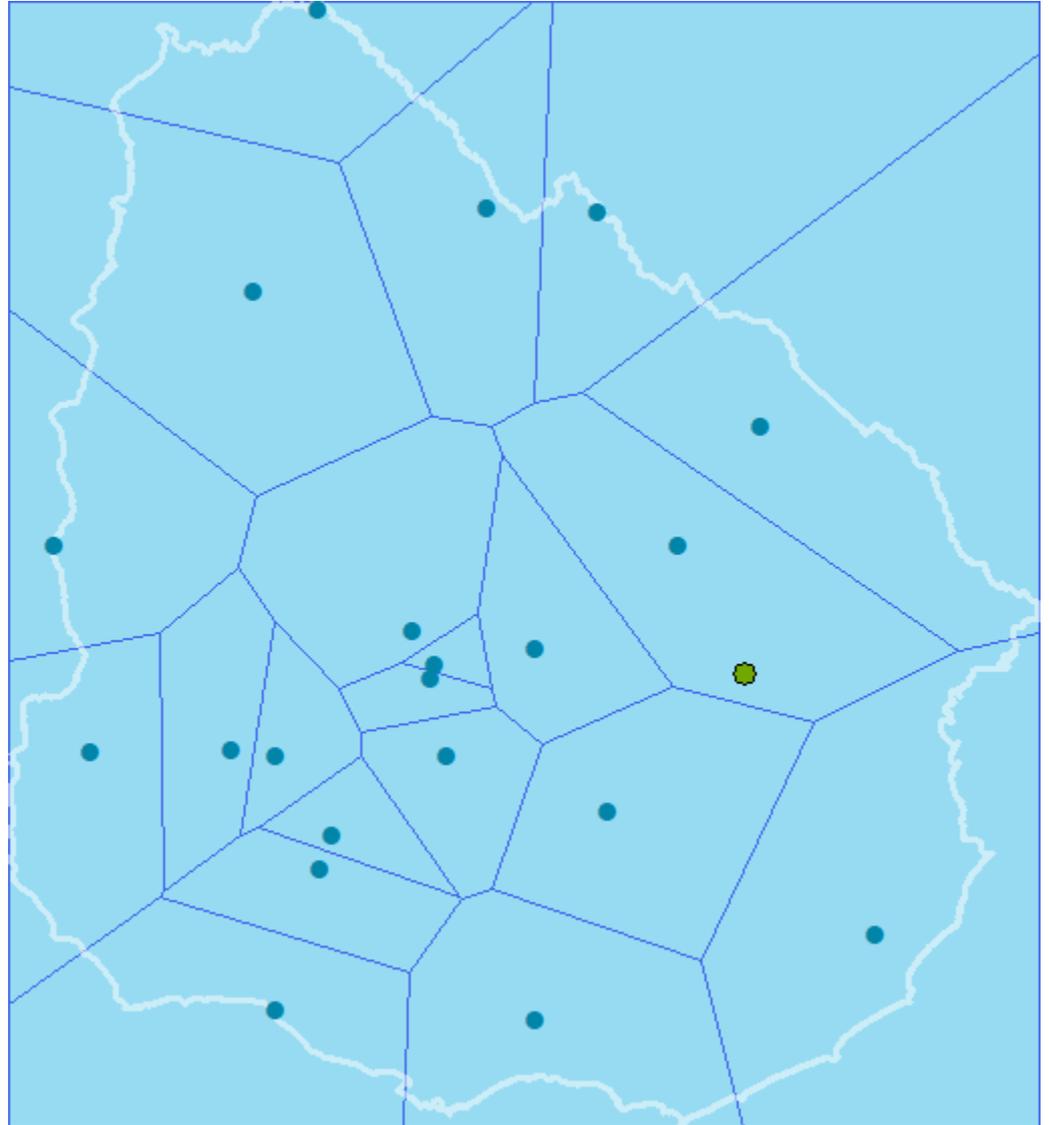
- Se construye un Diagrama de Voronoi de todos los puntos dados. Luego se crea un nuevo polígono de Voronoi alrededor del punto de interpolación.
- La proporción de superposición entre este polígono nuevo y los polígonos iniciales se utiliza como **ponderaciones**.



