



**UNIVERSIDAD SAN FRANCISCO DE QUITO**

**Colegio de Postgrados**

**Validación cruzada como método de investigación en la calidad de  
Modelos Digitales de Elevación**

**Julián Garzón Barrero**

**Richard Resl, Ph.Dc., Director de Tesis**

Tesis de grado presentada como requisito para la obtención del título de Magister en  
Sistemas de Información Geográfica

Quito, Abril 2013

**UNIVERSIDAD SAN FRANCISCO DE QUITO**

**Colegio de Postgrados**

## **HOJA DE APROBACIÓN DE TESIS**

**Validación cruzada como método de investigación en la calidad de  
Modelos Digitales de Elevación**

**Julián Garzón Barrero**

Richard Resl, PH.Dc.  
**Director de Tesis**

.....

Pablo Cabrera. MS  
**Miembro del Comité de Tesis**

.....

Richard Resl, PH.Dc.  
**Director de la Maestría en  
Sistemas de Información Geográfica**

.....

Stella de la Torre, PhD.  
**Decana del Colegio de  
Ciencias Biológicas y Ambientales**

.....

Víctor Viteri Breedy, PhD.  
**Decano del Colegio de Postgrados**

.....

Quito, Abril de 2013

© DERECHOS DE AUTOR

Por medio del presente documento certifico que he leído la Política de Propiedad Intelectual de la Universidad San Francisco de Quito y estoy de acuerdo con su contenido, por lo que los derechos de propiedad intelectual del presente trabajo de investigación quedan sujetos a lo dispuesto en la Política.

Asimismo, autorizo a la USFQ para que realice la digitalización y publicación de este trabajo de investigación en el repositorio virtual, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior.

Firma:

-----

Nombre: Julián Garzón Barrero

C. I.: 89007517

Fecha: abril de 2013

## **Dedicatoria:**

A Nancy, Sofía, Eunice y Adolia, por darme algo que nunca recuperarán: el tiempo que me cedieron y que debí dedicar a ellas (mi familia), para que yo pudiera subir este nuevo peldaño en mi formación académica.

## **Agradecimientos:**

A Dios por darme vida, disposición y fuerzas para estudiar.

A mi director de tesis y director del programa Richard Resl, por guiarme en la elaboración de este trabajo.

Al profesor Pablo Cabrera por darme las herramientas necesarias para hilvanar mi proyecto a través del espacio académico de estadística espacial.

A mi profesor, compañero y amigo Gonzalo Jiménez Cleves por darme la motivación (hace ya varios años) de mejorar el nivel académico de mi profesión.

Al profesor Hernando Hurtado, jefe del departamento de estadística de la Universidad del Quindío por las constantes y esclarecedoras asesorías.

## Resumen General

Los Modelos Digitales de Elevación (DEM) 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 la mayoría de las veces un comportamiento caótico; es aquí donde podría lograrse una mejor representación de dichas formas con la implementación de otros algoritmos como: *Inverse Distance Weight*, *Kriging*, *Natural Neighbor*, *Spline*, *Topo to Raster*; 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) extraídos de un DEM proveniente de una imagen de satélite ASTER, contando con aproximadamente 300.000 datos que serán remuestreados por la técnica de Validación Cruzada.

## **Abstract**

Digital Elevation Models (DEM) are defined as numeric data structures representing the spatial distribution of elevation above the earth's surface. In terms of representation, the solution that seems the most appropriate to the present has been the treatment of the terrain's surface through a spatial vector structure, called Triangle Irregular Network or TIN proposed by Poiker in 1968. By its nature, this algorithm follows a linear function, while the geometry of nature is not, showing mostly a chaotic behavior. It is suggested that a better representation of the terrain's surface can be achieved with the implementation of other algorithms such as *Inverse Distance Weight, Kriging, Natural Neighbor, Spline, Topo to Raster*, allowing analysis of statistical variation over different distances and directions, obtaining the minimum estimated error in elevation. This paper seeks to establish the relationship between surface geometry and interpolation algorithm through an analysis of variance (ANOVA) of the Root Mean Square Errors (RMSE) extracted from a DEM from a ASTER satellite image, with approximately 300.000 data to be resampled by the Cross-Validation technique.



## TABLA DE CONTENIDO

<b>Resumen General .....</b>	<b>7</b>
<b>Abstract.....</b>	<b>8</b>
<b>Capítulo 1: Introducción al problema .....</b>	<b>14</b>
1.1 Antecedentes .....	16
1.2 El problema .....	20
1.3 La hipótesis .....	22
1.4 La pregunta de investigación.....	24
1.5 Contexto y marco teórico .....	25
1.5.1 El propósito del estudio.....	26
1.5.2 El significado del estudio .....	26
1.6 Objetivos .....	28
1.7 Presunciones del autor.....	29
1.8 Supuestos del estudio .....	31
<b>Capítulo 2: Revisión de la literatura .....</b>	<b>32</b>
2.1 Géneros de literatura incluidos en la revisión .....	32
2.2 Proceso de revisión de la literatura .....	35
<b>Capítulo 3: Metodología y diseño de la investigación .....</b>	<b>37</b>
3.1 Justificación de la metodología seleccionada.....	45
3.2 Métodos de interpolación evaluados .....	47
3.2.1 IDW: método de ponderación inverso a la distancia ( <i>Inverse Distance Weighted</i> ) .....	48
3.2.2 KRG: método de Kriging .....	50
3.2.3 N.N: Vecinos naturales ( <i>Natural Neighbor</i> ) .....	54
3.2.4 SPL: método de mínima curvatura ( <i>Spline</i> ) .....	55
3.2.5 T2R: Topogrid ( <i>Topo to Raster</i> ) .....	56
3.2.6 TIN: Red de triángulos irregulares ( <i>Triangular Irregular Network</i> ) .....	57
3.3 Validación cruzada.....	59
3.4 Análisis de Varianza (ANOVA) .....	61
3.4.1 Hipótesis.....	62

3.5 Zona de estudio .....	63
<b>Capítulo 4: Análisis de datos .....</b>	<b>65</b>
4.1 Error en los valores de elevación .....	72
4.2 El error y la curvatura.....	74
4.3 El error y la pendiente .....	78
4.4 Características del error de elevación.....	82
<b>Capítulo 5: Conclusiones .....</b>	<b>87</b>
<b>5.1 Respuesta a la pregunta de investigación.....</b>	<b>90</b>
<b>Bibliografía .....</b>	<b>94</b>
<b>Anexo A. Esquema de geoprocesamiento.....</b>	<b>99</b>
<b>Anexo B. Raster de la zona de zona de estudio a distintos niveles de re-muestreo e interpolados por los algoritmos evaluados .....</b>	<b>100</b>
<b>Anexo C. Modelación de error en los distintos algoritmos evaluados .....</b>	<b>106</b>
<b>Anexo D. Pruebas de Bondad de Kolmogorov-Smirnov.....</b>	<b>107</b>
<b>Anexo E. Distribución de frecuencias de errores y gráficas cuantil - cuantil.....</b>	<b>108</b>

## LISTA DE TABLAS

Tabla 1. Re-muestreo de datos .....	40
Tabla 2. Análisis de varianza para EMC – suma de cuadrados tipo III .....	65
Tabla 3. Medias por Mínimos Cuadrados para el EMC con intervalos de confianza del 95.0% .....	69
Tabla 4. Pruebas de Múltiple Rangos para el EMC por el tipo de re-muestreo .....	70
Tabla 5. Pruebas de Múltiple Rangos para el EMC por el interpolador.....	70
Tabla 6. Estadísticas del error de elevación para los diferentes interpoladores .....	73
Tabla 7. Análisis de varianza para EMC por tipo de curvatura (cóncava o convexa).....	76
Tabla 8. Pruebas de Múltiple Rangos para el EMC por el tipo de curvatura (cóncava – convexa) .....	77
Tabla 9. Rangos de Pendiente (Ley, 1986) .....	79
Tabla 10. Valores de la curtosis estandarizada para la distribución de los errores de elevación .....	83
Tabla 11. Autocorrelación espacial del error en elevación usando <i>Moran's I</i> .....	84

## LISTA DE FIGURAS

Figura 1 - Kriging... ..	38
Figura 2- Spline.....	38
Figura 3- Topo to raster.....	38
Figura 4 - TIN.....	38
Figura 5 - Natural Neighbor .....	38
Figura 6 - IDW.....	38
Figura 7 - Diagrama de flujo metodología del proyecto. ....	44
Figura 8 - Cálculo de la diferencia cuadrada entre las ubicaciones asociadas. Fuente: Using ArcGIS® Geostatistical Analyst, 2003.....	51

Figura 9- Semivariograma y sus características. Fuente: Golden Software, 2002 .....	51
Figura 10 - Semivariograma empírico ajustado al experimental.....	52
Figura 11 - Tipos de semivariograma. Fuente: Miguel Ángel Sanz, Modelos Digitales de Terreno, 2008.....	53
Figura 12 - Polígonos de Thiessen. ....	54
Figura 13 - Red de Triángulos Irregulares. Fuente: Garzón, introducción a los Modelos Digitales de Elevación, 2012.....	57
Figura 14 - Triangulación de Delaunay. Fuente: Garzón, Introducción a los Modelos Digitales de Elevación, 2012.....	58
Figura 15 - Plano que pasa a través de tres puntos en el espacio. ....	58
Figura 16 - Resoluciones del re-muestreo.....	60
Figura 17 - Localización del área de estudio. (a) Localización en el continente sur Americano. (b) Localización entre los Departamentos de Cauca y Nariño. (c) Mapa de iluminación del DEM correspondiente al área de estudio. (d) Mapa con curvas de nivel equidistantes 50 metros. ....	64
Figura 18 - Distribución Fisher Snedecor, factor 1: interpoladores. ....	66
Figura 19 - Distribución Fisher Snedecor, factor 2: re-muestreo.....	67
Figura 20 - Dispersión por código de nivel.....	68
Figura 21 – Gráfico de medias, 95% de Fisher LSD. (a) EMC en función del algoritmo de interpolación. (b) EMC en función del tamaño de re-muestreo. ....	68
Figura 22 – interacción entre factores. ....	71
Figura 23 – Enumeración de celadas para cálculo de curvatura. Fuente: ArcGIS 10 Help. ....	74
Figura 24 – Patrón espacial de los errores en elevación. (a) Sobreestimación. (b) Subestimación.....	75
Figura 25 – (a) Gráfico de medias, 95% de Fisher LSD. EMC en función del tipo de curvatura. (b) Dispersión por código de nivel.....	77
Figura 26 – Mapa de pendientes de acuerdo a la clasificación de Ley, R (1986). ....	79
Figura 27 – Gráfico de medias, 95% de Fisher LSD. Error absoluto en función del tipo de terreno.....	80
Figura 28 – Mapa de errores absolutos para el interpolador Spline a resolución de 30 m.....	81
Figura 29 – Dispersión de errores absolutos por tipo de pendiente. ....	81

Figura 30 – Distribución de frecuencias del error del DEM. (a) para IDW con re-muestreo de 60x60. (b) para KRG re-muestreo de 90x90. (c) para NN re-muestro 120x120. (d) para SPL re-muestreo 150x150. (e) para TIN re-muestreo 180x180. (e) distribución para T2R re-muestreo 90x90. .... 82

Figura 31 – Gráfica cuantil-cuantil para error producido usando (a) interpolador Spline con re-muestro de 60x60. (b) interpolador Kriging con re-muestro 180x180..... 84

Figura 32 – Errores para algoritmo T2R (a) Distribución de frecuencias. (b) gráfica cuantil - cuantil. 85

## Capítulo 1: Introducción al problema

Actualmente los Modelos Digitales de Elevación (que para el presente documento se referirán bajo la sigla DEM) son la forma más efectiva de representación de la superficie terrestre y se han convertido en una herramienta y producto muy demandado para su uso en los ámbitos de la Geoinformática (p.e. análisis espaciales, proyectos viales, urbanismo, etc.) como de otras actividades relacionadas con el medio ambiente (p.e. estudios de impacto ocasionado por movimientos de tierra). Sin embargo, estos no han recibido en el ámbito geográfico una atención especial en cuanto a los procesos de evaluación de la calidad altimétrica, eligiendo el algoritmo de interpolación más por una apariencia visual que por conocimiento de las limitaciones que ellos presentan, causando grandes desviaciones altimétricas que ocasionan análisis espaciales erróneos.

