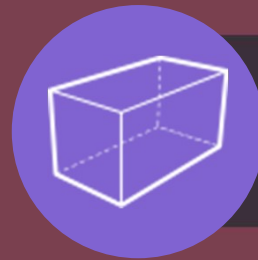


ISATIS 5

Análisis Estadístico y Geoestadístico de Datos
13 de Diciembre de 2019

Auxiliar: Fabián Soto F.
Profesor: Xavier Emery



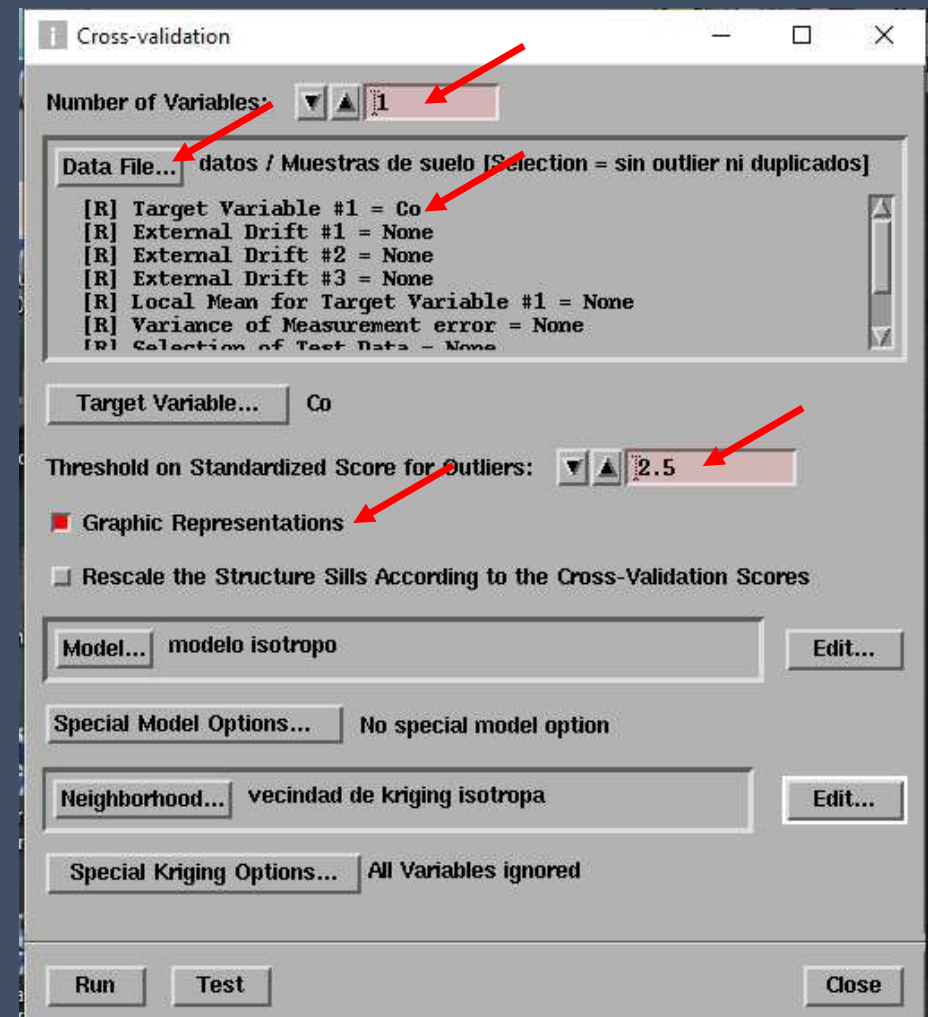


Validación cruzada

Requisito indispensable: tener un modelo de variograma para concentración de Co.

Objetivos: validar el modelo de variograma; definir y validar los parámetros de la vecindad de kriging.

- Ir a *Statistics > Cross Validation*
- Considerar 1 variable.
- Entrar los datos de muestras de suelo en *Data File*, con la variable cobalto (Co) y la selección *sin outlier ni duplicado*. Hay posibilidad de definir variables de output (valor estimado y error estandarizado). “Target variable” es para seleccionar la variable de interés en el caso multivariable.
- Dejar el umbral (threshold) en 2.5: de este modo, se pondrá de relieve a todos los datos para los cuales el error de kriging es mayor (en valor absoluto) a 2.5 veces la desviación estándar de error, o sea, los datos “mal” estimados. Típicamente, se espera que menos del 5% de los datos estén en esta condición.
- Activar las representaciones gráficas.



- No activar la opción de reescalar el modelo (“rescale...”), dado que Isatis cambiaría las mesetas del modelo de variograma de modo que haya 5% de datos con un error mayor a 2.5 veces la desviación estándar (no aconsejable, dado que el modelo de variograma podría dejar de ser un buen ajuste del variograma experimental).
- En “model”, entrar primero el **modelo de variograma isótropo**.
- En “Special model options”, no cambiar nada (podría usarse kriging simple con media conocida; por defecto, se aplica kriging ordinario de media desconocida).
- En “Neighborhood”, se define una vecindad llamada **“vecindad de kriging isotropa”** (entrar el nombre en la zona de texto, luego presionar “Add” y “OK”).
- Los parámetros de esta vecindad se definen al presionar el botón “Edit”.
- En “Special kriging option”, no se define ninguna opción particular (son opciones avanzadas, tales como filtrar errores de medición en los datos).

Cross-validation

Number of Variables: 1

Data File... datos / Muestras de suelo [Selection = sin outlier ni duplicados]

[R] Target Variable #1 = Co
 [R] External Drift #1 = None
 [R] External Drift #2 = None
 [R] External Drift #3 = None
 [R] Local Mean for Target Variable #1 = None
 [R] Variance of Measurement error = None
 [R] Selection of Test Data = None

Target Variable... Co

Threshold on Standardized Score for Outliers: 2.5

☒ Graphic Representations

☐ Rescale the Structure Sills According to the Cross-Validation Scores

Model... modelo isotropo Edit...

Special Model Options... No special model option

Neighborhood... vecindad de kriging isotropa Edit...

Special Kriging Options... All Variables ignored

Run Test Close

- “Neighborhood type”: podría ser una vecindad única (que considera todos los datos) o una vecindad móvil. La vecindad única es una opción posible en este caso porque hay relativamente pocos datos (359), pero en general uno dispone de varios miles o decenas de miles de datos y prefiere utilizar una vecindad móvil. Por esta razón, usaremos una vecindad de tipo “Moving”, la que corresponde a un elipse centrado en el punto a estimar.
- Se especifica la orientación del elipse en “Rotation”, según la anisotropía o isotropía de la variable a estimar. Como el variograma considerado es isótropo, se decide tomar una vecindad isótropa (en forma de círculo): no hay que especificar ninguna rotación.
- El radio de la vecindad se elige en 1000 metros (en cada dirección)
- Se deja la opción “use anisotropic distance” (en caso de una vecindad en forma de elipse, se considera una función de distancia anisótropa, es decir, las curvas de iso-distancia corresponden a elipses concéntricos).

Neighborhood Definition

Load... Default Reset

Neighborhood Type ☒ Moving

Parameters

Sectors Advanced Sorting

Rotation... No rotation

Search Ellipsoid: Maximum Distances in the System after Rotation:

U 1000 m V 1000 m W 1000 m

☒ Use Anisotropic Distances (According to the Search Ellipsoid)

Dimensions of the Proximity Rectangle (for Punctual Estimation Only):

X 0 m Y 0 m Z 0 m

Minimum Number of Samples: 4

Number of Angular Sectors: 8

Optimum Number of Samples per Sector: 5

Test Print OK Cancel

- Se dejan las dimensiones del rectángulo de proximidad en 0 (de encontrar un dato más cercano que estas dimensiones, el programa tomaría este dato como valor estimado, sin hacer kriging; al dejar las dimensiones en 0, esto significa que se hará kriging en cualquier circunstancia).
- Se elige el número mínimo de datos a utilizar (por ejemplo, 4), el número de sectores angulares para dividir la vecindad (por ejemplo, 8) y el número óptimo de datos a buscar dentro de cada sector (por ejemplo, 5, lo que totaliza $8 \times 5 = 40$ datos en la vecindad).
- Se valida los parámetros de la vecindad al presionar el botón “OK”.