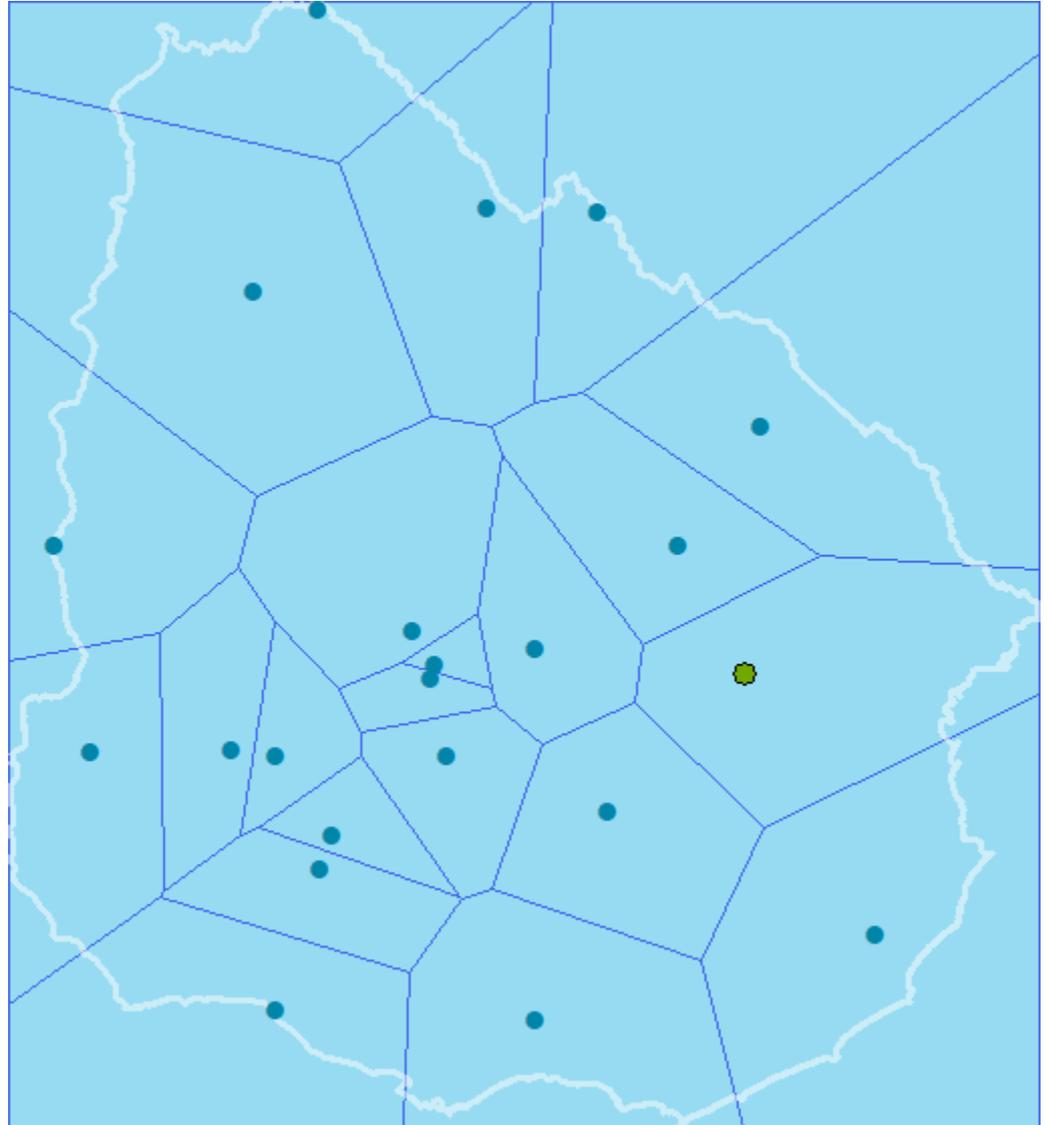
## Vecino natural (III)