La precisión de las Bases de Datos Geográficos (GBD) en los últimos años ha sido un tema de interés, dado el amplio espectro de la aplicación de DEM en los SIG, motivo por el cual existe una gran cantidad de literatura acerca de sus errores (Fisher y Tate, 2006). El incremento de la disponibilidad de DEM a escalas globales como ASTER, y de alta resolución como LIDAR (Wehr y Lohr, 1999) han hecho que sea muy importante contar con una buena metodología para la evaluación de su error y el entendimiento de su propagación a través de las operaciones aplicadas. Comúnmente la forma más utilizada de medir la calidad de los DEM es el Error Medio Cuadrático, el cual es calculado de la siguiente manera:

$$EMC_{elev} = \sqrt{\frac{\sum_{i=1}^n (Z_p - Z_o)^2}{n - 1}} \quad (1)$$

Donde  $Z_p$  corresponde al valor predicho de la elevación y  $Z_o$  el valor observado.

En el ámbito de los DEM el error se conoce como la desviación altimétrica del modelo con respecto a la superficie real; hay que tener en cuenta que la información siempre será imprecisa debido a la simplificación que está inmersa dentro de la modelación, pues la superficie real siempre tendrá un infinito número de puntos para modelarse con todas sus características morfométricas, es así como se hace un muestreo de los puntos más significativos que representen la superficie con la mayor cantidad de características posibles que hablen de su forma, o en el caso de las estructuras ráster una mejor resolución espacial.

Debido a que se estableció por Thomas K. Poiker, desde 1968 que la representación que genera el modelo TIN (*Triangular Irregular Network*) es la más adecuada para representar formas topográficas en un Sistema de Información Geográfica (SIG), y entendiendo que este algoritmo está basado en la ecuación de un plano que pasa a través de tres puntos no colineales en el espacio, se logra así determinar la elevación de un punto cualquiera que quede dentro de este triángulo: pero esto obedece a un sistema lineal de ecuaciones, mientras que las formas geométricas de una superficie natural no tiene comportamientos lineales sino más bien caóticos, es aquí donde “posiblemente” se puedan obtener mejores representaciones de las formas con otros algoritmos interpolantes como: *Inverse Distance Weight (IDW)*, *Kriging (KRG)*, *Natural Neighbor (N.N)*, *Spline (SPL)*, y *Topo to Raster (T2R)*.

## 1.1 Antecedentes

Desde mediados de los años setenta se han venido desarrollando estudios que buscan relacionar la creación de un DEM con su ajuste a la superficie que verdaderamente representa. En términos generales estos trabajos son coincidentes en que se hace necesario explorar otras posibilidades, lo que sin duda, representará un aumento significativo en el tiempo dedicado a una investigación. Algunos de los resultados de estos estudios insinúan de forma sutil en sus conclusiones que la calidad del modelo podría estar relacionada con la distinta orografía que se presente en el área de estudio, aunque no se ha demostrado que las formas de la superficie terrestre tengan incidencia decisiva sobre la calidad del modelo producido, mas bien esto depende de la resolución espacial de la información que se tenga de primera mano, del algoritmo de interpolación que se use y de la escala a la cual se analice.

Los avances en la estereorestitución digital de imágenes, la fotografía aérea y los datos producidos por sensores laser aerotransportados (LIDAR) ofrecen al usuario un volumen de información tal, que se podría pensar que más bien el problema radica en la selección de los datos adecuados para la generación del DEM, pese a ello, la información geográfica de calidad es costosa de adquirir y mantener, y con presupuestos en baja y exigencias en alza hay que explorar soluciones técnicas novedosas, es así como para proyectos de investigación es muy común el uso de imágenes ASTER GDEM (*ASTER Global Digital Elevation Model*), que es un sensor de imágenes construido por el Ministerio de Economía del Japón e instalado en el satélite Terra de la NASA, este permite obtener de forma libre un DEM de resolución espacial de 30 m, es así como ASTER GDEM se constituye en una fuente de información particularmente fuerte para los investigadores que pretenden generar sus propios modelos.



Debido al acceso creciente de los algoritmos y software para el modelamiento de un DEM dentro de las plataformas de desarrollo de los Sistemas de Información Geográfica, ha sucedido que su espectro de aplicación se amplíe en una forma considerable y deje de ser aplicación estricta de los profesionales de la topografía, teniendo una influencia notoria sobre numerosas variables que intervienen en la dinámica de los ecosistemas. Entre los usuarios de los DEM se encuentran: (a) la fotogrametría (Collins, 1981), (b) la ingeniería civil, (c) las aplicaciones militares, (d) las ciencias de la tierra (Bakose, 1990), (e) la planificación del manejo de recursos naturales (Fisher y Lattuada, 1996); es así como de manera específica se pueden identificar algunas de las principales tareas de los DEM (Burroughs, 1986; ESRI, 1996; 1997b; Golden Software, 1999; McCullagh, 1990; Weibel y Heller, 1991), entre las que se pueden destacar:

- Elaboración de cartografía topográfica.
- Simulación y análisis del paisaje.
- Determinación de pendiente, aspecto, rugosidad y sombreado del terreno.
- Estimación de volúmenes a remover o rellenar en obras de ingeniería.
- Estimación de áreas de inundación en proyectos hidroeléctricos.
- Análisis de intervisibilidad para definición de antenas en telecomunicaciones.
- Mapeo de procesos geomorfológicos y edafológicos.
- Modelación hidrológica de cuencas, el movimiento y la acción del agua y consecuentemente los numerosos procesos biológicos derivados de ellos.
- Corrección geométrica de imágenes satelitales en áreas montañosas.
- Estudios climáticos a escala local.

- 

De acuerdo con Eduard Imhof en su trabajo “Cartographic Relief Representation” (1982, p23), se define el origen del problema cuando refiere lo siguiente “...no es posible comparar lo que no puede ser exactamente determinado con lo que no puede estar completamente definido”; es claro que el valor de la elevación de la superficie no se puede determinar de forma exacta debido a que toda medición posee error (concepto fundamental de la teoría de errores de Gauss), así resulta entonces muy complejo comparar una superficie interpolada que por su naturaleza se encuentra incompletamente definida, contra la real que de ninguna forma puede ser exactamente determinada.

La precisión de los DEM está sujeta a muchos factores, como la calidad de la fuente de los datos, errores humanos en el muestreo, y errores en el modelo matemático usado para interpolar, Kumler (1994) desarrolló una metodología para estudiar la causa del error compilado por diferentes enfoques. Gao (1997) estudió la resolución y la precisión de la representación del terreno por medio de una retícula regular a una micro-escala, Rees (2000) estudió la precisión de los DEM interpolados a altas resoluciones y demostró que la interpolación bilineal simple produce resultados apropiado para aplicaciones DEM. Li et al. (2000) discuten sistemáticamente la teoría de los DEM especialmente en la precisión del análisis de los modelos y deducen la precisión de los modelos basados en estructuras ráster . Kinder (2003) argumentó que las técnicas de interpolación de orden superior fueron siempre más precisas que las generadas por la interpolación bilineal.

La elección óptima de cualquier método de interpolación debe estar regida por una evaluación de su eficiencia, lo cual exige el conocimiento de los factores que condicionan el resultado

final. Esto hace que la sistematización en este tipo de estudios no sea generalizada, sino mas bien, que se de en función de parámetros como: calidad de los datos fuente, escala, densidad, forma de medición del error, el descriptor de superficie a evaluar, los rangos de pendiente del territorio y el uso final que se le de al modelo elaborado.

Con los antecedentes mencionados, este documento busca analizar de alguna manera la relación que existe entre los DEM y los métodos de interpolación a través de la técnica de Validación Cruzada donde se analizarán los EMC en términos de su varianza.

## 1.2 El problema

Los DEM proporcionan información valiosa que gracias a su estructura permite ser almacenada y analizada sin necesidad de trabajar directamente sobre la superficie real, su precisión es impactada por la selección del algoritmo de interpolación y la calidad de los datos de entrada, es claro que la mayoría de los usuarios SIG que manipulan en sus bases de datos geográficos los DEM con distintos propósitos, sin tener en cuenta que la complejidad del fenómeno espacial captado (la superficie terrestre) tiene gran incidencia en el resultado final del producto, utilizan esta fuente de información de una manera bastante indiscriminada y sin el conocimiento necesario para la creación de dicho modelo, pues no se tiene conciencia de la utilización de elementos fundamentales como “breaklines”, zonas de recorte, puntos acotados, y lo más importante: los algoritmos de interpolación, haciendo una elección de éste, más por razones de presentación en términos de su apariencia visual, que por razones que obedezcan a la relación existente entre la geometría de la forma que se busca representar y su algoritmo de interpolación.

Eduard Imhof en su obra “Cartographic Relief Presentation” (1982, p.23) expone la dificultad de calcular la precisión de la cartografía topográfica: *“para calcular el error cartográfico se pretende comparar dos superficies diferentes y complejas, que no pueden ser expresadas mediante fórmulas matemáticas. Una de ellas es la actual superficie de la tierra, reducida mentalmente a la escala del mapa, la otra superficie es la formada por isolíneas que conforman el mapa bajo estudio. Como resultado de las deficiencias de la cartografía, ambas superficies no coinciden perfectamente, y se desea ser capaz de juzgar la naturaleza y extensión de tales variaciones. La primera superficie (la real) nunca es exactamente conocida,*

*ya que cualquier medición viene acompañada de errores. La segunda superficie, aparte de los errores cartográficos inherentes nunca está totalmente definida, ya que las isolíneas, debido a la equidistancia, dejan grandes espacios a los lados que no son cubiertos. Entre cada dos isolíneas la posición de la superficie es desconocida, aunque se presume que las une una línea recta, lo que es menos probable a medida que aumenta la equidistancia. Desde este punto de vista: ¿Cómo podemos comparar una superficie incompletamente definida (la real) con otra que no puede ser exactamente determinada (el mapa)? ”.*

El motivo principal entre estas diferencias (la realidad vs el modelo) se encuentra radicado en la evaluación subjetiva que tradicionalmente se hace al DEM, estos modelos son construidos con datos espaciales, que por su naturaleza poseen errores intrínsecos a las mediciones y su propagación en el modelo, tales errores deben ser tratados con un procedimiento geostadístico que permita cuantificar su error y buscar una respuesta a la incertidumbre que se presenta ante la tradicional evaluación subjetiva de un DEM.

El hecho de centrarse exclusivamente en el hallazgo de la relación entre la geometría de la superficie y el algoritmo de interpolación, posibilita el desarrollo de este trabajo de investigación de tipo explorativo y contemplado en un plazo adecuado a los propósitos que rigen la tesis de la Maestría en Sistemas de Información Geográfica.

### **1.3 La hipótesis**

La calidad de los datos utilizados en las herramientas de soporte SIG es una cuestión crítica, ya que la decisión que afectará a la ubicación de un área específica se debe ser efectiva en el tiempo y con la calidad adecuada. La precisión y la representatividad son propiedades del DEM que resultan determinantes en el sistema de apoyo sobre el cual se diseñan la evaluación de recursos renovables de energía, la previsión de inundaciones y la gestión de seguridad. Así mismo el análisis de idoneidad, el cálculo indicadores de monitoreo ambiental y la calidad del agua de una cuenca se basan en el uso de un DEM y las decisiones adoptadas tienen implicaciones financieras y legales. Por lo tanto, un requisito previo para la plena explotación de un DEM es la determinación de su calidad antes de ponerlo a disposición de la comunidad, de forma tal que cuente con la precisión suficiente para su uso en una gran variedad de aplicaciones en entornos geomáticos.

La complejidad espacial de las características del terreno en distintos paisajes impone un reto en el procedimiento automatizado de los DEM, pues se hace necesaria la aplicación de algún algoritmo de interpolación que permita hacer la conversión de una localización discreta a un espacio continuo, actualmente las nuevas fuentes de datos de alta resolución, como STRM, ASTER y LIDAR están siendo usadas por los usuarios SIG, pero su procesamiento no es sencillo y requiere mucha experiencia empírica y un profundo conocimiento, pues no parece haber un acuerdo unánime acerca de cuál método de interpolación produce los mejores resultados, es así como se enmarca la hipótesis de este trabajo definida en lo siguiente: La calidad de los DEM se ve afectada directamente por la elección del algoritmo de interpolación en aplicaciones propias de los Sistemas de Información Geográfica (SIG).

Como método de confrontación de datos se propone la Validación Cruzada que es una técnica estadística en la cual un DEM de alta resolución se re-muestrea a uno de baja resolución y luego es interpolado a su resolución original. En la ubicación de los puntos que fueron retirados en el re-muestreo, hubo un valor interpolado y su correspondiente valor verdadero. Esto produce un gran número de puntos en los cuales se conoce el error de interpolación, con una propagación igual a través del DEM.