Neighborhood Definition

Load... Default Reset

Neighborhood Type Moving ☐

Parameters

Sectors Advanced Sorting

Rotation... No rotation

Search Ellipsoid: Maximum Distances in the System after Rotation:

U m V m W m

☒ Use Anisotropic Distances (According to the Search Ellipsoid)

Dimensions of the Proximity Rectangle (for Punctual Estimation Only):

X m Y m Z m

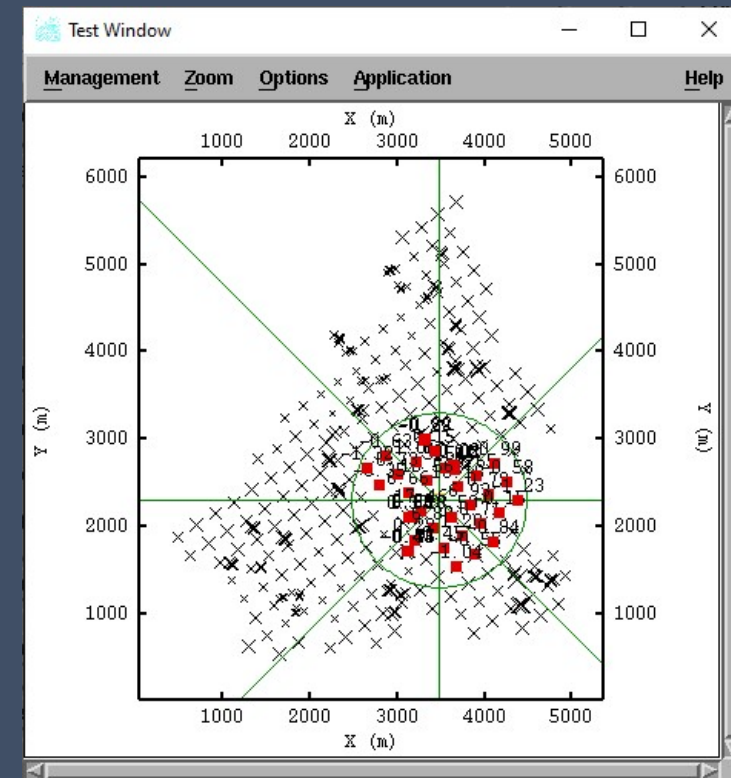
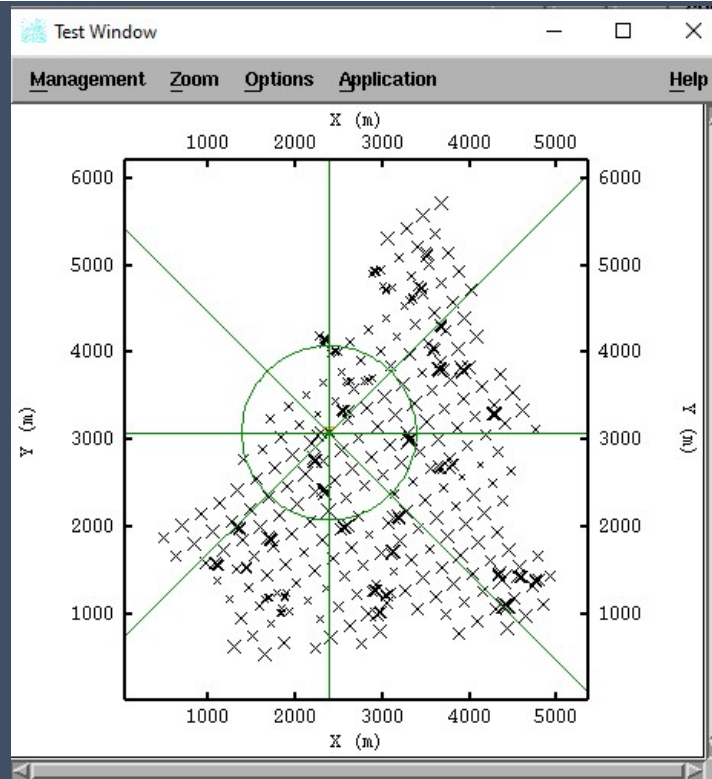
Minimum Number of Samples:

Number of Angular Sectors:

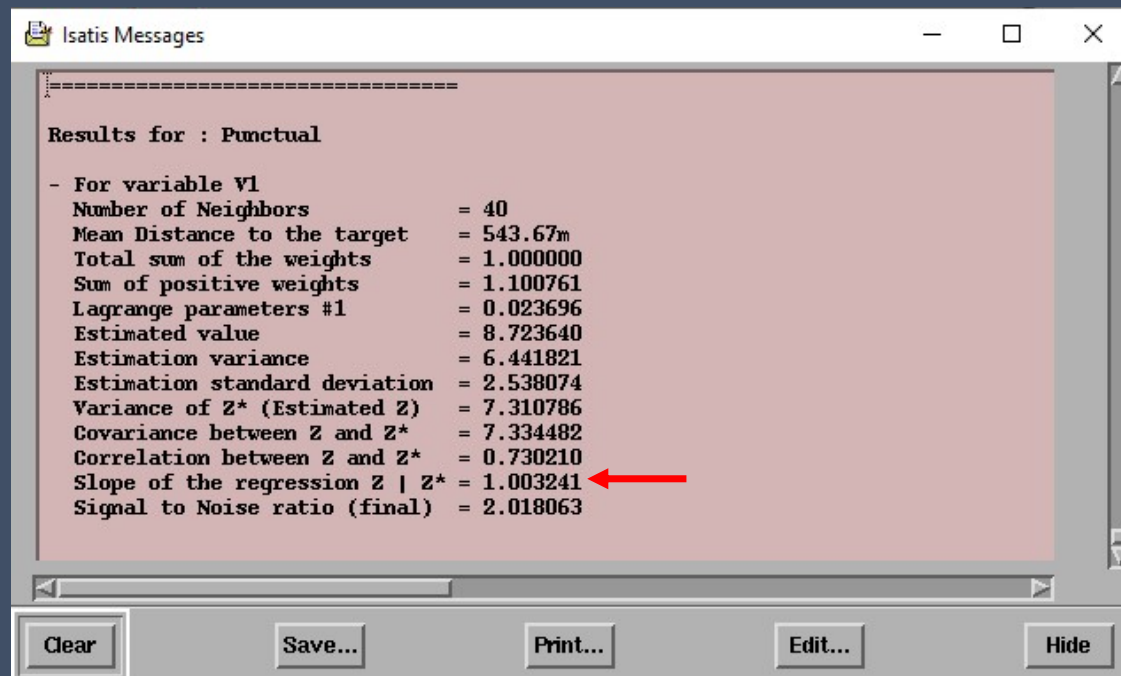
Optimum Number of Samples per Sector:

Test Print OK Cancel

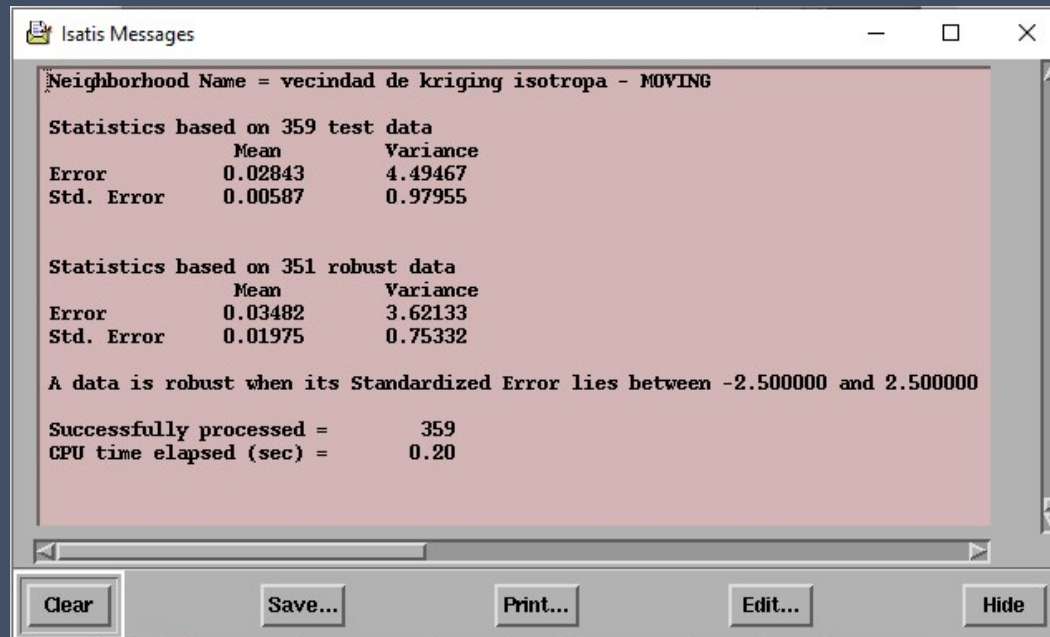
- Antes de ejecutar la validación cruzada, se hace una prueba con el botón “Test”.
- Aparece una ventana gráfica con la posición de los datos y un dibujo de la vecindad (circular, dividida en octantes). Permite ver si el tamaño de la vecindad es adecuado, así como probar el kriging para distintos datos.
- Al pinchar un dato con el mouse, se efectúa la validación de este dato, indicándose los datos vecinos seleccionados y sus ponderadores de kriging.



