



Encuentro Internacional de
Educación en Ingeniería ACOFI

Innovación en las facultades de ingeniería:
el cambio para la competitividad y la sostenibilidad

Centro de Convenciones Cartagena de Indias

4 al 7 de octubre de 2016



CÁLCULO DEL ERROR ALTIMÉTRICO DE MODELOS DIGITALES DE ELEVACIÓN A TRAVÉS MÉTODOS GEOESTADÍSTICOS

Julián Garzón Barrero, Gonzalo Jiménez Cleves

Universidad del Quindío, Armenia, Quindío

Resumen

Los Modelos Digitales de Elevación se definen como estructuras de datos numéricas que representan la distribución espacial de la elevación sobre la superficie terrestre. En términos de representación, la solución aparentemente más adecuada hasta el presente ha sido el tratamiento del relieve mediante una estructura vectorial espacial: la Red de Triángulos Irregulares o TIN siglas de Triangular Irregular Network propuesta por Poiker en el año 1968, pero por su naturaleza, este algoritmo obedece a una función lineal, mientras que la geometría de la naturaleza no lo es, mostrando un comportamiento caótico; es aquí donde podría lograrse una mejor representación de dichas formas con la implementación de algoritmos como Kriging, IDW, Natural Neighbor, Mínima Curvatura, Función de Base Radial y Shepard modificado, permitiendo analizar la variación estadística sobre diferentes distancias y direcciones, obteniendo el mínimo error en la elevación estimada; este trabajo busca establecer la relación entre la geometría de la superficie y el algoritmo de interpolación, a través de un análisis de varianzas (ANOVA) de los errores medios cuadráticos (EMC) de una muestra de 80 diferentes modelos clasificados de acuerdo a su rango de pendiente

Palabras clave: ANOVA; error medio cuadrático; geoestadística; Kriging; TIN

Abstract

Digital Elevation Models (DEM) are defined as numerical data structures that represent the spatial distribution of elevation above the surface. From the 70's was established the representation of surfaces generated by the TIN model as the most appropriate, but by its nature, this algorithm follows a linear function, while the geometry of nature is showing a chaotic behavior, for that reason could achieve a better representation of these forms with the implementation of algorithms like Kriging, IDW, Natural Neighbor, Minimum Curvature,

Radial Basis Function, Modified Shepard's Method, allowing to analyze the statistical variation over different distances and directions, causing the minimum error in estimated elevation; this work seeks to establish the relationship between surface geometry and the interpolation algorithm, through an analysis of variance (ANOVA) of root mean square Error (RMSE) of a sample of 80 different models ranked according to their slope.

Keywords: ANOVA; root mean square error; geostatistics; Kriging; TIN

1. Introducción

La superficie terrestre se puede modelar como una expresión simplificada de la realidad que contiene algunos de sus atributos, mediante una estructura numérica que representa la distribución espacial de la elevación conocida con el nombre de Modelo Digital de Elevaciones (DEM), debido a que la superficie de la tierra posee un número infinito de puntos, se hace un muestreo de aquellos que producen los rasgos más significativos de la geometría de la superficie, dichos puntos se procesan mediante algún sistema computarizado, y a partir de la utilización de algún algoritmo de interpolación se puede pasar de un variable matemática de tipo simple (elevación de puntos), a una de tipo continuo conocida como superficie.

El presente trabajo busca establecer la relación entre la geometría de la superficie a modelar y su algoritmo de interpolación. Se hace una revisión de los seis algoritmos en estudio para interpolar los datos de las superficies topográficas, clasificadas de acuerdo con los rangos de pendiente establecido por Ley en 1986, para ello se tiene una muestra significativa de 20 superficies por cada clasificación: plana, ondulada, escarpada y montañosa; cada una de las cuales será interpolada por todos los algoritmos mencionados.