El argumento aquí es que la metodología propuesta proporciona una forma alternativa de estudiar las características espaciales de los errores en un DEM. Es claro que esto es sólo análogo a los DEM derivados de puntos y no, por ejemplo, a aquellos derivados de la interpolación entre curvas de nivel y cotas de los mapas topográficos, pues aquí el proceso de interpolación es completamente diferente. También es relevante sólo para aquellas situaciones en que los datos originales son menos densos que el DEM final.

La comprobación de la hipótesis se plantea mediante un procedimiento estadístico conocido como Análisis de Varianza (ANOVA), que permite analizar resultados en procesos investigativos con diseños experimentales, a partir de los cuales se puede comparar sí los valores que toma un conjunto de datos que corresponden a variaciones de una misma variable dependiente, en este caso el Error Medio Cuadrático, se ve afectados por una o más variables de tipo independiente: El algoritmo de interpolación.

El Análisis de Varianza permitirá poner a prueba la nulidad de la hipótesis, pudiendo encontrar un porcentaje de probabilidad que defina la aceptación o rechazo de esta, identificando el mejor interpolador en función del menor Error Medio Cuadrático.

## **1.4 La pregunta de investigación**

¿Cómo la elección del algoritmo de interpolación incide sobre la representación de la superficie terrestre en un Modelo Digital de Elevación (DEM)?



## 1.5 Contexto y marco teórico

La interpolación es definida como un procedimiento que permite calcular el valor de una variable en una posición del espacio (punto no muestral con valor estimado), conociendo los valores de esa variable en otras posiciones del espacio (puntos muestrales con valores reales) (Bosque Sendra, 1992).

En los entornos DEM la interpolación se usa para predecir la elevación de las unidades en las que se encuentre subdividido el espacio, en el caso de la estructura vectorial se define por la ecuación de un plano que pasa a través de tres puntos no colineales en espacio, y en el caso de los ráster se calcula la elevación del centro geométrico de las celdas en que se divide el espacio geográfico. Esta división al encontrarse regularmente distanciada no quiere decir que los valores de la elevación en la superficie real sean igualmente homogéneos.

De acuerdo con McCullag, (1988), Burrough y McDonnell (1998) se deben cumplir una serie de condiciones para que el valor estimado de la interpolación sea fiable:

- Se debe crear una superficie continua a partir de datos distribuido discretamente en el espacio.
- Debe tener propiedades matemáticas.
- Debe ajustarse a los datos con un nivel de precisión determinado.

La confiabilidad de la predicción con cualquier algoritmo de interpolación está fuertemente relacionada con la calidad de los datos de la muestra inicial, una mala distribución de los mismos o una intermediación por falta de ellos, hacen que el resultado se aparte de la geometría que realmente tiene la zona representada mediante el DEM.

### **1.5.1 El propósito del estudio**

La integración de la información geoespacial y las herramientas informáticas establecen verdaderos sistemas que deben estar articulados de tal forma que los modelos producidos tengan validez en cuanto a la efectiva representación de la realidad. Bajo esta premisa operan los SIG, cuyas fortalezas radican en la posibilidad de obtener representaciones más certeras a medida que la información de entrada al sistema mejora, cualitativa y cuantitativamente. En la ejecución de este proyecto se busca hallar una luz que permita encontrar el patrón espacial del comportamiento del error, apoyado en el re-muestreo de un DEM interpolado por distintos algoritmos, hasta un punto en el cual los datos de baja densidad comienzan a generar errores claros en los DEM interpolados, es así como se tiene una expectativa en el sentido en que los algoritmos de interpolación dejen de ser cajas negras para el usuario, de tal forma que éste pueda tener información sobre la incertidumbre asociada de las estimaciones generadas durante la interpolación.

### **1.5.2 El significado del estudio**

La precisión de un DEM como estructura de todas las variables del medio biofísico, constituye una pauta a seguir en cuanto al resto de escalas de captura de datos, por extensión de este DEM se puede generalizar la explotación de cualquier otro atributo relacionado a la superficie terrestre como, redes de drenaje, índice de pluviometría, radiación solar o temperatura.

La determinación del error altimétrico de un DEM, en función de un algoritmo de interpolación, puede dar una idea general de la propagación del error, permitiendo usar esto a

favor de la modelación hidrológica, especialmente en la obtención de redes de drenaje y parámetros como el cálculo de caudales y las líneas de flujo que siguen las rutas de máxima pendiente que se definen a partir del DEM con la única limitación que supone la confiabilidad del algoritmo de interpolación inicial. Cada técnica de interpolación, en función de su algoritmo y los parámetro de decisión que maneja, genera una replica que intenta aproximarse a la realidad.

A pesar de que los múltiples autores citados a lo largo de este documento realizan aportes de mucha relevancia en temáticas relacionadas con la calidad de los modelos digitales de elevación, este trabajo busca esclarecer algo que parece no estar resuelto en su totalidad: la relación entre el algoritmo de interpolación y la geometría de la superficie que éste representa. Este hecho en conjunto con la metodología propuesta para la evaluación de la calidad de los algoritmos estudiados dan la posibilidad que el trabajo de investigación pueda tener la profundidad suficiente para obtener resultados confiables y completada en un plazo adecuado a la normatividad que exige una tesis de maestría.

## 1.6 Objetivos

General:

El presente trabajo busca establecer la relación entre la geometría de la superficie a modelar y su algoritmo de interpolación mediante el cálculo del error medio cuadrático.

Específicos:

- Hallar las diferencias en el error medio cuadrático alcanzado por los modelos generados por los distintos algoritmos de interpolación.
- Contrastar los distintos algoritmos de interpolación con los diferentes tipos de superficie clasificada de acuerdo a sus rangos de pendiente.
- Analizar la variabilidad del error de acuerdo con las diferentes separaciones que se proponen para el re-muestreo de datos.

## **1.7 Presunciones del autor**

Entre los sistemas satelitales que ofrecen datos topográficos, se encuentran plataformas cuya información está disponible de forma libre, de fácil acceso y con una disponibilidad a nivel global. Este es el caso del sistema radiométrico Japonés ASTER (Advanced Space Borne Thermal Emission and Reflection Radiometer), el cual es uno de los cinco instrumentos de observación terrestre que se encuentran al interior del satélite Norteamericano TERRA y que fue puesto en órbita en diciembre de 1999; en esfuerzos conjuntos entre la NASA y el Ministerio de Economía Comercio e Industria de Japón (METI) se ha logrado publicar el Modelo Digital de Elevaciones más completo del planeta: ASTER GDEM (ASTER Global Digital Elevation Model), elaborado a partir de cerca de 1,3 millones de imágenes estéreo capturadas por el radiómetro Japonés. De acuerdo al informe final producido por el equipo de validación de los datos del DEM global, su error vertical no sobrepasa los 20 metros, soportado en una confiabilidad del 95 %, destacándose por su alta resolución espacial (30 m) y disponibilidad global.

Hay que señalar que si se mira a escala más detallada, la confiabilidad de la información base esta dada en función de las condiciones locales del área de estudio y para ese caso no se tiene consideraciones de que tan precisos son los valores del DEM original. Cuando los valores interpolados son comparados con los valores originales, esto suministra una medida del error inducida por el proceso de interpolación pero no dice nada acerca de la precisión absoluta de los valores de elevación comparados con el valor real (el valor del punto en el terreno).

Con fines de tener confianza en los resultados de este estudio, se parte del principio que los errores del DEM son los indicados, de cualquier forma, estos resultados se encuentran limitados por las características de la fuente de información, sus errores inherentes, de modo que las conclusiones no puedan extrapolarse más allá de estos términos.

## 1.8 Supuestos del estudio

Al modelar el error de elevación en el DEM, siempre se asume que éste es espacialmente autocorrelacionado, esta es una suposición razonable dado que el terreno tiende a variar de forma suave en lugar de hacerlo de forma brusca, de modo que los pixeles vecinos probablemente tengan los mismos valores de altura, y dado el hecho de que la interpolación en sí misma probablemente produzca los mismos valores de los puntos vecinos.

La aplicación de la Validación Cruzada como método de re-muestreo de datos no es una prueba directa del nivel de error en un DEM en particular, a menos que haya sucedido que el DEM fuese generado por re-muestreo de una resolución muy baja. Sin embargo, la producción de un gran número de estimaciones de error de elevación, distribuidos sobre el DEM, permite explorar algunos cuestionamientos generales como la identificación de las características estadísticas del error del DEM y hasta qué punto ésta difiere entre los métodos de interpolación y la relación entre el EMC de la elevación y el EMC de la superficie.

Los hallazgos deben ser de particular relevancia para cualquier usuario SIG que utiliza un producto DEM, generado a partir de la reducción de escala de un modelo resultante en una resolución más fina.

En el siguiente capítulo del documento el lector encontrará la revisión de la literatura , que consiste en el repaso y reconstrucción de trabajos ya realizados por otros investigadores que tiene como fin detectar y consultar la bibliografía y otros materiales útiles en el propósito del estudio, posterior a ello se hallan: la metodología, el análisis de datos y las conclusiones del proyecto.

## Capítulo 2: Revisión de la literatura

En este capítulo se indican las principales fuentes bibliográficas, provenientes de libros fundamentales en el entorno de los DEM y casi de carácter obligatorio para sus procesos de investigación, así como algunas tesis doctorales en temas afines, y los artículos provenientes de revistas indexadas en el ámbito de la Geoinformática, que se constituyen en las publicaciones científicas que más se valoran como las editadas por Elsevier, considerada una editorial líder en el mundo en publicaciones científicas. También se ofrece una síntesis de las fuentes usadas para el desarrollo de la investigación: primarias, secundarias y otras fuentes. El objetivo de esta relación es facilitar a los lectores el área de información que puede resultar útil de cara al establecimiento de nuevas estancias de investigación.

### 2.1 Géneros de literatura incluidos en la revisión

Las fuentes primarias para el desarrollo de esta investigación son:

- la tesis doctoral *Modelos Digitales de Elevaciones, métodos de interpolación y parámetros morfométricos del relieve*, presentada por Joaquín Márquez Pérez en marzo de 2004, ante el departamento de geografía física y análisis geográfico regional, de la Universidad de Sevilla. Este magna obra es un compendio maravilloso donde se tratan temáticas que refieren los errores en los DEM a profundidad, además del invaluable estado del arte que presenta, su potencial se encuentra en la relación de los descriptores topográficos (productos derivados de un DEM) con el error de las modelación a partir de cartografía topográfica a escalas 1:50000 y 1: 10000.



- El libro *Digital Terrain Modeling, principles and methodology*, es una obra escrita por Zhilin Li, Qing Zhu, y Chistopher Gold, que fue publicada por la editorial CRC Press en 2005; es una obra que suministra de forma comprensible los desarrollos en el campo de los DEM, su temática incluye: análisis de superficies, estrategias de muestreo, metodologías en la adquisición de datos, principios en la modelación de superficies, técnicas de interpolación, control de calidad y adquisición de datos. Los autores presentan la información a un nivel de detalle técnico medio, ideal para uso de investigadores a nivel de maestría afín a las geociencias.
- El libro *Modelos Digitales de Terreno*, producto de tesis doctoral de Ángel Manuel Felicísimo, que desde su publicación en 1994 se ha convertido en referente obligado para los apasionados por el tema, y especialmente para aquellos que nos iniciamos en el camino de la investigación, esta obra se especializa en el tratamiento de los DEM y sus productos derivados, haciendo un aporte invaluable en relación al análisis y utilización concreta de este tipo de representación de la superficie terrestre.
- El libro *Digital Terrain Modelling, development and applications in a policy support environment*, que es una obra editada por R. J. Peckham y G. Jordan y publicada por la editorial Springer en enero de 2007, presenta una compilación de resultados de grupos internacionales de investigación, que describen los avances recientes en las teorías, fundamentación y aplicación de los DEM.

- El artículo *Cross-validation as a means of investigating DEM interpolation error*, escrito por el profesor Stephen Wise del departamento de geografía de la Universidad de Sheffield en el Reino Unido, y publicado por la editorial Elsevier con el aval de la revista indexada Canadiense: *Computers & Geosciences* en el año 2011, que brinda una interesante metodología para la determinación del error de los DEM apoyado en la técnica de Validación Cruzada, parte del diseño metodológico de este proyecto es adaptación de la metodología propuesta por Wise en este valioso documento.

Fuentes secundarias:

El libro *Digital Terrain Modelling*, escrito por el profesor Naser El-Sheimy del departamento de Ingeniería Geomática en la Universidad de Clagary (1999), que ofrece una clara fundamentación en términos de la generación, manipulación, interpretación, visualización y aplicación de los DEM a las ciencias de la tierra.