En *Application > Print Complete Information*, se tiene algunas estadísticas que permiten decidir si la vecindad es adecuada o no, en especial se puede apreciar si la pendiente de la regresión entre valor real (Z) y valor estimado (Z^*) (“Slope of the regression $Z | Z^*$ ” es cercana a 1. Una pendiente muy diferente de 1 puede significar que la vecindad es demasiado pequeña y no se encuentra suficientes datos para el kriging (habría entonces que aumentar el radio de búsqueda y/o aumentar el número de datos por sector angular). Para los datos periféricos, sólo se puede encontrar datos vecinos en un lado, por lo que la pendiente de la regresión puede alejarse de 1 (difícilmente mejorable).

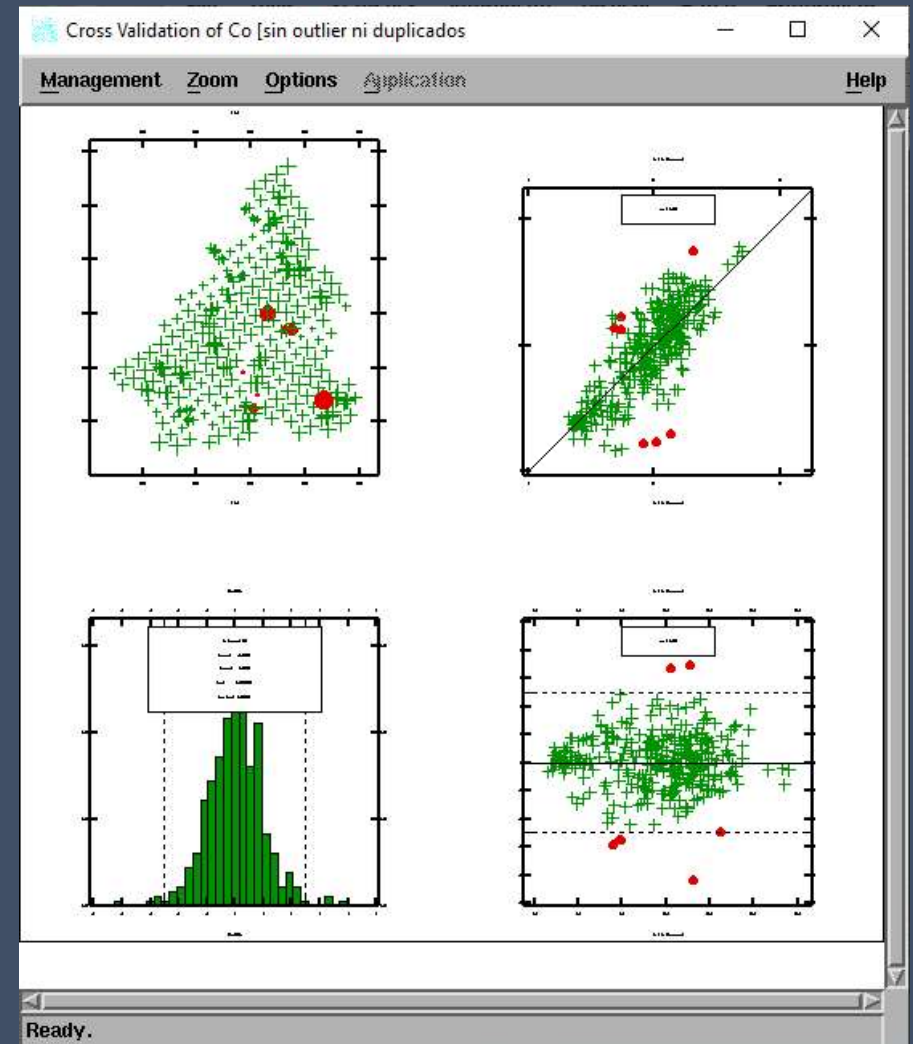


- Dado que la prueba es satisfactoria, se puede cerrar la ventana “Test”
- *Management > Delete Window* y ejecutar la validación cruzada (botón “Run”).
- Algunas estadísticas (media y varianza de los errores y errores estandarizados) aparecen en el buzón de mensajes. Idealmente, las medias de error deben ser cercanas a 0 (estimador sin sesgo), la varianza del error debe ser la más baja posible (estimador preciso) y la varianza del error estandarizado debe ser cercana a 1 (desviación estándar de kriging cuantifica adecuadamente la dispersión de los errores de estimación, lo que significa que el variograma modela adecuadamente la continuidad espacial de la variable, sin ser demasiado optimista – exagerando el alcance o subestimando el efecto pepita – o demasiado pesimista).



En este caso, la varianza de los errores de estimación es de 4.49.

- Mapa de ubicación de los datos donde se destaca los datos “mal” estimados (se puede cambiar el color a rojo en Preferences > Study Environment > Miscellaneous > Selected Outlier > Color). Deberían haber pocos datos mal estimados, ojalá no todos estén concentrados en la misma área del espacio.
- El histograma de los errores estandarizados. Debería ser centrado en 0 y poco disperso.
- La nube de correlación entre los valores reales y los valores estimados. Debería ser poco dispersa en torno a la línea diagonal, con una pendiente cercana a 1.
- La nube de correlación entre los errores estandarizados y los valores estimados. Debería ser poco dispersa en torno a la línea horizontal de ordenada 0, con una pendiente cercana a 0.



Se desea ahora probar el modelo de variograma anisótropo. En “Model”, se entra el **modelo de variograma anisotropo** y en “Neighborhood” se define una segunda vecindad llamada **“vecindad de kriging anisotropía”** (entrar el nombre en la zona de texto, luego presionar “Add” y “OK”). Los parámetros de esta vecindad se definen al presionar el botón “Edit”.

Cross-validation

Number of Variables: 1

Data File... datos / Muestras de suelo [Selection = sin outlier ni duplicados]

- [R] Target Variable #1 = Co
- [R] External Drift #1 = None
- [R] External Drift #2 = None
- [R] External Drift #3 = None
- [R] Local Mean for Target Variable #1 = None
- [R] Variance of Measurement error = None
- [R] Selection of Test Data = None

Target Variable... Co

Threshold on Standardized Score for Outliers: 2.5

☒ Graphic Representations

☐ Rescale the Structure Sills According to the Cross-Validation Scores

Model... modelo anisotropo Edit...

Special Model Options... No special model option

Neighborhood... vecindad de friging anisotropia Edit...

Special Kriging Options... All Variables ignored

Run Test Close

Neighborhood Definition

Load... Default Reset

Neighborhood Type **Moving**

Parameters

Sectors Advanced Sorting

Rotation... Az=30.00

Search Ellipsoid: Maximum Distances in the System after Rotation:

U 1500 m V 1000 m W 1000 m

☒ Use Anisotropic Distances (According to the Search Ellipsoid)

Dimensions of the Proximity Rectangle (for Punctual Estimation Only):

