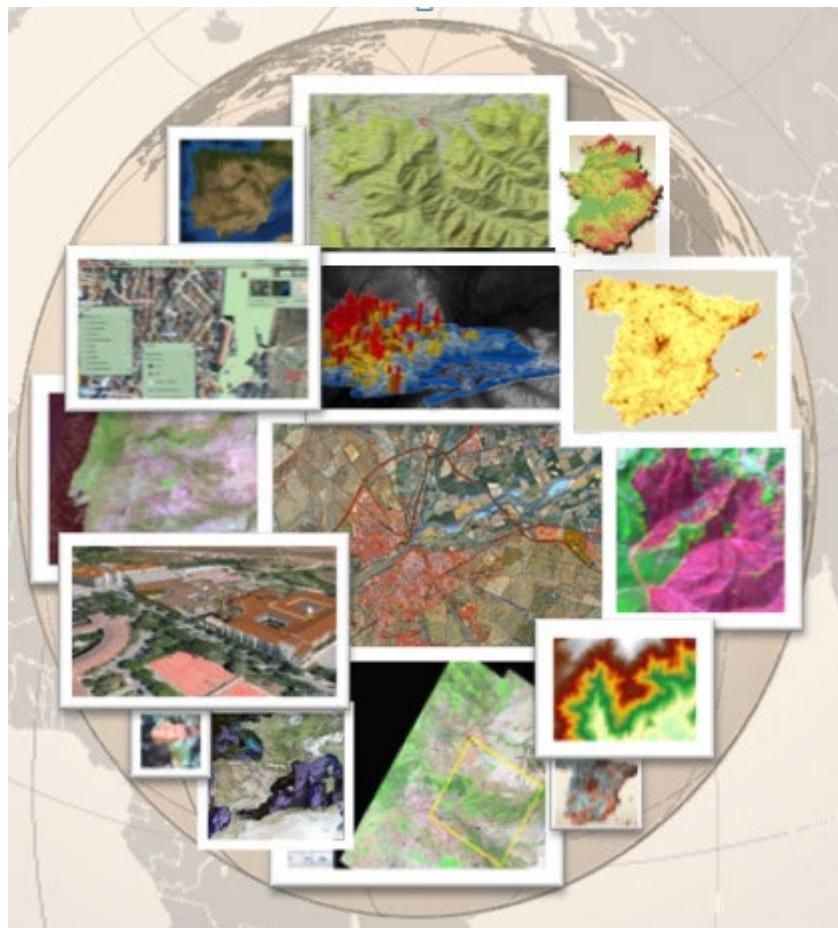


Máster en Tecnologías de la Información Geográfica: SIG y Teledetección



Geoestadística y calidad de la información

Tema 3. Métodos de interpolación

Contenido

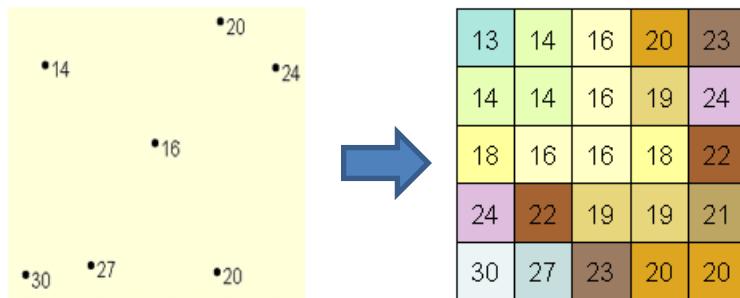
1. Introducción	2
1.1 ¿Qué es la interpolación?.....	2
1.2 Métodos de interpolación.....	3
1.3 Interpolación Geoestadística.....	4
1.3.1. <i>Método de interpolación IDW</i>	5
1.3.2. <i>Método de interpolación Spline</i>	6
1.3.3. <i>Método de Vecino natural</i>	7
1.3.3. <i>Interpolación Kriging</i>	7
1.4 Otros parámetros de la interpolación	12
1.5 Elección del método de interpolación	12
2. Interpolación en ArcGIS	13
2.1 Interpolación IDW.....	19
2.2 Splines.....	21
2.3 Vecino natural	22
2.4 Kriging	23
3. Bibliografía	42

1. Introducción

La geoestadística es una técnica estadística utilizada para la estimación, predicción y simulación de datos correlados espacialmente el arte de modelar datos espaciales.

Permite describir la continuidad espacial de las variables y estimar valores muy cercanos a los reales en puntos desconocidos.

La necesidad de interpolar surge de la imposibilidad de medir en los infinitos puntos del espacio.



1.1 ¿Qué es la interpolación?

En términos generales, un método de interpolación es una herramienta que permite el cálculo del valor de una variable en una coordenada para la cual dicho valor no es conocido a partir de los valores conocidos para otra serie de coordenadas.

En el caso particular de la creación de una capa ráster, las coordenadas x,y donde han de calcularse los valores desconocidos son los centros de las celdas sin dato recogido.

En general, cualquier variable recogida mediante muestreo puede ser el punto de partida para la aplicación de métodos de interpolación.

En algunos casos, los valores en las celdas guardan una dependencia no solo con los puntos que contienen la variable interpolada, sino también con otras variables de las que puede o no disponerse de información:

- La temperatura, se encuentra influenciada por la elevación.
- Las características del suelo tendrán relación con parámetros del relieve tales como la pendiente o índices relacionados con la humedad topográfica.

Estas variables, que denominamos variables de apoyo o predictores pueden incorporarse como datos de partida a algunos métodos de interpolación, aumentando así su exactitud.

En general, se requiere que dichas variables de apoyo estén recogidas en formato ráster.

También, pueden plantearse análisis de regresión mediante los cuales, y sin necesidad de utilizar la componente espacial, puedan estimarse los valores en las celdas problema a partir de los valores de la variable en los puntos conocidos y los valores de los predictores tanto en dichos puntos como en las celdas a llenar.

CASO SENCILLO:

	20				30
		?			
	20				30

La lógica nos indica que el valor que buscamos ha de ser inferior a 30 y más cercano a 20 ya que la celda problema se sitúa más cerca de ese valor medio.

Razonando de este modo, aplicamos el hecho de que la proximidad incrementa la semejanza de valores. Es decir, que existe autocorrelación espacial para la variable interpolada.

CASO REAL

	28				15
		?			
	3				17

No parece tan sencillo “adivinar” el valor que corresponde:

La autocorrelación espacial de la variable, no resulta tan obvia.

Utilizando distintos métodos de interpolación, puede ser que obtengamos valores distintos.

1.2 Métodos de interpolación

Existen diferentes clasificaciones para distinguir los distintos métodos de interpolación:

- Según los puntos considerados para el cálculo:
 - Métodos Globales: consideran que todos los puntos de los que disponemos tienen influencia sobre el valor a calcular.
 - Métodos Locales: sólo consideran un conjunto restringido de los datos.
- Según el valor en los puntos de partida:
 - Métodos Exactos: los valores que la interpolación asigna a los valores de los puntos departidas son exactamente los mismos.
 - Métodos Aproximados: los valores que la interpolación les asigna es el que corresponde con el mejor ajuste del método.
- Según la inclusión o no de elementos probabilísticos:
 - Métodos Simples.
 - Métodos Geoestadísticos.

Si aludimos a esta última clasificación, hablaremos de **interpolación simple** cuando la interpolación se basa en un principio de las ciencias naturales del cual se deriva la continuidad de los datos donde la razón de cambio es constante. El ejemplo más claro es el de la interpolación de elevaciones del terreno.



Por el contrario, la **interpolación Geoestadística** se utiliza cuando hay variación irregular en los datos. Algunos métodos geoestadísticos proporcionan estimaciones probabilísticas de la calidad de la interpolación (validación cruzada) y una herramienta (semivariograma o variograma) que permite explorar los datos.

1.3 Interpolación Geoestadística

La interpolación Geoestadística se basa en los siguientes supuestos:

✓ **Estacionariedad de la variable:**

Se refiere a que la estructura probabilística del campo aleatorio es similar en diferentes partes del área de estudio, es decir que el proceso alcanza un estado de equilibrio. Esto ocurre cuando las coordenadas absolutas no muestran ninguna influencia sobre la ocurrencia de la variable.

✓ **Hipótesis intrínseca:**

Una función aleatoria $f(x)$ se dice intrínseca cuando:

- a) Su esperanza matemática existe y no depende de la localización.
- b) Para todo vector h el incremento $[f(x+h) - f(x)]$ tiene varianza finita y no depende de la localización x_i .

✓ **Distribución de probabilidad de los datos:**

Si la variable modelada se comporta como una distribución normal, el método será más exacto.

La interpolación está basada en el principio de la autocorrelación espacial, que mide el grado de relación entre los objetos cercanos y lejanos.

La **autocorrelación** espacial determina si los valores están interrelacionados. Si lo están, determina si hay algún patrón espacial que defina dicha relación.

La correlación se utiliza para medir:

- Similaridad de objetos dentro de un área.
- El grado en el que un fenómeno espacial es correlado consigo mismo en el espacio.
- El nivel de interdependencia entre las variables.
- La naturaleza y fuerza de la interdependencia.

Cada método de interpolación producirá resultados diferentes.

Los métodos de interpolación geoestadística pueden ser:

- **Determinísticos:**

Métodos estáticos donde las entradas producen con certeza salidas únicas. Su planteamiento está basado en la interacción de variables independientes que determinan el comportamiento de una o algunas variables dependientes.

Entre ellos se encuentran el *IDW*, *Spline* y *Vecinos naturales*.

- **Estocásticos:**

Son métodos que describen la forma que adquiere un conjunto de datos observados cuyo comportamiento se supone aleatorio. Los métodos estocásticos, representan tanto fenómenos compuestos por eventos perfectamente predecibles como por otros eminentemente aleatorios.

El método por excelencia de este grupo es el *Kriging*.

1.3.1. Método de interpolación IDW

La interpolación IDW (Ponderación inversa de la distancia) es un método exacto de interpolación que estima los valores de la celda promediando los valores de los datos a partir de los valores de sus vecinos. Parte de la afirmación de que los puntos más cercanos tienen más influencia que los más lejanos.

$$\hat{z}_j = \sum_{i=1}^n k_{ij} \cdot z_i$$

Donde:

\hat{z}_j es el valor estimado de la magnitud a interpolar para el punto j .

n es el número de puntos usados en la interpolación.

z_i el valor en el punto i -ésimo.

k_{ij} el peso asociado al dato i en el cálculo del nodo j .

Los pesos k varían entre 0 y 1 para cada dato, y la suma total de ellos es la unidad.

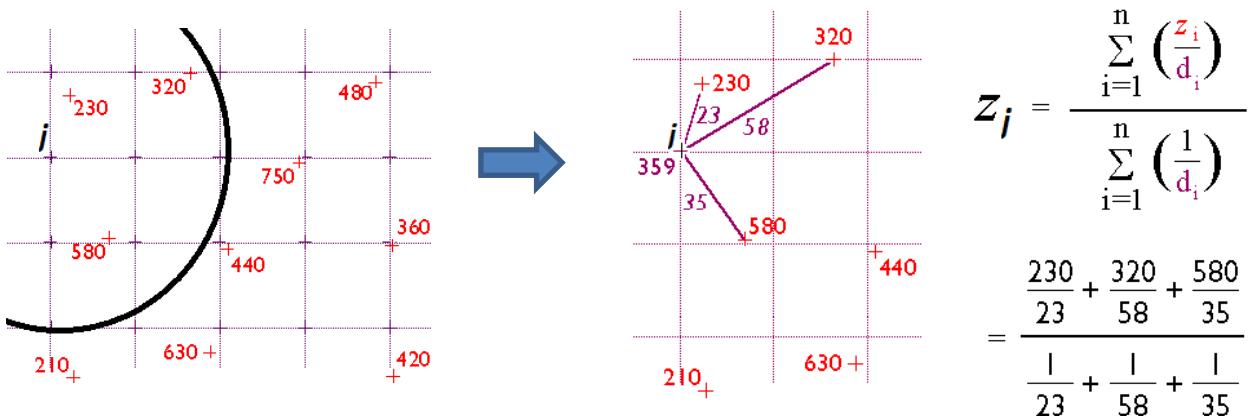
Para establecer una función de proporcionalidad entre el peso y la distancia, la fórmula general queda como sigue:

$$\hat{z}_j = \frac{\sum_i \frac{z_i}{d_{ij}^\beta}}{\sum_i \frac{1}{d_{ij}^\beta}}$$

Donde:

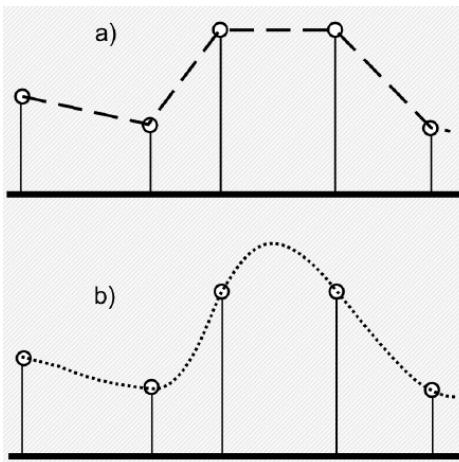
$\hat{z}_j = 1/d_{ij}^\beta$ y β es un exponente de ponderación que controla la forma en la que el peso disminuye con la distancia.

Un ejemplo:



La interpolación IDW permite la generación de información ráster de una forma rápida y simple. Sin embargo, se trata esencialmente de una media ponderada y, por tanto, el resultado se encuentra siempre incluido dentro del rango de variación de los datos.

De no recogerse en la muestra los valores extremos (tales como cimas o valles), estos desaparecen y son sustituidos por tramos llanos. El resultado deriva en un «aplanamiento» de la superficie y la aparición de falsas terrazas.



El método IDW solo tiene en cuenta el alejamiento, pero no la posición. Es decir, un punto situado a una distancia d hacia el Norte tiene la misma influencia que uno situado a esa misma distancia d , pero hacia el Oeste.

La calidad del resultado de interpolación puede disminuir si la distribución de los puntos de datos de la muestra es desigual. La interpolación IDW debe ser usada cuando el conjunto de puntos es lo suficientemente denso para cubrir toda la zona.

1.3.2. Método de interpolación Spline

Los métodos Spline estiman los valores utilizando una función matemática que minimiza la curvatura total de la superficie. Como resultado se obtiene una superficie suave que pasa exactamente por los puntos por los que se forma (Método Exacto).

Puede predecir valles y crestas en los datos, siendo un método idóneo para representar variaciones suaves de datos tales como temperaturas.

Desde un punto de vista matemático, los splines son funciones polinómicas por tramos, de tal modo que, en lugar de emplear un único polinomio para ajustar a todo un intervalo, se emplea un polinomio distinto para cada tramo.

Este tipo de interpolación puede alcanzar valores fuera del rango definido por los puntos de partida.

No obstante, en zonas con cambios bruscos de valores (como, por ejemplo, dos puntos de entrada cercanos, pero con valores muy diferentes), pueden presentarse oscilaciones artificiales significativas. Para solucionar este problema, una solución es el empleo de splines con tensión.

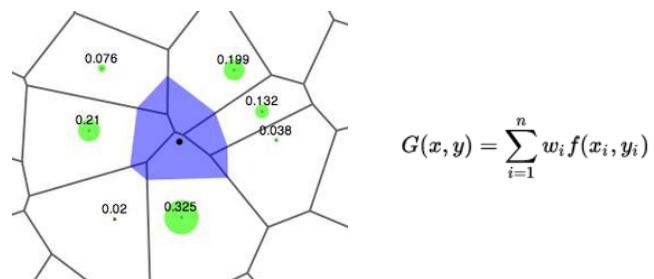
Existen variaciones del método tales como B-splines (Basic splines) o Plate-splines que se asemeja a una flexión que tiene una hoja delgada de metal, de tal manera que el ajuste se resiste más a la flexión penalizando más la suavidad de la superficie.

1.3.3. Método de Vecino natural

La interpolación por vecindad no es adecuada para el trabajo con variables continuas, pero sí para variables categóricas.

Por ejemplo, para un conjunto de puntos cada uno de los cuales esté identificado con un código numérico, la interpolación por vecindad de ese valor da como resultado una capa donde los valores de las celdas identifican el punto más cercano. Esto puede utilizarse para calcular la influencia de cada uno de ellos en el espacio representado.

Se basa en la teselación de Voronoi, que consiste en unir los puntos entre sí, trazando las mediatrices de los segmentos de unión. Las intersecciones de estas mediatrices determinan una serie de polígonos en un espacio bidimensional alrededor de un conjunto de puntos de control, de manera que el perímetro de los polígonos generados sea equidistante a los puntos vecinos y designan su área de influencia.



Los denominados polígonos de Thiessen/Voronoi son de uso habitual en el análisis climatológico/hidrológico asociando a cada zona los valores de la estación meteorológica más cercana.

El resultado es una capa con saltos abruptos (tanto como lo sean las diferencias entre los valores de puntos cercanos), con un aspecto «atarrazado».

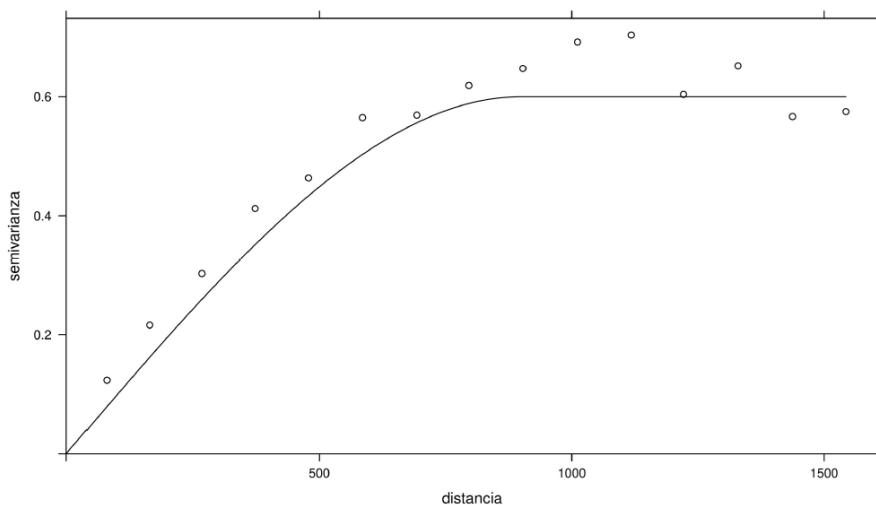
El conjunto de celdas con el mismo valor (dentro de la misma terraza) representa el lugar geométrico de las celdas cuyo punto más cercano de entre los de partida es uno dado.

1.3.3. Interpolación Kriging

Es un método de interpolación muy potente que se utiliza en diversas aplicaciones como ciencias de la salud, ciencias ambientales, etc.

Asume que la distancia o dirección entre los puntos de muestra refleja una correlación espacial que puede ser utilizada para explicar la variación de la superficie total.

Ajusta una función a un número de puntos específicos o a todos los puntos dentro de un radio específico para determinar el valor de la salida para cada localización.



Este método es más apropiado cuando se conoce la distancia de correlación espacial o su dirección, por ello es muy utilizado en ciencias de la tierra y geología.

Los valores pronosticados no tienen por qué estar entre los valores de las muestras y su superficie no pasa por los valores de la muestra.

Junto con la superficie interpolada, el kriging genera asimismo superficies con medidas del error de interpolación (validación cruzada) que pueden emplearse para conocer la bondad de esta en las distintas zonas.

La interpolación óptima debe cumplir los siguientes requisitos, que son cubiertos por el método kriging.

