

Selección óptima del método de interpolación en el tratamiento de las bases datos: una aplicación en la construcción de cartas magnéticas

Eduardo Pérez⁽¹⁾, Lourdes Ma. García⁽¹⁾, Rosa Álvarez⁽¹⁾, Martín Díaz⁽¹⁾

⁽¹⁾ Instituto de Geofísica y Astronomía. Calle 212. No.2906 entre 29 y 31, Rpto. La Coronela. La Lisa, Ciudad de La Habana. Cuba

Recibido: enero-octubre, 2000	Aceptado: noviembre, 2000
-------------------------------	---------------------------

Resumen

El presente estudio está encaminado a evaluar la efectividad de dos métodos de interpolación, el Kriging y el Inverso de la Distancia (ID), aplicados sobre diferentes conjuntos de datos correspondientes al valor de la componente anómala T_a en dos regiones del Levantamiento magnético vectorial realizado en la República de Cuba en el período 1985-1988. Para comprobar la eficiencia de estos interpoladores se determinó el error medio cuadrático (EMC) por el método de la validación cruzada, el EMC producto de las diferencias entre los valores reales e interpolados en la red y el coeficiente de correlación entre los valores reales y estimados. En el caso del ID es muy importante la selección correcta de la potencia, ya que esta influye considerablemente en la ponderación. Para estos datos es adecuado seleccionar elevadas potencias con pocos números de puntos vecinos. La validación cruzada no constituye un criterio concluyente acerca de la calidad de la interpolación. El método de interpolación que ofreció los mejores resultados fue el ID y por consiguiente fue utilizado en la confección de las cartas magnéticas. La elección óptima de cualquier método de interpolación debe estar regido por una evaluación de su eficiencia.

Palabras clave: cartas magnéticas, magnetismo.

Abstract

The aim of this study is to evaluate the effectiveness of two spatial interpolation methods (Kriging and Inverse to the Distance - ID) applied over different groups of magnetic data. In order to assess the efficiency of these interpolators different parameters were estimated such as: the Mean Square Error (MSE) from the results of the Crossed Validation method, the MSE from the differences between the real and interpolated values over a grid dataset and the correlation coefficient between the real and estimated values. For these dataset good results were achieved using the ID method with high powers and a few numbers of neighboring points, therefore it was used to produce the final grids of the magnetic data. The Crossed Validation method doesn't constitute a conclusive approach to assessing the quality of the interpolation procedure. The suitable selection of any interpolation method should be guided by an evaluation of its efficiency.

Key words: magnetic charts, magnetism.

1. Introducción

En la construcción de mapas de campos potenciales, magnéticos, gravimétricos y de isolíneas en general, juegan un papel importante los métodos de interpolación; en muchos casos se hace imprescindible la utilización de estos, cuando se cuenta con redes no homogéneas y se necesita realizar transformaciones al campo estudiado, aplicar métodos de interpretación tanto cualitativos como cuantitativos, procesamiento de imágenes, entre otros, que requieren de redes homogéneas para su aplicación.

El problema de evaluar la calidad entre diferentes interpoladores espaciales ha sido objeto de estudio por distintos autores (Hewlett, 1964; Knudsen et al., 1978; Raymond, 1979; Raymond, 1982). Recientemente se ha demostrado que de forma general, la calidad de la interpolación está condicionada a las características propias de cada base de datos (Weber y Englund, 1993), a la aplicación de diferentes metodologías en el procesamiento de los datos que puedan ser llevadas a cabo por distintos investigadores (Englund, 1990; Weber, Englund, 1992), así como del método usado para evaluar la misma.

En el presente trabajo se evalúa el efecto de dos métodos de interpolación, el Kriging Ordinario Puntual (KOP) y el Inverso de la Distancia (ID) con sus variantes, los cuales son aplicados sobre dos regiones que difieren entre sí estadísticamente, con el objetivo final de proporcionar una herramienta que ofrezca la posibilidad de establecer la eficiencia de los mismos con vistas a una mayor precisión en la construcción de cartas magnéticas.