El artículo *Analysis of errors of derived slope and aspect related to DEM data properties*, editado por Elsevier y publicado en el año 2004 por la revista indexada *Computers & Geosciences*, presentado por los profesores Qiming Zhou, y Xuejun Liu del departamento de geografía de la Universidad de Baptist en Hong Kong; que entregan los resultados de una investigación de la precisión de los algoritmos en la medición de la pendiente y el aspecto de un DEM con estructura ráster.

Otras fuentes:

Conferencia presentada por J. Morillo Barragán et al. en el XIV Congreso Internacional de Ingeniería Gráfica en Santander España en el año 2002, titulada: *análisis de calidad de un modelo digital de elevaciones generado con distintas técnicas de interpolación*, donde su objetivo fue el cálculo del error y la fiabilidad de un DEM con distintos algoritmos de interpolación, y su contraste en función del número y la calidad de los puntos de control.

El artículo publicado en la revista indexada: *Investigaciones Geográficas* en el año 2008, *Evaluación de los errores de modelos digitales de elevación obtenidos por cuatro métodos de interpolación*, presentado por Azucena Pérez Vega, y Jean Francois Mas, profesores de la Universidad Nacional Autónoma de México, donde expresan que la resolución, el método de interpolación y los insumos (curvas de nivel solas o con datos de escurrimientos y puntos acotados) influyen de manera importante en la magnitud de los errores generados en el DEM.

## **2.2 Proceso de revisión de la literatura**

Entre los paradigmas dominantes en el campo de la investigación, la revisión literaria comprende el marco fundamental del desarrollo metodológico del proyecto; la selección, clasificación y depuración de la bibliografía a utilizar hace que se obtengan los documentos de mayor interés para el sustento fuerte del un documento de investigación bien fundamentado, así autores reconocidos en el área de los DEM como Ángel Manuel Felicísimo, Naser El-Sheimy, Zhilin Li, Chistopher Gold, Burroughs, y Mc Donnell son los de mayor apetencia entre los investigadores relacionados con la modelación de la superficie terrestre.

La revisión literaria (en este caso particular) no exigió búsquedas muy especializadas pues desde hace ya algunos años y sin pensar en este objetivo, he venido engrosando las filas de mi biblioteca personal con los documentos citados en las fuentes primarias, secundarias y otras en el apartado anterior. Además de los documentos ya citados, cuento con gran extensión de artículos (algunos más interesantes que otros) producto de búsquedas especializadas con bases de datos científicas como Scopus, ERIC, y Science Direct que me han brindado un bagaje importante a la hora de realizar la selección literaria.

### **Capítulo 3: Metodología y diseño de la investigación**

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

La metodología propuesta es adaptada de la que usaron Rees (2000) y Kidner (2003), donde re-muestrearon un DEM que originalmente tiene una resolución espacial de 50 m, a una serie de resoluciones mayores con los que se generaron nuevos DEM, los cuales fueron interpolados nuevamente a la resolución inicial (50 m), para poder comparar los valores interpolados con el DEM original.

Las figuras 1,2,3,4,5, y 6 corresponden a modelos de interpolación propuestos en este estudio pertenecientes a un mismo conjunto de datos muestrales, dejando percibir a simple vista que se pueden generar diferencias altimétricas (que se traducen en error) como consecuencia de la selección del algoritmo. Es evidente que si se toman datos de una superficie geométrica, como por ejemplo un paraboloides, y se adopta la hipótesis fundamental de que la interpolación con paraboloides es adecuada, los resultados serán óptimos. Sin embargo, la naturaleza general de los datos reales puede discrepar bastante de las hipótesis que justifican cada técnica de interpolación y, a pesar de dichas discrepancias, obtener interpolaciones razonables.

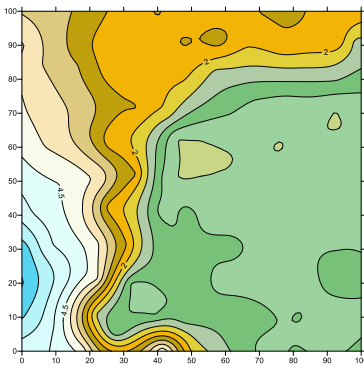


Figura 1 - Kriging

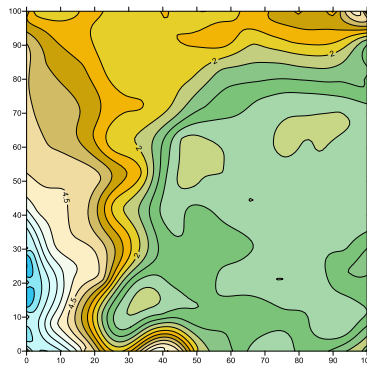


Figura 2- Spline

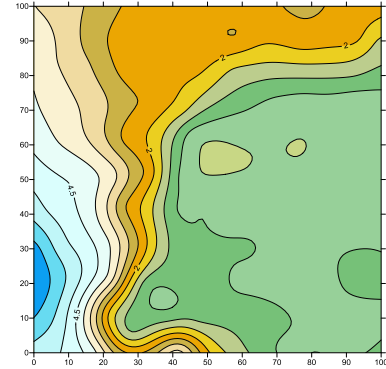


Figura 3- Topo to raster

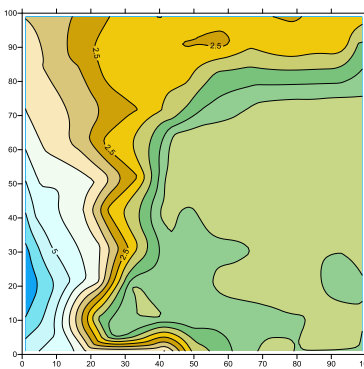


Figura 4- TIN

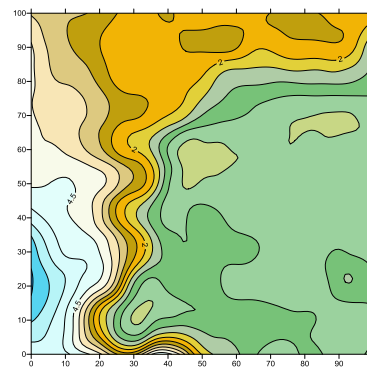


Figura 5 - Natural Neighbor

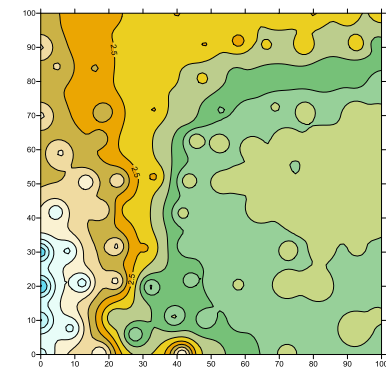


Figura 6 - IDW

El objeto de la metodología propuesta es la implementación de un modelo experimental que permita en función de una muestra considerable de elevaciones provenientes de una imagen ASTER GDEM, determinar qué algoritmo (s) representa (n) mejor la superficie del área de estudio, sin comprometer la calidad altimétrica de la representación de la superficie en los DEM evaluados. Para lograr este objetivo se propone el siguiente protocolo:

## **a) Obtención del universo de datos**

Se entiende por universo la totalidad de datos que conforman el ámbito de estudio, que para este caso en particular consiste en la extracción de las elevaciones de una zona correspondiente a imagen de satélite ASTER GDEM, con una cobertura aproximada de 16 Km x 16 Km, y resolución espacial de 30 m, es así como el universo de datos tendrá aproximadamente 300.000 muestras.

## **b) Re-muestreo**

El re-muestreo es el procedimiento con el cual se pueden determinar nuevos valores de las celdas de un ráster en función de un tamaño determinado.

La intención de este procedimiento es re-muestrear el DEM original hasta un punto tal que los datos de baja densidad comiencen a generar errores claros en los DEM interpolados. Es importante que en todos los casos las cuatro esquinas de la zona de estudio se incluyan como puntos del conjunto de datos del re-muestreo, de lo contrario algunos puntos serán estimados mediante extrapolación en lugar de interpolación. Con el fin de evitar esto se seleccionó un área de estudio con casi 300.000 de puntos en X y Y, lo que permite que el DEM pueda ser re-muestreado por potencias de 3 y conserve un conjunto razonable de datos con los cuales interpolar. Los valores utilizados, el espaciamiento y el número de puntos en las retículas re-muestreadas se muestran en la tabla 1

Tabla 1. Re-muestreo de datos

Re-muestreo	Tamaño de celda	Número de datos
0	30m x 30m	292.681
1	60m x 60m	73.442
2	90m x 120m	32.761
3	120m x 120m	18.496
4	150m x 150m	11.881
5	180m x 180m	8.281

### c) Interpolación

Interpolarse cada uno de los DEM re-muestreados por los siguientes métodos: IDW, Kriging, Natural Neighbor, spline, Topo to Raster, y TIN.

### d) Cálculo del error

De acuerdo a la teoría de errores desarrollada por Gauss, se ha adoptado el Error Medio Cuadrático como el modelo matemático más exigente para calcular el error de un DEM, por lo que su uso facilita la comparación de los valores obtenidos; lo que se pretende es establecer y cuantificar las diferencias más significativas encontradas entre la aplicación de estos algoritmos usados para la modelación de superficies.

$$EMC_{elev} = \sqrt{\frac{\sum_{i=1}^n (Z_p - Z_o)^2}{n - 1}} \quad (2)$$

Para el cálculo del EMC el procedimiento a seguir desde el punto de vista teórico es:



- En primer lugar hay que calcular, para cada punto de control, la diferencia altimétrica existente entre el modelo interpolado y su correspondiente posición con el dato de elevación original del DEM.
- Luego se elevan dichos residuos al cuadrado  $(Z_p - Z_o)^2$  para eliminar el valor del signo.
- Después se obtiene el promedio de las diferencias para todo el conjunto de puntos de control  $\frac{\sum_{i=1}^n (Z_p - Z_o)^2}{n-1}$
- Por último se halla la raíz cuadrada de dicho promedio obteniendo el EMC.

El resultado es la cobertura puntual que contiene los registros que sirven para caracterizar los modelos del área de estudio en función de su EMC.

Es así como se pretende extraer de los DEM producidos los valores de elevación interpolados a la resolución espacial inicial, es decir 30 m. Con lo cual se obtendrá el resultado del valor interpolado para poder compararlo con el valor original de la muestra, obteniendo como producto final una matriz de errores.

### **e) Análisis estadístico**

Una vez obtenida la matriz de errores se hace necesario un procesamiento estadístico que permita darle rigor científico al estudio, es así como se plantea el Análisis de Varianza (ANOVA) que permite calcular la medida de la dispersión de un conjunto de datos alrededor de una magnitud promedio, determinando la magnitud del error que posee cada algoritmo.

La validez del Análisis de Varianza está supedita al cumplimiento de algunos requisitos de tipo matemático y probabilístico en las observaciones. Se dice que el modelo ANOVA es apto para comparar efectos de variables en un fenómeno si los datos experimentales se ajustan al modelo, es decir, si las condiciones subyacentes al modelo están implícitas en los datos. Para este caso específico se realizará un ANOVA de dos factores, lo que quiere decir que se permite comparar varios grupos de una variable cuantitativa, se definieron como variables nominales: a) los cinco re-muestreos y b) el algoritmo de interpolación; como la variable cuantitativa, es decir, en la cual se desean comparar los grupos: el Error Medio Cuadrático de cada re-muestreo, resumiendo se tiene:

Variables independiente: re-muestreos (5), algoritmos de interpolación (6)

Variables dependientes: Error Medio Cuadrático

La hipótesis que se pone a prueba en el ANOVA de dos factores es aquella donde las medias poblacionales (las medias del EMC en re-muestreo) son iguales. Si las medias poblacionales son iguales, esto significa que los grupos no difieren en el Error Medio Cuadrático, y que en consecuencia, el tamaño del re-muestreo no tiene ninguna relación con el Error Medio Cuadrático, ni con el algoritmo de interpolación.

$$h_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_a \quad (\text{hipótesis nula}) \quad (3)$$

$$h_1: \text{No es cierto } h_0 \quad (\text{hipótesis alternativa}) \quad (4)$$

La estrategia para poner a prueba la hipótesis nula  $h_0$  de igualdad de las medias consiste en obtener el valor F (de la distribución de Fisher) que refleja el grado de parecido existente entre las medias que se están comparando. Si las medias de la muestra son parecidas el coeficiente F

tomará un valor cercano a 1. Si las medias muestrales son distintas dicho coeficiente tomará un valor mayor que 1, y cuanto más diferentes sean las medias, mayor será el valor de F. Si el nivel crítico asociado al valor F es menor que 0,05 se rechaza la hipótesis de igualdad de medias y se concluye que no todas las medias poblacionales son iguales. En el caso contrario no se puede rechazar la hipótesis de igualdad y no se puede afirmar que los grupos que se están comparando difieran en sus promedios poblacionales.