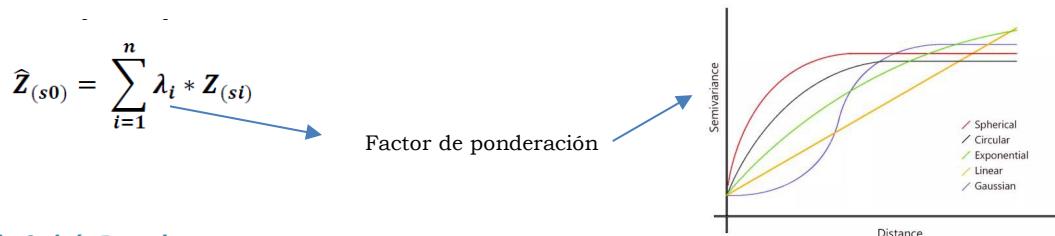
- El error de predicción debe ser mínimo.
- Los puntos cercanos deben tener pesos mayores que los lejanos.
- La presencia de un punto cercano en una dirección dada debe restar influencia (enmascarar) a puntos en la misma dirección, pero más lejanos.
- Puntos muy cercanos con valores muy similares deben «agruparse» de tal forma que no aparezca sesgo por sobremuestreo.
- La estimación del error debe hacerse en función de la estructura de los puntos no de los valores.

En su expresión fundamental, el kriging es semejante a un método basado en ponderación por distancia.

$$\hat{Z}_{(s0)} = \sum_{i=1}^n \lambda_i * Z_{(si)}$$

El método de Kriging supone que, cuando disponemos de un punto con un valor conocido para una determinada variable, la variación de ésta será menor en el entorno más cercano a dicho punto, sin embargo, conforme las distancias se incrementen respecto al punto conocido, la variación de la variable será mucho mayor.

El cálculo de la ponderación se realiza de forma más compleja que en la ponderación por distancia de IDW, ya que en lugar de utilizar dichas distancias se acude al análisis de la autocorrelación a través del variograma teórico (Variogramas).



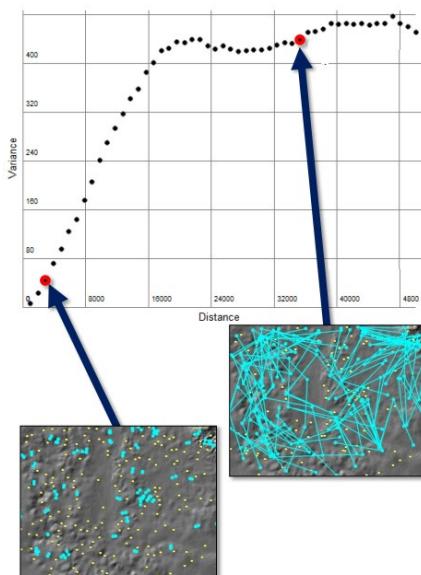
Por ello se requiere, asimismo, que exista un número suficiente de puntos (mayor de 50) para estimar correctamente el variograma.

A partir de los valores del variograma, se estima un vector de pesos que, multiplicado por el vector de valores de los puntos de influencia, da el valor estimado.

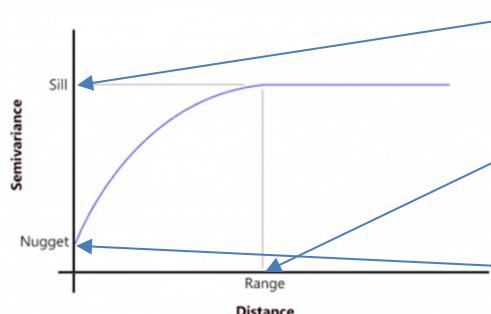
Variograma: es una función matemática que mide la variabilidad, es decir, la desemejanza de una variable, cuando sus valores se observan en puntos que están separados una distancia concreta.

Además de como expresión matemática, se puede representar su gráfica generando las parejas de puntos con los datos reales separados en 1, 2, 3, etc. unidades de distancia a lo largo de una dirección. En el caso de un variograma, las cosas más cercanas son más predecibles y tienen menos variabilidad. Mientras que las cosas distantes son menos predecibles y están menos relacionadas (correlación espacial).

Para comprender la dependencia espacial, los variogramas toman 2 ubicaciones de muestra y miden la distancia entre ambos puntos. En el eje X, se traza la distancia en retardos (lags), que no son más que distancias agrupadas. Tomando cada conjunto de 2 ubicaciones de muestra, mide la varianza entre la variable que estamos midiendo y la representa en el eje Y. En teoría tendríamos que considerar todas las distancias entre 2 muestras y su variabilidad (combinación de todos los puntos tomados de dos en dos). Para no representarlos todos, se hacen agrupaciones con características similares.



En puntos de muestreo con distancias cercanas, la diferencia de valores entre puntos tiende a ser pequeña. En otras palabras, la semivarianza es pequeña. Pero cuando las distancias de los puntos de muestreo son más alejadas, es menos probable que sean similares. Esto significa que la semivarianza se hace grande. A medida que aumenta la distancia entre los puntos de muestra, deja de haber relación entre ellos. Su varianza empieza a aplanarse y los valores de las muestras no están relacionados entre sí.



Meseta: valor en el que el modelo se aplana por primera vez. Coincide con la varianza estadística de la variable.

Alcance: distancia a la que el modelo se aplana por primera vez. Coincide con la distancia máxima sobre el eje donde existe autocorrelación espacial.

Nugget o pepita: valor en el que el semivariograma (casi) intercepta el valor Y.

Cuando se tienen dos puntos de muestra en el mismo lugar, se puede esperar que tengan el mismo valor, por lo que la pepita debería ser cero. A veces no es así y esto añade aleatoriedad. Pero antes de que el gráfico empiece a nivelarse, estos valores están espacialmente autocorrelacionados.

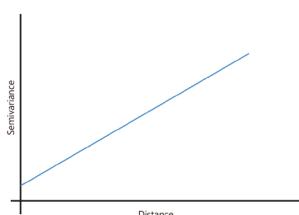
Como era de esperar, cuando aumenta la distancia, aumenta la varianza. Hay menos pares de puntos separados por distancias lejanías, de ahí la menor correlación entre los puntos de la muestra.

Pero como se indica en el variograma, con la meseta y el alcance, empieza a alcanzar su nivel plano, asintótico. Es entonces cuando se intenta ajustar una función para modelizar este comportamiento.

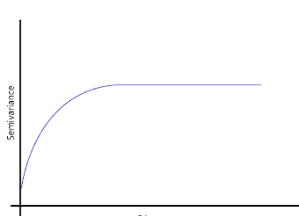
Es importante acotar que la definición de un variograma se ajusta siempre a una dirección específica.

- Si el variograma es similar en diferentes direcciones se puede llamar un variograma isotrópico.
- Si se detectan variaciones bruscas en la definición del variograma en diferentes direcciones corresponde entonces a un variograma anisotrópico en el cual además de la distancia, la dirección influye en la autocorrelación espacial de una variable.

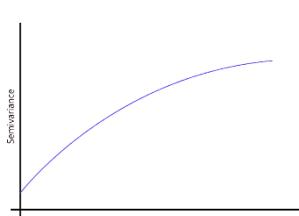
La forma correcta de proceder es seleccionar el tipo de modelo que mejor se ajuste a los datos y que proporcionará una función matemática con relación entre valores y distancias. Habitualmente se utilizan funciones como la exponencial, la lineal, la esférica, la gaussiana, etc.:



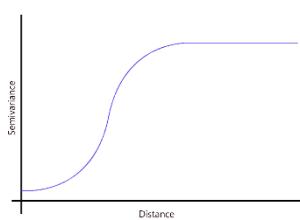
Modelo Lineal: Un modelo lineal significa que la variabilidad espacial aumenta linealmente con la distancia. Es el tipo de modelo más simple, sin meseta, lo que significa que el usuario tiene que seleccionar arbitrariamente el umbral y el rango.



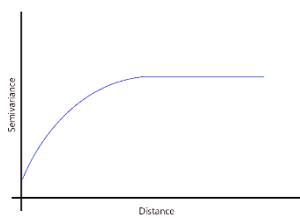
Modelo Esférico: El modelo esférico es uno de los modelos más comunes que utilizamos en la modelización de variogramas. Se trata de una ecuación cuadrática modificada en la que la dependencia espacial se aplana con el umbral y el rango.



Modelo Exponencial: El modelo exponencial se parece al modelo esférico en que la variabilidad espacial alcanza el umbral gradualmente. La relación entre dos puntos de muestra decae gradualmente, mientras que a una distancia infinita la dependencia espacial se disipa.



Modelo Gaussiano: La función gaussiana utiliza una curva de distribución de probabilidad normal. Este tipo de modelo es útil cuando los fenómenos son similares a distancias cortas debido a su progresivo ascenso por el eje Y.



Modelo Circular: Este tipo de modelo de predicción utiliza una función circular para ajustar la variabilidad espacial en un variograma. Se asemeja a la función del modelo esférico, donde la dependencia espacial se desvanece en su nivel asintótico.

Los variogramas constituyen un paso preliminar útil para comprender la naturaleza de los datos. Cada fenómeno tiene su propio variograma y su propia función matemática. El usuario descubre la relación entre valores y distancias y, a continuación, elige el modelo que mejor se ajusta. Aunque los variogramas son útiles para comprender la variación con la distancia, el modelo que se elige a partir de los variogramas suele utilizarse en el kriging.

Existen dos tipologías de kriging:

- **Kriging ordinario**

La aplicación del kriging ordinario implica la asunción de una serie de características de los datos:

- La variación es aleatoria.
- Existe dependencia espacial.
- El proceso espacial subyacente es intrínsecamente estacionario con media constante y varianza que depende solo de la separación en distancia entre los sitios y no de su posición.

- **Kriging universal**

La suposición de estacionariedad intrínseca no se cumple cuando existen tendencias geográficas pronunciadas de naturaleza sistemática y no aleatoria. La tendencia puede ser regional, es decir, una variación sistemática en toda la región de interés o local de un punto a otro dentro de la región estudiada. La existencia de tendencias puede ser explorada graficando los datos de la variable analizada en función a la variable que se supone genera la tendencia espacial. La tendencia también se manifiesta en los variogramas experimentales con un incremento de la semivarianza con la distancia que no tiene límites.

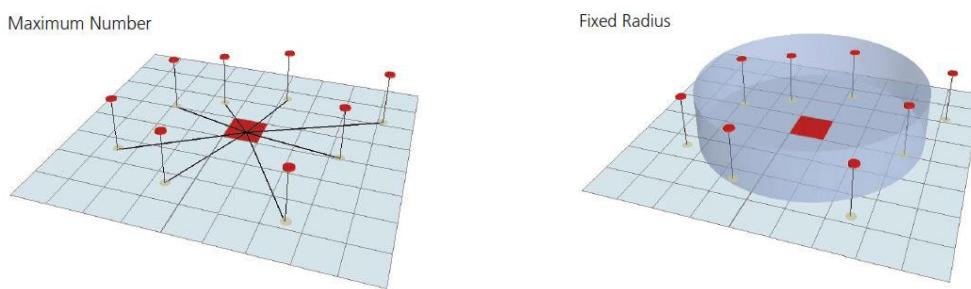
La exactitud del kriging depende de varios factores:

- El número de muestras tomadas.
- La calidad de la medición en cada punto.
- Las ubicaciones de las muestras en la zona si las muestras son igualmente espaciadas se alcanza una mejor cobertura.
- Las distancias entre las muestras para la predicción es mejor usar muestras vecinas que muestras distantes esto es, la precisión mejora cuando la cercanía de las muestras aumenta, y se deteriora cuando esta disminuye.

- La continuidad espacial de la variable o atributo en estudio es más fácil estimar el valor de una variable bastante regular en una región que una que presenta grandes fluctuaciones.

1.4 Otros parámetros de la interpolación

Existen otros factores que pueden influir en el resultado final de la interpolación, tales como el número de puntos utilizados en la interpolación de cada localización. Así podemos limitar en función de un máximo número de puntos o en función de una distancia dada. Esta última suele ser la más utilizada, cuando se sabe el alcance de la variable a interpolar. Por ejemplo, se sabe que los granos de polen no viajan más de ciertos km, ese límite será muy importante a la hora de configurar la interpolación, porque se sabe de antemano que la estación con datos no puede recoger polen de localizaciones más alejadas de esa distancia.



Otro parámetro que puede constreñir en gran medida los resultados es la existencia o no de barreras en la interpolación. Aunque las barreras parecen ser más usuales en la interpolación de elevaciones del terreno, también podrían ser aplicadas a otra serie de variables a estimar.



Las barreras crean discontinuidades lineales en las superficies. Los métodos IDW y Splines soportan el uso de barreras.

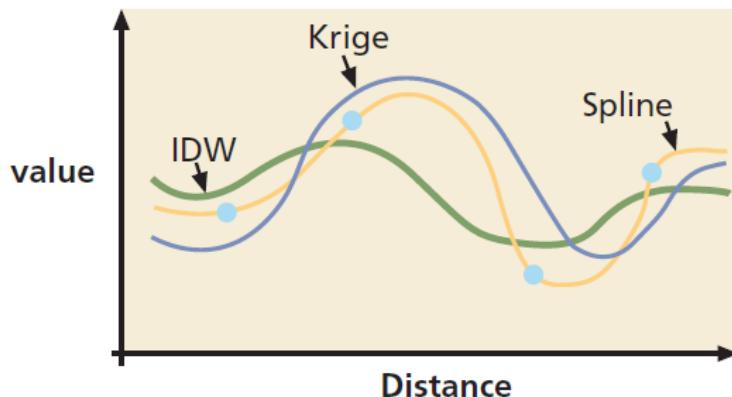
1.5 Elección del método de interpolación

No existe un método universalmente establecido como más adecuado en todas situaciones y la elección se ha de fundamentar en diversos factores. Al mismo tiempo, un método puede ofrecer resultados muy distintos en función de los parámetros de ajuste, con lo que no solo se ha de elegir el método adecuado, sino también la forma de usarlo.

Los factores a tener en cuenta para la elección son:

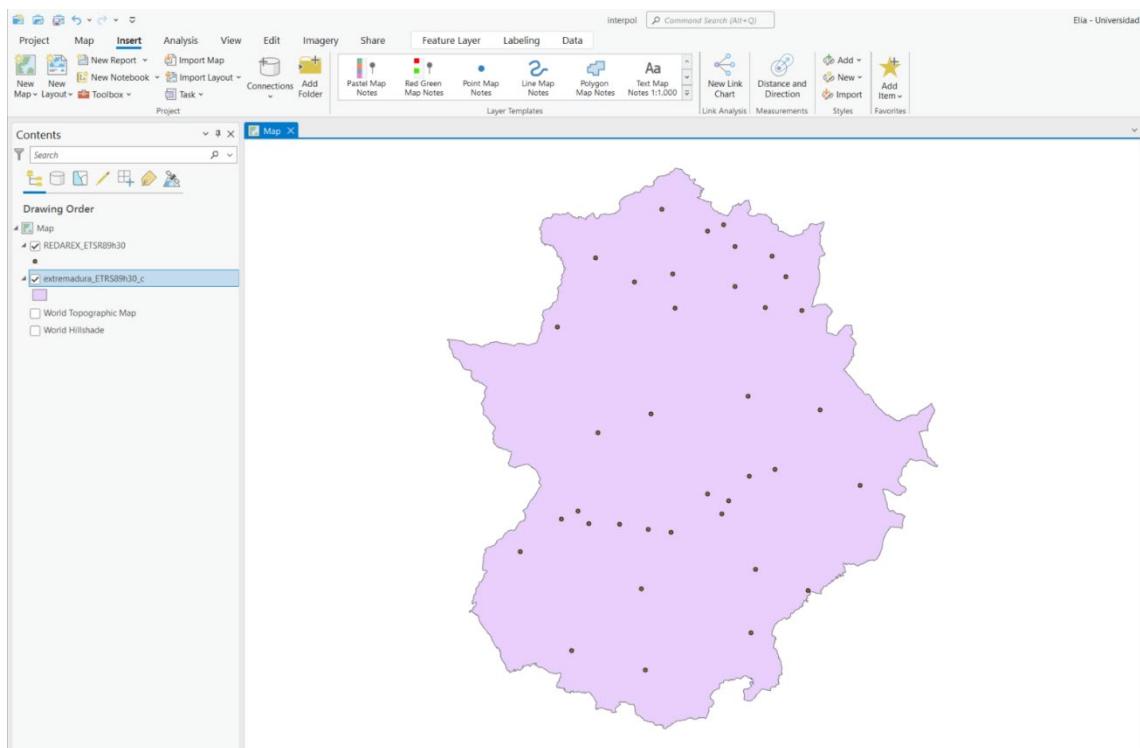
- Las características de la variable a interpolar: Si interpolamos valores de precipitación máxima anual, no es adecuado utilizar aquellos métodos que suavicen excesivamente.
- Las características de la superficie a interpolar: Si necesitamos introducir barreras, Kriging no las contempla.
- La calidad de los datos de partida: Cuando los datos de partida son de gran exactitud, los métodos exactos pueden tener más interés.

- El rendimiento de los algoritmos: IDW es el más simple y rápido y Kriging el más sofisticado y lento.
- El conocimiento de los métodos: Kriging exige una sólida base de conceptos geoestadísticos para su aplicación.

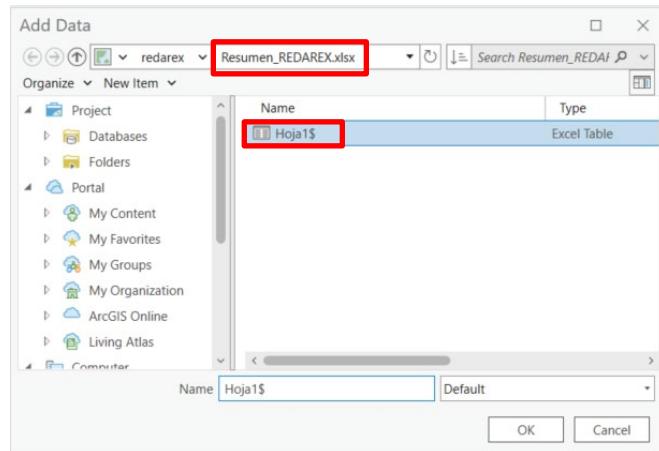


2. Interpolación en ArcGIS

En primer lugar, cargaremos las capas del contorno de Extremadura y las estaciones de REDAREX.



Ahora, añadiremos la tabla con los datos climatológicos de cada estación, para hacer la unión.