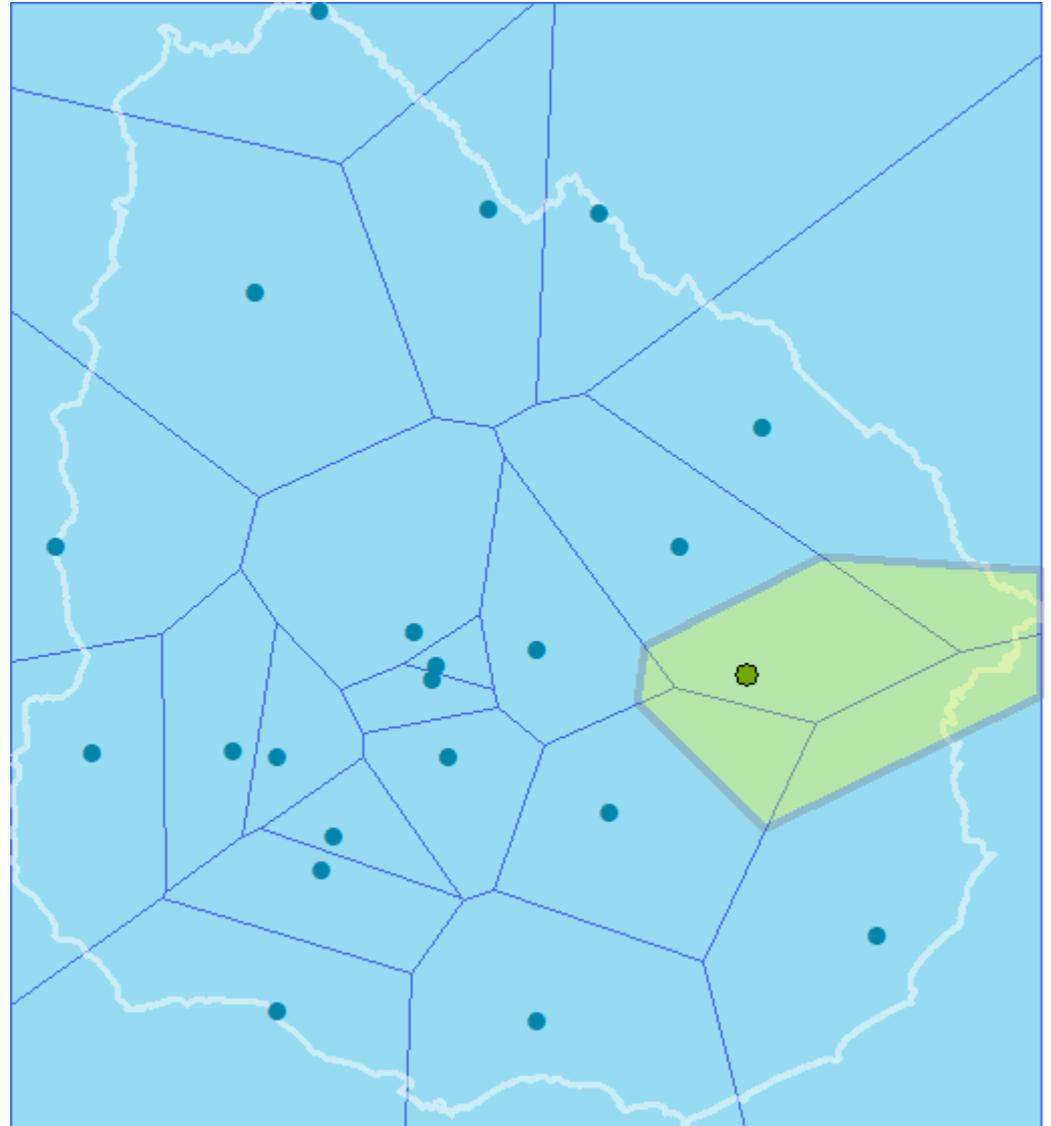
## Vecino natural (III)