2. Materiales de partida

Las bases de datos utilizadas corresponden con la componente total anómala (T_a) del campo geomagnético, correspondientes al Levantamiento Vectorial a escala 1:1 000 000 realizado en la República de Cuba en el período de 1985 a 1988, por el Instituto de Geofísica y Astronomía. En la figura 1 se muestra la representación tridimensional de T_a correspondiente a la Región Occidental (ROc), con un total de 290 puntos distribuidos con una densidad de 4 puntos por cada 100 km² como promedio; la figura 2 representa la Región Oriental (ROr), cuya red contiene 602 puntos con una densidad similar a la anterior. En las figuras 3a y 3b se encuentran sus respectivos histogramas. Ambos conjuntos de datos se diferencian ostensiblemente, en cuanto a las características del campo que describen (tabla I), siendo la ROc algo más estable que la ROr, garantizando un espectro más amplio de las distintas realidades que puedan estar presente.

Tabla I: Estadígrafos de la región Occidental y Oriental

Estadígrafos	Reg. Occ.	Reg. Ori.
# de Puntos	290	602
Media (nT)	11.17	-16.77
Varianza (nT ²)	18315.89	55230.93
Desv. Est. (nT)	135.34	235.01
Mínimo (nT)	-338	-965
Máximo (nT)	520	1137
1 ^{er} cuartil (nT)	-70	-145
Mediana (nT)	-2	-27
3 ^{er} cuartil (nT)	72	74
Simetría	.75	.85
Curtosis	1.9	6.7

2. Aspectos teóricos.

Estimadores

Los métodos de interpolación empleados fueron el KOP y el ID. De forma general la estimación en un punto es calculada teniendo en cuenta los valores de los puntos más cercanos a él, a través de la expresión:

$$Z_e = \sum_{i=1}^n \alpha_i Z_i \quad (1)$$

n: # de puntos cercanos seleccionados para la estimación de Z_e (punto a estimar).

Z_i : valores de los puntos seleccionados para la estimación.

μ_i : pesos determinados para cada Z_i los cuales son calculados indistintamente para ambos interpoladores.

El ID es un interpolador de promedios ponderados, donde la potencia controla el factor de ponderación y los μ_i son calculados:

$$\alpha_i = \frac{\frac{1}{r_i^p}}{\sum_{i=1}^n \frac{1}{r_i^p}} \quad (2)$$

r_i : distancia entre el punto a estimar y los i -puntos vecinos. p : potencia.

El Kriging es considerado un estimador lineal insesgado cuyos μ_i son calculados de modo tal que el estimador cumpla con la condición de insesgadez y que la varianza de la estimación s^2_e sea mínima, a partir de estas condiciones es planteado un sistema de ecuaciones que en el caso del KOP es:

$$\sum_{i=1}^n \alpha_i = 1$$

$$\sum_{i=1}^n \alpha_i \sigma_{ji} - \mu_i = \sigma_{jx_0} \quad (3)$$

m : Multiplicador de Lagrange. s_{ji} : Función de covarianza. (x_0 , punto a estimar)

cuya solución da como resultado estos μ_i . La teoría al respecto se encuentra descrita en (Journel and Huijbregts, 1978; Isaaks and Srivastava, 1989).

Estadígrafos

Dentro de los estadígrafos utilizados para evaluar la calidad de los interpoladores se encuentran:

- Error medio cuadrático producto de la validación cruzada, EVC
- Error medio cuadrático producto de la interpolación, EI.
- Función estandarizada, Z_{std}
- Coeficiente de correlación entre los valores reales y estimados, para las validaciones cruzadas y las interpolaciones.

La validación cruzada no es más que realizar la estimación de los puntos medidos mediante el interpolador seleccionado, empleando para ello los valores muestrales en una vecindad dada, exceptuando el valor del punto a estimar. El EVC se expresa a través de los valores reales y estimados como:

$$EVC = \frac{\sum_{i=1}^N (Z_{r_i} - Z_{e_i})^2}{N - 1} \quad (4)$$

N : número total de puntos a estimar Z_{r_i} : valor real del punto i -ésimo

Z_{e_i} : valor estimado para el punto i -ésimo

El segundo estadígrafo utilizado se determina cuando a partir de una distribución no homogénea de puntos muestrales, es confeccionada una red regular según un criterio de interpolación, y para cada punto real el valor

estimado es el resultado de la interpolación entre los puntos de la red obtenida. De este modo EI es calculado de forma similar a (4). Con vistas a establecer una comparación mucho más clara, los EVC y EI fueron estandarizados Zstd, para cada región.