Mediante la metodología propuesta se busca evaluar la calidad de cada modelo a partir del cálculo del Error Medio Cuadrático, en función del número de puntos de control que dependen del modelo a evaluar; posteriormente se calcula el error cualitativa y cuantitativamente mediante de un Análisis de Varianza ANOVA que determina la calidad de un conjunto de medidas en términos de las medias de sus varianzas, definiendo la significancia o no de la variable dependiente (el error) en función de una posible afectación causada por dos variable de tipo independiente (algoritmo de interpolación, y tipo de terreno).

2. Metodología

La utilización de técnicas diferentes sobre los mismos conjuntos de datos permite determinar si existe mayor o menor grado de concordancia entre datos e hipótesis y también cómo este varía irregularmente con la distribución espacial de la información.

Los modelos de interpolación propuestos en este estudio pertenecientes a un mismo conjunto de datos muestrales, dejando percibir a simple vista que se pueden generar errores como consecuencia de la aplicación del algoritmo.

El objeto de la metodología propuesta es la implementación de un modelo experimental que permita en función de una muestra considerable de superficies (80 planos topográficos), determinar en qué rangos de pendiente es posible utilizar alguno o algunos de los algoritmos de interpolación en estudio, sin comprometer la calidad altimétrica de la representación de la superficie en los Modelos Digitales de Elevación. Para lograr este objetivo se propone el siguiente protocolo:

- Obtención del universo de datos
- Clasificación de las superficies en términos de su pendiente
- Interpolación de las superficies con algoritmo TIN
- Interpolación de todas las superficies con todos los algoritmos
- Determinación del número de puntos para evaluación de cada muestra
- Implementación del algoritmo de Box Müller para distribuir de manera normal los puntos a evaluar sobre los DEM
- Cálculo de los errores medios cuadráticos (EMC) de superficie
- Análisis de varianza (ANOVA)

3. Resultados y discusión

Este procedimiento ejecuta un análisis de varianza de varios factores para la variable dependiente *Error Medio Cuadrático*. Así mismo, realiza varias pruebas y gráficas para determinar qué factores tienen un efecto estadísticamente significativo sobre el *EMC*. También evalúa la significancia de las interacciones entre los factores, si es que hay suficientes datos. Las pruebas-F en la tabla ANOVA permitirán identificar los factores significativos. Para cada factor significativo, las *Pruebas de Rangos Múltiples* darán informe de cuales medias son significativamente diferentes de otras. La Gráfica de Medias y la Gráfica de Interacciones ayudarán a interpretar los efectos significativos.

ANOVA Multifactorial

Variable dependiente: *Error Medio Cuadrático*

Factores:

1. Tipo de terreno (plano, ondulado, escarpado, montañoso)
2. Método de interpolación (Kriging, vecinos naturales, Shepard modificado, inverso de la distancia, mínima curvatura, función de base radial). Número de casos completos: 479

Tabla II: Análisis de Varianza para Error - Suma de Cuadrados Tipo III

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
EFFECTOS PRINCIPALES					
A:Terreno	635242,	3	211747,	610,05	0,0000
B:Metodo	331833,	5	66366,7	191,21	0,0000
INTERACCIONES					
AB	291229,	15	19415,2	55,94	0,0000
RESIDUOS	157929,	455	347,096		
TOTAL (CORREGIDO)	1,40199E6	478			

Todas las razones-F se basan en el cuadrado medio del error residual

La tabla ANOVA descompone la variabilidad de Error en contribuciones debidas a varios factores. Puesto que se ha escogido la suma de cuadrados Tipo III (por omisión), la contribución de cada factor se mide eliminando los efectos de los demás factores. Los valores-P prueban la significancia estadística de cada uno de los factores. Puesto que 3 valores-P son menores que 0,05, estos factores tienen un efecto estadísticamente significativo sobre el *Error Medio Cuadrático* con un 95,0% de nivel de confianza, por lo tanto se debe rechazar la hipótesis nula y adoptar como cierta la hipótesis alternativa, es decir, el *Error Medio Cuadrático* Sí es dependiente del *algoritmo de interpolación* y del *tipo de terreno* sobre el cual se aplique.