## Vecino natural (III)



## Vecino natural (III)



# Kriging (I)

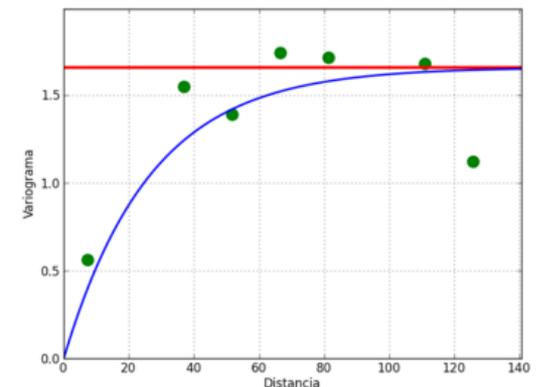
- Conocido como **Método de Interpolación de Kriging, Kigreado o Kigreja**.
- Es un **método geoestadístico** muy avanzado de estimación de puntos.
- A diferencia de otros métodos de interpolación, Kriging **implica una investigación interactiva del comportamiento espacial del fenómeno representado por los valores a interpolar** antes de seleccionar el mejor método de estimación para generar la superficie de salida.

## Kriging (II)

- Los **métodos de interpolación geoestadísticos** están basados en modelos estadísticos que incluyen la **autocorrelación** (las relaciones estadísticas entre los puntos medidos).
- Por esto, las técnicas de estadística geográfica no solo tienen la capacidad de producir una superficie de predicción sino que **también proporcionan alguna medida de certeza o precisión de las predicciones.**
- La interpolación de Kriging presupone que la distancia o la dirección entre los puntos de muestra reflejan una correlación espacial que puede utilizarse para explicar la variación en la superficie.

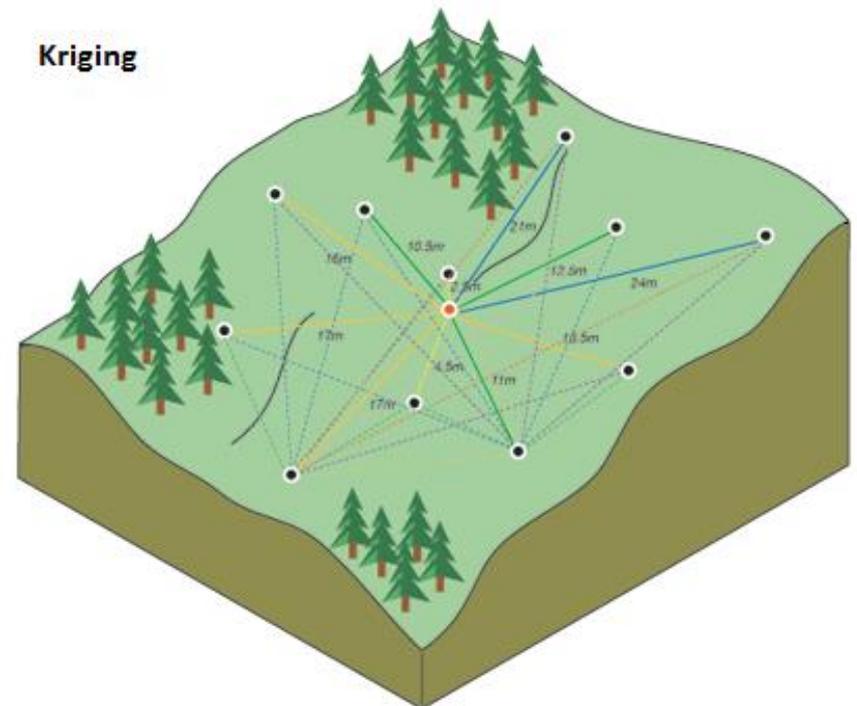
# Kriging (III)

- El método de Kriging **ajusta una función matemática** a una cantidad especificada de puntos o a todos los puntos dentro de un radio específico para determinar el valor de salida para cada ubicación.
- Kriging tiene tres pasos: 1) **análisis estadístico** exploratorio de los datos y el **modelado de variogramas**; 2) la **creación de la superficie** (resultado de la interpolación); 3) la **exploración de la superficie de varianza** (opcional).