X 0 m Y 0 m Z 0 m

Minimum Number of Samples: 4

Number of Angular Sectors: 8

Optimum Number of Samples per Sector: 5

Test Print OK Cancel

3D Rotation Definition

☐ Horizontal Rotation

Convention **Geologist**

Rotation

	Axis	Angle [-180 +180]
1	A	60 deg
2	+X	0 deg
3	-Z	0 deg

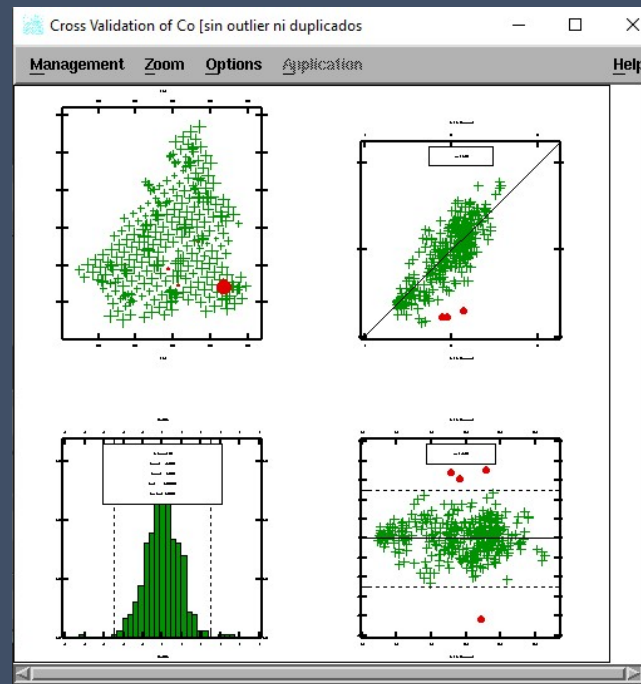
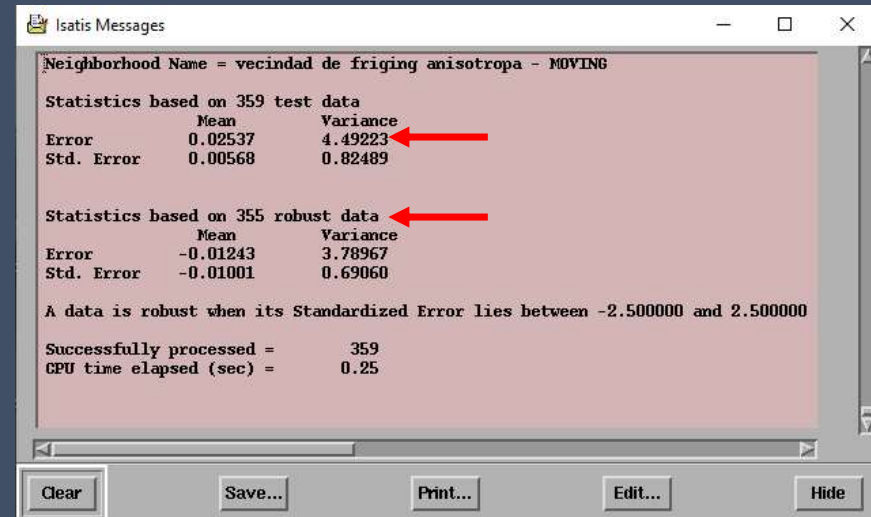
No Rotation

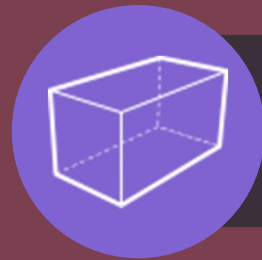
OK Reset

Nuevamente, se puede hacer pruebas preliminares con el botón “Test”.

Luego, se ejecuta la validación cruzada (botón “Run”), obteniendo los siguientes resultados.

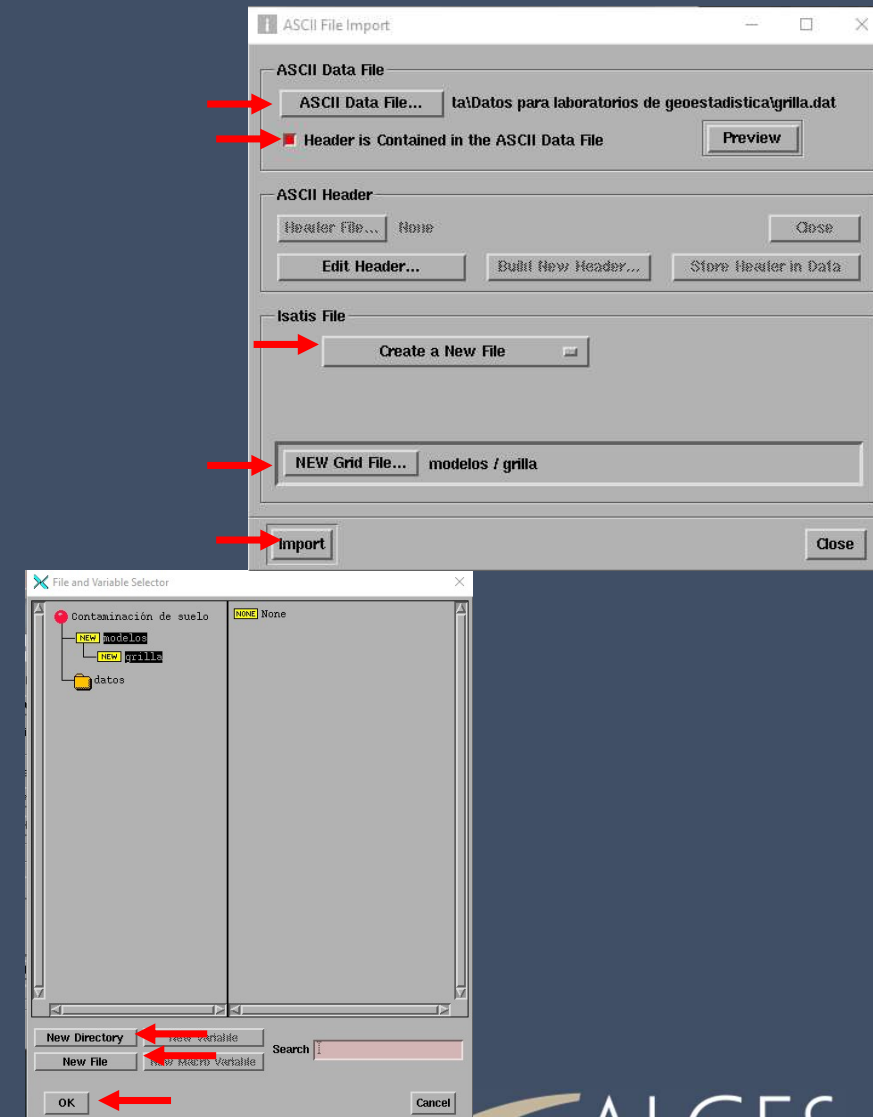
En este caso, la varianza de los errores de estimación es de 4.49 (lo mismo que con el modelo de variograma isótropo), pero hay un poco más de datos “bien estimados” (355 en lugar de 351). En resumen, los dos modelos de variograma (isótropo y anisótropo) son igualmente válidos.

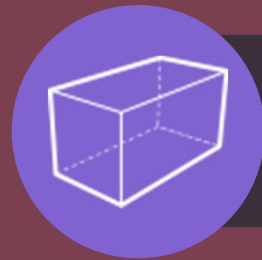




Importación de la grilla para interpolar la
concentración de cobalto

- *File > Import > ASCII*
- En “Ascii data file”, entrar el archivo “Grilla.dat” provisto junto con el archivo de datos. El encabezado está contenido en el archivo (dejar activada la opción “Header is contained in the Ascii data file”).
- El archivo contiene una sola columna que indica si el nodo de la grilla pertenece (código 1) o no pertenece (código 0) a la zona de interés. Nota: La grilla fue creada con el menú *File > Create Grid File*, con los siguientes parámetros: origen (0.315km ; 0.315km); malla (0.03km ; 0.03km); número de nodos (162,184). La zona fue delimitada con el menú *File > Selection > Geographic*, cargando la grilla, superponiendo un mapa de los datos (*Management > Background*) y utilizando la opción “Mask from polygon”
- Opción: “Create a new file”.
- New grid file: pinchar “New directory” y entrar un nuevo nombre (por ejemplo, “Modelos”), luego “New file” y entrar un nombre de archivo (por ejemplo, “Grilla”). Validar (botón OK).
- Importar la grilla (botón “Import”).