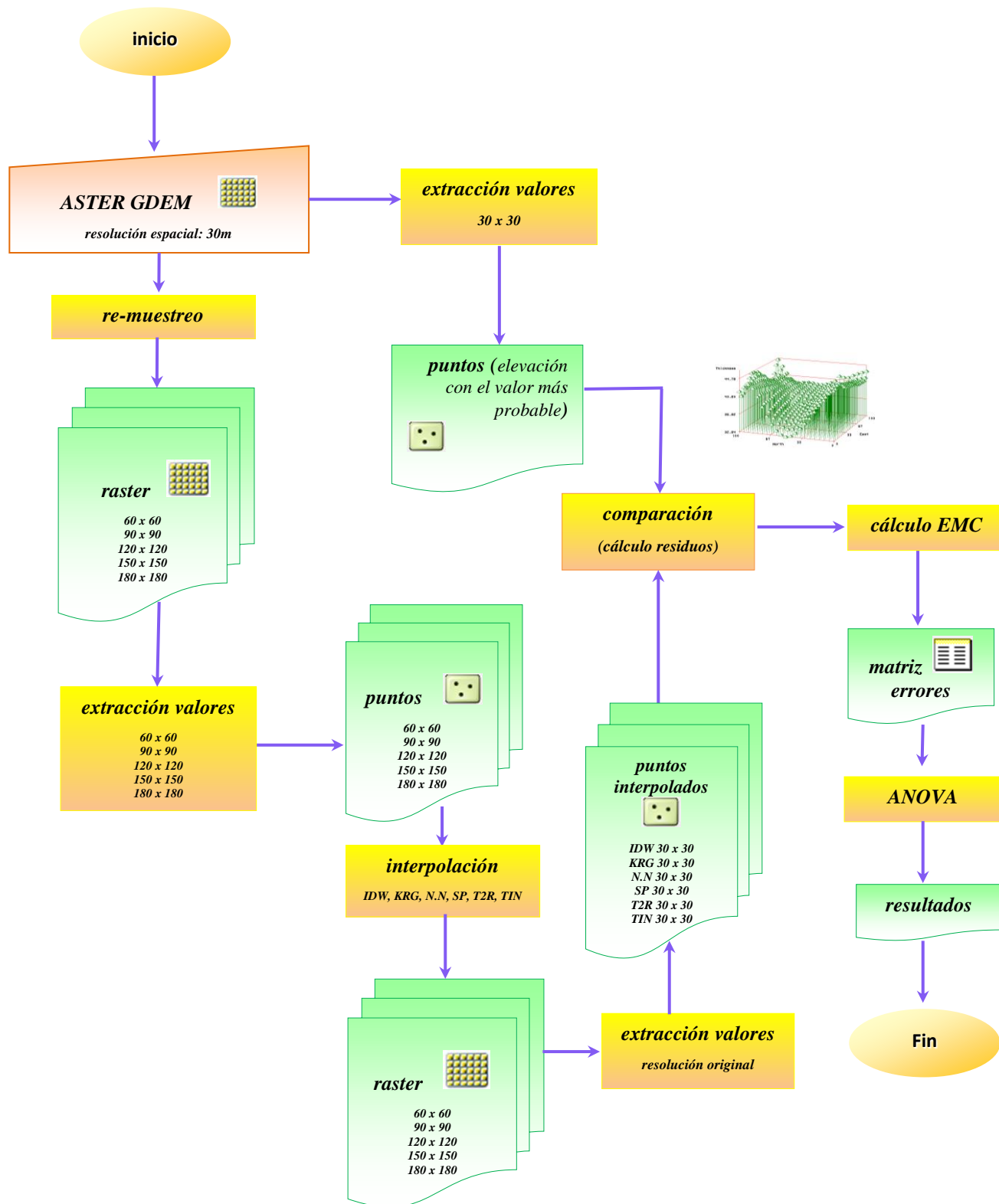


Figura 7 - Diagrama de flujo metodología del proyecto.

### 3.1 Justificación de la metodología seleccionada

En la modelación digital de superficies siempre se compone la realidad a partir de una muestra, con esa información como dato de entrada, y mediante procedimientos informáticos se reconstruye la superficie en formato digital de modo que el resultado obtenido conserve la mayor semejanza con el objeto real. Las fuentes de error en los DEM están estrechamente relacionadas con la calidad de la fuente de los datos, los instrumentos de su captura, los métodos de transformación de los puntos de control, la resolución espacial, y el modelo matemático para la construcción de la superficie, es decir: el algoritmo de interpolación.

La interpolación es el mecanismo que permite estimar la elevación en zonas donde no se han capturado datos altimétricos, se basa en el principio de autocorrelación espacial, el cual mide el grado de relación o dependencia entre las elevaciones cercanas y las distantes; en términos de representación de superficies la estructura de tipo ráster se considera una superficie funcional puesto que para cualquier posición  $x$  y  $y$  es almacenado solo un valor de  $z$ .

La metodología de esta tesis posee una estructura sencilla pero recursiva<sup>1</sup> que produce información derivada a distintas resoluciones de una misma fuente de datos que permite determinar la diferencia entre dos alturas pertenecientes a un mismo punto (la del DEM original y la del DEM interpolado), calculando el error de interpolación, y su incidencia sobre la representación de la superficie terrestre con los diferentes algoritmos de interpolación evaluados.

---

<sup>1</sup> Es la forma en la cual se especifica un procedimiento basado en su propia definición.

La información procesada siempre será imprecisa debido al procesos de simplificación que supone la modelación, es necesario conocer y controlar esta incertidumbre a cerca del error, con objeto de que se pueda conocer la fiabilidad de los resultados obtenidos, en esta tesis y con el ánimo de producir gran cantidad de datos derivados de un mismo DEM se ha optado por hacer el re-muestreo del ráster a diferentes resoluciones hasta que la densidad de los puntos produzca errores claramente identificables en los procesos de interpolación, es allí donde la rigidez de la estructura de la metodología aplica una técnica estadística muy apropiada para diseños experimentales: ANOVA que sirve para comprobar la hipótesis a través del análisis de la variabilidad donde ésta se divide en componentes asociados a una fuente específica de variación, estimando la magnitud con la que contribuye cada una de esas fuentes a la variación total, en función de esto, el análisis de la varianza se utiliza para dos propósitos, uno: probar la hipótesis respecto a las varianzas poblacionales, y dos: estimar y probar las hipótesis respecto a las medias poblacionales.

## 3.2 Métodos de interpolación evaluados

Este proyecto utiliza los algoritmos de interpolación que están disponibles en la mayoría los paquetes de software comercial, hay que destacar que éstos no son los mejores métodos para la producción de los DEM en los que se utiliza una serie de técnicas para modelar el terreno adecuadamente y reducir la incidencia de los algoritmos de interpolación (Sharif y Makarovic 1989, Podobnikar 2005). Sin embargo existen métodos típicos usados para producir muchos de los DEM hoy en día y que seguirán siendo utilizados por los no especialistas para su creación desde sus propias fuentes de datos. Los métodos que se eligieron se encuentran disponibles en ArcGIS 10, como indica la siguiente lista. Entre paréntesis se encuentran las abreviaturas que se usan en las figuras y tablas:

1. *Inverse Distance Weighted*: Método de ponderación inverso a la distancia (IDW)
2. *Kriging* (KRG)
3. *Natural Neighbor*: Vecinos naturales (N.N)
4. *Spline*: método de mínima curvatura (SP)
5. *Topo to Raster*: Topogrid (T2R)
6. *Triangular Irregular Network*: Red de Triángulos Irregulares (TIN)

El método de interpolación de *ponderación inverso a la distancia* (IDW) fue elegido por su masivo uso en los entornos SIG y especialmente por su querencia en los procedimientos de reducción de escala. El método de interpolación *Spline* (SPL) se considera de uso general y al igual que el método de *Vecinos Naturales* (N.N) se presume que pueden ser de mucha utilidad en situaciones en las que los datos originales son escasos, pues se puede argumentar que no es necesario forzar la superficie interpolada a través de los puntos, especialmente si se

considera que los datos contienen errores, además de que han sido utilizados en trabajos anteriores del autor (Garzón, 2012).

El interpolador *Topo to Raster* (T2R) se eligió porque fue específicamente diseñado para crear DEM derivando modelos hidrológicamente correctos, convirtiéndose en uno de los principales tributarios de la demanda de modelación de superficies terrestres, *Kriging* es considerado el algoritmo geoestadístico por excelencia, puede utilizarse con casi cualquier tipo de dato debido a que su estructura es muy flexible, se considera uno de los mejores métodos de interpolación ya que produce estimaciones insesgadas y de mínima varianza, y el por último, la inclusión del algoritmo TIN que desde 1968 se considera como el más apropiado para representar las formas topográficas en los entornos SIG.

En todos los casos los algoritmos necesitan valores de los parámetros alternativos que controlan la interpolación, y que generalmente modifican el producto final, su elección depende de las características propias de los datos utilizados: cantidad, densidad y distribución espacial, para este estudio y con el ánimo de evitar confusiones lógicas derivada de la multiplicada generación de los distintos modelos, todos los parámetros se utilizaron por defecto conforme como los sugiere ArcGIS 10.

### **3.2.1 IDW: método de ponderación inverso a la distancia (Inverse *Distance Weighted*)**

Este algoritmo calcula los valores de las celdas por medio de una composición ponderada de los datos de la muestra, el modelo sugiere que el resultado predicho reduce su incidencia en la medida en que aumenta la separación entre el punto a evaluar y los puntos de su entorno. Es



así como se acepta que los puntos más cercanos al centroide poseen un mayor peso en el cálculo del valor definitivo de la elevación para dicha celda, su modelo matemático es:

$$z = \left( \sum \left( \frac{Z_i}{d^n} \right) \right) / \left( \sum \left( \frac{1}{d^n} \right) \right) \quad (5)$$

donde,

*Z*: es el valor a interpolar

*Z<sub>i</sub>*: son los valores de elevación de los puntos vecinos

*d*: distancia entre los puntos vecinos y el punto a predecir

*n*: es la potencia a la cual se eleva la distancia, y oscila entre 0,5 y 4 (ESRI, 1993)

Este procedimiento de interpolación tiene en cuenta:

- La definición de la zona circundante del punto a interpolar (delimitación de radio de acción, forma de hacer el re-muestreo y la orientación de los puntos).
- La cantidad de puntos que se utilicen al interior de la zona circundante.
- La elección de la función matemática que se ajuste a la variación sobre la cantidad de puntos seleccionados.

La función que se use en la ponderación y la cantidad de puntos que se usen como vecinos en el área circundante, son parámetros definitivos en el resultado de la interpolación. La potencia *n* permite controlar la significancia de los puntos conocidos en los valores interpolados basándose en la distancia desde el punto de salida, su valor predeterminado es 2. (ESRI,2010)

### 3.2.2 KRG: método de Kriging

Es una técnica de interpolación geoestadística que determina el valor de la elevación apoyado en el cálculo del promedio de las elevaciones más próximas. Su cómputo se apoya en la teoría de la variable regionalizada, que presupone que la variación espacial de una variable está dada en función de tres grandes módulos:

- Un módulo estructural que tiene una tendencia constante (componente determinístico).
- Un módulo aleatorio que se conoce como variable regionalizada y que se encuentra espacialmente corelacionado.
- Y un módulo aleatorio no correlacionado de error no explicable conocido como error residual.

Este modelo busca minimizar la varianza del error, y llevar la media del error de los valores predecidos a cero, para que no se presenten sobre ni subestimaciones, su expresión matemática es:

$$Z_{(x_0)} = \lambda_1 \cdot Z_{(x_1)} + \lambda_2 \cdot Z_{(x_2)} + \lambda_3 \cdot Z_{(x_3)} + \dots + \lambda_n \cdot Z_{(x_n)} \quad (6)$$

donde,

$Z_{(x_0)}$ : Valor estimado de  $Z(x)$ , en el punto  $X_0$ .

$\lambda_i$ : Peso asignado a la  $i$  – ésima medición, para realizar la predicción del valor  $Z_{(x_0)}$

$Z_{(x_n)}$ : elevación de los puntos vecinos que intervienen en el procedimiento.