## Kriging (IV)

- Este método es más adecuado cuando se sabe que hay una influencia direccional o de la distancia correlacionada espacialmente en los datos. Se utiliza a menudo en la ciencia del suelo y la geología.



# Kriging (V)

## Fórmula General de Kriging:

$$\hat{Z}(s_0) = \sum_{i=1}^N \lambda_i Z(s_i)$$

$Z(s_i)$  = el valor medido en la ubicación  $i$

$\lambda_i$  = una ponderación desconocida para el valor medido en la ubicación  $i$

$s_0$  = la ubicación de la predicción

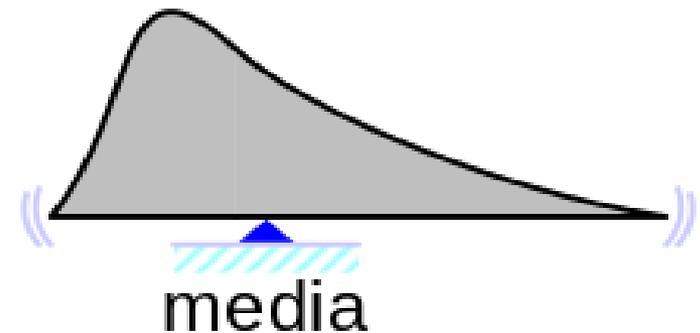
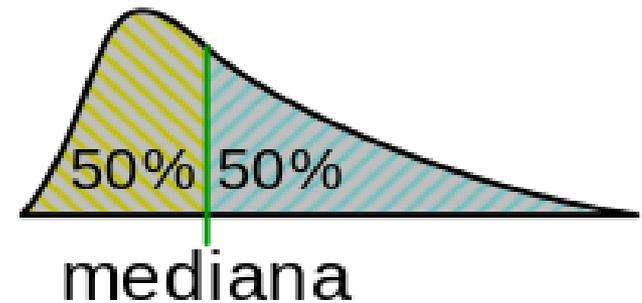
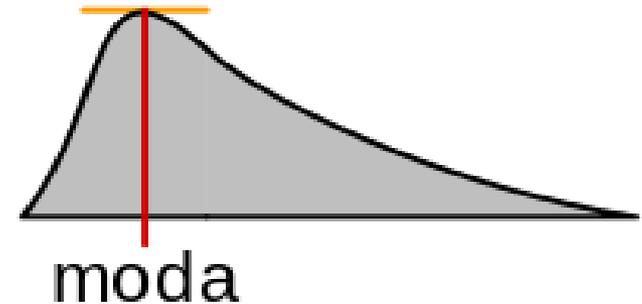
$N$  = la cantidad de valores medidos

## Kriging (VI)

- En IDW, la ponderación,  $\lambda_i$ , depende sólo de la distancia a la ubicación de la predicción.
- En el Método Kriging, las **ponderaciones están basadas no solo en la distancia entre los puntos medidos y la ubicación de la predicción, sino también en la disposición espacial general de los puntos medidos.**
- **Variograma:** herramienta que permite **analizar el comportamiento espacial de una variable sobre un área definida.** Se obtiene como resultado un variograma experimental que refleja la distancia máxima y la forma en que un punto tiene influencia sobre otro punto a diferentes distancias.

# Repaso (muy breve) de estadística (I)

- **Moda:** La moda de un conjunto de números es el número que aparece más a menudo.
- **Mediana:** Es el número medio en el conjunto o, si hay un número par de datos, la mediana es el promedio de los dos números medios.
- **Media:** La media de un conjunto de números es la suma de los datos dividida entre el número total de datos.