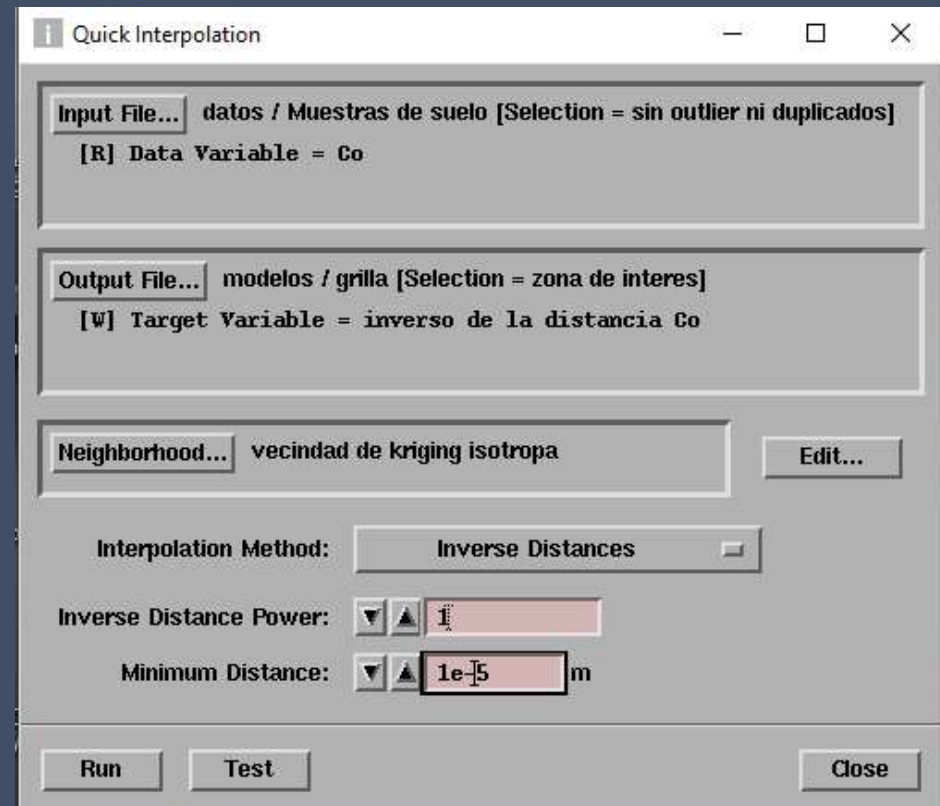


Estimación de las concentraciones de cobalto

Empezaremos con una **interpolación por el método de inverso de la distancia**.

Ir a *Interpolate > Interpolation > Quick Interpolation*

- “Input file”: Datos > Muestras de suelo > Sin outlier ni duplicado
- “Data variable”: Co
- “Output file”: Modelos > Grilla > Zona de interes
- “Target variable”: new variable. Entrar nuevo nombre “Inverso de la distancia Co” y validar (OK)
- “Neighborhood”: usaremos la vecindad de kriging isotropa en este caso
- Interpolation method: Inverse distances
- Inverse distance power: 1.0
- Minimum distance: $1e-5$
- Ejecutar (Run)

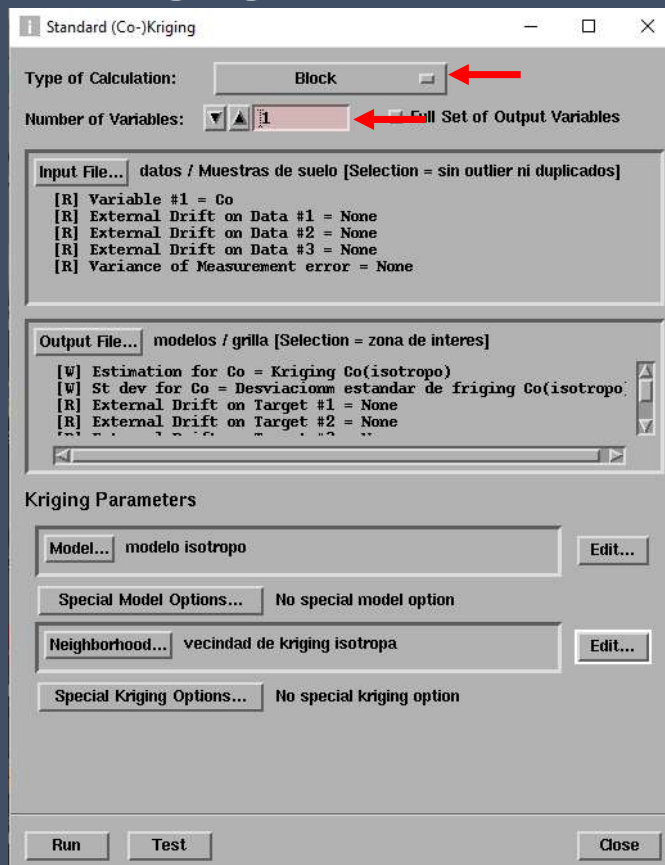


Seguimos con una **interpolación por kriging ordinario**.

Ir a Interpolate > Estimation > (Co)kriging

El co-kriging es la versión multivariable del kriging; ambos se hacen en el mismo menú.

1) Kriging con el modelo de variograma isótropo



“Calculation”: *block* (para estimar la concentración promedio de cobalto en cada bloque de la grilla, en lugar de la concentración puntual en el centro de cada bloque)

- Number of variables: 1
- No es necesario activar la opción “Full set of output variables” (permite tener outputs avanzados, tales como el valor del multiplicado de Lagrange del sistema de kriging ordinario, la suma de los ponderadores positivos, el ponderador de la media en el kriging simple, etc.)

Standard (Co-)Kriging

Type of Calculation:

Number of Variables: ☐ Full Set of Output Variables

Input File... datos / Muestras de suelo [Selection = sin outlier ni duplicados]

[R] Variable #1 = Co
[R] External Drift on Data #1 = None
[R] External Drift on Data #2 = None
[R] External Drift on Data #3 = None
[R] Variance of Measurement error = None

Output File... modelos / grilla [Selection = zona de interes]

[W] Estimation for Co = Kriging Co(isotropo)
[W] St dev for Co = Desviacionm estandar de friging Co(isotropo)
[R] External Drift on Target #1 = None
[R] External Drift on Target #2 = None

Kriging Parameters

Model... modelo isotropo

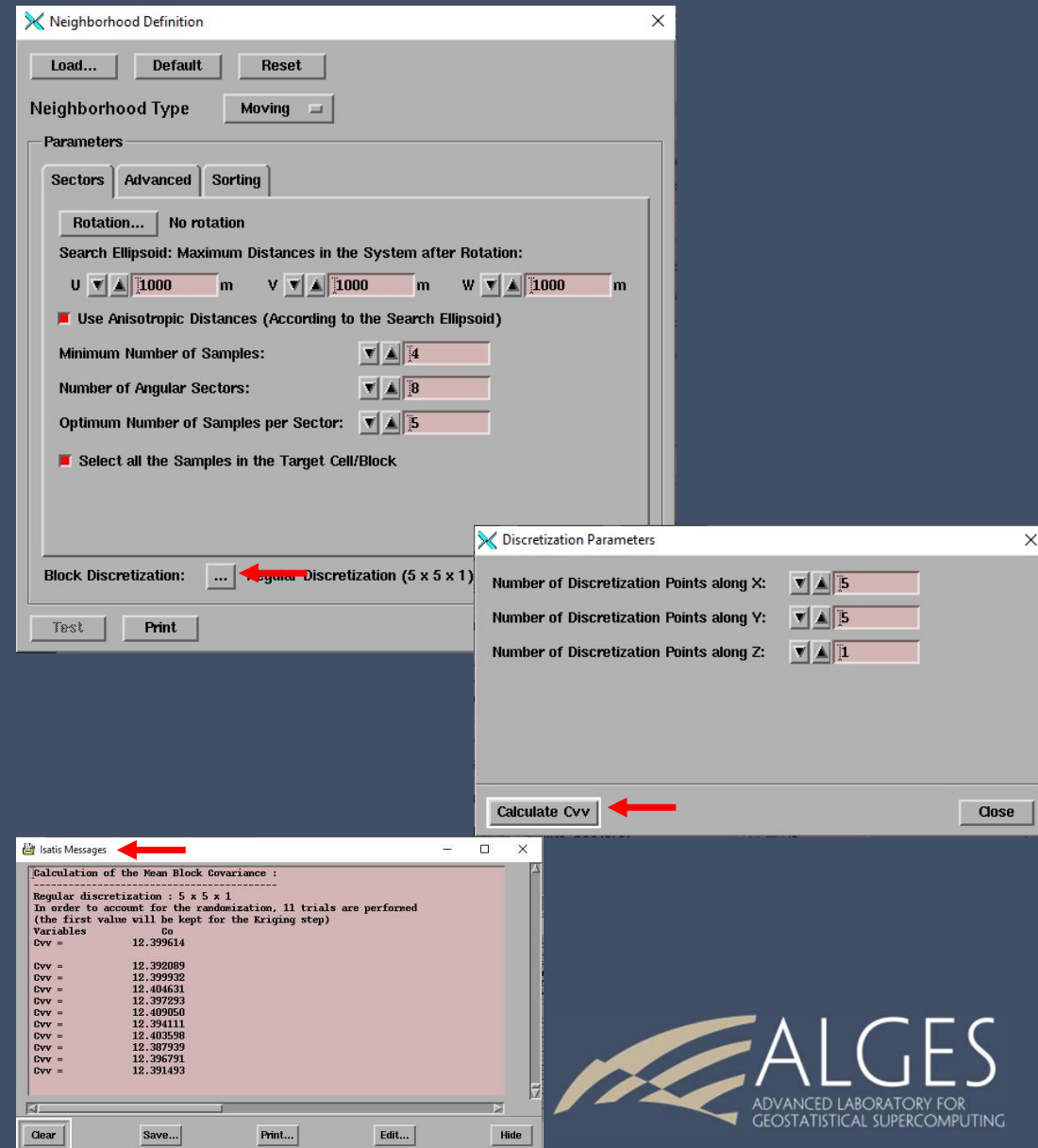
Special Model Options... No special model option