Se tiene entonces:

$$Z(x_0) = \sum_{i=0}^n \lambda_i \cdot Z(x_i) \quad (7)$$

En la siguiente figura se puede apreciar la relación de un punto (color rojo) con el resto de puntos que tendrán incidencia en la predicción:

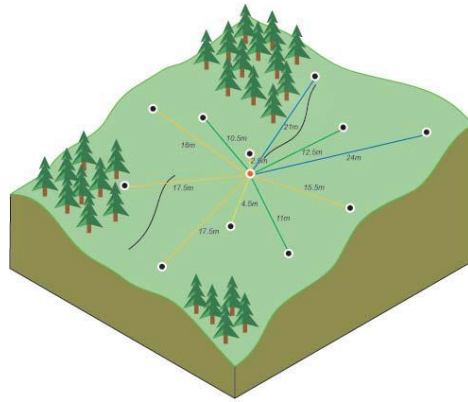


Figura 8 - Cálculo de la diferencia cuadrada entre las ubicaciones asociadas. Fuente: Using ArcGIS® Geostatistical Analyst, 2003.

La representación de las distancias de cada punto al resto de ellos es un tanto complicada de interpretar en el gráfico anterior, en su lugar se utiliza el semivariograma empírico como el modelo que asume la dirección y distancia entre los puntos, indicando una autocorrelación espacial que se puede usar para explicar la continuidad de la superficie, esto se conoce como variografía y se refleja mediante un semivariograma.

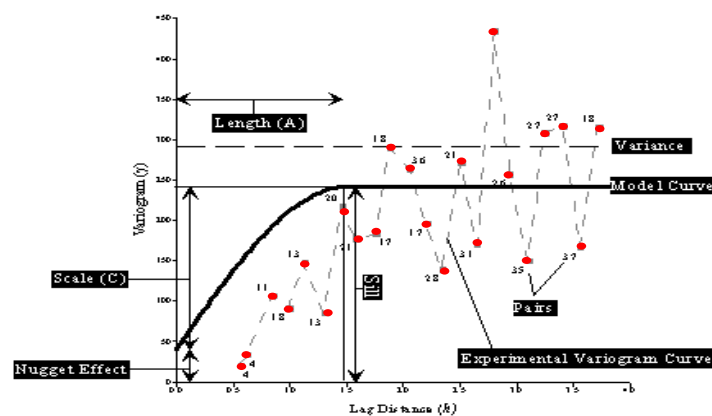


Figura 9- Semivariograma y sus características. Fuente: Golden Software, 2002

El interpolador refleja la variación espacial de la elevación en diferentes direcciones del espacio geográfico haciendo el cálculo del semivariograma empírico (puntos de color rojo),

deduciendo que tanta similaridad existe entre los puntos a medida que aumenta su separación, y se calcula de la siguiente forma:

$$\gamma(h) = \sum_{i=1}^{n(h)} \frac{(Z(x_1) - Z(x_1 + h))^2}{2n(h)} \quad (8)$$

donde,

$\gamma(h)$ : es la semivarianza para todas las muestras distanciadas por el intervalo de separación  $h$

$n(h)$ : número total de pares de la muestra

$Z(x_1)$ : es el valor de la muestra en una posición  $x_1$

$Z(x_1 + h)$ : es el valor de la muestra a la distancia del intervalo  $h$  desde  $x$

Para poder realizar la predicción del valor es necesario ajustar el semivariograma empírico a un semivariograma experimental para cuantificar el grado y escala de variación espacial, siendo esto similar a un análisis de regresión donde se ajustan un conjunto de puntos a una línea o curva continua, como se puede apreciar en el siguiente gráfico (línea de color azul):

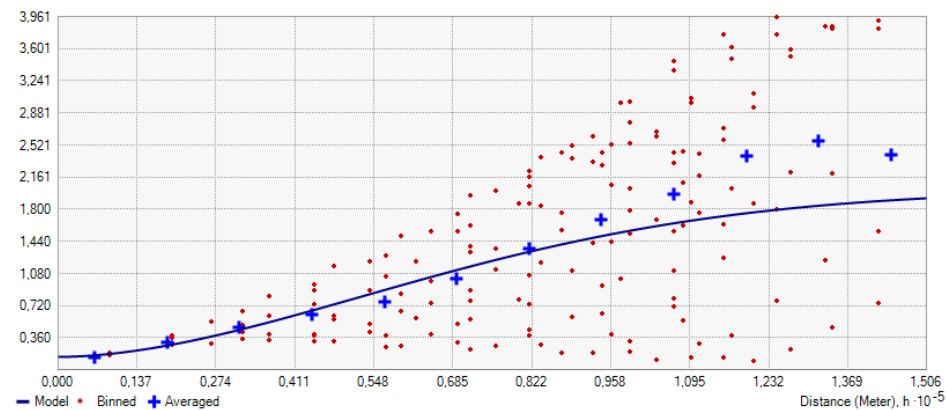


Figura 10 - Semivariograma empírico ajustado al experimental.

Con este ajuste se pueden extraer los parámetros que se utilizan para realizar la predicción:

- Rango (*length*): muestra la distancia donde la separación entre puntos indica una autocorrelación significativa. Indica la amplitud del radio de búsqueda de los vecinos locales.
- Efecto pepita (*nugget effect*): es el atribuido a los errores de medición.
- meseta (*sill*): máxima semivarianza hallada entre un par de puntos, que es encargada de encontrar la incertidumbre en el momento de la interpolación.
- Lag (*h*): son las distintas distancias a las que se mide la semivarianza, es como un círculo cuyo radio se va ampliando de forma regular.

Existen muchos tipos de semivariograma, que de acuerdo con el tipo, calidad, distribución y representación de los datos se debe elegir para realizar el ajuste, los principales son:

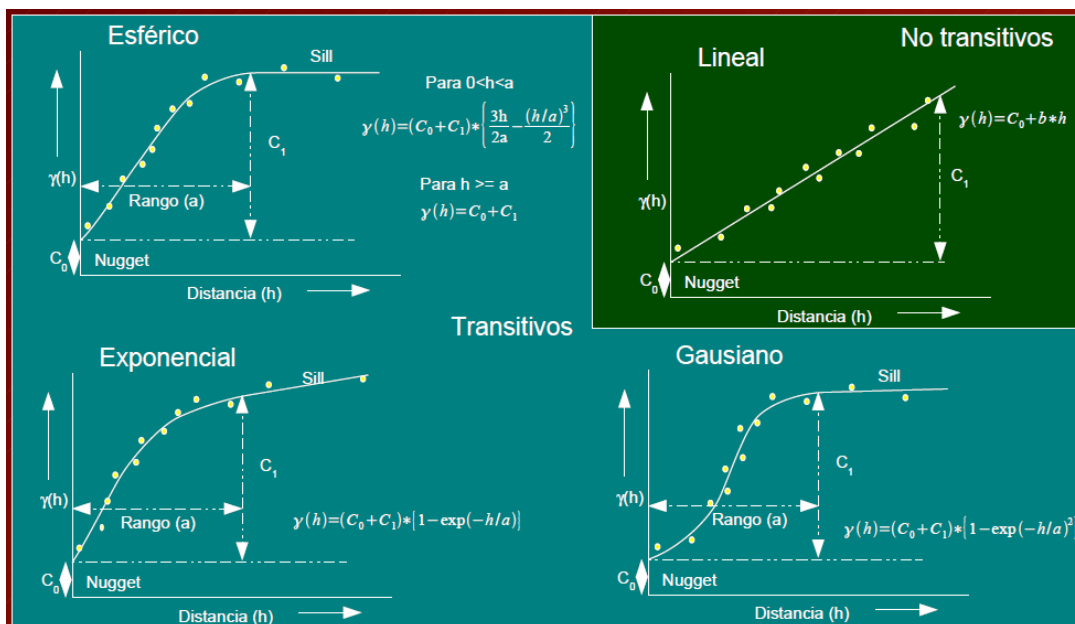


Figura 11 - Tipos de semivariograma. Fuente: Miguel Ángel Sanz, Modelos Digitales de Terreno, 2008

### 3.2.3 N.N: Vecinos naturales (*Natural Neighbor*)

Es una técnica de interpolación que utiliza las regiones naturales locales generadas alrededor de cada punto para elegir y ponderar la elevación de los puntos vecinos. Su ecuación básica es:

$$Z_{(x,y)} = \sum_{i=1}^n w_i \cdot f_{(x_i,y_i)} \quad (9)$$

donde,

$Z_{(x,y)}$ : es el valor de la estimación en la posición  $(x, y)$

$f_{(x_i,y_i)}$ : es el valor de la elevación observada en la posición  $(x_i, y_i)$

$w_i$ : es el peso asociado a  $f_{(x_i,y_i)}$

Esta ecuación es similar a la usada en el interpolador IDW, la diferencia está en el método utilizado para calcular los pesos, para ello, el algoritmo se apoya en la generación de los polígonos de *Thiessen* alrededor de los puntos dados, luego se crea un nuevo polígono de *Thiessen* alrededor del punto donde se desea conocer la elevación (estrella), y finalmente la sobreposición entre los polígonos se usa para otorgar los pesos.

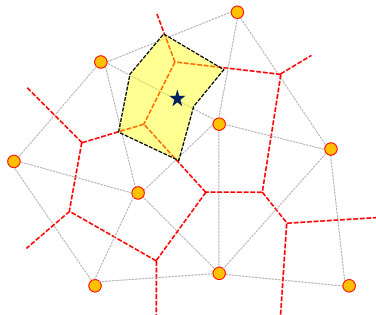


Figura 12 - Polígonos de Thiessen.

Esta técnica de interpolación tiene la ventaja de que no necesita la introducción de ningún parámetro adicional, y los resultados son aceptables en puntos muestrales con una distribución irregular. Esta eventual comodidad para el proceso, puede ser un inconveniente, si se tiene en cuenta el escaso número de parámetros de decisión que emplea la técnica.

### 3.2.4 SPL: método de mínima curvatura (*Spline*)

Es un algoritmo que busca hacer el ajuste de una función a una limitada cantidad de puntos, generando una línea que pasa exactamente por las muestras originales, asegurando así la continuidad en la unión de las distintas curvas. Para interpolar superficies se usa un tipo especial de *Spline* llamado de *mínima curvatura* que pasa exactamente por los puntos muestrales, consiguiendo ajustar el terreno a la forma de una membrana elástica. El modelo matemático del interpolador es:

$$Z_{(x,y)} = T_{(x,y)} + \sum_{i=1}^n \lambda_j \cdot R_{(r_i)} \quad (10)$$

donde,

$Z_{(x,y)}$  es el valor de la estimación en la posición  $(x, y)$

$n$  es la cantidad de puntos

$\lambda_j$ : son los coeficientes determinandos para la solución de un sistema lineal de ecuaciones

$r_i$ : es la distancia del punto  $x, y$  al punto  $i$

$T_{(x,y)}$  y  $R_{(r_i)}$ : se definen de acuerdo si se usa el algoritmo regularizado o por tensión

Este método se encuentra muy bien recomendado cuando los cambios en superficie tienen variaciones suaves, no es apropiado si existe gran cantidad de cambios en distancias horizontales muy cortas porque se pueden sobrepasar los valores estimados.

### **3.2.5 T2R: Topogrid (*Topo to Raster*)**

Este algoritmo está diseñado para generar superficies hidrológicamente conectadas y correctas, lo hace a partir de curvas de nivel o de entidades tipo punto, su característica principal es la preservación las líneas partidoras y las redes de transmisión de curvas de nivel. Este algoritmo es basado en el software ANUDEM desarrollado por Michael Hutchinson entre 1988 y 1989. El método usa una técnica de interpolación diferencial iterativa finita. Ha sido optimizado para lograr la eficiencia computacional de los métodos de interpolación locales como IDW sin perder la continuidad de la superficie de los métodos de interpolación global como Kriging y Spline. Es esencialmente una fina técnica de discretización en la cual se modifica la rugosidad para permitir que el proceso de ajuste del DEM siga los cambios abruptos en el terreno tales como líneas colectoras y partidoras.

El agua es la principal fuerza erosiva que determina la forma general de la mayoría de paisaje. Por este motivo las estructuras orográficas poseen hacia las partes altas las llamadas zonas de colina (máximos locales) y hacia las zonas bajas los sumideros naturales (mínimos locales), dando como resultado en su forma natural una estructura de drenaje conectada, el algoritmo usa este principio sobre toda la superficie e impone restricciones en el proceso de interpolación que dan como resultado una estructura que define correctamente las líneas partidoras y las colectoras. Esta condición impuesta por las redes de drenaje produce superficies más precisas con una menor cantidad de datos. La condición de drenaje global elimina cualquier necesidad de edición o de post-proceso para disminuir errores en la superficie generada.