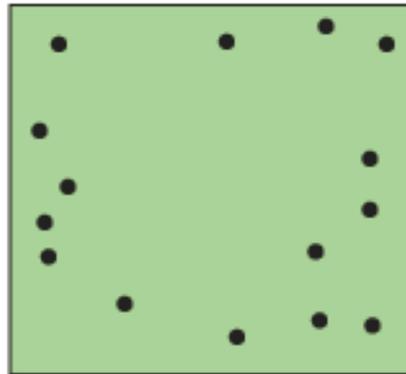


# Centro Medio (I)

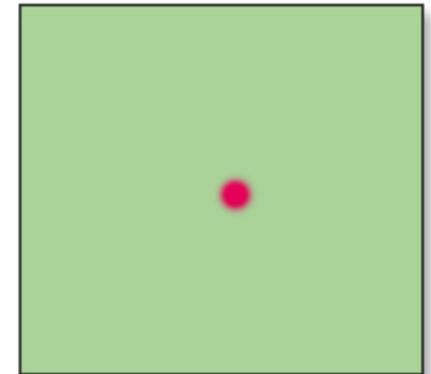
- Permite identificar el centro geográfico (o el centro de concentración) de un grupo de entidades.
- El **centro medio** es un punto construido desde los valores  $X$  e  $Y$  promedios para los **centroides** de entidad de entrada.

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$$

$$\bar{Y} = \frac{\sum_{i=1}^n y_i}{n}$$



INPUT



OUTPUT

# Centro Mediano (I)

- **Permite identificar la ubicación que minimiza la distancia euclidiana total a las entidades en un set de datos.**
- Mientras que el centro medio devuelve un punto en las coordenadas X promedio e Y promedio, el centro mediano utiliza un algoritmo iterativo para **buscar el punto que reduce la distancia euclidiana a todas las entidades de entrada.**

$$d_i^t = \sqrt{(X_i - X^t)^2 + (Y_i - Y^t)^2}$$

- Tanto el centro medio como el centro mediano son mediciones de la tendencia central, pero el **centro mediano está menos influenciado por valores atípicos de datos (outliers).**

## Distancia Standard (I)

- Permite medir la compactación de una distribución proporcionando un valor simple que representa la dispersión de las entidades alrededor de un centro.
- El valor es una distancia, de modo que la compactación de un conjunto de entidades se puede representar en un mapa mediante el dibujo de un círculo que tenga el radio igual al valor de la distancia estándar.

$$SD_w = \sqrt{\frac{\sum_{i=1}^n w_i (x_i - \bar{X}_w)^2}{\sum_{i=1}^n w_i} + \frac{\sum_{i=1}^n w_i (y_i - \bar{Y}_w)^2}{\sum_{i=1}^n w_i}}$$