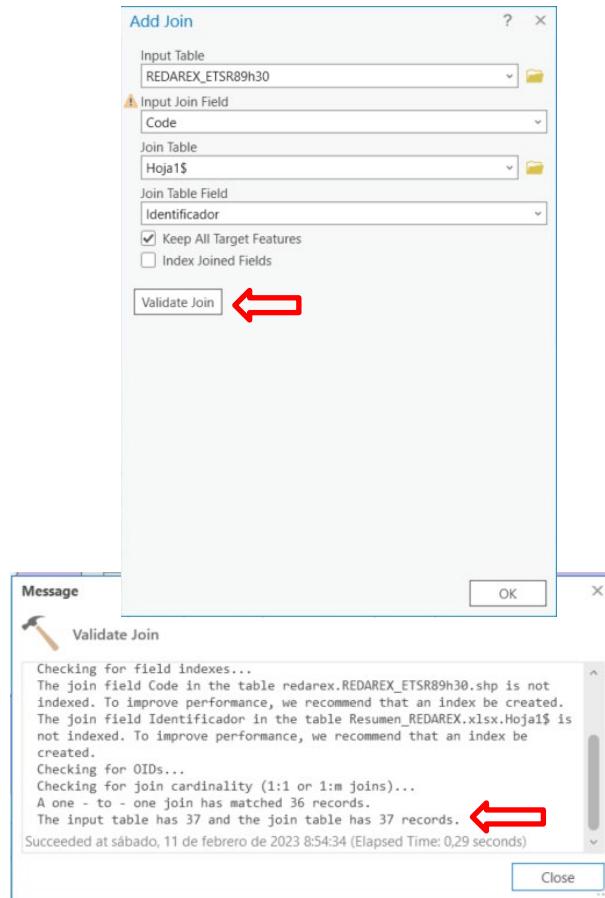


Abrimos la tabla de atributos y comprobamos que la importación se ha realizado correctamente:

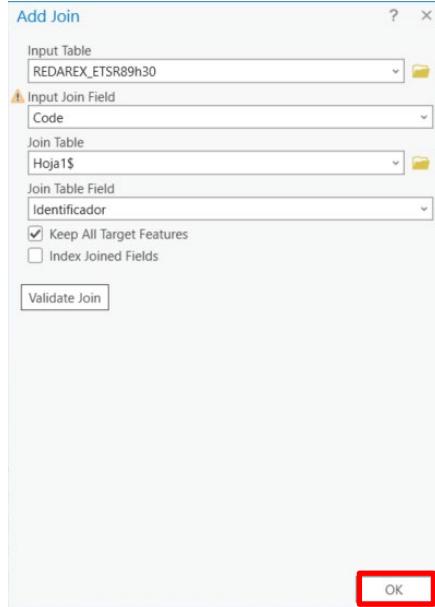
	F1	Estación	Identificador	t_max	t_med	t_min	preci	oscial_temp	corr_preci-tempo	R2	F11	F12	ObjectID
1	1	Zalamea de la Serena	BA01	24.49	18.07	11.68	307.07	12.82	-0.191885	0.03682	<Null>	<Null>	1
2	2	Monterrubio de la Serena	BA02	24.24	17.58	10.99	357.41	13.25	-0.202694	0.041085	<Null>	<Null>	2
3	3	Villagonzalo	BA04	25.02	17.96	11.08	456.92	13.94	-0.092851	0.008621	<Null>	<Null>	3
4	4	Jerez de los Caballeros	BA05	24.97	16.7	8.75	410.68	16.22	-0.135374	0.018326	<Null>	<Null>	4
5	5	Olivenza	BA06	24.42	16.95	9.55	424.91	14.87	-0.133556	0.017837	<Null>	<Null>	5
6	6	Don Benito	BA08	24.76	17.72	10.87	316.76	13.89	-0.105943	0.011224	<Null>	<Null>	6
7	7	Villafranca de los Barros	BA09	24.52	17.84	11.4	261.59	13.12	-0.180854	0.032708	<Null>	<Null>	7
8	8	Mérida	BA101	24.73	18.27	12.05	362.4	12.68	-0.143735	0.02066	<Null>	<Null>	8
9	9	Azuaga	BA102	23.16	16.04	8.77	312.6	14.39	-0.175569	0.030825	<Null>	<Null>	9
10	10	Puebla de Alcocer	BA103	23.6	18.13	12.83	629.69	10.77	-0.206423	0.042611	<Null>	<Null>	10
11	11	Fuente de Cantos	BA104	22.73	16.92	11.36	320.46	11.37	-0.187023	0.034978	<Null>	<Null>	11
12	12	Don Benito-EFA	BA105	24.87	17.23	10.09	422.75	14.78	-0.097717	0.009549	<Null>	<Null>	12
13	13	Santa Amalia	BA106	24.86	17.81	11.24	443.2	13.62	-0.113991	0.012994	<Null>	<Null>	13

Por tanto, podremos ya hacer la unión:

Identificador	t_max	t_med	t_min
Zalamea de la Serena	BA01	24.49	18.07
Monterrubio de la Serena	BA02	24.24	17.58



Una vez, comprobado que la unión se va a realizar correctamente, procedemos a realizarla.



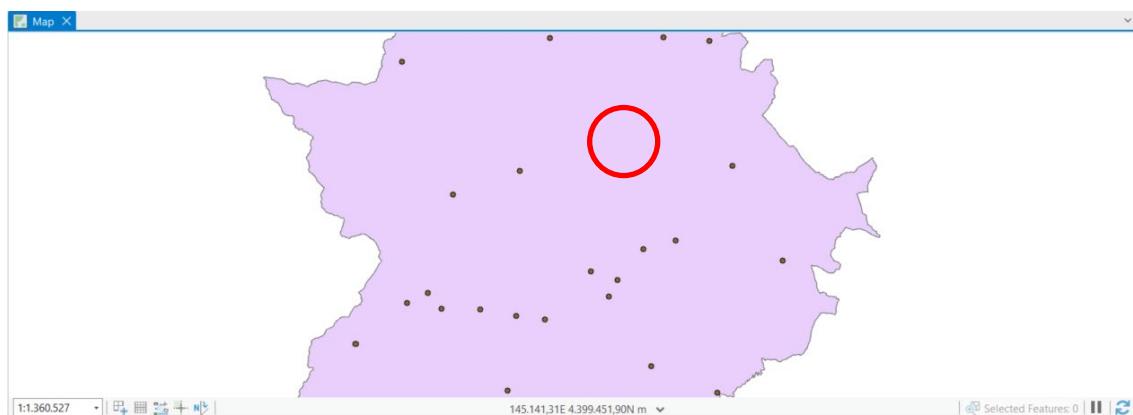
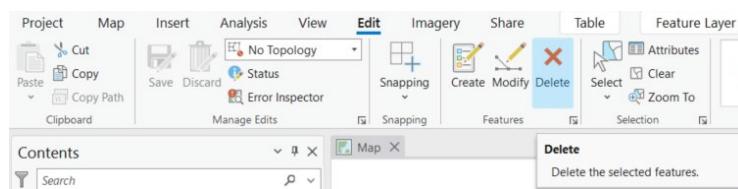
Al comprobar la nueva tabla de atributos del shp de las estaciones, vemos como hay una estación, la de madroñera, que no se ha hecho la unión.

REDAREX_ETRS89h30 X Hoja15

Field: Add Calculate Selection: Select By Attributes Zoom To Switch Clear Delete Copy

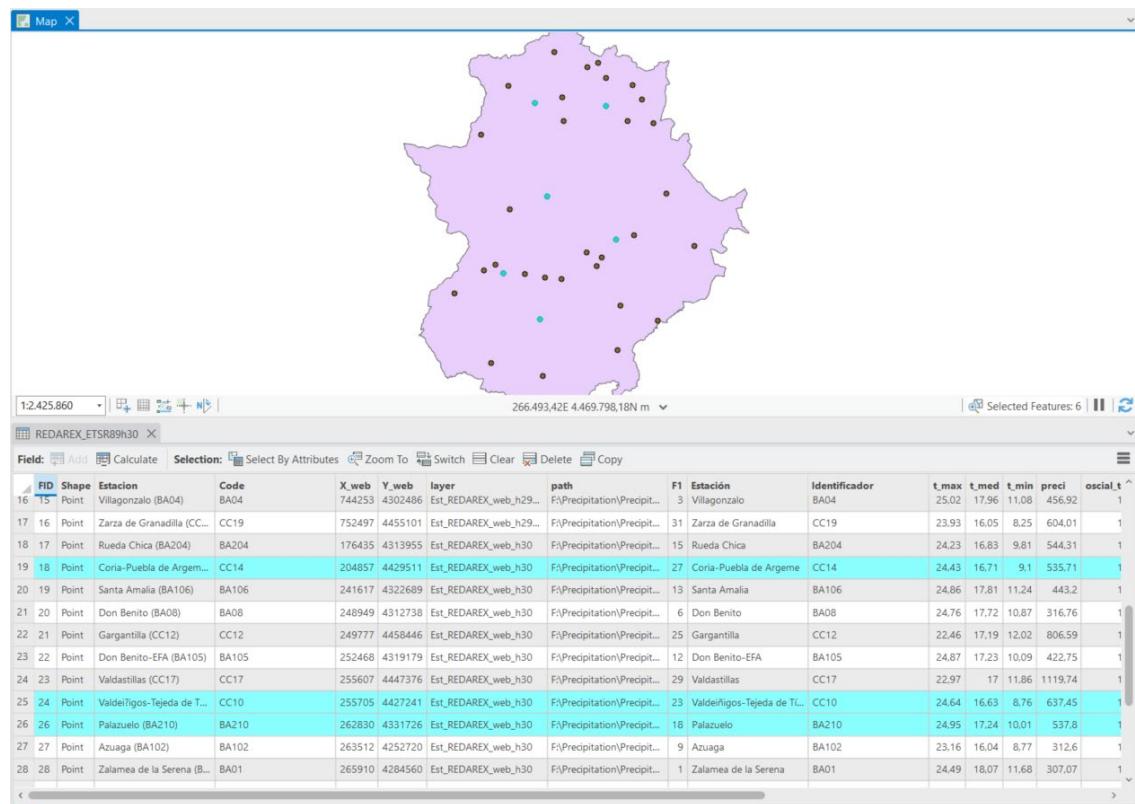
FID	Shape	Estación	Code	X_web	Y_web	layer	path	F1	Estación	Identificador	t_max	t_med	t_min	preci	c
17	16 Point	Zarza de Granadilla (CC...	CC19	752497	4455101	Est_REDAREX_web_h29...	F:\Precipitation\Precipit...	31	Zarza de Granadilla	CC19	23,93	16,05	8,21	604,01	
18	17 Point	Rueda Chica (BA204)	BA204	176435	4313955	Est_REDAREX_web_h30	F:\Precipitation\Precipit...	15	Rueda Chica	BA204	24,23	16,83	9,81	544,31	
19	18 Point	Coria-Puebla de Argem...	CC14	204857	4429511	Est_REDAREX_web_h30	F:\Precipitation\Precipit...	27	Coria-Puebla de Argeme...	CC14	24,43	16,71	9,1	535,71	
20	19 Point	Santa Amalia (BA106)	BA106	241617	4322689	Est_REDAREX_web_h30	F:\Precipitation\Precipit...	13	Santa Amalia	BA106	24,66	17,81	11,24	443,2	
21	20 Point	Don Benito (BA08)	BA08	248949	4312738	Est_REDAREX_web_h30	F:\Precipitation\Precipit...	6	Don Benito	BA08	24,76	17,72	10,87	316,76	
22	21 Point	Gargantilla (CC12)	CC12	249777	4458446	Est_REDAREX_web_h30	F:\Precipitation\Precipit...	25	Gargantilla	CC12	22,46	17,19	12,02	806,59	
23	22 Point	Don Benito-EFA (BA105)	BA105	252468	4319179	Est_REDAREX_web_h30	F:\Precipitation\Precipit...	12	Don Benito-EFA	BA105	24,87	17,23	10,09	422,75	
24	23 Point	Valdastillas (CC17)	CC17	255607	4447376	Est_REDAREX_web_h30	F:\Precipitation\Precipit...	29	Valdastillas	CC17	22,97	17	11,86	1119,74	
25	24 Point	Valdeiglesias-Tejeda de T...	CC10	255705	4427241	Est_REDAREX_web_h30	F:\Precipitation\Precipit...	23	Valdeiglesias-Tejeda de T...	CC10	24,64	16,63	8,76	637,45	
26	25 Point	Madroñera (CC102)	CC102	262355	4372008	Est_REDAREX_web_h30	F:\Precipitation\Precipit...	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	
27	26 Point	Palazuelo (BA210)	BA210	262830	4331726	Est_REDAREX_web_h30	F:\Precipitation\Precipit...	18	Palazuelo	BA210	24,95	17,24	10,01	537,8	
28	27 Point	Azuaga (BA102)	BA102	263512	4252720	Est_REDAREX_web_h30	F:\Precipitation\Precipit...	9	Azuaga	BA102	23,16	16,04	8,77	312,6	
29	28 Point	Zalamea de la Serena (B...	BA01	265910	4284560	Est_REDAREX_web_h30	F:\Precipitation\Precipit...	1	Zalamea de la Serena	BA01	24,49	18,07	11,68	307,07	

La razón es porque no se hicieron los cálculos estadísticos ya que su serie temporal no era tan extensa como el resto de estaciones. Por tanto, procedemos a borrar ese punto en concreto.

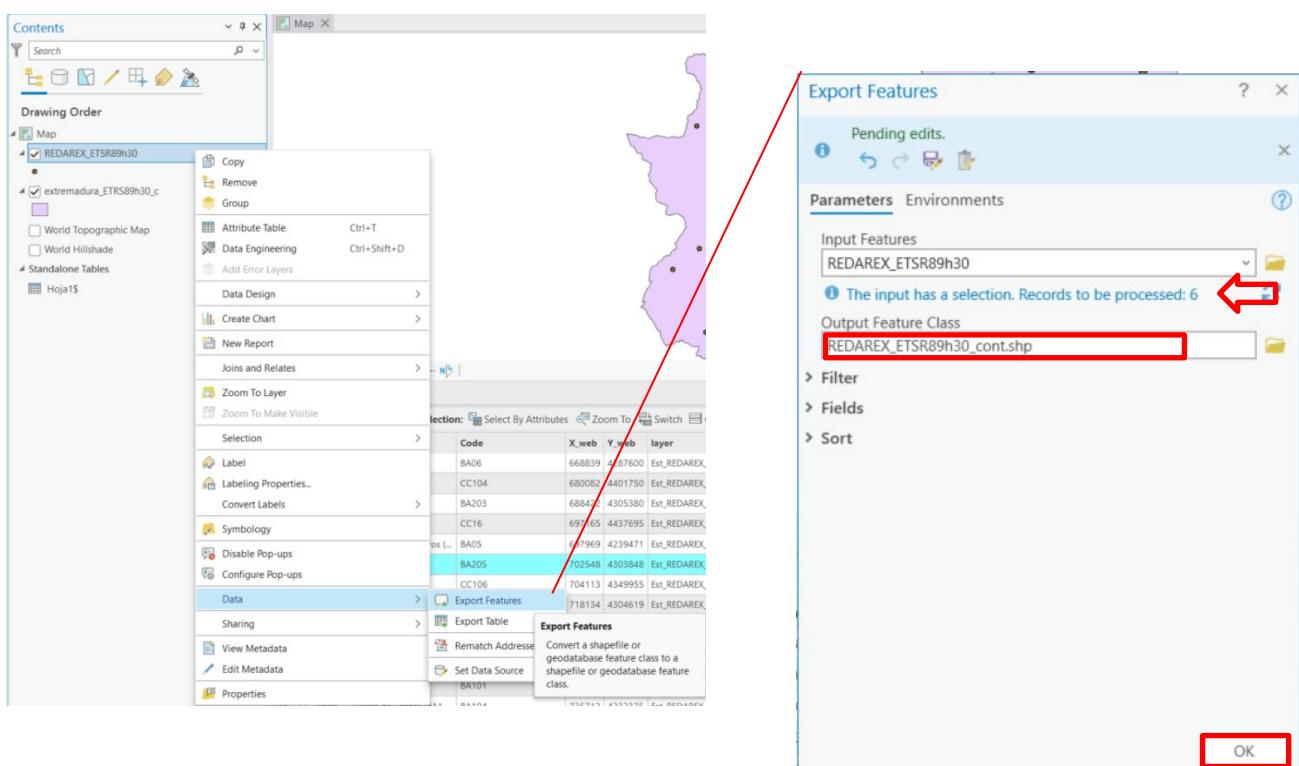


Ahora consolidaremos la unión exportando el archivo como un nuevo shp. Aprovecharemos y haremos una selección aleatoria, para guardar ciertos puntos y hacer con ellos los controles de error.

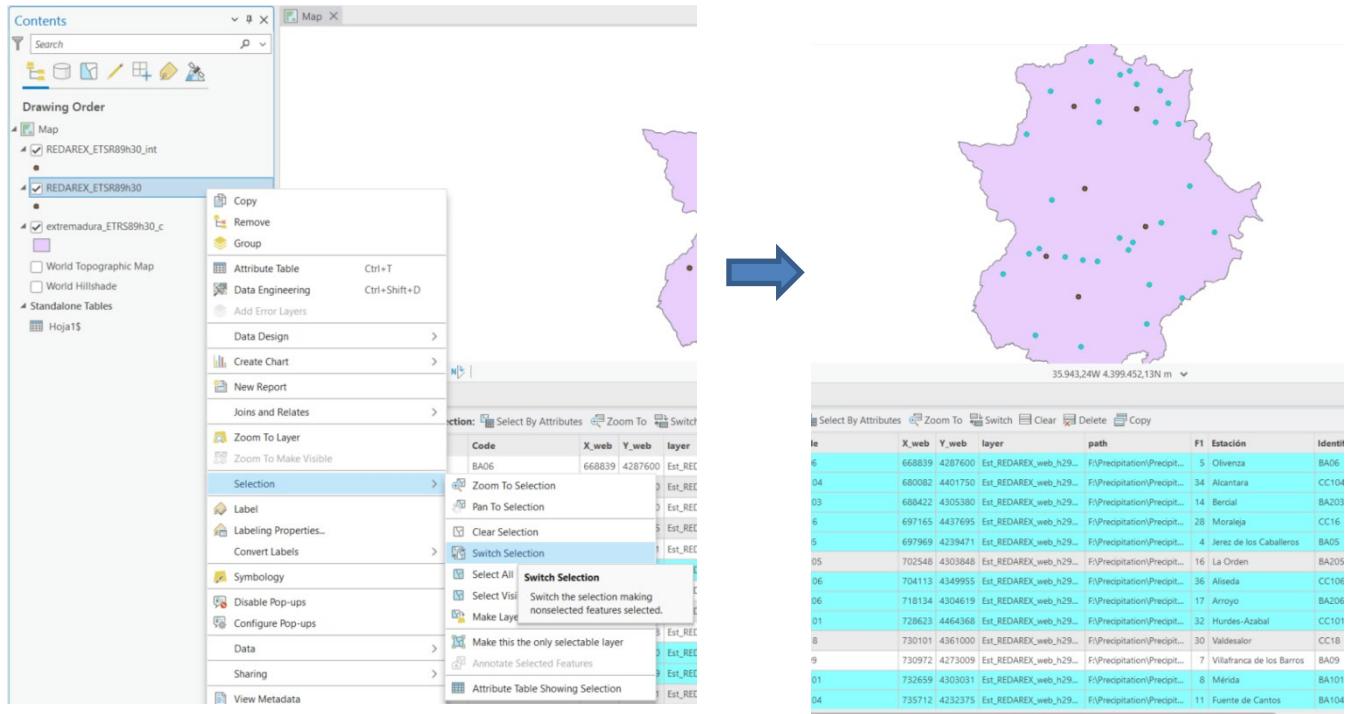
Para ello, vamos a seleccionar a mano (**cada estudiante seleccionará distintos puntos**) 3 puntos en la provincia de Cáceres y 3 en la de Badajoz, cumpliendo sólo la premisa de no estar en los límites geográficos.