### 3.2.6 TIN: Red de triángulos irregulares (*Triangular Irregular Network*)

Este algoritmo se apoya en la creación de una red conformada por triángulos irregulares cuyos vértices son los puntos muestrales obtenidos para la definición del terreno. Consiste en generar sobre un plano horizontal una red de triángulos, la cual resulta de la conexión, por medio de segmentos de rectas, de las proyecciones de los puntos del modelo sobre el plano. La conformación de estos triángulos, conocida como triangulación, debe hacerse de forma que:

- Los triángulos sean, en lo posible, equiláteros y pequeños,
- Se incluyan todos los puntos del modelo
- El resultado del proceso esté libre de ambigüedad (sea inequívoco)

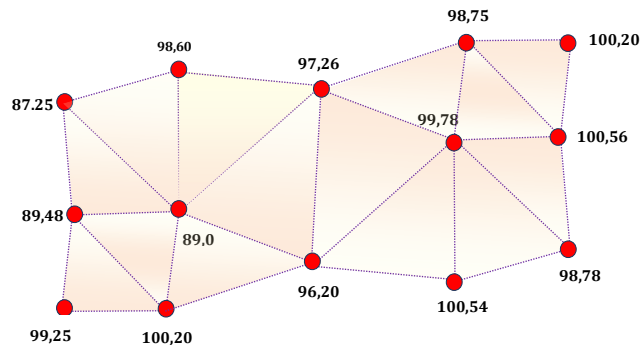


Figura 13 - Red de Triángulos Irregulares. Fuente: Garzón, introducción a los Modelos Digitales de Elevación, 2012

Para poder sistematizar el procedimiento existen varios algoritmos que seleccionan la mejor forma de hacer la triangulación, el más conocido es el de *Delaunay* el cual se apoya en los polígonos de *Thiessen*, ésta condición dice que la circunferencia circunscrita de cada triángulo de la red no debe contener ningún vértice de otro triángulo. Dos puntos son vecinos naturales si se encuentran en el mismo círculo

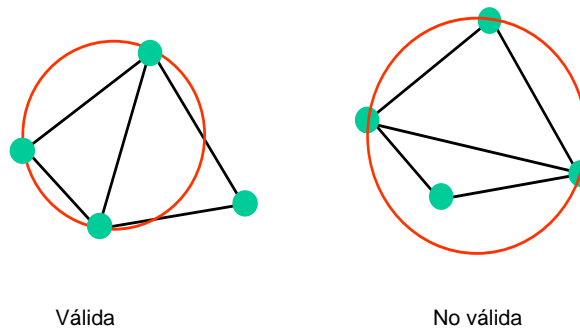


Figura 14 - Triangulación de Delaunay. Fuente: Garzón, Introducción a los Modelos Digitales de Elevación, 2012

Una vez se ha generado la triangulación, el siguiente paso es calcular la ecuación del plano que pasa a través de tres puntos en el espacio mediante el siguiente modelo:

$$A_x + B_y + C_z + D = 0 \quad (11)$$

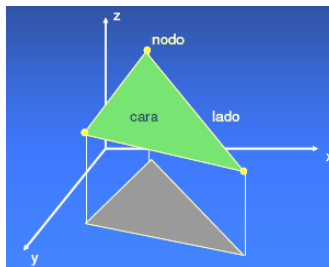


Figura 15 - Plano que pasa a través de tres puntos en el espacio.

Donde los coeficientes, A, B, C y D se calculan sabiendo que este plano pasa por los puntos  $P_1(x_1, y_1)$ ,  $P_2(x_2, y_2)$ ,  $P_3(x_3, y_3)$ , se soluciona entonces un sistema de tres ecuaciones con tres incógnitas, los coeficientes se obtienen con las siguientes expresiones:

$$A = y_1(z_2 - z_3) + y_2(z_3 - z_1) + y_3(z_1 - z_2) \quad (12)$$

$$B = z_1(x_2 - x_3) + z_2(x_3 - x_1) + z_3(x_1 - x_2) \quad (13)$$

$$C = x_1(y_2 - y_3) + x_2(y_3 - y_1) + x_3(y_1 - y_2) \quad (14)$$

$$D = Ax_1 - By_1 - Cz_1 \quad (15)$$

En definitiva la función interpolante es:

$$z = Z(x, y) = -\frac{A}{C}x - \frac{B}{C}y - \frac{D}{C} \quad (16)$$

La cual se utilizará para interpolar la cota del punto  $p_x$  al interior del triángulo que define al plano dado por la expresión anterior (11).

### 3.3 Validación cruzada

Existe una manera de corroborar el resultado de las predicciones realizadas por los algoritmos de interpolación espacial, se conoce con el nombre de *Validación Cruzada*, y consiste en eliminar un valor de elevación, ejecutar el algoritmo de interpolación y estimar el valor eliminado, si esto se hace a distintos niveles de re-muestreo del ráster original, se podrán representar todos los valores interpolados frente a sus valores reales, la diferencia entre estos dos valores se conoce como *residual de la validación* y se usa para calcular el EMC de cada uno de los algoritmos evaluados a las diferentes resoluciones propuestas.

En la técnica de validación cruzada los procesos de re-muestreo son de uso generalizado. Para poder calcular los valores de los puntos que se desean evaluar se utiliza esta técnica, que es un proceso de determinación de nuevos valores de las celdas en una retícula que resulta luego de aplicar alguna transformación geométrica al *grid* de entrada. Esta transformación puede ser en un diferente sistema coordenado, a una resolución distinta, o puede ser una rotación con respecto al *grid* inicial. Para este caso en particular se optó por hacer cambios de resolución como lo indica la tabla 1.

Tabla 1. Re-muestreo de datos

Re-muestreo	Tamaño de celda	Número de datos
0	30m x 30m	292.681
1	60m x 60m	73.442
2	90m x 120m	32.761
3	120m x 120m	18.496
4	150m x 150m	11.881
5	180m x 180m	8.281

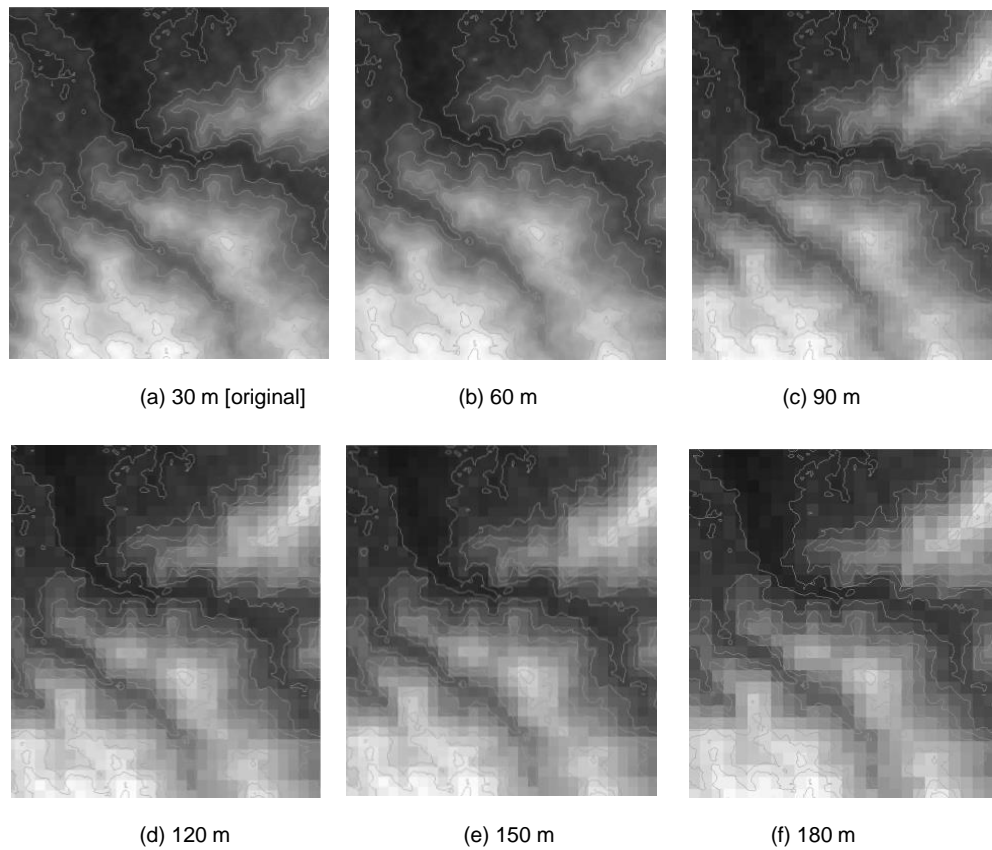


Figura 16 - Resoluciones del re-muestreo.

Como se aprecia en la figura anterior, se han hecho cinco niveles de re-muestreos con resoluciones más gruesas a partir del conjunto de datos original, con los cuales se pretende percibir dónde se genera los errores más prominentes en los algoritmos de interpolación en una superficie topográfica asociada a distintos rangos de pendiente.

### 3.4 Análisis de Varianza (ANOVA)

La Varianza se define como la media aritmética de la suma de los cuadrados de las desviaciones de una variable con respecto a su media. Por lo tanto, cuanto mayor sea este valor, menos representativa de la realidad será la media de dicha variable.

ANOVA es la sigla para *ANalysis Of VAriance*, y se constituye como una técnica estadística normalmente utilizada para analizar resultados en procesos investigativos de diseños experimentales, con ella se puede comparar si los valores que toma un conjunto de datos que corresponden a variaciones de una misma variable dependiente, se ven afectados por una o más variables de tipo independiente.

El modelo tiene la siguiente expresión:

*donde,*

$$Y_{ij} = \mu + \tau_j + \varepsilon_{ij} \quad (17)$$

$Y_{ij}$  es la variable de estudio

$\mu$  es una constante e indica la respuesta media de todos los niveles

$\tau_j$  es el efecto diferencial del nivel  $j$  recoge la importancia de cada tratamiento y es el objetivo

del análisis dado que los  $\tau_j$  son efectos diferenciales sobre  $\mu$  tenemos que  $\sum \tau_i = 0$

$\varepsilon_{ij}$  es un término de error considerado como variable aleatoria

La validez del Análisis de Varianza está supedita al cumplimiento de algunos requisitos de tipo matemático y probabilístico en las observaciones. Se dice que el modelo ANOVA es apto para comparar efectos de variables en un fenómeno si los datos experimentales se ajustan al modelo, es decir, si las condiciones subyacentes al modelo están implícitas en los datos. Para este caso en particular se plantea un ANOVA de dos factores, lo que quiere decir que se

permite comparar varios grupos de una variable cuantitativa, para ello se definieron como variables nominales: a) el tipo de superficie (plano, ondulado, escarpado y montañoso), el algoritmo de interpolación y el re-muestreo (60, 90, 120, 150, 180) y b) como la variable cuantitativa, es decir, en la cual se desean comparar los grupos se tomó el Error Medio Cuadrático de cada una de las distintos DEM generados, resumiendo se tiene:

*Variable dependiente:* Error Medio Cuadrático.

*Variables independientes:* método de interpolación, tipo de superficie, re-muestreos

### 3.4.1 Hipótesis

La hipótesis que se pone a prueba en el ANOVA de dos factores, es aquella donde las medias poblacionales (las medias del EMC en cada nivel del tipo de superficie) son iguales. Si las medias poblacionales son iguales, esto significa que los grupos no difieren en el Error Medio Cuadrático, y que en consecuencia, el tipo de superficie, el tamaño del re-muestreo y el método de interpolación son independientes del Error Medio Cuadrático.

$$h_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_a \quad (\text{hipótesis nula}) \quad (18)$$

$$h_1: \text{No es cierto } h_0 \quad (\text{hipótesis alternativa}) \quad (19)$$

La estrategia para poner a prueba la hipótesis nula  $h_0$  de igualdad de las medias consiste en obtener el valor F (distribución de Fisher) que indica el grado de semejanza que pueda haber entre las medias que se están comparando. Esta distribución permite hacer cálculos sobre las

varianzas estableciendo si las diferencias calculadas son representativas y por lo tanto aplicables a cambios significativos en la conducta de las poblaciones en estudio. Si las medias de la muestra son parecidas el coeficiente F tomará un valor cercano a 1. Si estas son diferentes dicho coeficiente tomará un valor mayor que 1, y en la medida en que estas sean distintas mayor será el valor de F. Si el nivel crítico asociado al valor F es menor que 0,05 se rechaza la hipótesis de igualdad de medias y se concluye que no todas las medias poblacionales son iguales. En el caso contrario no se puede rechazar la hipótesis de igualdad y no se puede afirmar que los grupos que se están comparando difieran en sus promedios poblacionales.

$$\text{modelo de error} = \mu + t_i + m_{i,j} + e_{i,j,k} \quad (20)$$

*Donde,*

$\mu$ : error propio del modelo

$t_i$ : error debido al re – muestreo

$m_{i,j}$ : error debido al modelo de interpolación

$e_{i,j,k}$ : error aleatorio

### **3.5 Zona de estudio**

En la actualidad el DEM de uso libre de más fácil acceso y mayor resolución espacial es ASTER GDEM, muchos estudios relacionados con el tema han indicado que existe una relación entre el error y el tipo de terreno, debido a ello se seleccionó un área que contuviese distintos rangos de pendiente (tabla 9), dicha zona se encuentra en Colombia y más específicamente en territorio limítrofe entre de los departamentos de Nariño y Cauca.