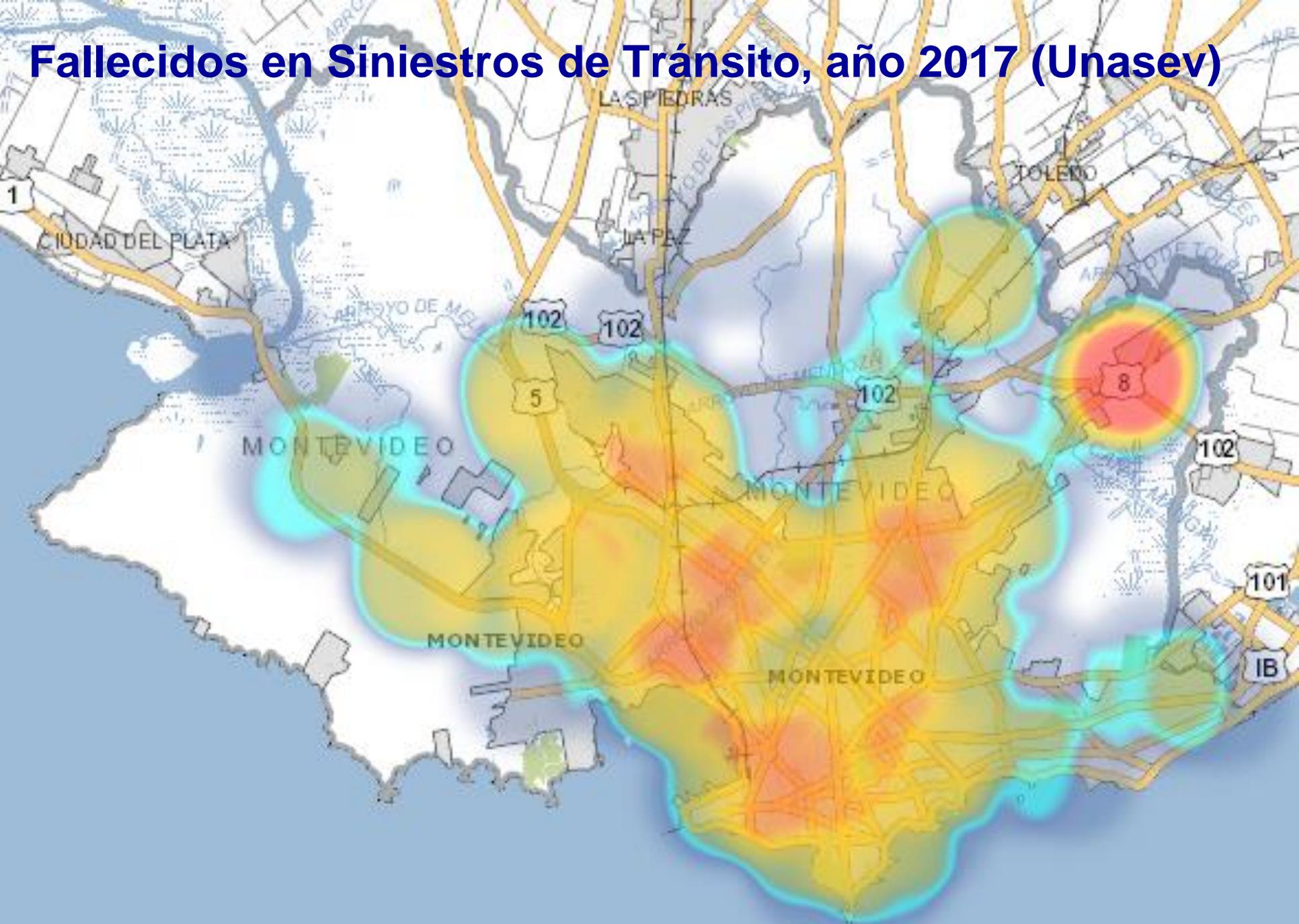
## Distancia Standard (II)

- **Podemos utilizar los valores de dos o más distribuciones para comparar las muestras.**
- A modo de ejemplo, un analista podría comparar la **compactación de asaltos** y de **robos de automóviles**. Conocer cómo se distribuyen los diferentes tipos de delitos puede ayudar a desarrollar estrategias para combatir el delito. *(Si la distribución en un área en particular es compacta, estacionar un auto cerca del centro del área puede ser suficiente. Si la distribución es dispersa, el patrullaje de varios coches de la policía en el área puede ser más efectivo para responder a los delitos.)*

## Análisis de Densidad (II)

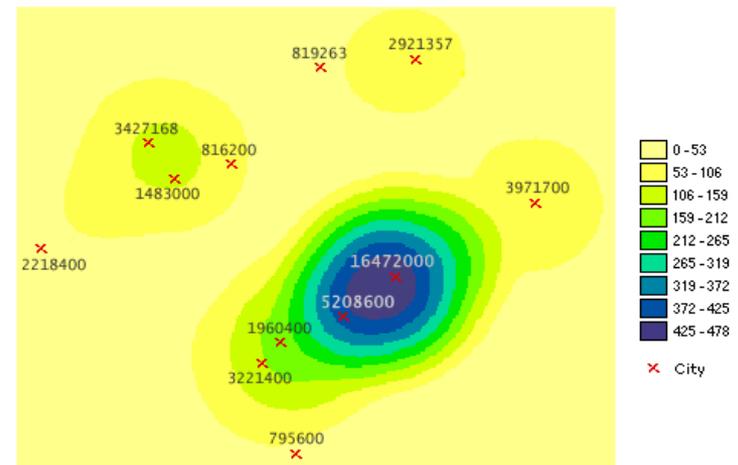
- Las áreas donde se concentran determinados tipos de eventos son llamadas **hot spots (punto caliente)** y tienen como elementos definatorios la **intensidad** y la **extensión**.
- Un **hot spot** es un área donde ocurren más eventos en comparación con el promedio.