$$ME = \frac{1}{Ne} \sum_{i=1}^{Ne} E_i \quad (5)$$

$$SD = \frac{1}{Ne} \sum_{i=1}^{Ne} (E_i - ME)^2 \quad (6)$$

$$Zstd_i = -\frac{E_i - ME}{SD} \quad (7)$$

ME: media de todos los errores. SD: desviación estándar de los errores.

Ne: número de interpoladores utilizados. Ei: errores medios cuadráticos (EVC y EI).

de modo que el Zstdi de mayor valor positivo, será el de menor error y por tanto el mejor aproximado; y el más negativo, el de mayor error. También como evaluador importante se utilizó el coeficiente de correlación entre los valores reales y estimados.

3. Procesamiento de los datos

Antes de comenzar la evaluación se hizo un exhaustivo estudio para la determinación de los posibles outliers espaciales y/o distribucionales que estuvieran presentes en los datos y que pudieran influenciar negativamente en el proceso de interpolación en general.

Para obtener una caracterización de la estructura espacial del campo magnético y con vistas a realizar el Kriging se obtuvo el semivariograma experimental, que teóricamente es la función que mejor refleja las características espaciales de la variable en estudio.

Fueron seleccionados los modelos que mejor se ajustaban para la ROc y RO_r, figuras 4a y 4b, basándonos para ello, en el criterio de información de Akaike (Akaike, 1973). Este criterio nos da la bondad de ajuste del modelo seleccionado al semivariograma experimental.

Después se procedió con el sistema GeoEAS y tomando como modelos del semivariograma de entrada los seleccionados anteriormente, debido a que en el GeoEAS la obtención del variograma se hace trabajoso y en ocasiones imposible, se realizó la validación cruzada variando el número de vecinos en cuatro categorías, siempre cuatro puntos como mínimos y 4,10,16,24 puntos como máximos (tabla II).

Estas vecindades fueron elegidas bajo el criterio, de que todos los puntos cercanos estuvieran incluidos dentro del radio de correlación o alcance de la muestra. Más tarde fueron confeccionadas las redes correspondientes con el KOP.

Tabla II: Errores medios cuadráticos de la validación cruzada: Región Oriental.

Vecindad	4x4	10x4	16x4	24x4	radio (km)
GeoEAS	47508.59	45132.70	45147.80	45315.59	40
ID ₁	47806.23	45269.52	45556.66	46386.80	40
ID ₂	48984.51	46285.63	45991.34	46048.01	40
ID ₃	50385.03	48072.95	47683.66	47507.80	40

Concluida la etapa con el Kriging se realizó el procesamiento con el ID. Para este último no hace falta modelos de semivariogramas, por lo que directamente se realizó la validación cruzada para distintas potencias, con las vecindades antes mencionadas y finalmente fueron construidas sus respectivas redes.