Neighborhood... vecindad de kriging isotropa

Special Kriging Options... No special kriging option

- “Input file”: Datos > Muestras de suelo > Sin outlier ni duplicado
- “Variable #1”: Co
- “Output file”: Modelos > Grilla > Zona de interes
- “Estimation for Co”: new variable. Entrar nuevo nombre “Kriging Co (isotropo)” y validar (OK)
- “St dev for Co”: new variable. Entrar nuevo nombre “Desviacion estandar de kriging Co (isotropo)” y validar (OK)
- “Model”: modelo de variograma isotropo
- “Special model options”: nada
- “Neighborhood” : vecindad de kriging isótropa
- “Special kriging options”: nada

En “*Edit*”, se define una discretización de bloque (*Block discretization*), dado que el kriging tiene que hacerse para estimar un valor promedio en cada bloque de la grilla. Por defecto, se tiene una discretización regular de $5 \times 5 \times 1$, es decir, cada bloque se discretiza en 25 puntos. Esta discretización del bloque se necesita para calcular el miembro de la derecha del sistema de kriging de bloques. Para validarla, se puede ir en *Block discretization* > *Calculate Cvv*, que permite calcular la varianza de la concentración de cobalto al soporte de bloque, usando el modelo de variograma y la discretización definida (Isatis entrega 11 intentos con cambios en los puntos que discretizan el bloque: si la discretización es suficiente, deberían haber pocos cambios entre los 11 intentos). Validar la discretización de $5 \times 5 \times 1$ (botón “close” y “OK”)



- Si hay tiempo, se puede usar el botón *Test* antes de *Run*. Tal como para la validación cruzada, permite ver si el tamaño de la vecindad es adecuado, así como probar el kriging para distintos datos. Aparecen los datos seleccionados con sus ponderadores de kriging. En *Application -> Print Weights and Results*, se tiene algunas estadísticas, en particular, la pendiente de la regresión entre valor real y valor estimado.
- Run

2) kriging con el modelo de variograma anisótropo

“Calculation”: block

- Number of variables: 1
- “Input file”: Datos > Muestras de suelo > Sin outlier ni duplicado
- “Variable #1”: Co
- “Output file”: Modelos > Grilla > Zona de interes
- “Estimation for Co”: new variable

Entrar nuevo nombre “Kriging Co (anisotropo)” y validar (OK)

- “St dev for Co”: new variable

Entrar nuevo nombre “Desviacion estandar de kriging Co (anisotropo)” y validar (OK)

- “Model”: modelo de variograma anisotropo
- “Special model options”: nada
- “Neighborhood” : vecindad de kriging anisotropa
Dejar la discretización de bloques por defecto (5*5*1)
- “Special kriging options”: nada
- Run

Standard (Co-)Kriging

Type of Calculation: Block

Number of Variables: 1 ☐ Full Set of Output Variables

Input File... datos / Muestras de suelo [Selection = sin outlier ni duplicados]

[R] Variable #1 = Co
[R] External Drift on Data #1 = None
[R] External Drift on Data #2 = None
[R] External Drift on Data #3 = None
[R] Variance of Measurement error = None

Output File... modelos / grilla [Selection = zona de interes]

[W] Estimation for Co = Kriging Co (anisotropo)
[W] St dev for Co = Desviacion estandar de kriging Co (anisotropo)
[R] External Drift on Target #1 = None
[R] External Drift on Target #2 = None
[R] External Drift on Target #3 = None
[R] Collocated Variable = None

Kriging Parameters

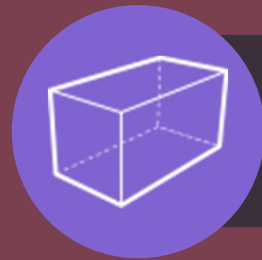
Model... modelo anisotropo Edit...

Special Model Options... No special model option

Neighborhood... vecindad de friging anisotropa Edit...

Special Kriging Options... No special kriging option

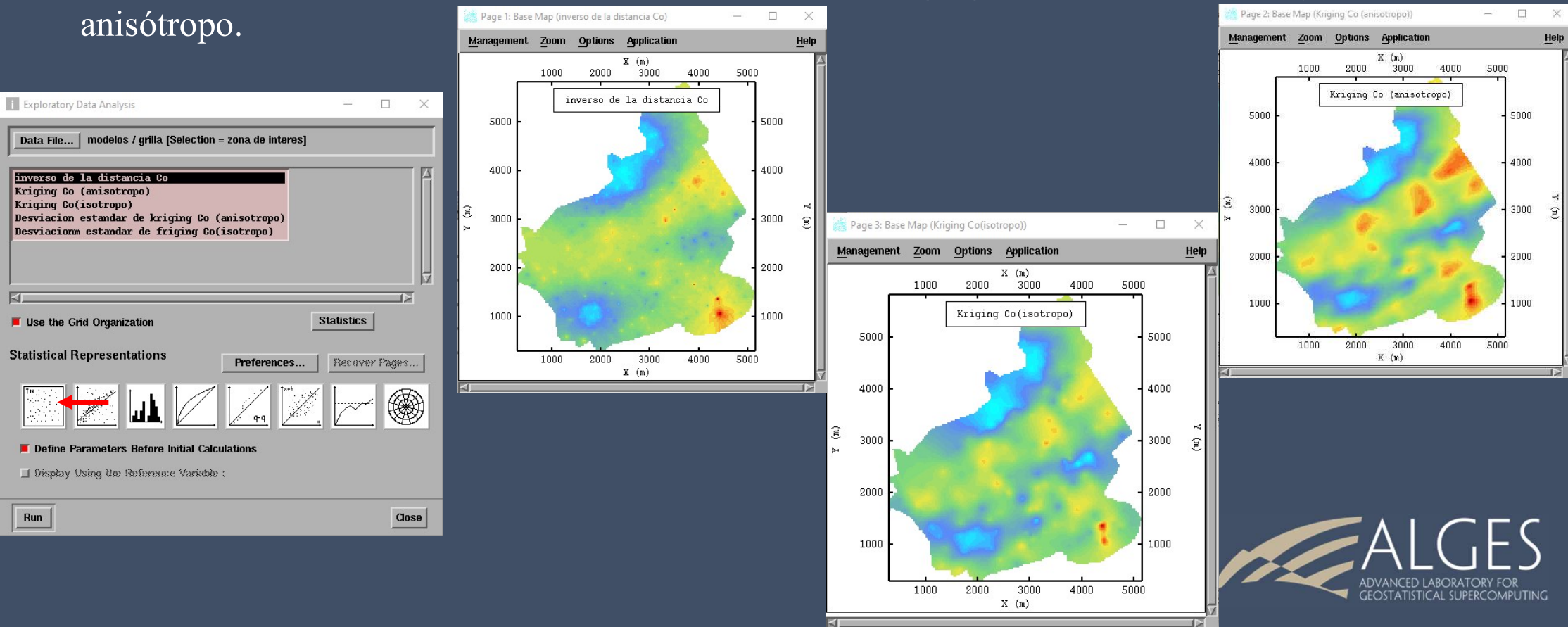
Run Test Close



Comparación de resultados

Los resultados de las interpolaciones pueden verse en el menú *Statistics > Exploratory Data Analysis*

- **Mapas** de los valores estimados. En todos los casos, se aprecia el suavizamiento producido por los métodos de interpolación, con algunos artefactos para el método de inverso de la distancia (los datos tienden a formar extremos locales). El kriging permite tomar en cuenta la continuidad espacial de la variable, tal como se aprecia al comparar los mapas de kriging con modelo isótropo y con modelo anisótropo.