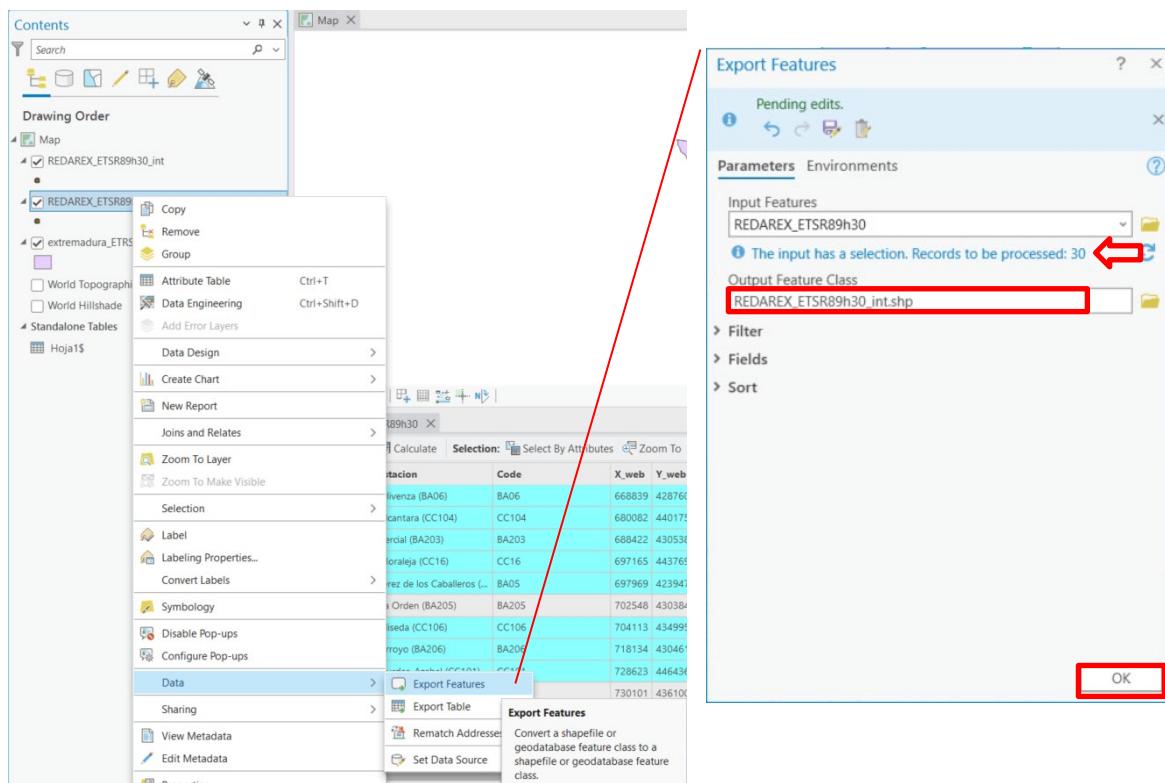
Ahora exportaremos esos puntos que nos servirán de control a un shp independiente:



Ahora invertiremos la selección y el resto de los puntos los guardaremos para hacer las interpolaciones.

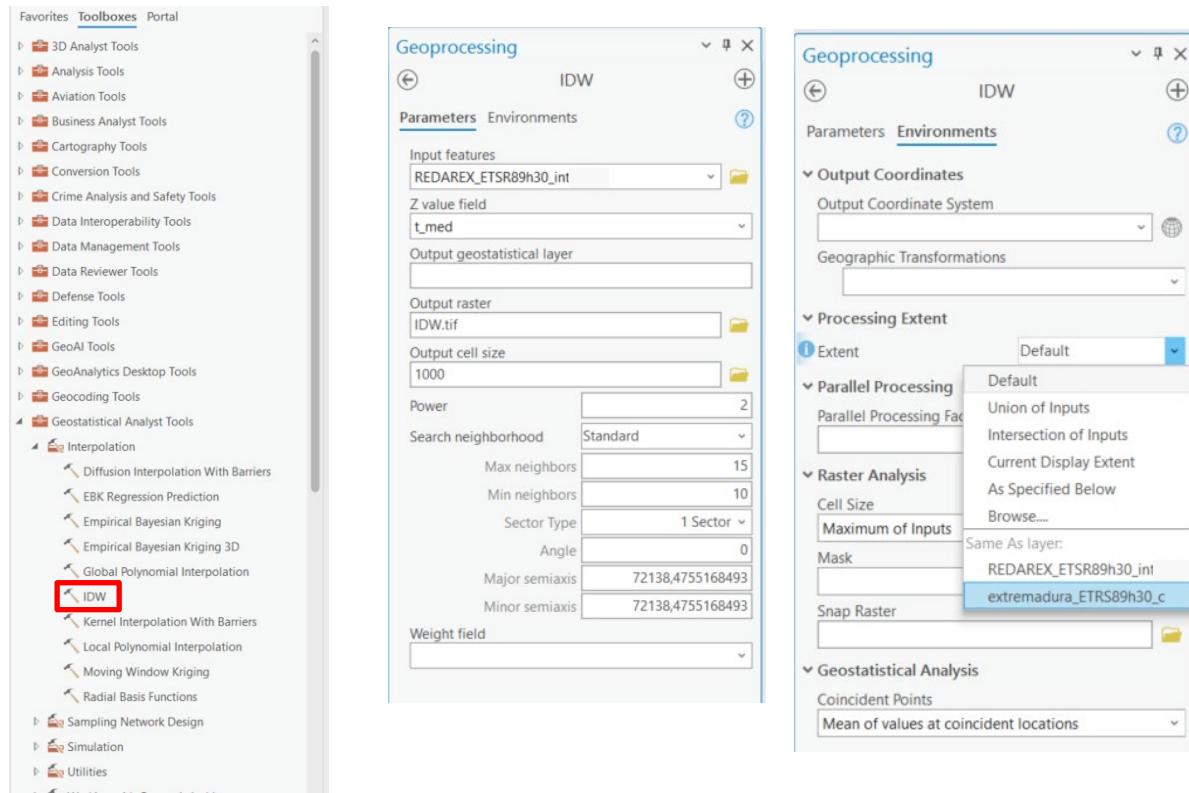


Y ahora exportamos los puntos para hacer las interpolaciones.

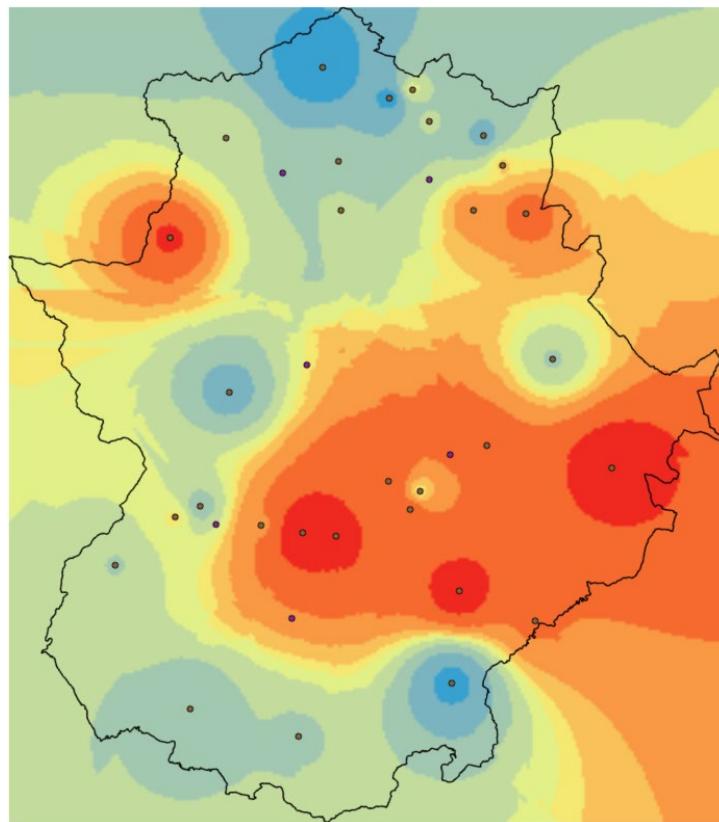


2.1 Interpolación IDW

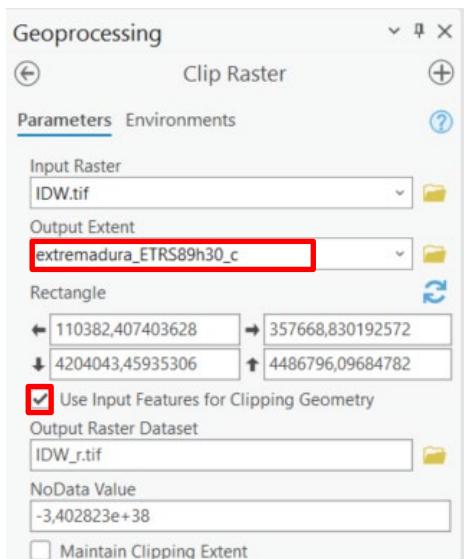
Para realizar una interpolación IDW, iremos la caja de herramientas de geoprocесamiento y en la caja de interpolación, ejecutaremos la herramienta IDW con los siguientes parámetros.



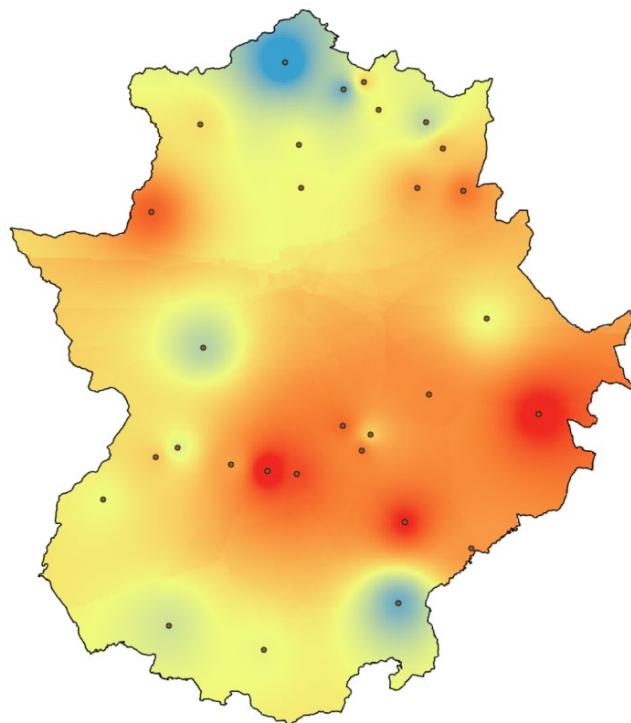
Obteniendo el siguiente resultado:



Procedemos a recortarlo por el contorno de Extremadura.

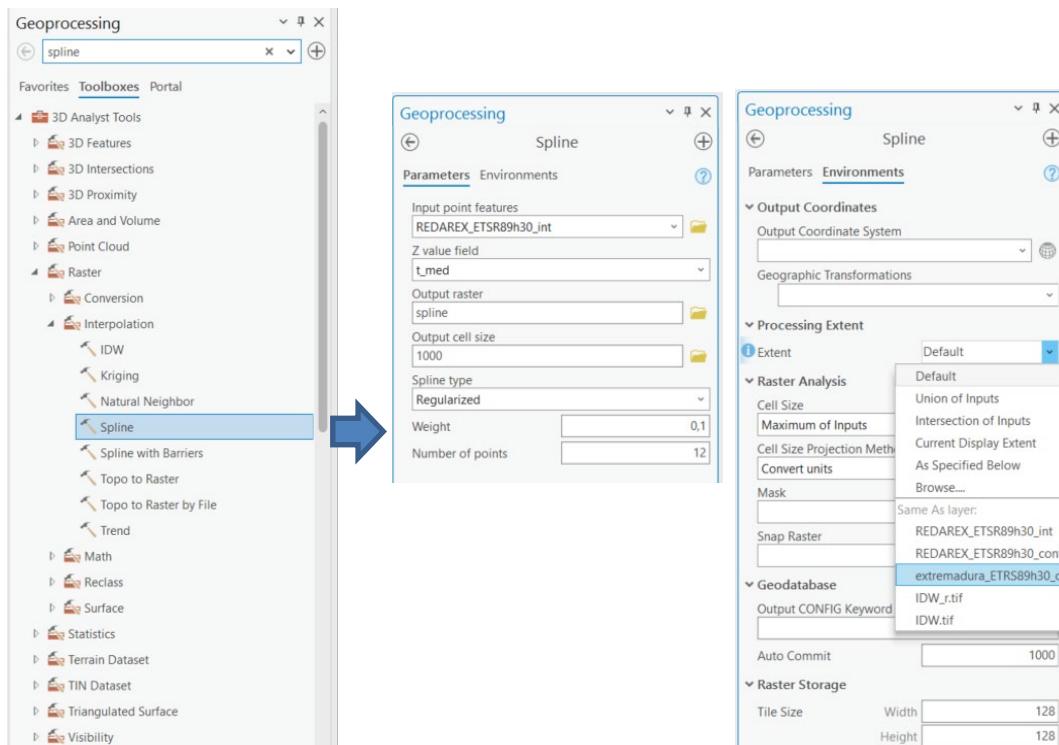


Y lo guardamos con el nombre IDW_r.img

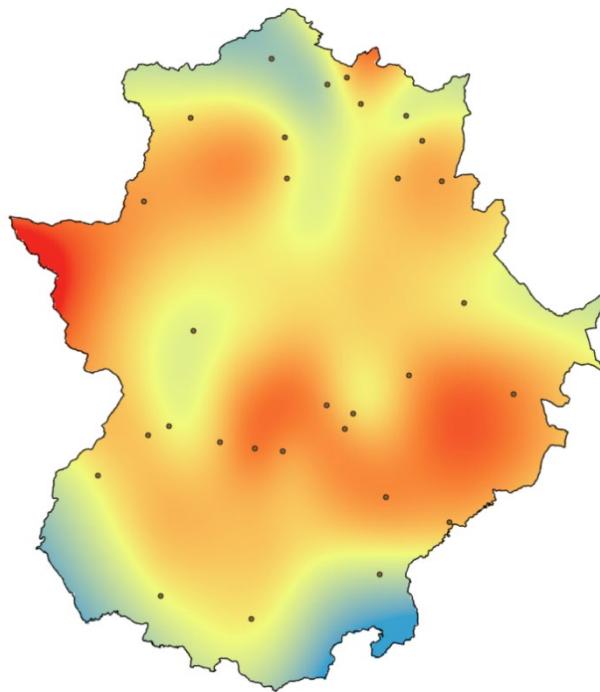


2.2 Splines

Esta vez iremos a la caja de herramientas de análisis 3D y, dentro del cajón de Ráster y de Interpolación, ejecutaremos la herramienta Spline.

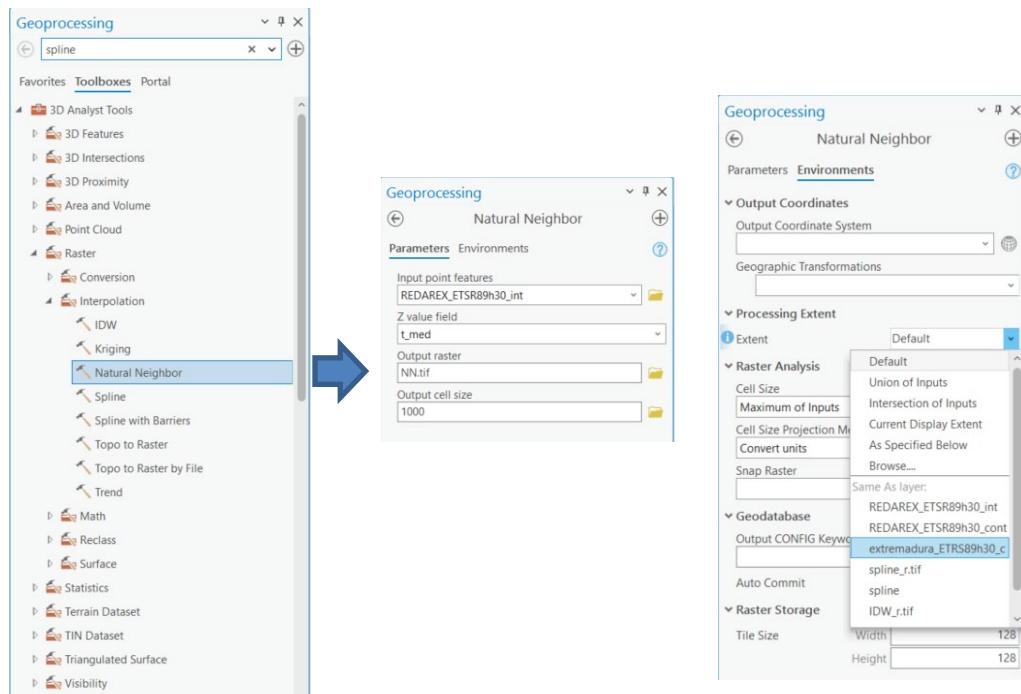


Obtendremos como resultado, una vez recortado, el siguiente ráster:

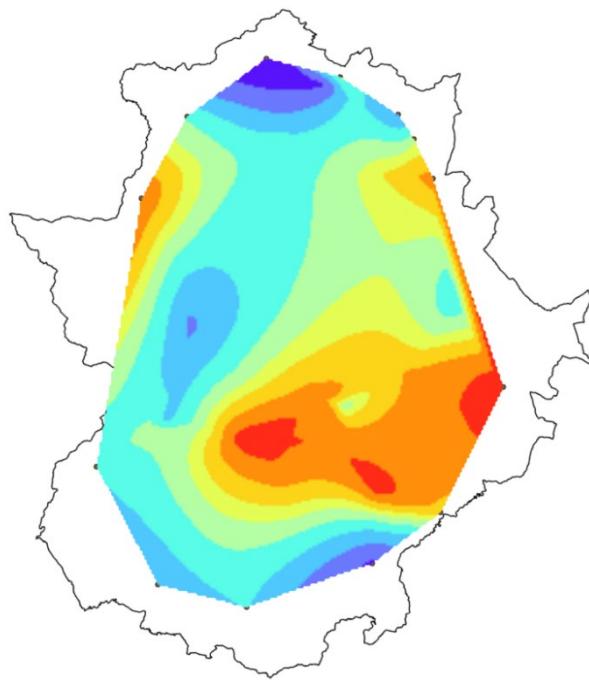


2.3 Vecino natural

Lo configuraremos de la siguiente forma:

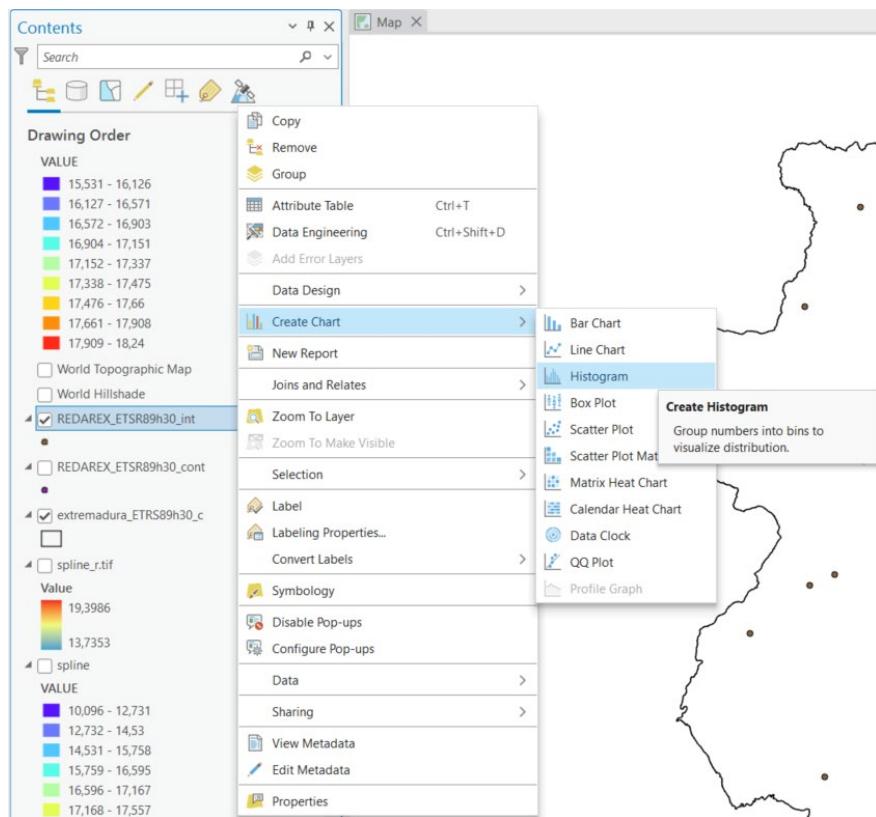


A pesar de haberle dado a una extrapolación, forzando que rellenara todo el contorno de Extremadura, no lo realiza:

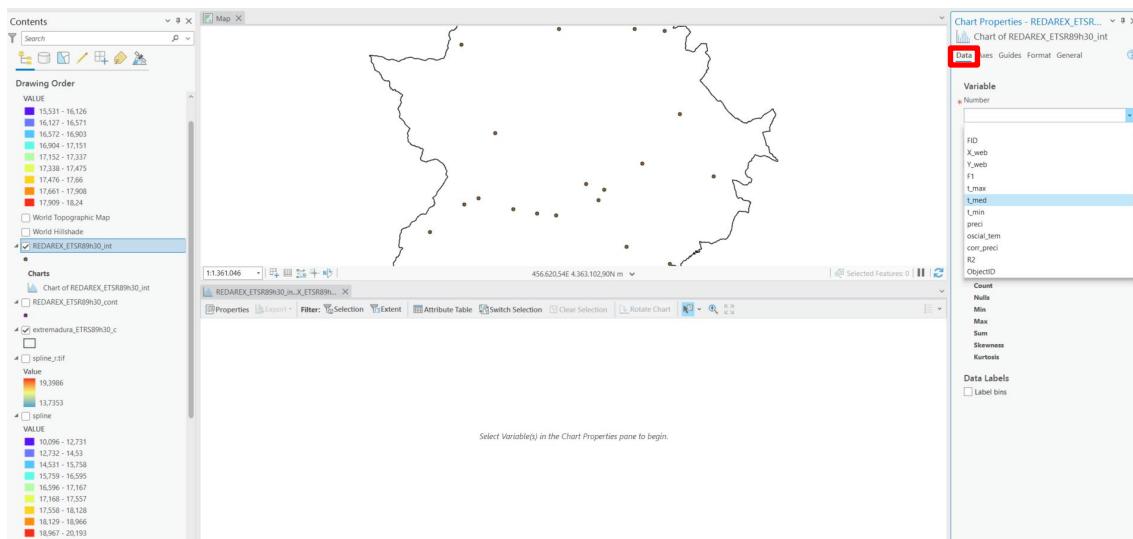


2.4 Kriging

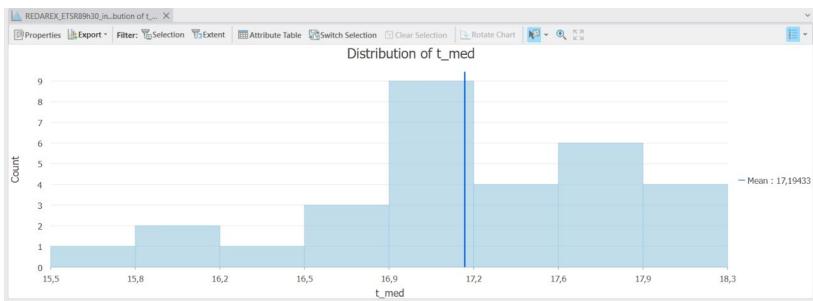
Antes de comenzar, debemos comprobar cuál es la distribución de la variable a interpolar (si es normal o no). Para ello, generaremos el histograma correspondiente:



Tendremos que seleccionar la variable t_med para hacer el histograma



Obteniendo entonces el siguiente histograma:

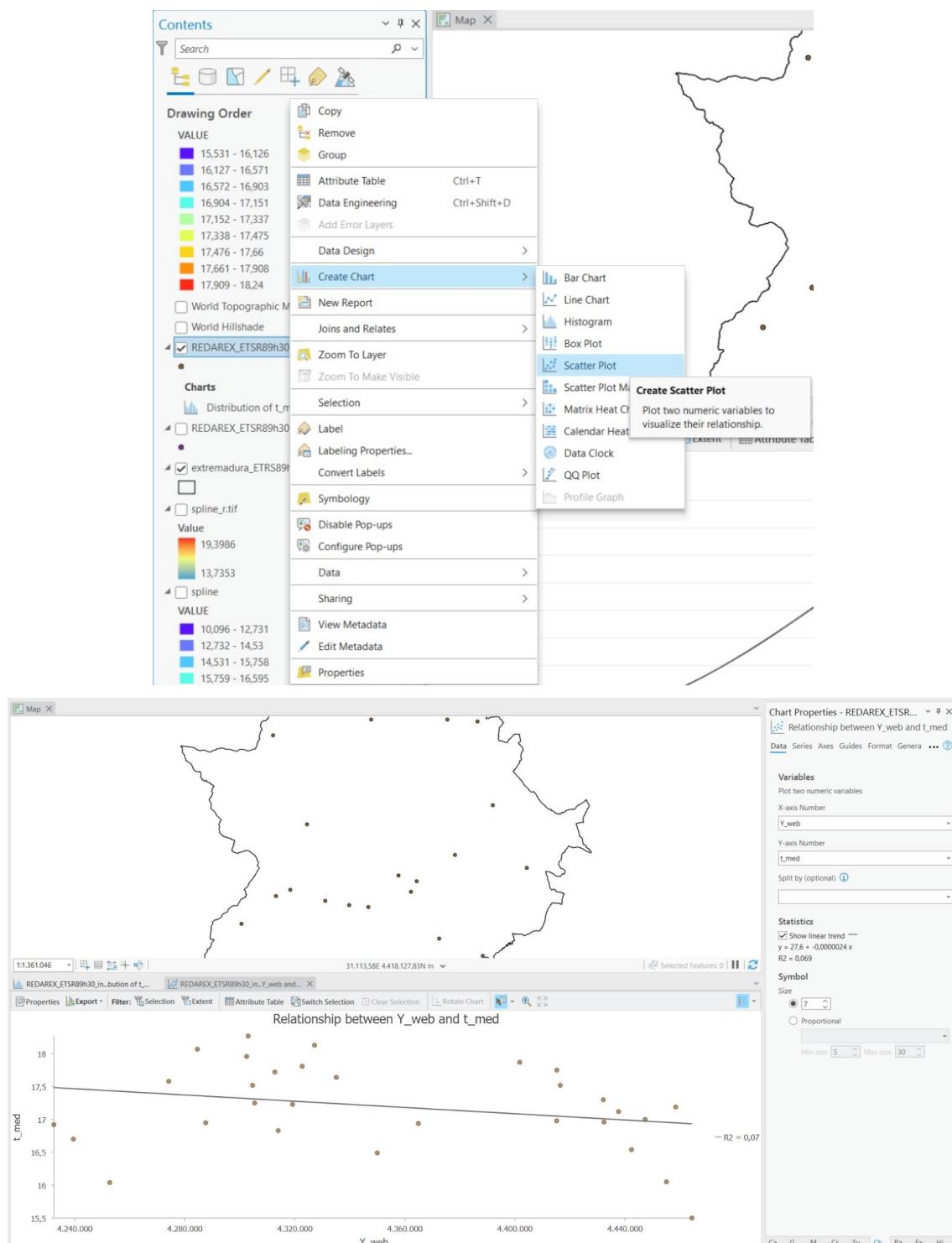


Si activamos la casilla para mostrar la distribución normal, se nos dibujará encima del gráfico.



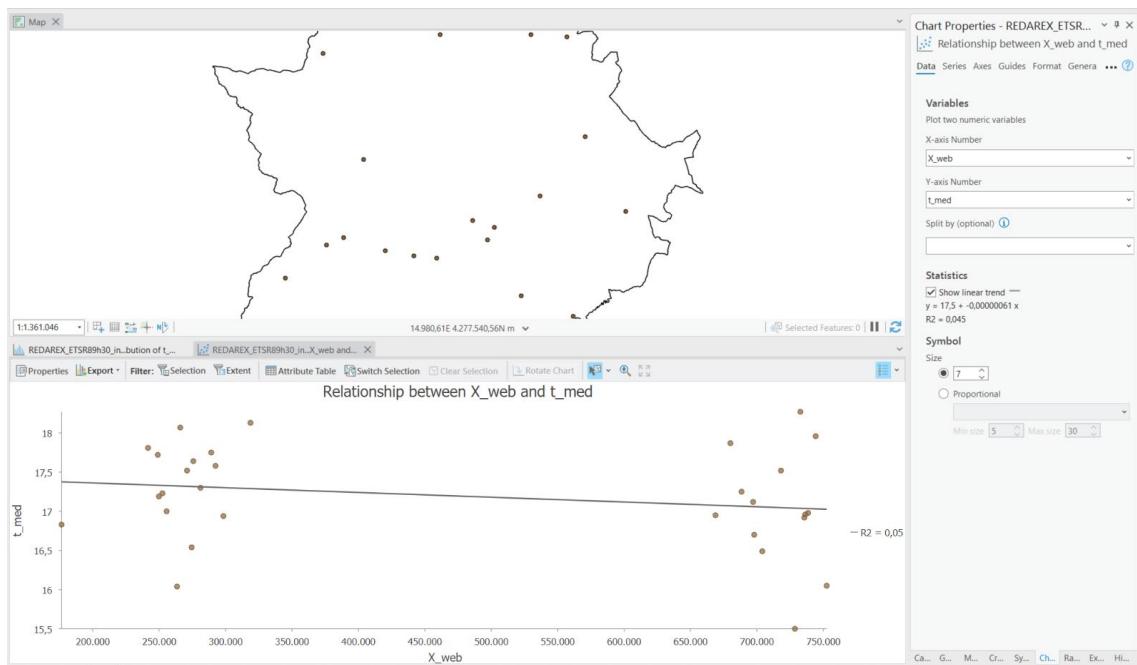
Una vez comprobada que la variable a interpolar tiene una distribución más o menos normal, vamos a buscar si existe alguna tendencia espacial.

Para ello haremos un estudio de correlación de la variable t_{med} con las coordenadas X e Y (para ver si hay tendencias en las direcciones Oeste-Este o Norte-Sur).



Si observamos este primer gráfico, en la dirección N-S, existe una pequeña tendencia a disminuir la temperatura media a medida que aumentan la coordenada Y (hacia el norte), pero la pendiente es tan pequeña ($-0,0000024$), que no merece la pena ser tenida en cuenta.

Si representamos ahora las coordenadas X podremos ver si existe correlación en la dirección W-E:

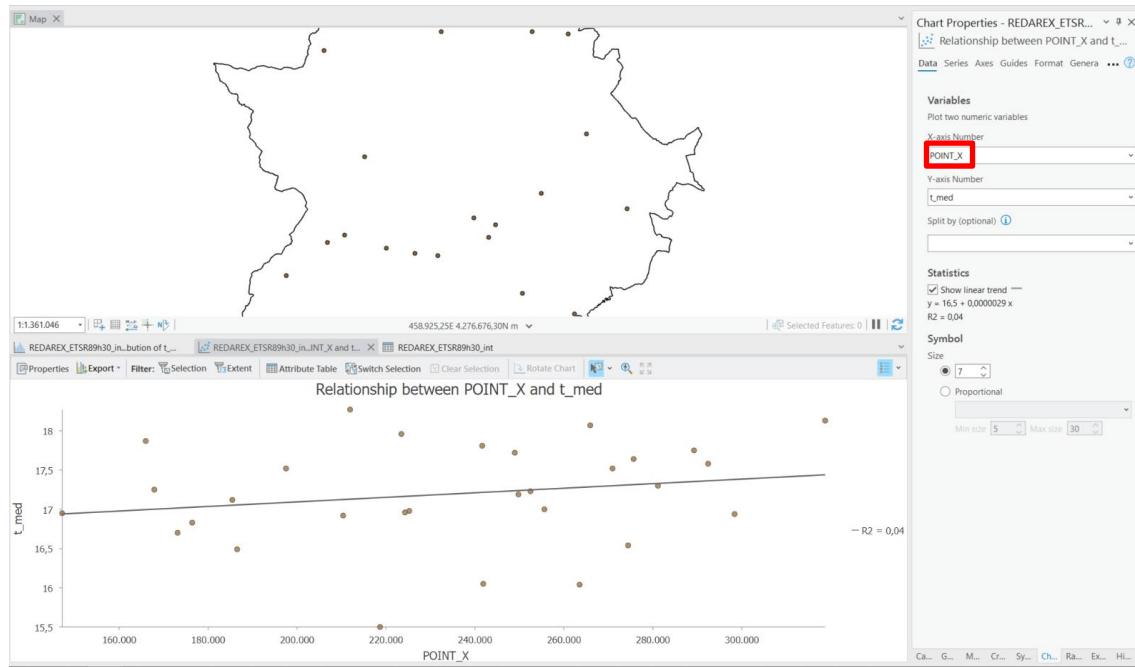


Podemos ver dos grupos de coordenadas, ya que existen dos grupos de estaciones, cada uno en un huso.

Vamos a obtener todas las coordenadas en un único huso y ver si hay alteraciones. Para ello, añadiremos las coordenadas en el sistema de referencia del proyecto (ETRS89-h30), como dos nuevas columnas en la tabla de atributos de las estaciones a interpolar.

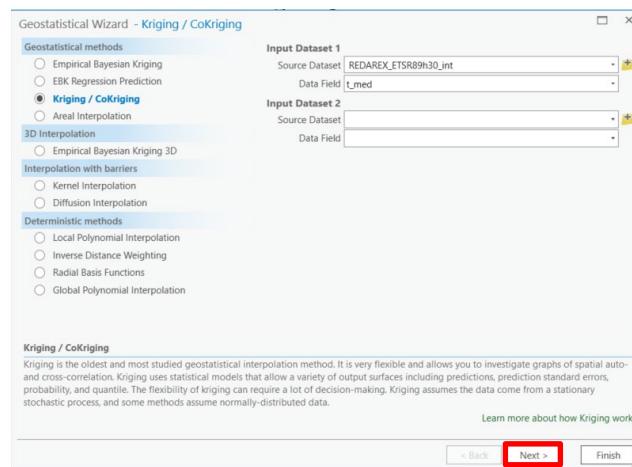
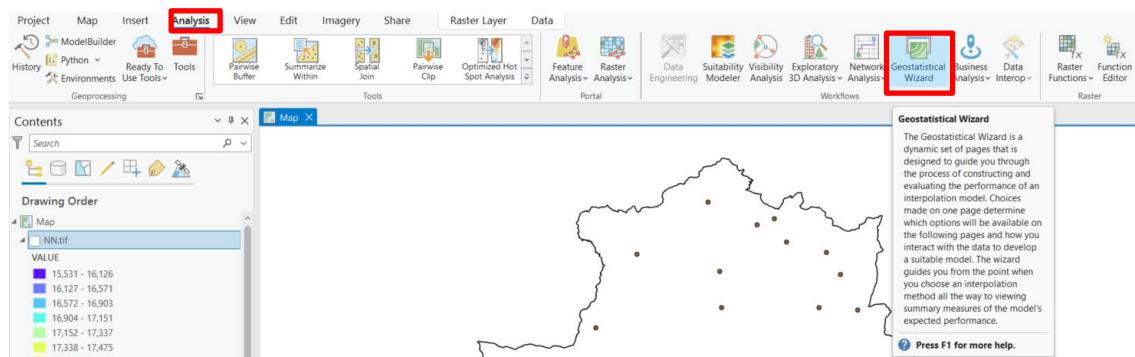
The screenshot illustrates the workflow for adding XY coordinates to station data. On the left, the Geoprocessing pane shows the 'Add XY Coordinates' tool selected. In the center, a table titled 'REDAREX_ETRS89h30_int' displays station attributes including 'Identifica', 't_max', 't_med', 't_min', 'preci', 'oscial_tem', 'corr_preci', 'R2', 'F11', 'F12', 'ObjectID', 'POINT_X', and 'POINT_Y'. A red box highlights the 'POINT_X' and 'POINT_Y' columns. On the right, the 'Add XY Coordinates' tool dialog box is open, showing the 'Input Features' parameter set to 'REDAREX_ETRS89h30_int'.

Y ahora volvemos a hacer el gráfico de correlación:

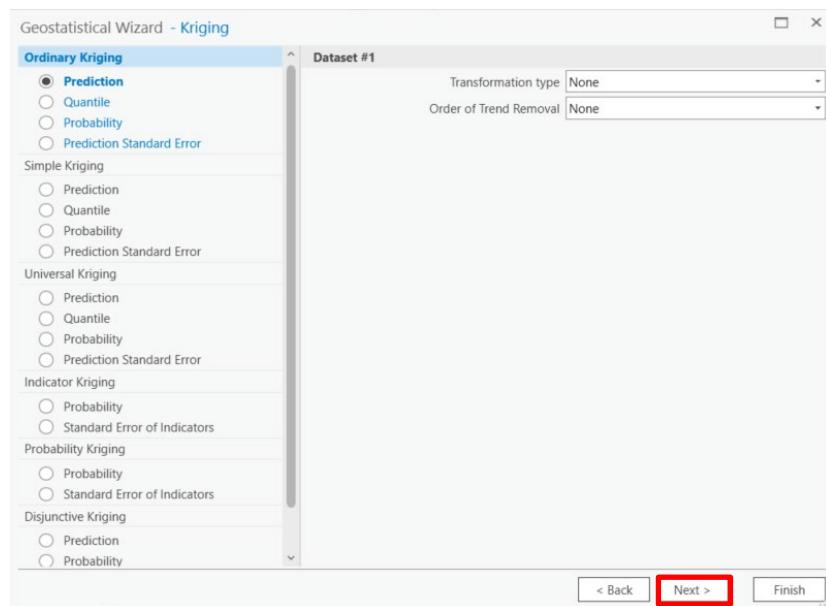


Vemos como es significativamente distinto, pero, aunque hay también cierta tendencia a un aumento de la temperatura media hacia el este (aumentando las X), la pendiente también es muy pequeña ($y = 16,5 + 0,0000029 x$).

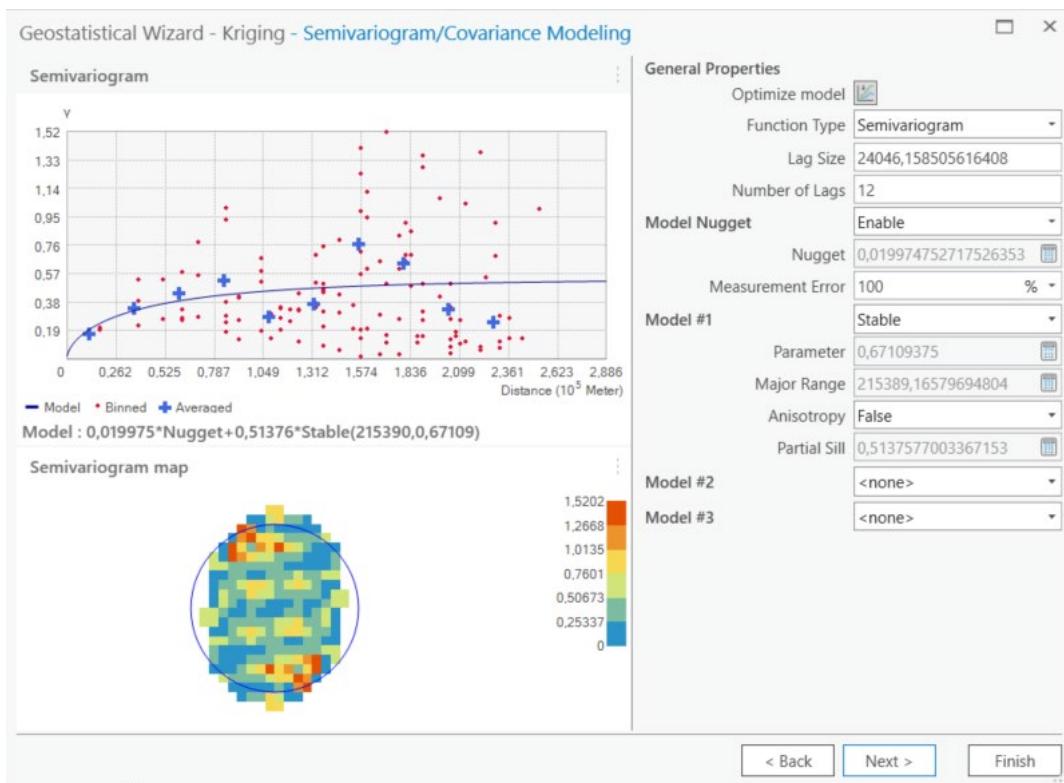
Una vez realizadas estas comprobaciones, procederemos a utilizar el asistente geoestadístico de ArcGIS para hacer esta interpolación:



Como no hemos encontrado ninguna tendencia destacable y la distribución de la variable es normal, elegiremos el kriging ordinario sin ningún tipo de transformación.