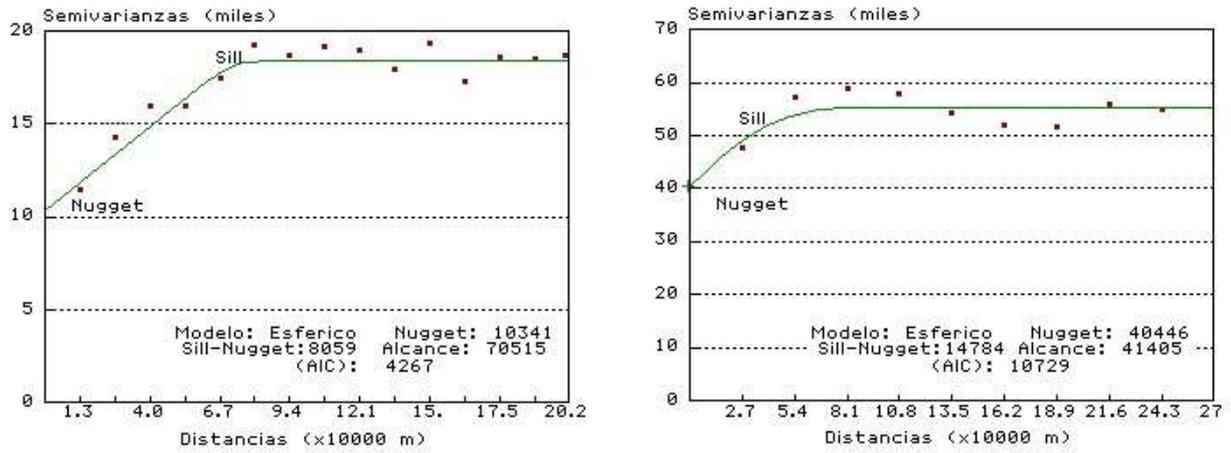


Figura 4: a) Semivariograma de la Región Occidental. b) Semivariograma de la Región Oriental.

Con esta información, para ambas regiones procesadas, se determinaron los siguientes parámetros:

- EVC, tabla II (Región Oriental).
- EI, tabla III (Región Oriental).
- Zstd, figura 6 (Región Oriental).
- Coeficientes de correlación de las validaciones cruzadas, CVC y de las interpolaciones, CI.

Tabla III: Errores medios cuadráticos de la Interpolación: Región Oriental.

Vecindad	4x4	10x4	16x4	24x4	radio (km)
GeoEAS	24885.04	31555.17	32299.45	35281.29	40
Id1	8589.47	13533.40	16251.97	19102.13	40
Id2	6100.30	7217.18	7692.76	8162.57	40
Id3	5678.70	5960.56	6036.82	6103.43	40

Finalmente, fueron construidos diferentes mapas utilizando algunos de los interpoladores estudiados, con vista a comparar los resultados matemáticos con los geólogos-geofísicos.

4. Discusión de los resultados

Del análisis estructural de los datos se puede apreciar que la correlación espacial para la ROc es de 70 Km y 40 Km para la ROr. En ambos casos los altos valores del nugget (ver figuras 4a y 4b) alcanzan un 56% para la ROc y hasta un 73% en la ROr de sus varianzas respectivas, reflejando los grandes gradientes del campo. De este análisis se observó que no existía tendencia en ninguna dirección por lo que el kriging ordinario podría ofrecer buenos resultados, ya que entre sus requerimientos de aplicación está la no existencia de la misma en los datos.

En el caso de la validación cruzada queda claro que ella no constituye un buen criterio de selección del mejor interpolador.

En la figura 5 vemos que para la validación cruzada los Zstd son negativos en comparación con los de la interpolación, además como se observa en la tabla II el GeoEAS(10x4) tiene el menor valor EVC y sin embargo es el que más desviado está de los valores reales, de ahí que la utilización de esta técnica sea tal vez necesaria pero no suficiente. El EI es mucho más adecuado y para nuestros datos responde con mayor claridad.