# Fallecidos en Siniestros de Tránsito, año 2017 (Unasev)



# Densidad de Kernel (I)

- Este tipo de análisis toma cantidades conocidas de algunos fenómenos y las expande a través de la geografía basándose en la cantidad que se mide en cada ubicación y la relación espacial de las ubicaciones de las cantidades medidas.
- **¿Para qué?** Las superficies de densidad muestran las concentraciones.

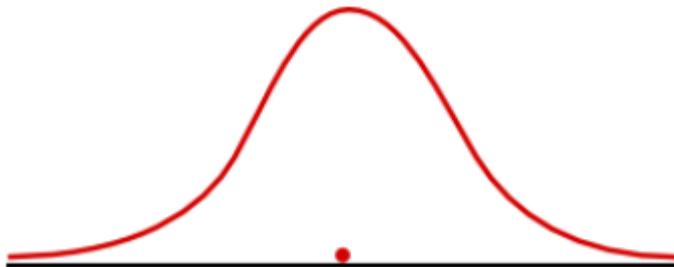


## Densidad de Kernel (II)

Hay dos factores fundamentales para este tema:

- **Función de Probabilidad:** Función que asignamos para la distribución en el espacio.
- **Ancho de Banda:** Radio predeterminado de búsqueda (**h**).

$$g(x_j) = \sum \left\{ [W_i I_i] \frac{1}{h^2 2\pi} e^{-\frac{d_{ij}^2}{2h^2}} \right\}$$



Gaussiana

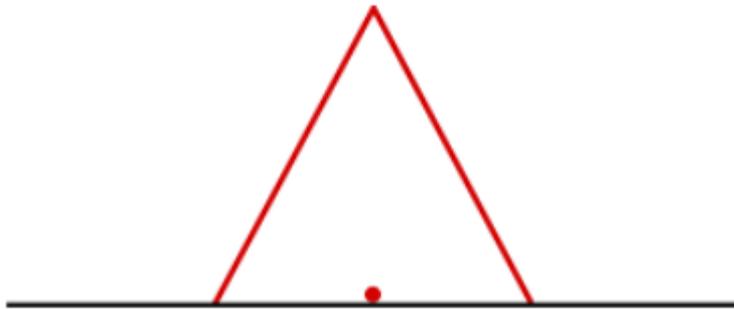
$$g(x_j) = \sum \left\{ [W_i I_i] \frac{3}{h^2 \pi} \left[ 1 - \frac{d_{ij}^2}{h^2} \right] \right\}$$



Esférica

# Densidad de Kernel (II)

$$g(x_j) = \sum \left\{ K - \left( \frac{K}{h} \right) \right\} d_{ij}$$



Triangular

$$g(x_j) = \sum A e^{K \cdot d_{ij}}$$



Exponencial Negativa

$$g(x_i) = \sum K$$



Uniforme

# Densidad de Kernel (V)

Para determinar el radio de búsqueda:

1. Se calcula el **Centro Medio** de los puntos de entrada.
2. Se calcula la **Distancia** desde el centro medio a todos los puntos de entrada.
3. Se calcula la **Distancia Media** de esas distancias ( $D_m$ ).
4. Se calcula el valor de la **Distancia Standard (SD)**.
5. Se aplica la fórmula:

$$\text{RadioBusqueda} = 0.9 * \min \left( SD, \sqrt{\frac{1}{\ln(2)}} * D_m \right) * n^{-0.2}$$