3.1 contraste de la hipótesis

Contraste de la existencia de interacción:

Para establecer la región de criterio de contraste, es necesario hallar el estadístico F-Snedecor:

$$F_{0,05;15;455} = 1,69$$

$$F = \frac{\text{interacción media cuadrados}}{\text{error media de los cuadrados}} = \frac{1941,5}{347,096} = 55,94$$

Estimado: $55,94 > 1,69 = F_{0,05;15;455}$ *Estadístico*

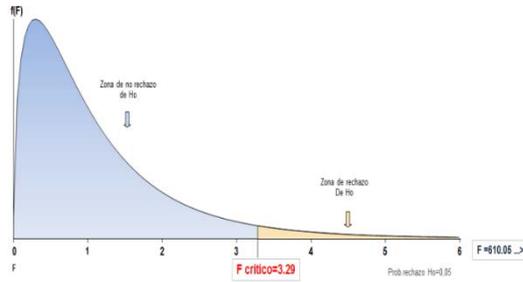


Figura 1 - Distribución Fisher-Snedecor

Por lo tanto, se rechaza la hipótesis nula H_0 considerando que el valor del estimador calculado $F = 55,93$ cae fuera de la región de aceptación o contraste. Por lo tanto, SI existe interacción entre el *Error Medio Cuadrático* y el *tipo de terreno*.

Contraste del factor terreno:

$$F_{0,05;3;15} = 3,287$$

$$\text{Estimado: } 610,05 > 3,287 = F_{0,05;3;15} \text{ Estadístico}$$

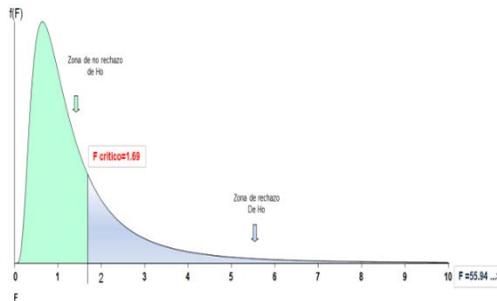


Figura 2 - Distribución Fisher-Snedecor

Contraste del factor método de interpolación:

$$F_{0,05;5;15} = 2,901$$

$$\text{Estimado: } 191,21 > 2,901 = F_{0,05;5;15} \text{ Estadístico}$$

Al igual que en el contraste del factor *Terreno* se rechaza la hipótesis nula H_0 considerando que el valor del estimador calculado $F = 191,21$ cae fuera de la región de aceptación o contraste. Por lo tanto, SI existe interacción entre el *Error Medio Cuadrático* y el *algoritmo de interpolación*.

Se ha graficado la dispersión del error respecto de cada uno de los tipos de superficie (Figura # 2), donde se corrobora de forma gráfica la significancia entre el conjunto de datos, así, los menores errores

independientemente del algoritmo de interpolación se concentran en las superficies de tipo plano, y los errores más predominantes se encuentran en las superficies de tipo montañoso.

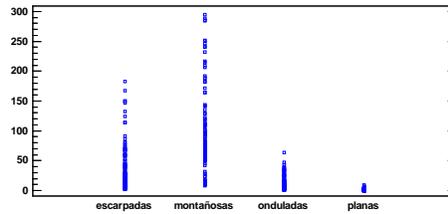


Figura 3 - Dispersión por tipo de terreno

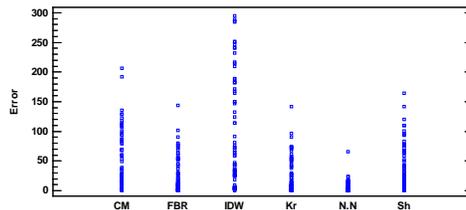


Figura 4 - Dispersión por algoritmo de interpolación

La gráfica 4 representa la dispersión del *Error Medio Cuadrático* en función a los distintos *algoritmos de interpolación* dejando percibir que el algoritmo que mejor comportamiento presenta es de Vecinos Naturales (N.N), identificando valores entre 0 y 20 m, también se puede notar que el algoritmo de mayor dispersión de error es IDW, esta dispersión justifica el peso que se le puede dar a la variación particular del valor de un punto del muestreo sobre los que están a su alrededor.