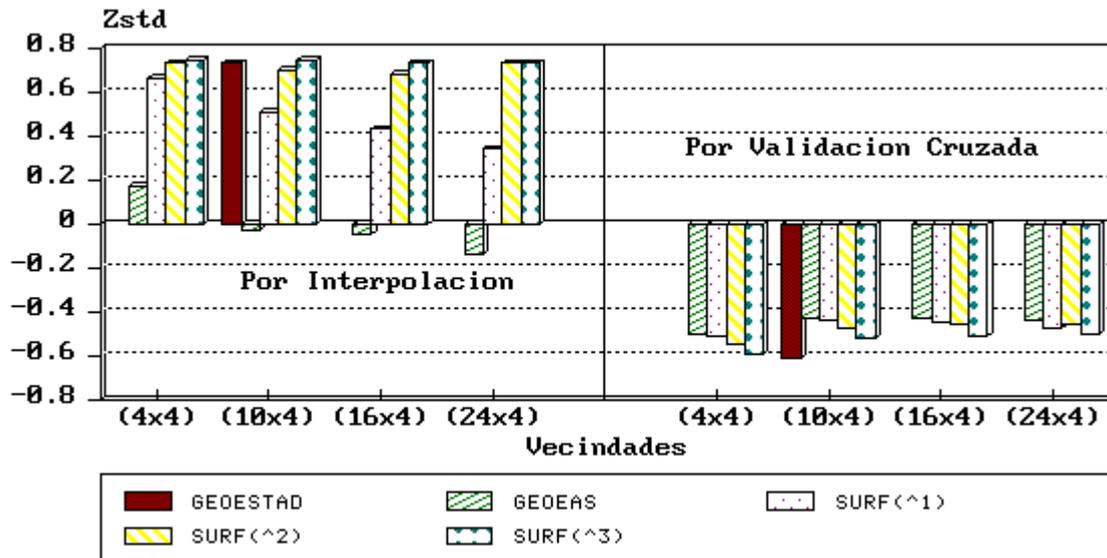


Figura 5: Función estandarizada Zstd, Región Oriental.

El KOP de GeoEAS presenta un pobre comportamiento, esto se explica por la gran varianza que presentan los datos; al tener un nugget elevado la estimación es menos confiable ya que el Kriging se comporta como un "interpolador suavizador" (comparar figuras 6 y 7).

Se puede apreciar en la tabla III que en el ID la selección de diferentes vecindades para altas potencias, no ofrece resultados interesantes no siendo así para pequeñas potencias en el cual a medida que aumentan las vecindades el EI es mayor, por tanto Zstd disminuye (ver figura 5). Esto se debe a que el aumento o disminución de la potencia es un reflejo del peso que tendrán los puntos vecinos durante la estimación de un punto cualquiera; significando que para grandes potencias, dentro de los puntos vecinos, sólo los que están mucho más cerca del punto a estimar son los que aportan un mayor peso, el resto casi es despreciable. Todo esto se puede ver claramente en la figura 8, con pocos puntos vecinos y potencia 1 son relativamente similares los resultados en comparación con 10 puntos y potencia 8.

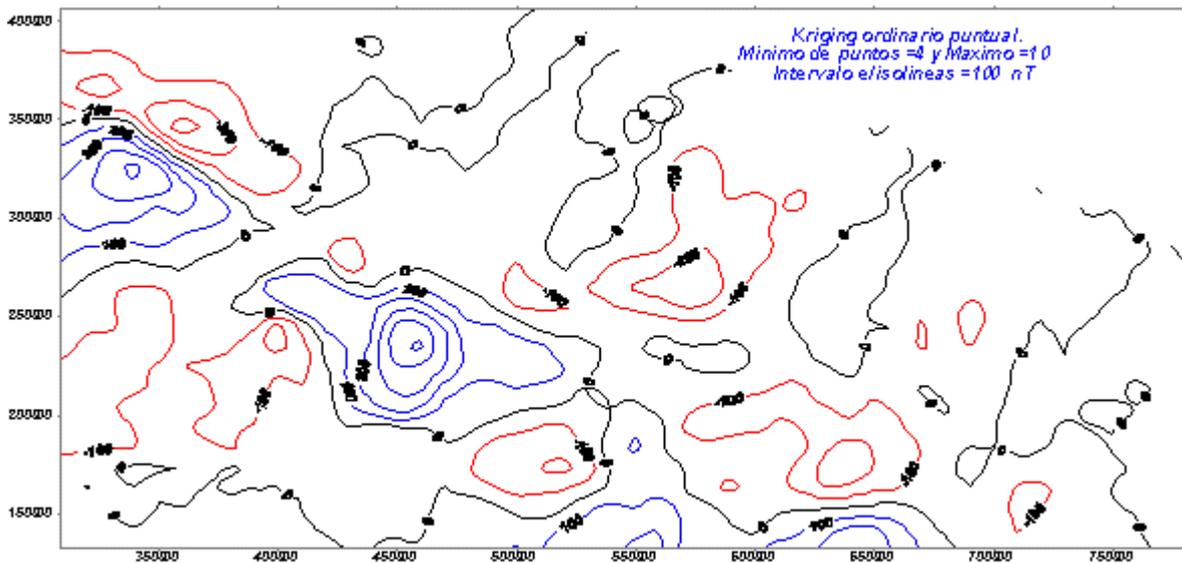


Figura 6: Mapa de isolíneas de la Región Oriental utilizando el Kriging Ordinario Puntual.