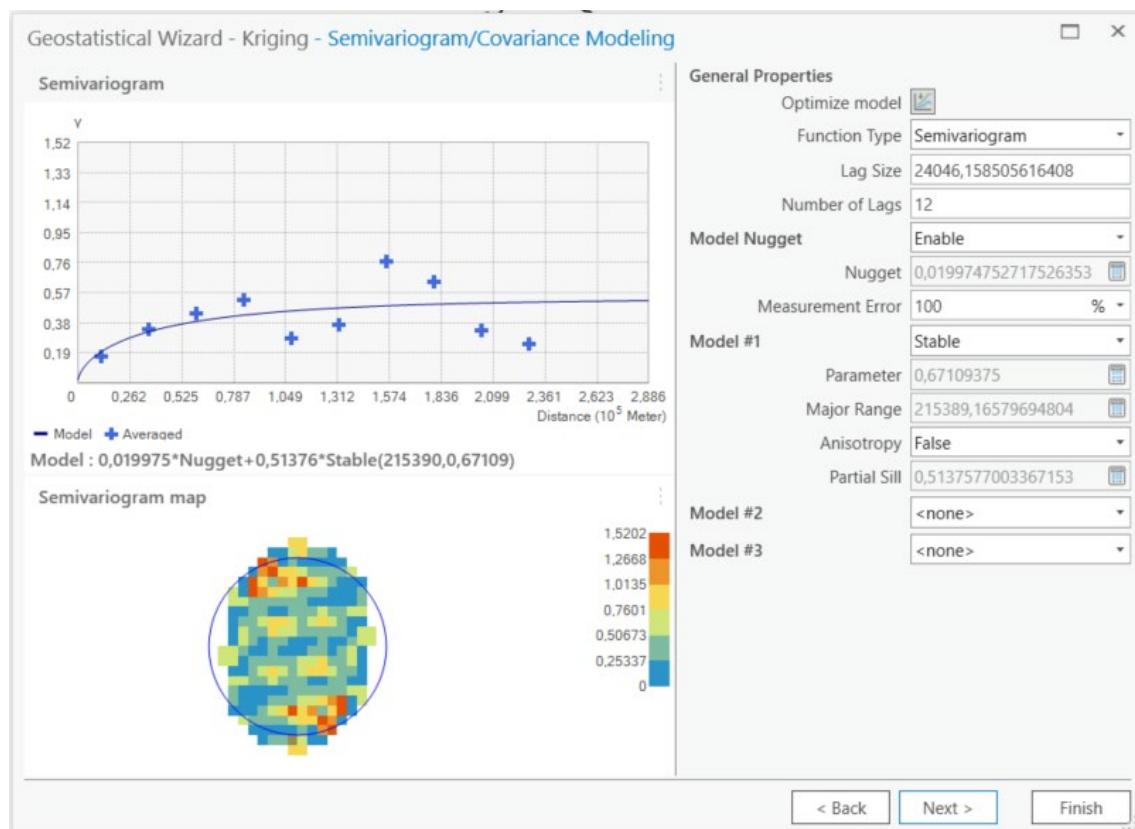
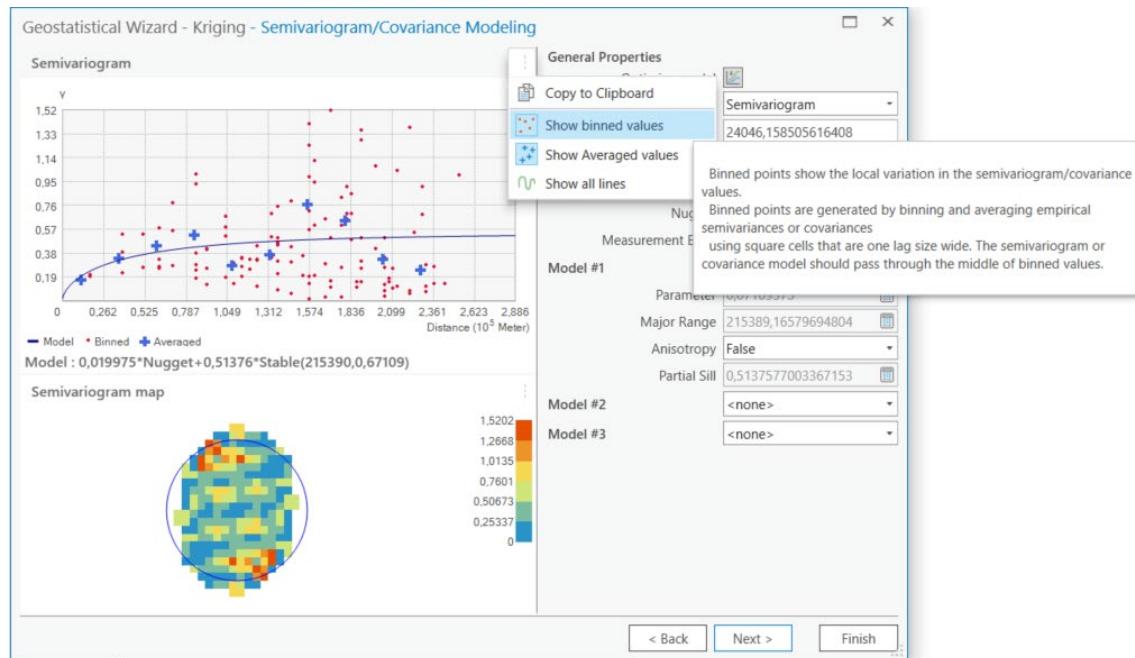
Ahora aparece ya un variograma o semivariograma con nuestros datos.



El asistente geoestadístico proporciona tres vistas diferentes de los valores empíricos del semivariograma. Los valores agrupados se muestran como puntos rojos y se generan agrupando (binning) puntos empíricos de semivariograma utilizando celdas cuadradas de un lag de ancho. Los puntos medios se muestran como cruces azules y se generan agrupando (binning) puntos empíricos de semivariograma que caen dentro de sectores angulares. Los puntos rojos muestran una variación local de los valores de semivariograma, mientras que las cruces muestran una variación suave de los valores de semivariograma. En muchos casos es más fácil ajustar un modelo a los valores promediados, ya que ofrecen una visión menos desordenada de la autocorrelación.

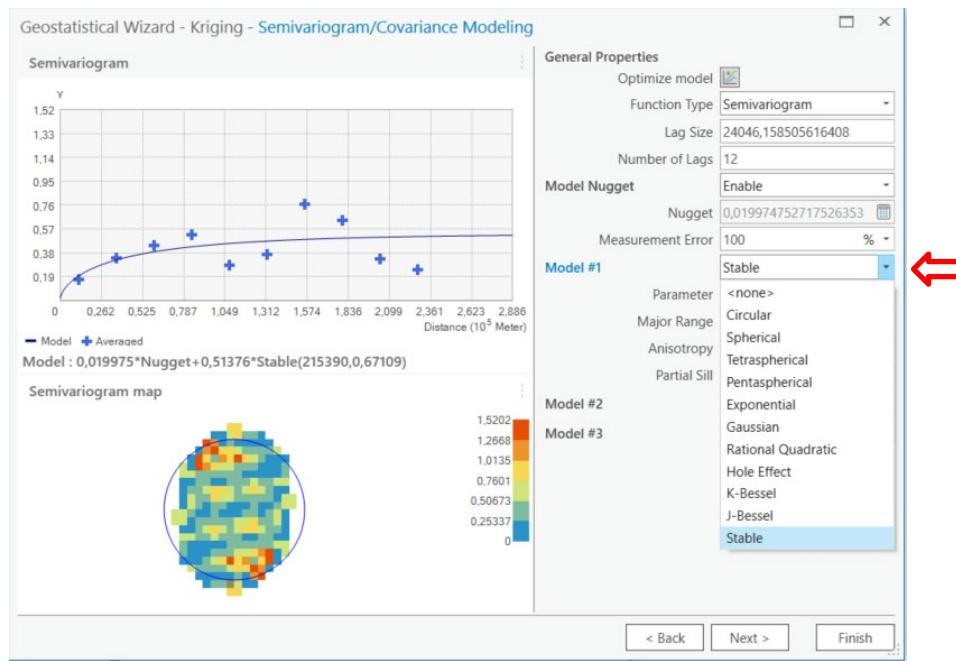
espacial en los datos y muestran cambios más suaves en los valores del semivariograma que los puntos agrupados.

Por tanto, quitaremos de la visualización los puntos agrupados y dejaremos solo los promediados:



Ahora ya podremos buscar el modelo que mejor se adapte a la variable.

Por defecto ofrece el stable, pero si observamos el menú desplegable, existen múltiples modelos:



Las fórmulas matemáticas de cada uno de los modelos son las siguientes:

Nugget effect

The semivariogram model is

$$\gamma(\mathbf{h}; \boldsymbol{\theta}) = \begin{cases} 0 & \text{for } \mathbf{h} = 0 \\ \theta_s & \text{for } \mathbf{h} \neq 0 \end{cases}$$

where $\theta_s \geq 0$.

Circular

The semivariogram model is,

$$\gamma(\mathbf{h}; \boldsymbol{\theta}) = \begin{cases} \frac{2\theta_s}{\pi} \left[\frac{\|\mathbf{h}\|}{\theta_r} \sqrt{1 - \left(\frac{\|\mathbf{h}\|}{\theta_r} \right)^2} + \arcsin \frac{\|\mathbf{h}\|}{\theta_r} \right] & \text{for } 0 \leq \|\mathbf{h}\| \leq \theta_r \\ \theta_s & \text{for } \theta_r < \|\mathbf{h}\| \end{cases}$$

where $\theta_s \geq 0$ is the partial sill parameter and $\theta_r \geq 0$ is the range parameter.

Spherical

The semivariogram model is

$$\gamma(\mathbf{h}; \boldsymbol{\theta}) = \begin{cases} \theta_s \left[\frac{3}{2} \frac{\|\mathbf{h}\|}{\theta_r} - \frac{1}{2} \left(\frac{\|\mathbf{h}\|}{\theta_r} \right)^3 \right] & \text{for } 0 \leq \|\mathbf{h}\| \leq \theta_r \\ \theta_s & \text{for } \theta_r < \|\mathbf{h}\| \end{cases}$$

where $\theta_s \geq 0$ is the partial sill parameter and $\theta_r \geq 0$ is the range parameter.

Tetraspherical

The semivariogram model is

$$\gamma(\mathbf{h}; \boldsymbol{\theta}) = \begin{cases} \frac{2\theta_s}{\pi} \left(\arcsin \left(\frac{\|\mathbf{h}\|}{\theta_r} \right) + \frac{\|\mathbf{h}\|}{\theta_r} \sqrt{1 - \left(\frac{\|\mathbf{h}\|}{\theta_r} \right)^2} + \frac{2}{3} \frac{\|\mathbf{h}\|}{\theta_r} \left(1 - \left(\frac{\|\mathbf{h}\|}{\theta_r} \right)^2 \right)^{\frac{3}{2}} \right) & \text{for } 0 \leq \|\mathbf{h}\| \leq \theta_r \\ \theta_s & \text{for } \theta_r < \|\mathbf{h}\| \end{cases}$$

where $\theta_s \geq 0$ is the partial sill parameter and $\theta_r \geq 0$ is the range parameter.

Pentaspherical

The semivariogram model is

$$\gamma(\mathbf{h}; \boldsymbol{\theta}) = \begin{cases} \theta_s \left[\frac{15}{8} \frac{\|\mathbf{h}\|}{\theta_r} - \frac{5}{4} \left(\frac{\|\mathbf{h}\|}{\theta_r} \right)^3 + \frac{3}{8} \left(\frac{\|\mathbf{h}\|}{\theta_r} \right)^5 \right] & \text{for } 0 \leq \|\mathbf{h}\| \leq \theta_r \\ \theta_s & \text{for } \theta_r < \|\mathbf{h}\| \end{cases}$$

where $\theta_s \geq 0$ is the partial sill parameter and $\theta_r \geq 0$ is the range parameter.

Exponential

The semivariogram model is

$$\gamma(\mathbf{h}; \boldsymbol{\theta}) = \theta_s \left[1 - \exp\left(-\frac{3\|\mathbf{h}\|}{\theta_r}\right) \right] \text{ for all } \mathbf{h},$$

where $\theta_s \geq 0$ is the partial sill parameter and $\theta_r \geq 0$ is the range parameter.

Gaussian

The semivariogram model is

$$\gamma(\mathbf{h}; \boldsymbol{\theta}) = \theta_s \left[1 - \exp\left(-3\left(\frac{\|\mathbf{h}\|}{\theta_r}\right)^2\right) \right] \text{ for all } \mathbf{h},$$

where $\theta_s \geq 0$ is the partial sill parameter and $\theta_r \geq 0$ is the range parameter. Because this model has unstable behavior without nugget, by default the Geostatistical Analyst adds a small nugget to the model, equal to 1/1000 of the sample variance computed for the data.

Rational quadratic

The semivariogram model is

$$\gamma(\mathbf{h}; \boldsymbol{\theta}) = \theta_s \frac{19\left(\frac{\|\mathbf{h}\|}{\theta_r}\right)^2}{1 + 19\left(\frac{\|\mathbf{h}\|}{\theta_r}\right)^2} \text{ for all } \mathbf{h},$$

J-Bessel

The semivariogram model is

$$\gamma(\mathbf{h}; \boldsymbol{\theta}) = \theta_s \left[1 - \frac{2^{\theta_d} \Gamma(\theta_d + 1)}{(\Omega_{\theta_d} \|\mathbf{h}\| / \theta_r)^{\theta_d}} J_{\theta_d} (\Omega_{\theta_d} \|\mathbf{h}\| / \theta_r) \right] \text{ for all } \mathbf{h}$$

where $\theta_s \geq 0$, $\theta_r \geq 0$, $\theta_d \geq 0$, Ω_{θ_d} must satisfy,

$$\min_{B>0, \gamma(B)=\theta_s, \gamma'(B)<0} B = \theta_r$$

$\Gamma(\theta_k)$ is the gamma function,

$$\Gamma(y) = \int_0^\infty x^{y-1} \exp(-x) dx$$

and $J_{\theta_d}(\bullet)$ is the J-Bessel function (Abramowitz and Stegun, 1965, p. 358).

Stable

The semivariogram model is

$$\gamma(\mathbf{h}; \boldsymbol{\theta}) = \theta_s \left[1 - \exp\left(-3\left(\frac{\|\mathbf{h}\|}{\theta_r}\right)^{\theta_e}\right) \right] \text{ for all } \mathbf{h},$$

where $\theta_s \geq 0$ and $0 \leq \theta_e \leq 2$. Because this model has unstable behavior without nugget, by default the Geostatistical Analyst adds a small nugget to the model, equal to 1/1000 of the sample variance computed from the data.

where $\theta_s \geq 0$ is the partial sill parameter and $\theta_r \geq 0$ is the range parameter.

Hole effect

The semivariogram model is

$$\gamma(\mathbf{h}; \boldsymbol{\theta}) = \begin{cases} 0 & \text{for } \mathbf{h} = \mathbf{0} \\ \theta_s \frac{1 - \sin(2\pi\|\mathbf{h}\|/\theta_r)}{\sin(2\pi\|\mathbf{h}\|/\theta_r)} & \text{for } \mathbf{h} \neq \mathbf{0} \end{cases}$$

where $\theta_s \geq 0$.

K-Bessel

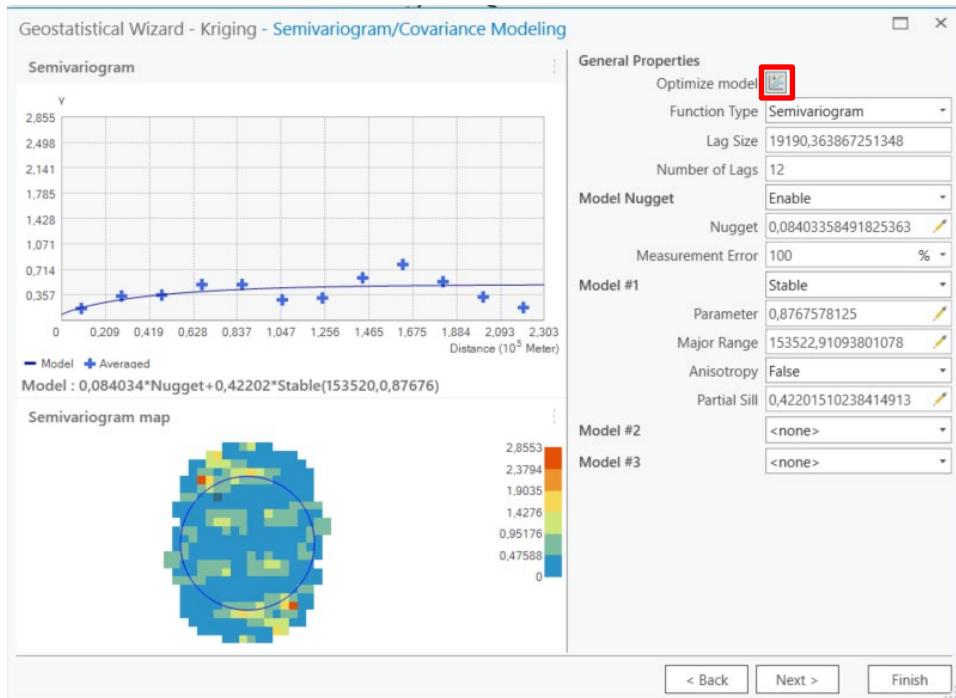
The semivariogram model is

$$\gamma(\mathbf{h}; \boldsymbol{\theta}) = \theta_s \left[1 - \frac{(\Omega_{\theta_k} \|\mathbf{h}\| / \theta_r)^{\theta_k}}{2^{\theta_k-1} \Gamma(\theta_k)} K_{\theta_k} (\Omega_{\theta_k} \|\mathbf{h}\| / \theta_r) \right] \text{ for all } \mathbf{h},$$