El mejor algoritmo de representación de la superficie evaluado en este estudio será aquel que menor *Error Medio Cuadrático* presente, a la vista del gráfico de medias (Figura 4) se puede entender fácilmente que el algoritmo *Vecinos Naturales* es el que se encuentra más cercano al valor cero de error, también se puede inferir que no existe una diferencia estadísticamente significativa entre el algoritmo de *Kriging* y *Función de base radial*, pues el valor de error máximo en *Kriging* se encuentra contenido en el rango de error del algoritmo de *función de base radial*, a pesar de ser estos dos los algoritmos más cercanos al de menor error, no se puede decir que sus errores sean iguales pues el máximo valor que toma el error del algoritmo *Vecinos Naturales* no alcanza a proyectarse en el valor mínimo de aquel que se encuentre más cercano (en términos de su representación), es así, como se ratifica el rechazo de la hipótesis nula. Siendo consecuentes con la gráfica de dispersión, aquí también se puede observar que el algoritmo *IDW* es de peor comportamiento presentando errores alrededor de los 90 m, lo cual hace que no se pueda utilizar para la representación de superficies topográficas en este estudio.

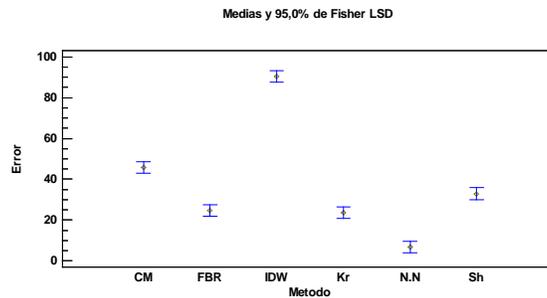


Figura 5 - Gráfico de medias del EMC en función del algoritmo de interpolación

3.2 Gráfico de interacciones

En la figura 5 se permite visualizar la interacción que existe entre factores (tipo de terreno y algoritmo de interpolación) permitiendo identificar un comportamiento convergente desde los territorios montañosos hacia los territorios planos, se puede inferir que el efecto de cada factor depende los niveles de los demás factores. Así mismo se puede deducir que cada vez que aumenta la pendiente del terreno que se está representando también lo hace su error, permitiendo así definir la calidad de los algoritmos de interpolación en función de su *Error Medio Cuadrático* y el *tipo de terreno* que representan.

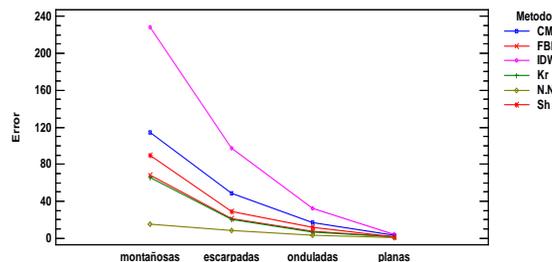


Figura 6 - interacción entre factores

A partir de la interpretación de esta gráfica se puede observar el comportamiento del error para cada algoritmo en los diferentes tipos de terreno, pudiendo determinar que para superficies categorizadas como planas solo el algoritmo *Vecinos Naturales* presenta un comportamiento aceptable con errores bajos tendientes a cero (0,66), el resto de algoritmos tienen un error medio de representación superior a un metro, lo cual es considerado alto.

Ya en las superficies de tipo ondulado el mejor comportamiento es el del algoritmo de *Vecinos Naturales*, sin que esto signifique que pueda ser recomendado para representar superficies topográficas, pues el valor de su error: 3,08 m hace que se aleje mucho de una representación que se pueda considerar como precisa, así mismo los algoritmos *Kriging* y *Función de Base Radial*, *Shepard modificado*, *Mínima Curvatura* e *IDW* ya empiezan a presentar errores de mucha magnitud, que sugieren que este para este tipo de superficies no se interpole con alguno de ellos. En las superficies categorizadas como *escarpadas* el algoritmo de mejor comportamiento sigue siendo el de *Vecinos Naturales* con valor de 8,37 m, lo cual indica que es demasiado grande para tomarse como interpolador de superficies para este tipo de terreno. En las superficies de tipo *montañoso* se pueden observar diferencias muy amplias entre cada uno de los algoritmos evaluados mostrando al *IDW* como el algoritmo de

mayor error para este tipo de superficie, y nuevamente posicionando a *Vecinos Naturales* como el de menor error dentro del grupo de interpoladores evaluados.