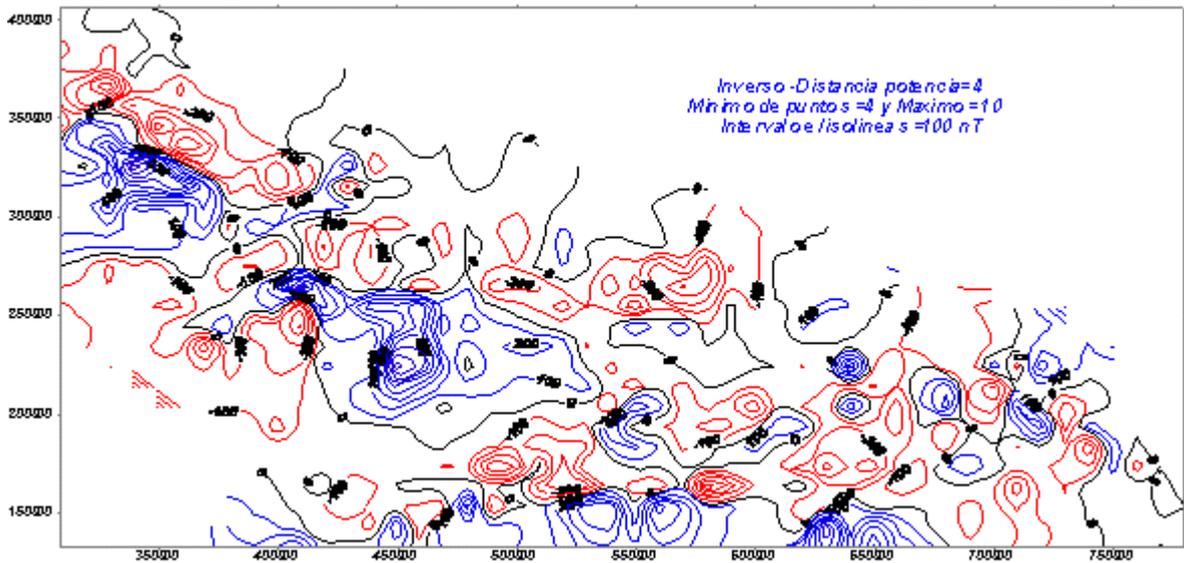


Figura 7: Mapa de isólinas de la Región Oriental utilizando el Inverso de la Distancia.

La varianza nos da un criterio de la eficiencia de la interpolación, para la ROc esta es de 18315 y el mínimo EI es de 1600, mientras que para la ROr es igual a 55230 y el error de 5679, esto quiere decir que a mayor varianza la estimación es menos precisa. Empíricamente y utilizando diferentes conjuntos de datos de campo magnético anómalo procedentes del mismo levantamiento vectorial encontramos una relación entre la varianza de la muestra s^2 y el mínimo error de interpolación EI (min.) según esta metodología:

$$EI(\min) \approx s^2 / 10.67 \quad (8)$$

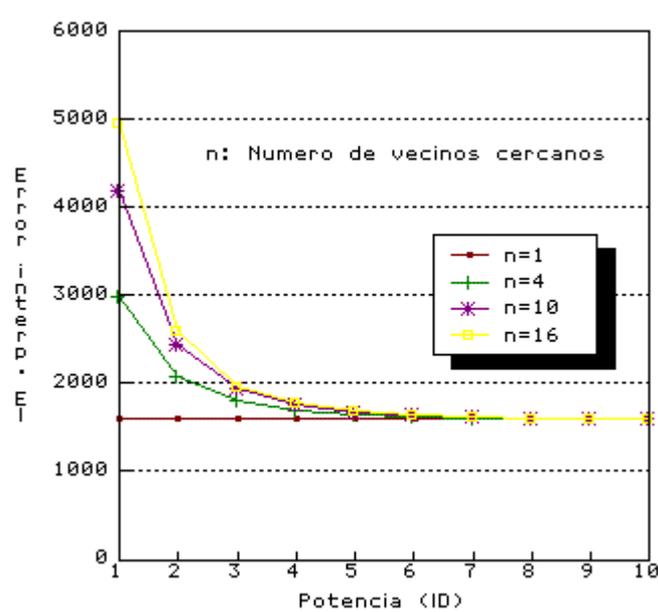


Figura 8: EI vs Potencia (ID).