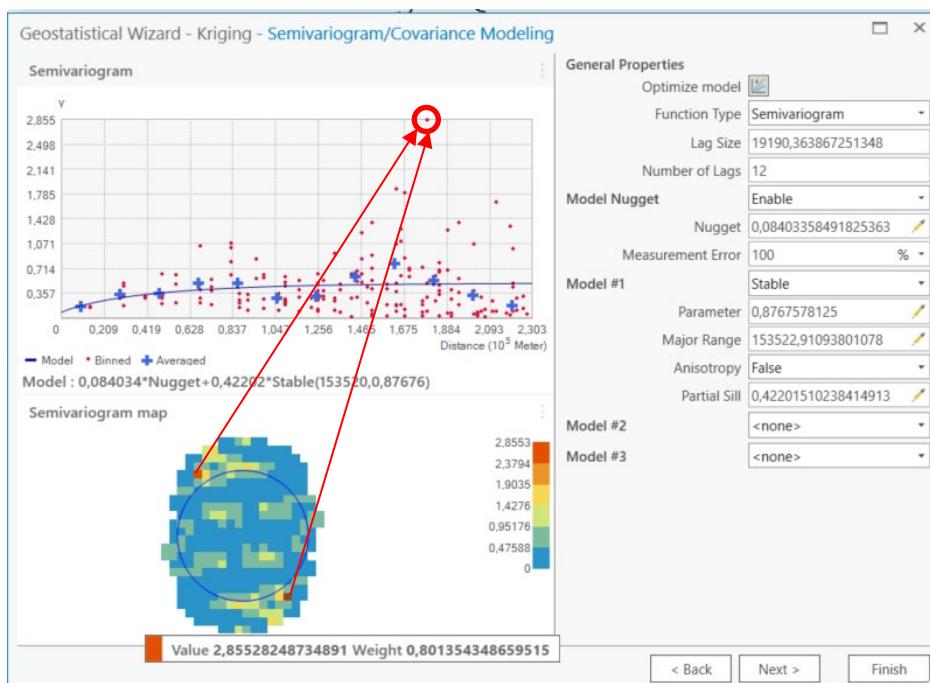
where $\theta_s \geq 0$, $\theta_r \geq 0$, $\theta_k \geq 0$, Ω_{θ_k} is a value found numerically so that $\gamma(\theta_r) = 0.95$ θ_s for any θ_k , $\Gamma(\theta_k)$ is the gamma function,

$$\Gamma(y) = \int_0^\infty x^{y-1} \exp(-x) dx$$

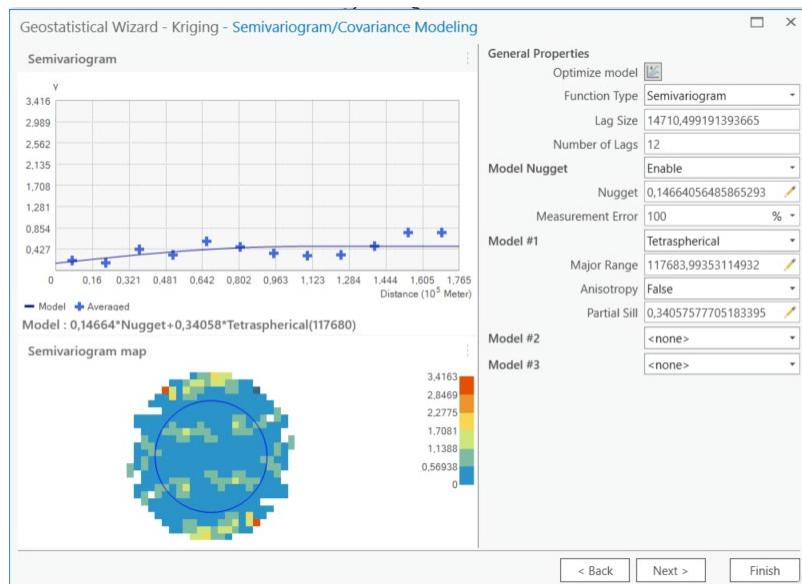
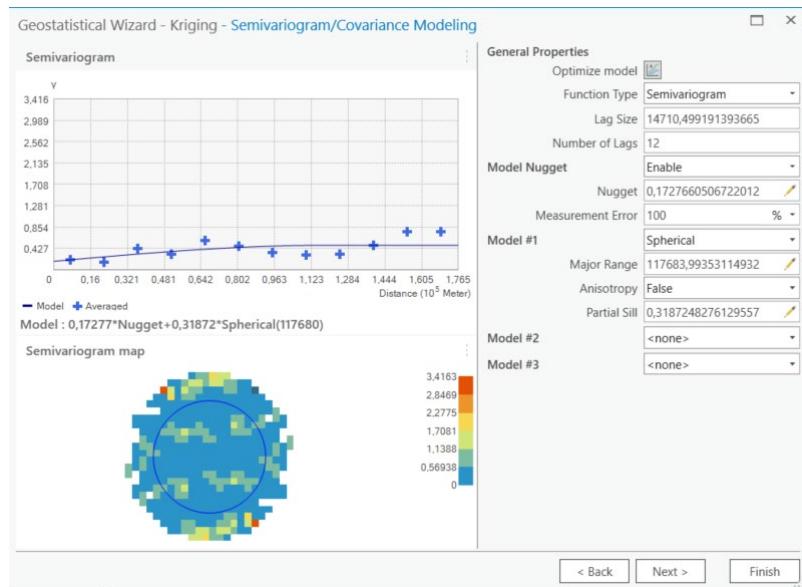
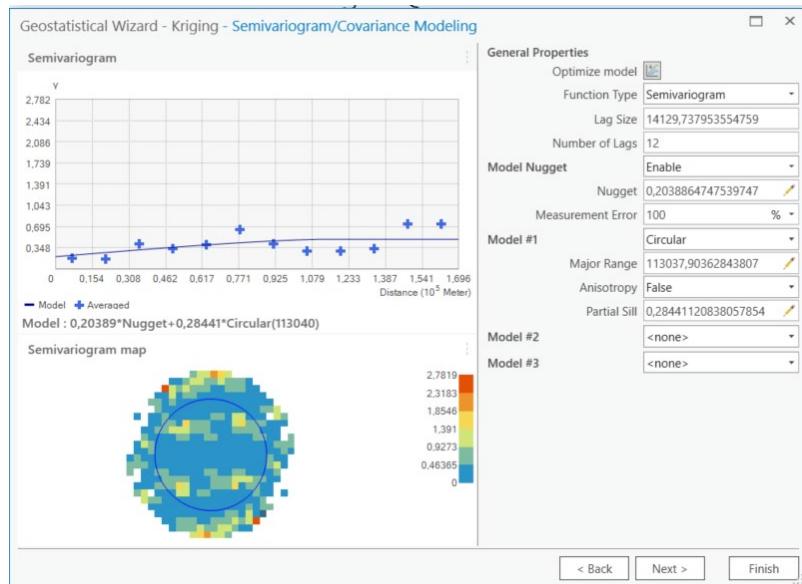
and $K_{\theta_k}(\bullet)$ is the modified Bessel function of the second kind of order θ_k (Abramowitz and Stegun, 1965, p. 374).

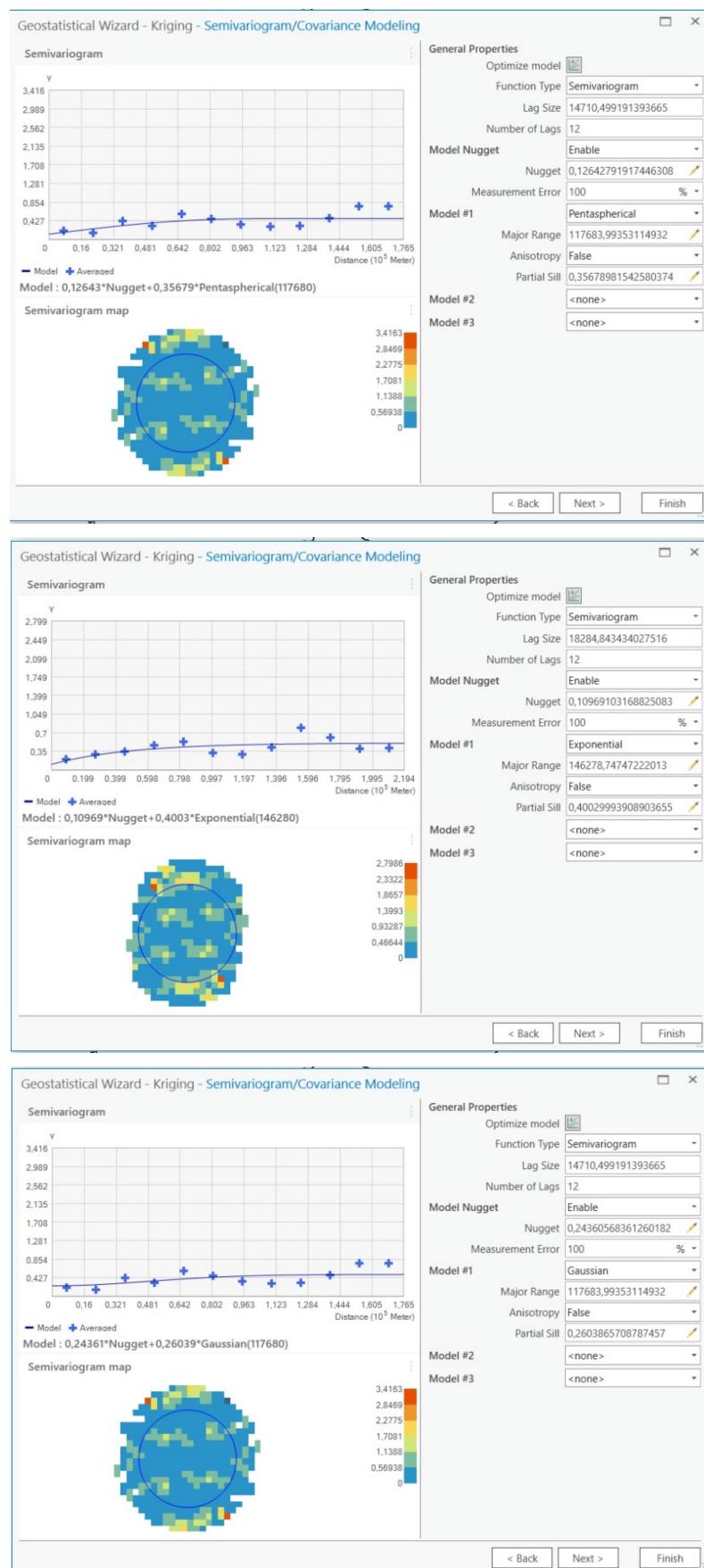


En la parte inferior aparece la localización de los puntos del semivariograma. Si encendemos de nuevo los puntos agrupados podemos ver cómo hay una pareja que lleva a un valor de varianza de 2.8553 y su localización geográfica en el mapa.

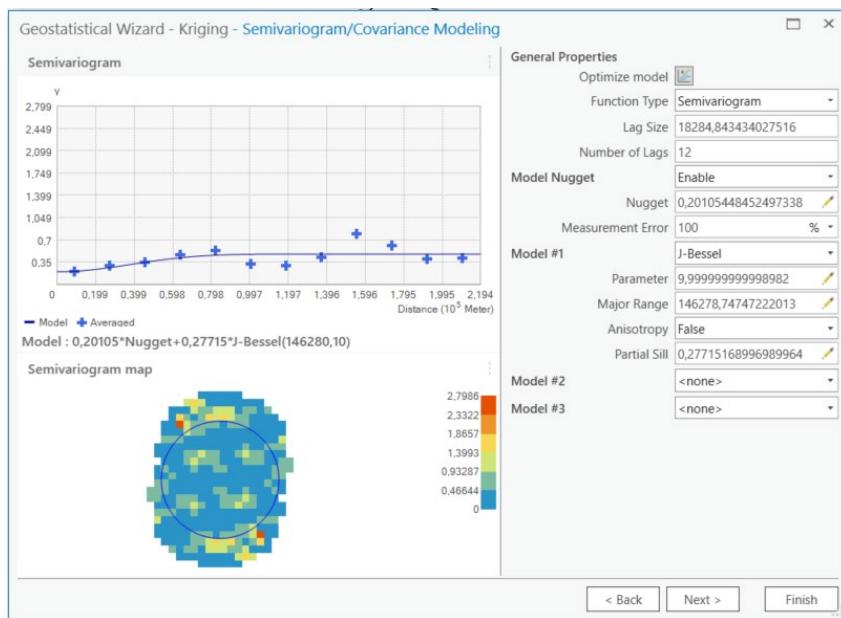


Probamos todas las posibles funciones y vemos el mejor ajuste.

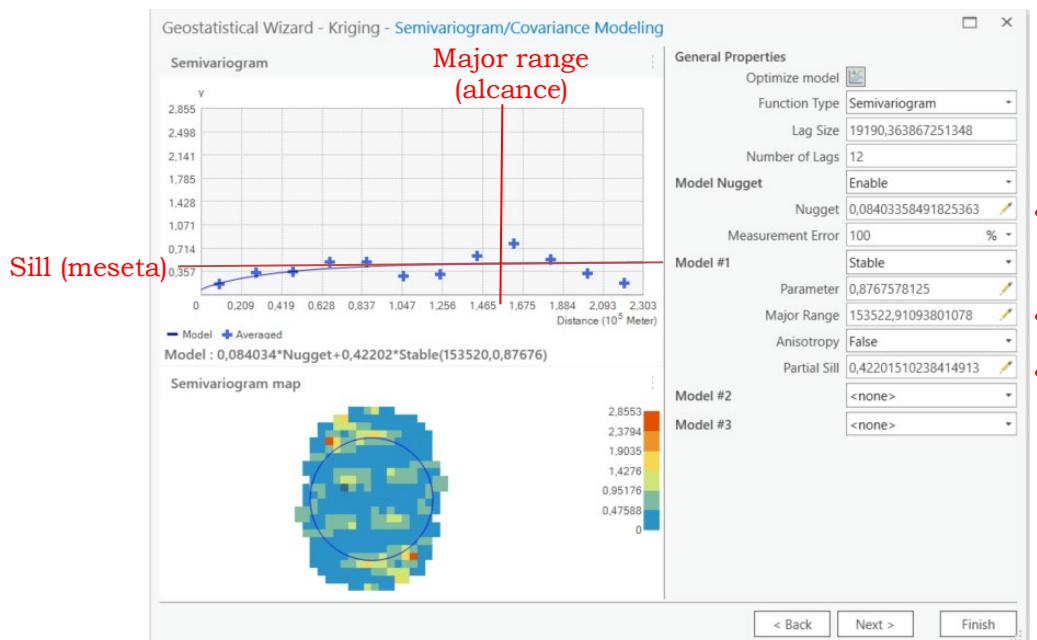








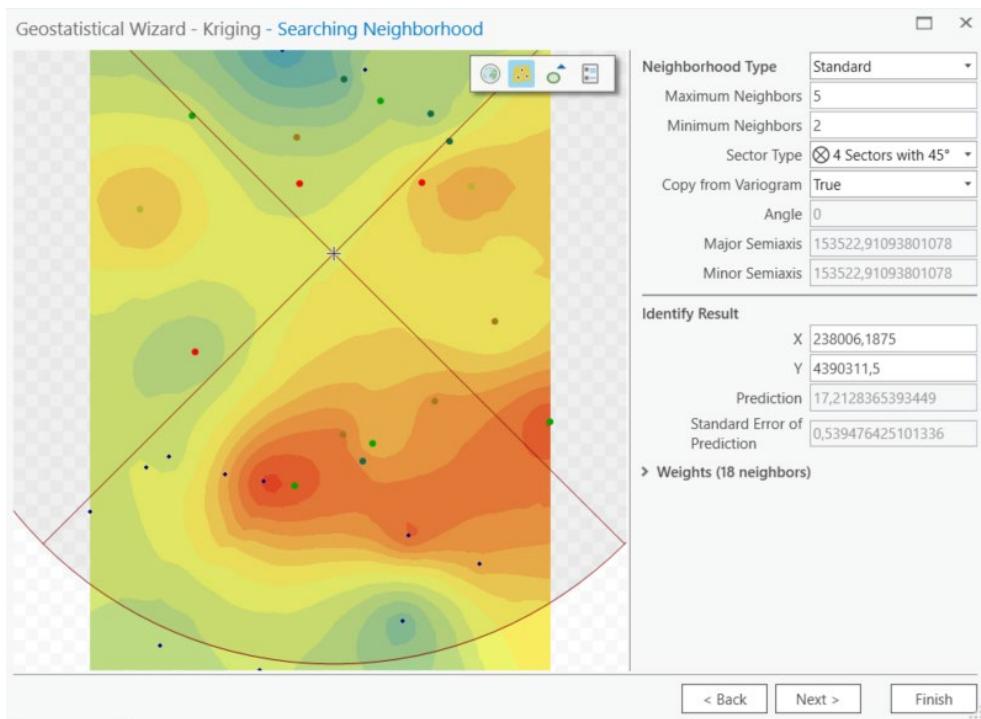
Definitivamente nos quedamos con el modelo stable, que es el que produce un nugget más pequeño.



Si echamos un vistazo al valor del Major Range, vemos que es 153523 m que es la distancia límite a partir de la cual ya no hay correlación.

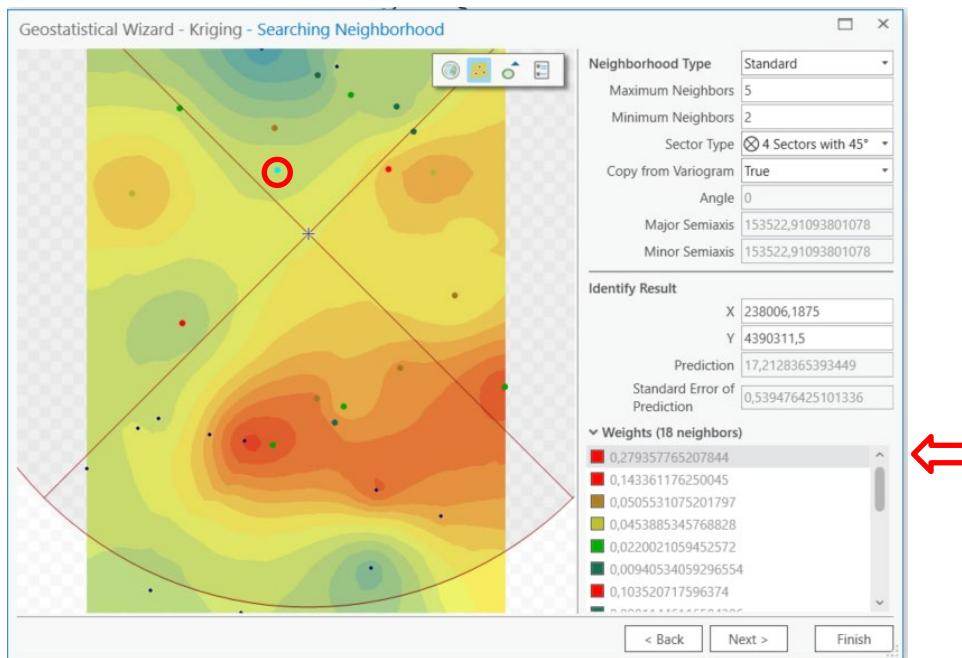
Finalmente, destacar que tenemos configurado como no anisotrópico nuestro modelo, ya que la correlación entre dos puntos no depende de la orientación.

En la siguiente ventana configuraremos el número de estaciones utilizadas definiendo un círculo (o elipse) que encierre los puntos que se utilizan para predecir valores en lugares no medidos.