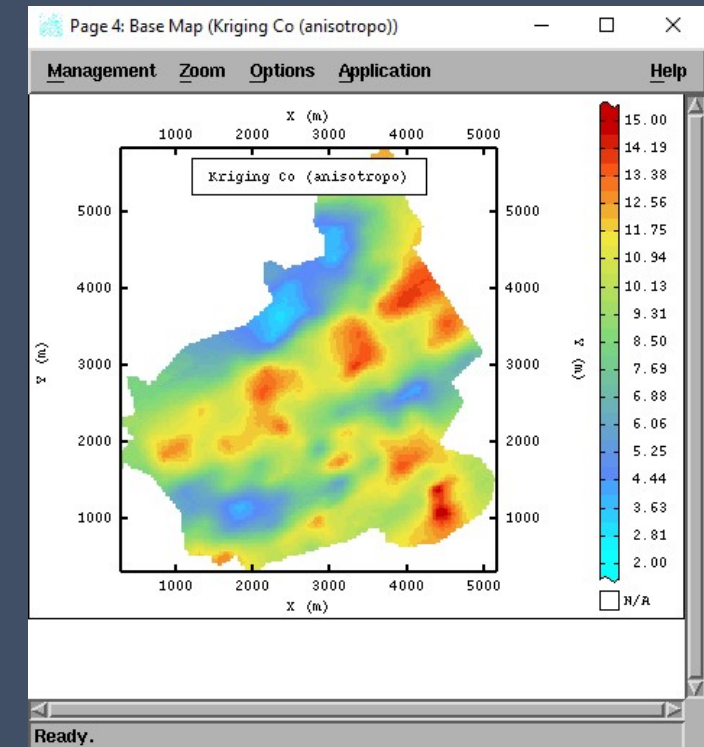
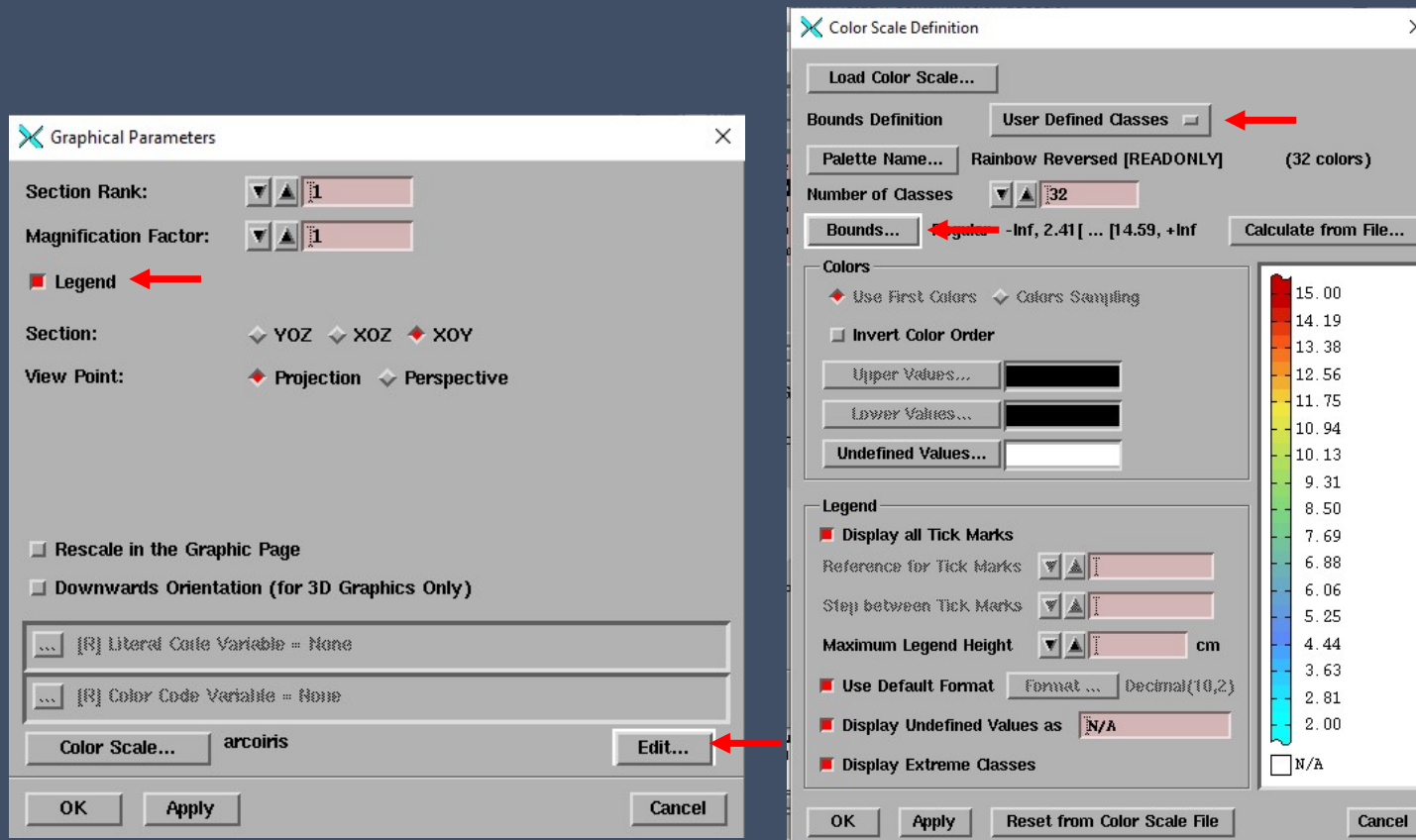
La escala de colores se puede desplegar y modificar en *Application > Graphic Specific Parameters*

Despliegue: activar la opción “Legend”

Modificar la escala de color: ir a Color Scale > Edit:

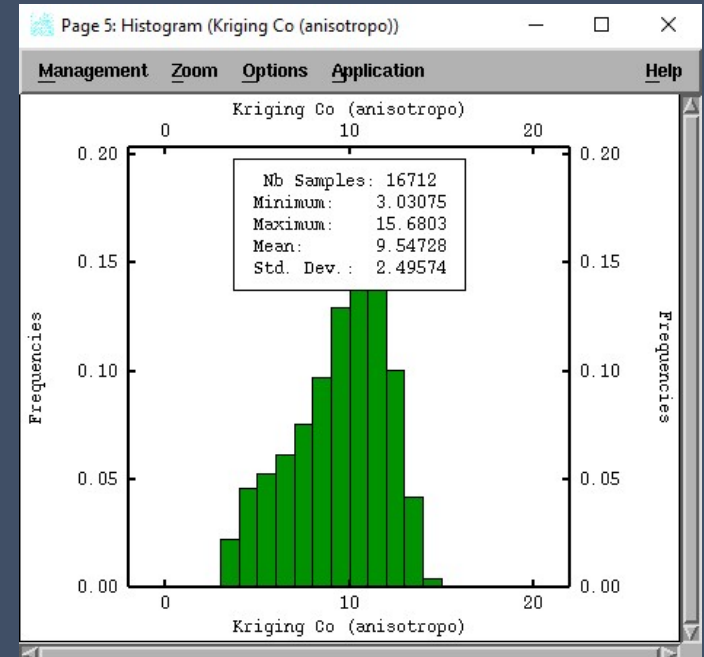
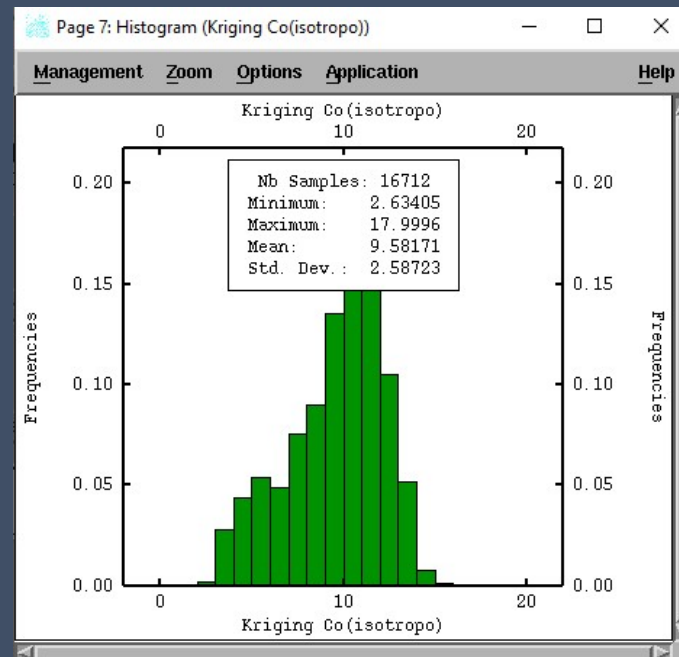
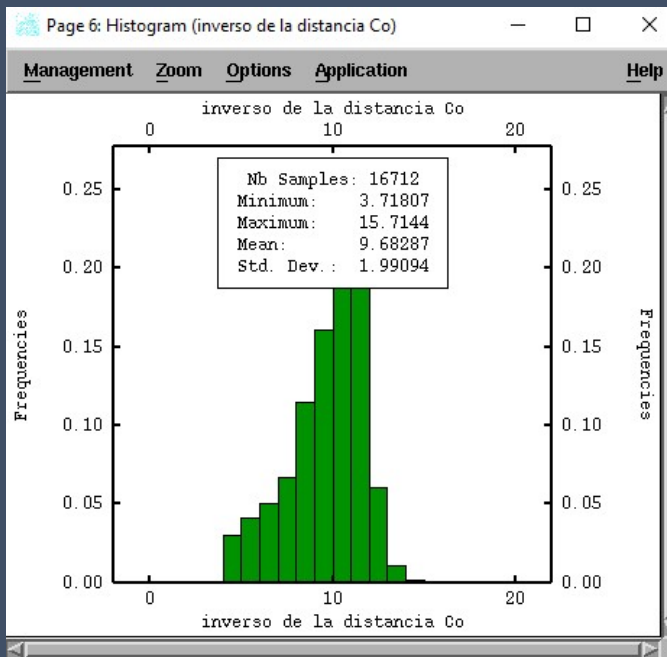
Bounds definition = “User defined classes”