El coeficiente de correlación, CVC, calculado por validación cruzada, y el obtenido finalmente a partir de las redes, CI, para una misma región e interpolador, refleja igualmente lo antes expuesto en cuanto a la validación cruzada, además el CI es menos sensible a cambios con respecto al EI.

La interpolación por el Kriging requiere de varios pasos tales como el cálculo del variograma, ajuste del modelo y confección de la red y el ID solamente requiere de la confección de la red, de esta comparación se dedujo que si los dos métodos dan resultados similares e incluso en muchas ocasiones el ID los ofrece mejores, entonces desde el punto de vista práctico y tiempo de ejecución podría ser obviado el Kriging como método de interpolación para estos datos.

Conclusiones y recomendaciones

La validación cruzada no constituye una técnica apropiada a partir de la cual podríamos tomar un criterio concluyente de la calidad de la interpolación.

Cuando los datos presentan una varianza elevada los métodos de interpolación conducen a resultados menos precisos, siendo más crítico para el Kriging, mientras que para el ID esto es menos drástico.

Para este caso es más factible la utilización del ID porque ofrece buenos resultados y conlleva menos tiempo de ejecución.

Es importante para el ID la selección correcta de la potencia, ya que esta influye considerablemente en la ponderación. Para estos datos es adecuado seleccionar elevadas potencias o potencias medias con pocos números de puntos vecinos.

El EI y el CI son buenos estadígrafos para la evaluación de la interpolación.

Seleccionamos finalmente para estos datos, el ID con potencia 8 y 4 puntos vecinos como interpolador óptimo y fueron utilizados en la confección de los mapas anómalos del campo geomagnético para el territorio cubano a escala 1:1 000 000. Es recomendable la confección de mapas de errores que muestren el grado de confiabilidad o distorsión de la estimación en regiones determinadas, permitiendo al usuario de dicho mapa conocer hasta que punto puede ser usada esa información, lo cual es de gran importancia ya que la utilización incorrecta de un interpolador puede traer consigo grandes divergencias entre la realidad y la interpolada por estos métodos, alterando así su posterior interpretación.

Los resultados que han sido expuestos, responden a los datos utilizados y cualquier conclusión en cuanto a la elección de un interpolador u otro dependerá en gran medida de cada base de datos, por tanto el uso de algún método de interpolación debe estar regido por una evaluación de su eficiencia.

Bibliografía

- Akaike, H., 1973. Information Theory and an extension of Maximum likelihood principle., in Petrov, B.N. and Csaki, F. (ed.) Second International Symposium on information theory (Academia Kiadó, Budapest).
- Englund, E.J., 1990. A Variance of Geostatisticians. Mathematical Geology, Vol 22.
- Hewlett, R.F., 1964. Comparison of the triangular, polygonal and a statistical method of computing grade and tonnage of ore for the Silver Bell oxide porphyry copper deposit: U.S. Bureau of Mines Report of investigation 7331.
- Isaaks, E.H., R.M. Srivastava, 1989. An Introduction to Applied Geostatistics. Oxford University Press, New York.
- Journel, A.G., CH.J. Huijbregts, 1978. Mining Geostatistics. Academic Press, London.
- Knusen, H.P., Y.C. Kim, E. Muller, 1978. Comparative study of geostatistical ore reserve method over the conventional methods. Mining Engineering.
- Raymond, G.F., 1979. Ore estimation problems in an erratically mineralized orebody. CIM Bulletin.
- Raymond, G.F., 1982. Geostatistical production grade estimation in Mount Isa's copper orebodies. Proc. Australas. Inst. Min. Metall, No.284.
- Weber, D.D., E.J. Englund, 1992. Evaluation and Comparison of Spatial Interpolators I. Mathematical Geology. Vol.24, No.4
- Weber, D.D., E.J. Englund, 1993. Evaluation and Comparison of Spatial Interpolators II. Mathematical Geology