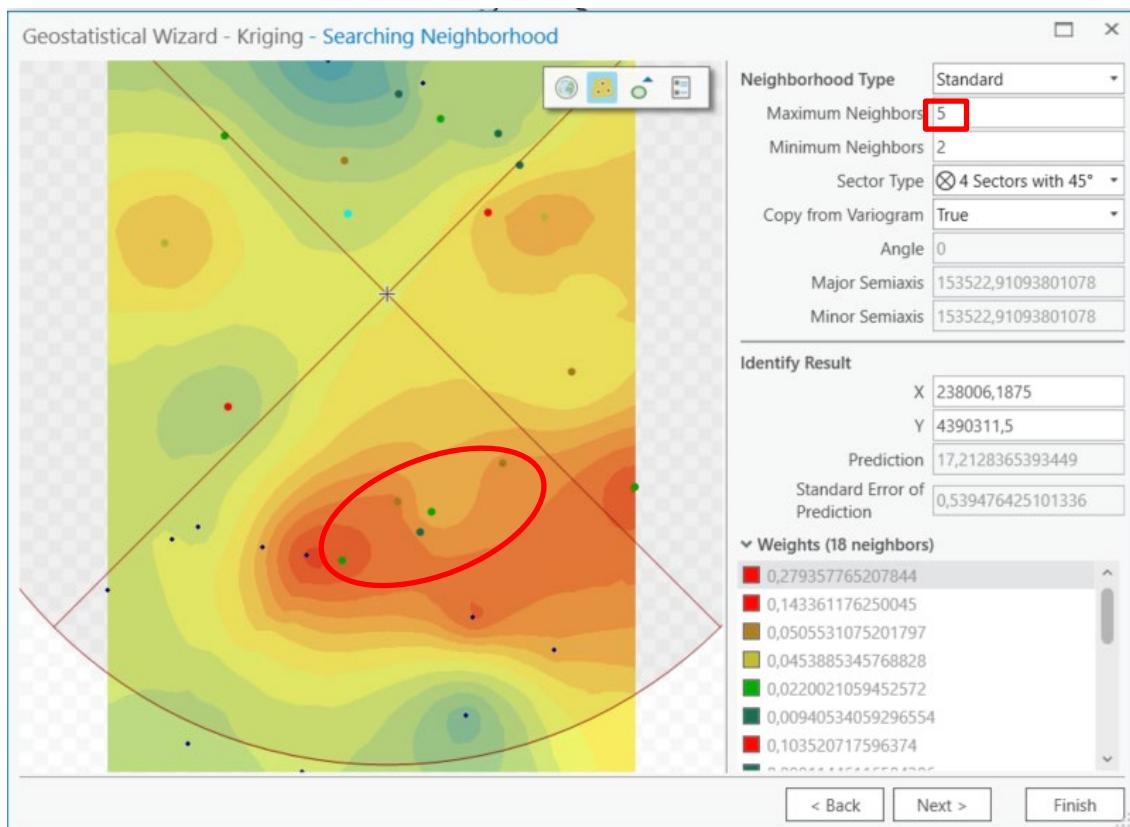
Además, para evitar sesgos en una dirección concreta, el círculo (o elipse) puede dividirse en sectores de los que se selecciona un número igual de puntos. Se puede especificar el número de puntos (un máximo de 200), el radio (o eje mayor/menor) y el número de sectores del círculo (o elipse) que se utilizarán para la predicción.

Los puntos resaltados en la ventana de vista de datos dan idea de los pesos que se asociarán a cada localización en la predicción de valores desconocidos.

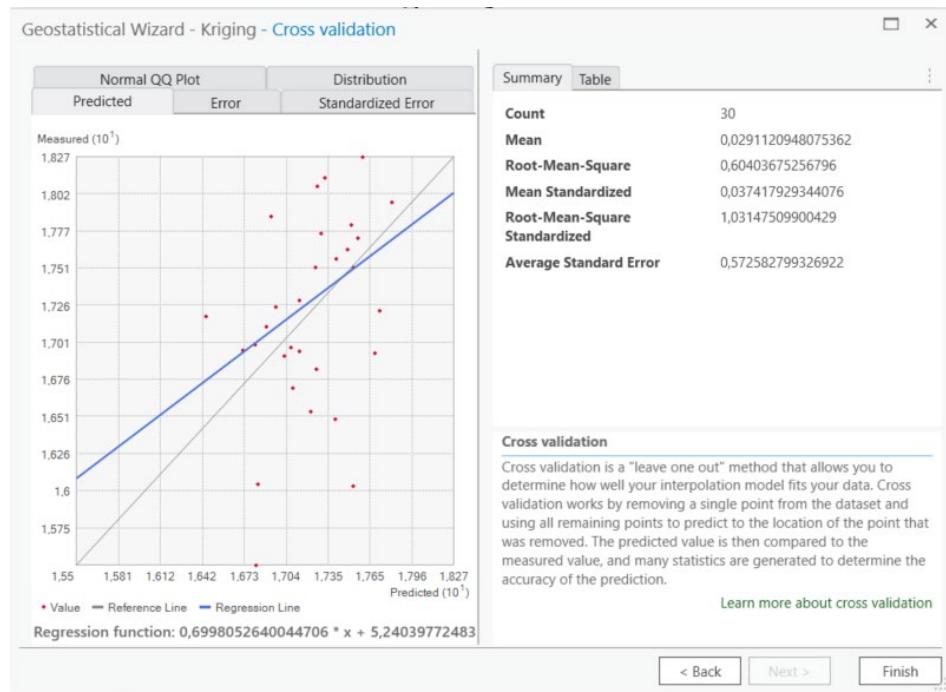


Por ejemplo, el punto marcado tendrá un peso de 0.28 en la predicción.

Además, vemos qué 5 puntos cogería para predecir del pixel localizado en el centro del círculo, por ejemplo en el sector sur, porque quedan resaltados:



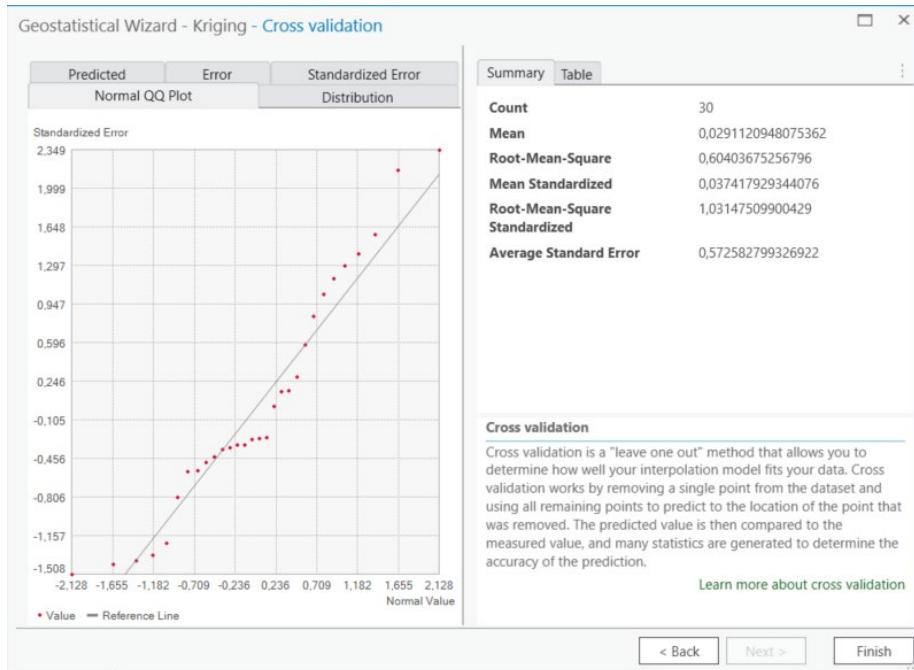
La última sección será la que nos muestre los errores estimados en la predicción utilizando validación cruzada.



En la primera pestaña nos muestra los estadísticos de error estimado, tales como el RMSE el EMS y otros varios. La predicción se realiza mediante el método de “dejar uno fuera”. El método utiliza primero todos los puntos de entrada para estimar los parámetros del modelo de interpolación. A continuación, elimina un único punto de entrada y utiliza los puntos restantes para predecir el valor en la ubicación del punto oculto, y el valor predicho se compara con el valor medido. A continuación, se vuelve a

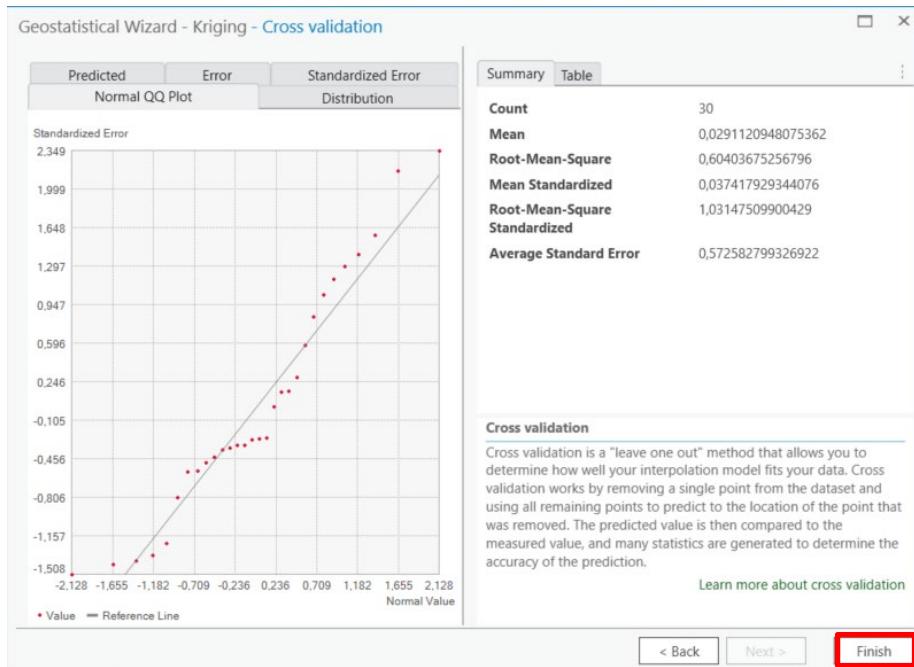
añadir el punto oculto al conjunto de datos y se oculta y predice un punto diferente. Este proceso se repite con todos los puntos de entrada.

En nuestro gráfico de correlación entre los valores de temperatura predichos y los medidos por validación cruzada, se observa una desviación de la línea teórica (línea de referencia gris).

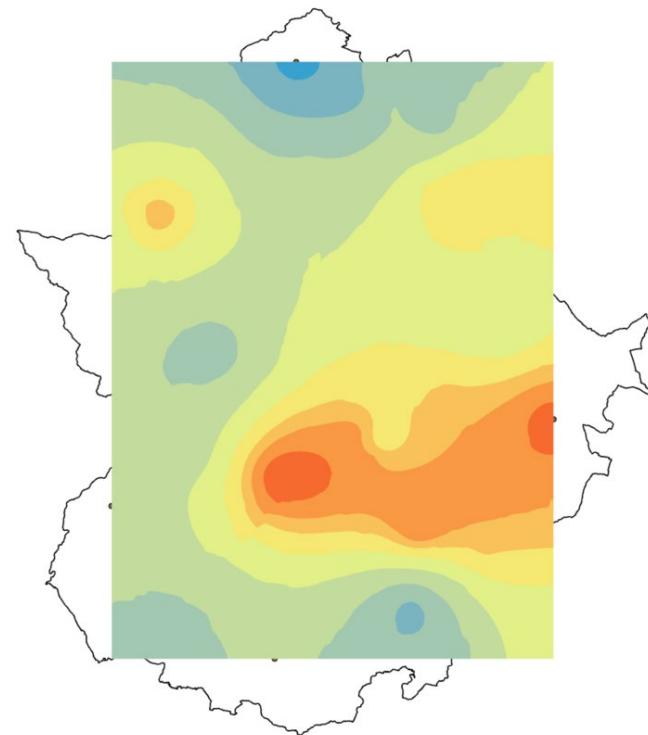
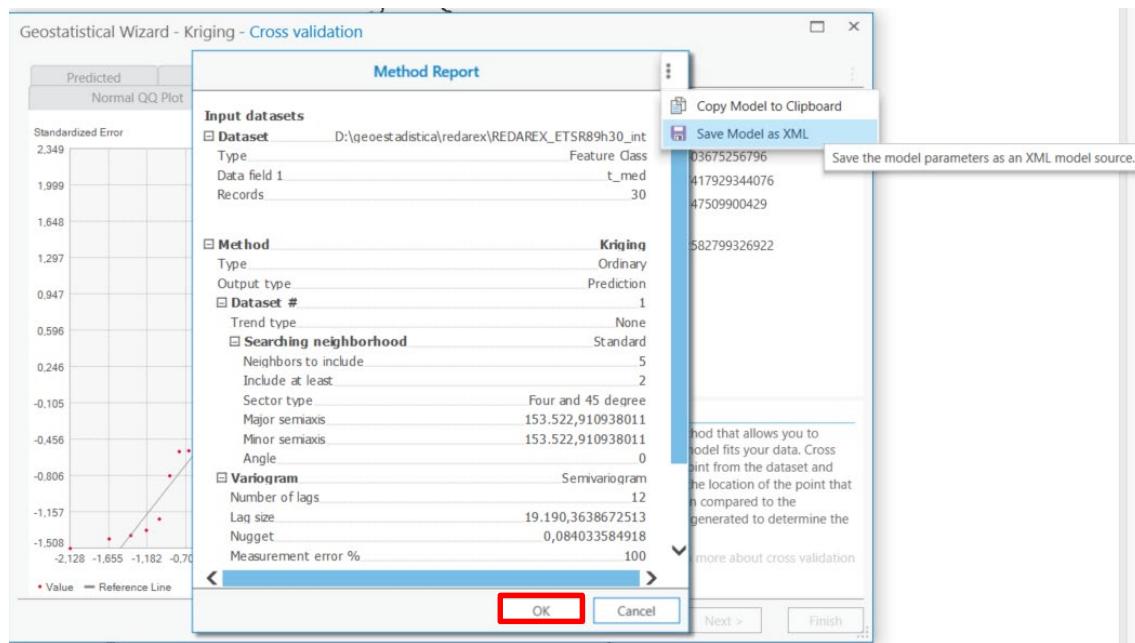


En la solapa del gráfico QQ-Normal podemos ver la distribución del error estándar, en el que todos los puntos debían estar alineados con la línea de referencia gris.

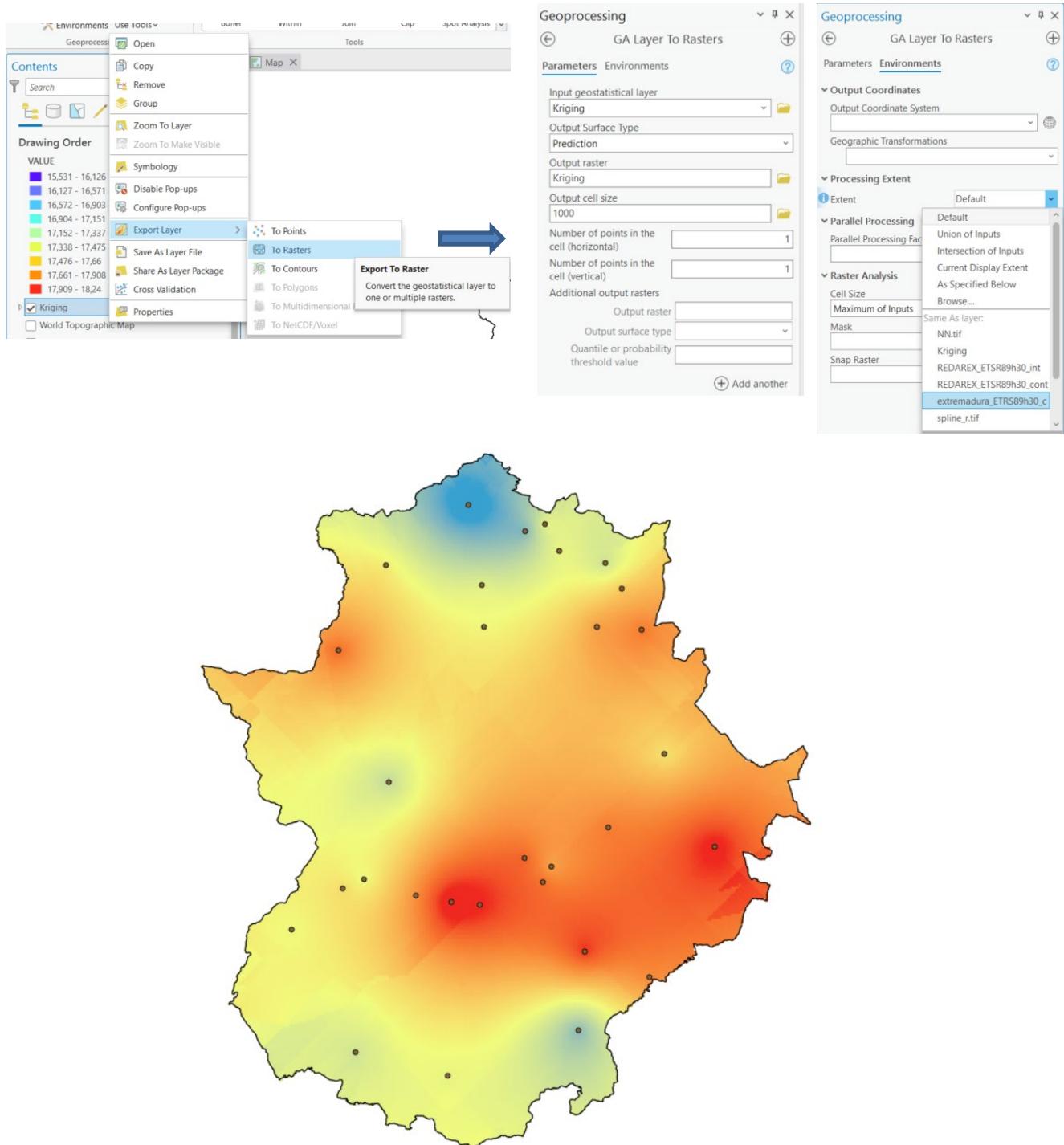
Aceptamos ya todos los parámetros con los que hemos configurado el Kriging y obtendremos la interpolación.



Se nos abrirá entonces un informe resumen de todo, que podremos exportar.



Como en ningún momento de la configuración hemos indicado la extensión de la interpolación, podemos hacerlo ahora para que nos haga una extrapolación a toda la región de Extremadura.

**Ejercicio propuesto 1:**

Calcular el error de todas las interpolaciones utilizando los puntos de control.

3. Bibliografía

- Barnes, R. (2003). Variogram tutorial. Golden, CO: Golden. Software. Available online at <http://www.goldensoftware.com/.variogramTutorial.pdf>.
- Childs, C. (2004). Interpolating surfaces in ArcGIS spatial analyst. ArcUser, July-September, 3235(569), 32-35.
- Johnston, K., Ver Hoef, J. M., Krivoruchko, K., & Lucas, N. (2001). Using ArcGIS geostatistical analyst (Vol. 380). Redlands: Esri.
- Lloyd, C. (2010). Spatial data analysis: an introduction for GIS users. Oxford university press.
- Matínez, D. (2015). Comparación Estadística de Métodos Interpolación Determinísticos y Estocásticos para la generación de Modelos Digitales del Terreno a partir de datos LIDAR, en la parroquia de Tumbabiro, cantón San Miguel de Urcuquí, provincia de Imbabura (Doctoral dissertation, Tesis de Licenciatura, Quito: USFQ).
- Felicísimo, Á. M. (1994). Modelos digitales del terreno. Oviedo: Pentalfa.
- Giraldo, R. (2011). Introducción a la geoestadística. Universidad Nacional de Colombia.
- Matheron, G. (1969). Curso de geoestadística. Escuela de Minas de París.
- Mitchell, A. (1999). ESRI Manual for GIS analysis. Volume, 1, 190.
- Olaya, V. (2009). Sistemas de información geográfica. Cuadernos internacionales de tecnología para el desarrollo humano, (8), 15.