Bounds: minimum = 2; maximum = 15

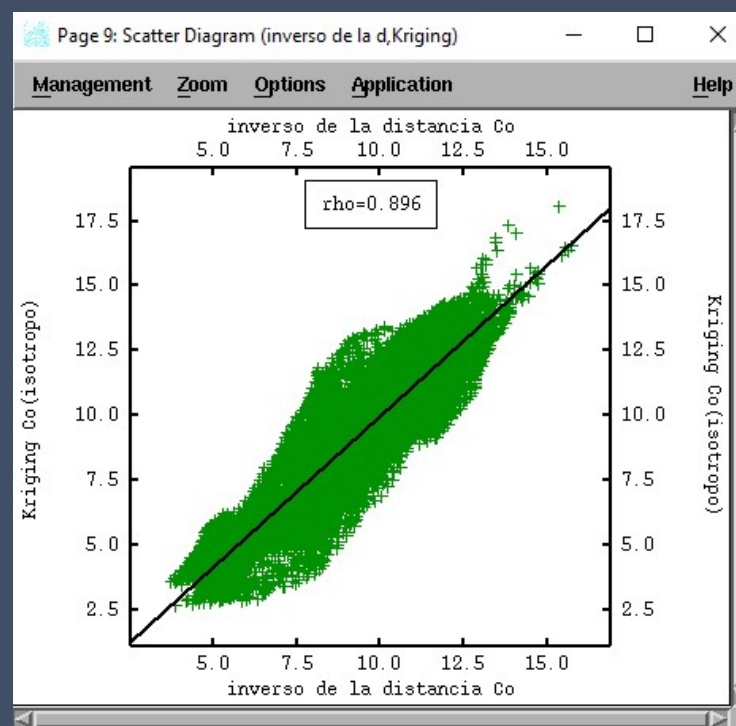
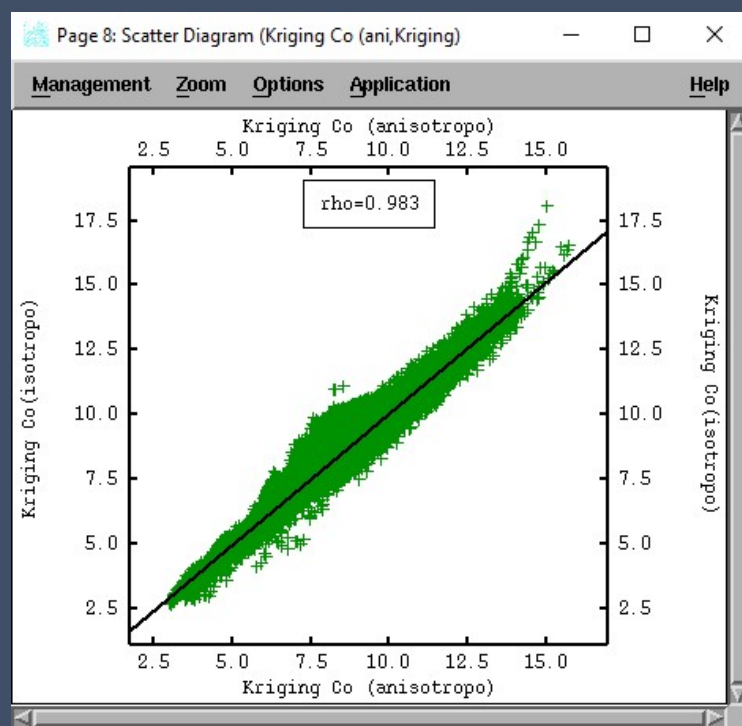


Histogramas y estadísticas de los valores estimados

Los parámetros de los histogramas se pueden modificar en *Application > Calculation Parameters*. Por ejemplo, mínimo = 0, máximo = 20, número de clases = 20. Notar que los histogramas de valores estimados son más estrechos que el histograma de los datos, con aproximadamente la misma media, pero menos varianza. Esto refleja la propiedad de suavizamiento.



Comparación de los valores estimados por kriging con modelo isótropo y con modelo anisótropo, por medio de una nube de dispersión. Las estimaciones son relativamente similares (nube de puntos cercana a la línea a 45°), no así al comparar kriging con inverso de la distancia.



Mapas de las desviaciones estándares de kriging. Se aprecia que esta desviación es pequeña en la cercanía de los datos y aumenta al alejarse de ellos, especialmente en la zona periférica oeste-noroeste. También, la desviación del error aparece como independiente del valor de los datos (sólo depende de la configuración espacial de los datos y puntos a estimar).

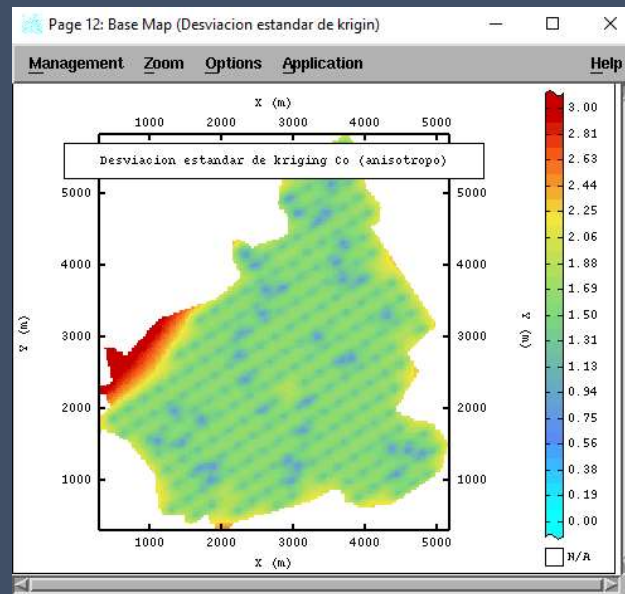
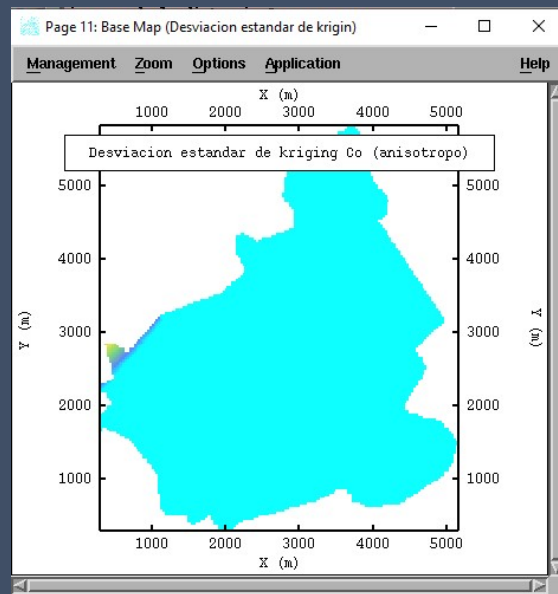
Para el despliegue, se sugiere modificar la escala de colores en *Application > Graphic Specific Parameters*

Activar la opción “Legend”

Modificar la escala de color: ir a Color Scale > Edit:

Bounds definition = “User defined classes”

Bounds: minimum = 0; maximum = 3



ISATIS 5

Análisis Estadístico y Geoestadístico de Datos
13 de Diciembre de 2019

Auxiliar: Fabián Soto F.
Profesor: Xavier Emery