Conclusiones

De acuerdo con el resultado de rechazo de hipótesis nula de la ausencia de diferencias entre las varianzas de las superficies *Onduladas*, *Escarpadas* y *Montañosas* se identifican los algoritmos: *Vecinos Naturales*, *IDW*, y *Curvatura Mínima* como aquellos que mayor error indican en términos de la representación topográfica, por lo tanto, se sugiere no utilizarlos para tal fin.

El error de representación de la superficie es directamente proporcional con el aumento de la pendiente del terreno.

El algoritmo de interpolación con mejor comportamiento para todos los tipos de superficie, es *Vecinos Naturales*, a pesar de esta afirmación, se sugiere no usar este interpolador para superficies diferentes a las *planas*, pues sus errores de representación son superiores a un metro, causando esto cambios significativos en la geometría de las formas de la superficie terrestre. Este estudio es limitado por las características de la fuente de información, sus errores inherentes y el software utilizado para su desarrollo, de modo que las conclusiones no puedan extrapolarse más allá de estos términos.

Los resultados emitidos en este proyecto son limitados a los valores por defecto que sugiere el software SURFER V.10.1 para cada uno de los algoritmos evaluados; a futuro sería de mucho interés evaluar el comportamiento de cada uno de los algoritmos planteados de forma independiente pues al interior de cada uno de ellos existen parámetros de ajuste (como los semivariogramas) que son susceptibles de ajuste en función de los parámetros de entrada.

Referencias

Libros

- Barredo, J.I.: "Aplicación de técnicas de análisis espacial integrando evaluación multi-criterio y Sistemas de información geográfica para la realización de estudios de localización-/asignación de actividades" Tesis doctoral. Universidad de Alcalá. Texto mecanografiado. 1995.
- Barredo, J.I. y Bosque, J.: "Delimitación de unidades homogéneas del relieve a partir de un modelo digital de elevaciones". Estudios geográficos, nº 225, 1996, pp. 615-643. 1996.
- Bosque, J.: "Sistemas de Información Geográfica". Ediciones Rialp, Madrid, 1992, 452 pp. 1992.
- Eastman, R.; Kyen, P.; Toledano, J.; Jin, W.: "GIS an Decision Making. Exploration en Geographic Information Systems Technology". Vol. 4. Unitar. Geneva, Switzerland, pp. 110. 1993.

- Felicísimo Ángel M.: "Modelos Digitales del Terreno. Introducción y aplicaciones en las Ciencias ambientales". Oviedo, Pentalfa, 222 p. 1994.
- Garzón, J.; Jiménez, G.; Vila, J. "Introducción a los Modelos Digitales de Elevación en Topografía". Saarbrücken, Editorial académica española. 113p. 2012.
- Sircar, J.K y Cebrian, J.A.: "Creación de Modelos Topográficos Digitales a partir de curvas de nivel rasterizadas" Anales de Geografía de la Universidad Complutense, nº 10, pp. 13-36. 1990.

Sobre los autores

- **Julián Garzón Barrero:** Ingeniero de sistemas, Tecnólogo en topografía, Especialista en Geomática, Magister en sistemas de Información Geográfica. Profesor Asistente. juliangarzonb@uniquindio.edu.co
- **Gonzalo Jiménez Cleves:** Topógrafo, Especialista en Computación, magister en Ingeniería de sistema. Profesor Asociado. gjcleves@uniquindio.edu.co

Los puntos de vista expresados en este artículo no reflejan necesariamente la opinión de la Asociación Colombiana de Facultades de Ingeniería.

Copyright © 2016 Asociación Colombiana de Facultades de Ingeniería (ACOFI)