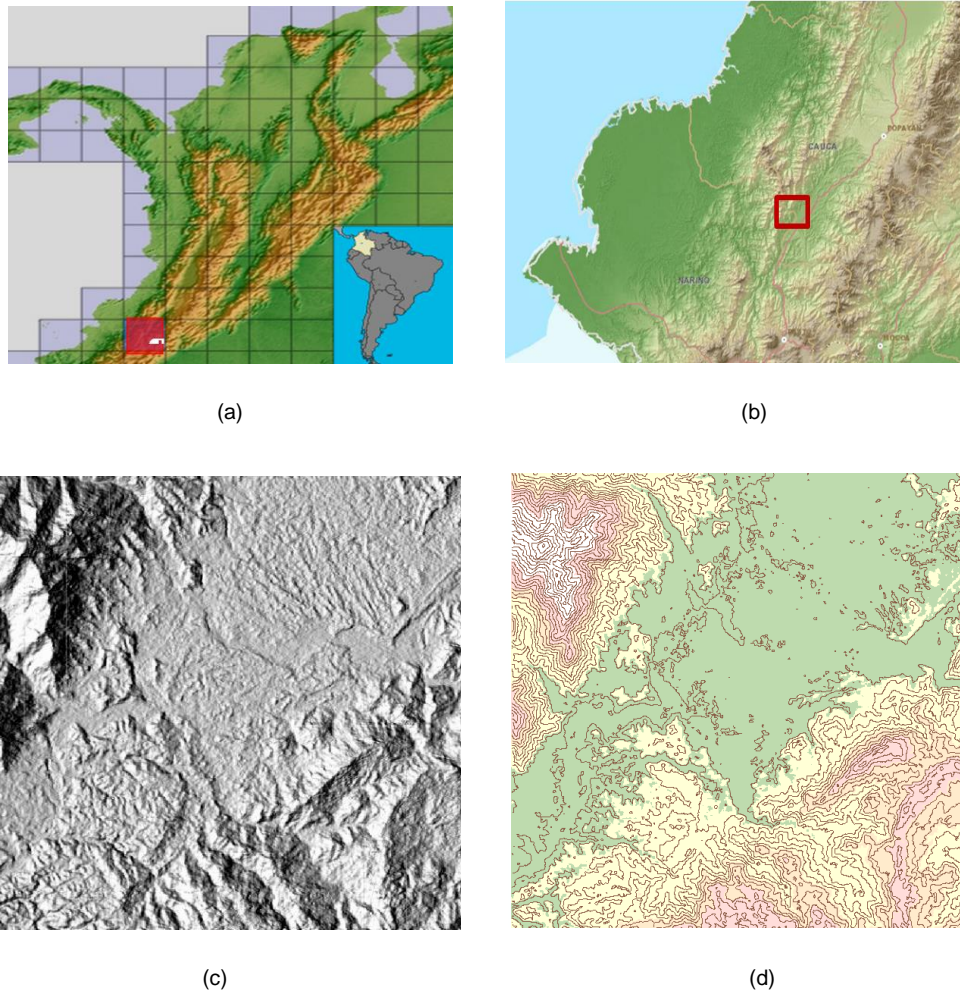


Figura 17 - Localización del área de estudio. (a) Localización en el continente sur Americano. (b) Localización entre los Departamentos de Cauca y Nariño. (c) Mapa de iluminación del DEM correspondiente al área de estudio. (d) Mapa con curvas de nivel equidistantes 50 metros.

Este DEM tiene una resolución espacial de 30 m, con 541 filas por 541 columnas generando una cobertura de 263 Km<sup>2</sup>, y cuenta con elevaciones que van entre los 500 a 1600 metros sobre el nivel medio del mar. La imagen de la cual se extrajo el área de estudio se encuentra enmarcada entre las latitudes 01° y 02° N, y las longitudes 77° y 78° W, esta zona se caracteriza por ser fuertemente hídrica para el Departamento del Cauca, además de albergar vegetación xerofítica y ecosistemas ricos en flora y fauna declarados como reserva de la biósfera por las UNESCO.



## Capítulo 4: Análisis de datos

Como se estableció en la metodología, el procesamiento estadístico de los datos se hará mediante un *análisis de varianza* de varios factores para la variable dependiente: *Error Medio Cuadrático*. De igual forma se realizarán varias pruebas y gráficas para determinar qué factores tienen un efecto estadísticamente significativo sobre el EMC. Este análisis también evaluará la significancia de las interacciones entre los factores (variables independientes). Las pruebas-F en la tabla ANOVA permitirán identificar los factores significativos. Para cada factor significativo, las *Pruebas de Rangos Múltiples* darán informe de cuáles medias son significativamente diferentes de otras. La *Gráfica de Medias* y la *Gráfica de Interacciones* ayudarán a interpretar los efectos significativos. Las *Gráficas de Residuos* servirán para juzgar si los datos han violado los supuestos subyacentes al análisis de varianza.

*Variable dependiente:* Error Medio Cuadrático

*Factores:*

1. Re-muestreo (60, 90, 120, 150, 180)
2. Método de interpolación (IDW, KRG, NN, SP, T2R, TIN)

Tabla 2. Análisis de varianza para EMC – suma de cuadrados tipo III

Fuente	Suma de Cuadrados	GI	Cuadrado Medio	Razón-F	Valor-P
<b>EFFECTOS PRINCIPALES</b>					
A: interpolador	15,9981	5	3,19962	48,69	0,0000
B: re-muestreo	147,872	4	36,968	562,50	0,0000
<b>RESIDUOS</b>	1,31441	20	0,0657207		
TOTAL (CORREGIDO)	165,185	29			

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

La tabla ANOVA descompone la variabilidad del EMC 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 2 valores-P son menores que 0,05, estos factores tienen un efecto estadísticamente significativo sobre el EMC 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 tamaño del re-muestreo del modelo ráster. Para hacer el contraste de la hipótesis del primer factor (algoritmos de interpolación), se hace necesario establecer la región de criterio de contraste, hallando el estadístico F-Snedecor:

$$F_{0,05(5;20)} = 2,71 \quad (21)$$

$$\text{Estimado: } 48,69 > 2,71 = F_{0,05(5;20)} \quad \text{Estadístico} \quad (22)$$

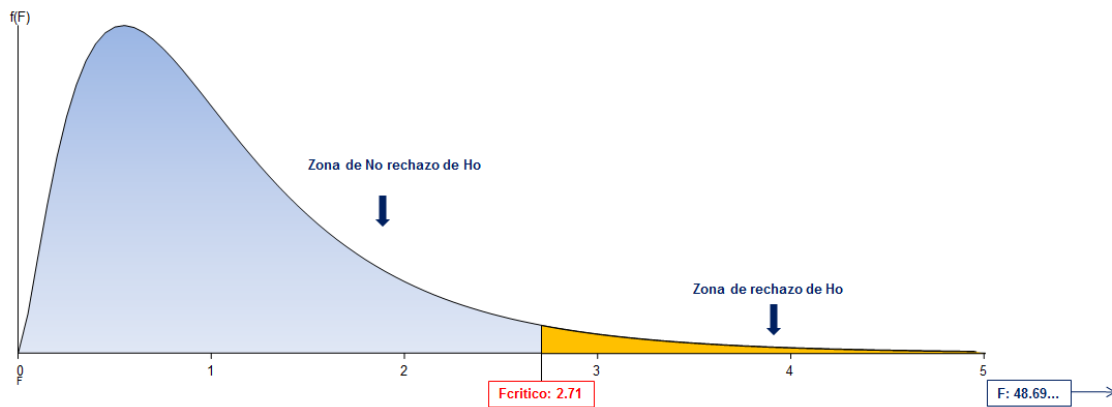


Figura 18 - Distribución Fisher Snedecor, factor 1: interpoladores.

El contraste del segundo factor: *re-muestreo*, presenta los siguientes resultados:

$$F_{0,05(4;20)} = 2,87 \quad (23)$$

$$\text{Estimado: } 562,50 > 2,87 = F_{0,05(4;20)} \text{ Estadístico} \quad (24)$$

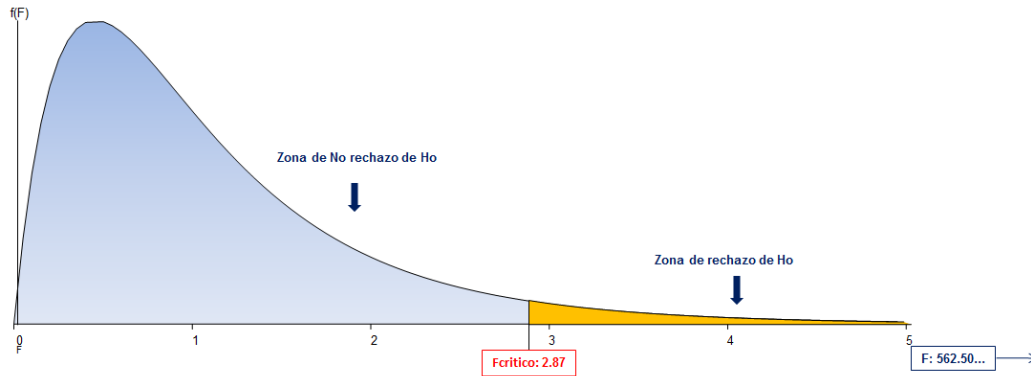


Figura 19 - Distribución Fisher Snedecor, factor 2: re-muestreo.

Al igual que en el contraste del factor *interpolador* se rechaza la hipótesis nula  $H_0$  considerando que el valor del estimador calculado  $F = 562,5$  cae fuera de la región de aceptación o contraste. Por lo tanto SI existe interacción entre el *Error Medio Cuadrático* y el nivel de *re-muestreo* usado en los ráster.

Se ha graficado la dispersión del error respecto de cada uno de los re-muestreos del ráster original (Figura 20), donde se corrobora algo que intuitivamente se presumía: la significancia entre el conjunto de datos, así, los menores errores independientemente del algoritmo de interpolación se concentran en los re-muestreos de valores más bajos, y los errores más predominantes se encuentran en los re-muestreos de mayor amplitud.

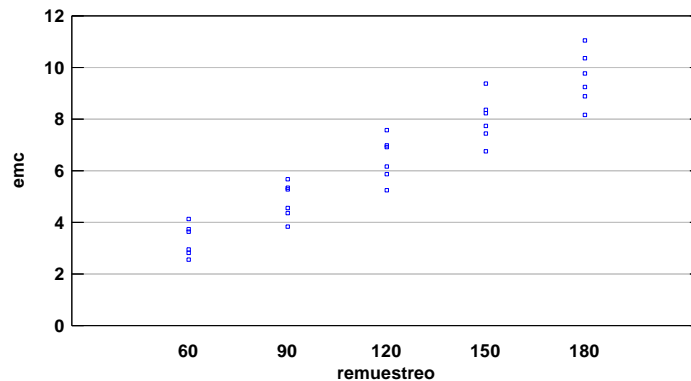


Figura 20 - Dispersión por código de nivel.

El mejor algoritmo de representación de la superficie evaluado en este estudio será aquel que en conjunto (diferentes órdenes de re-muestreo) presente el menor *Error Medio Cuadrático*, a la vista del gráfico de medias (Figura 21 (a)) se puede entender fácilmente que el algoritmo *Spline* es el que se encuentra más cercano al menor error, también se puede inferir que no existe una diferencia estadísticamente significativa entre los algoritmos: *Topo to Raster* y *Kriging*, pues el valor de error máximo en *Kriging* se encuentra contenido en el rango de error del interpolador *Topo to Raster*, de esta misma manera sucede entre los interpoladores: *TIN* y *Natural Neighbor*, aunque estos presentan menores deformaciones alimétricas que los dos anteriores. Es de aclarar en la interpretación de la figura 21 (a) que ésta indica las medias para el conjunto de la totalidad de los re-muestreos en cada uno de los interpoladores.

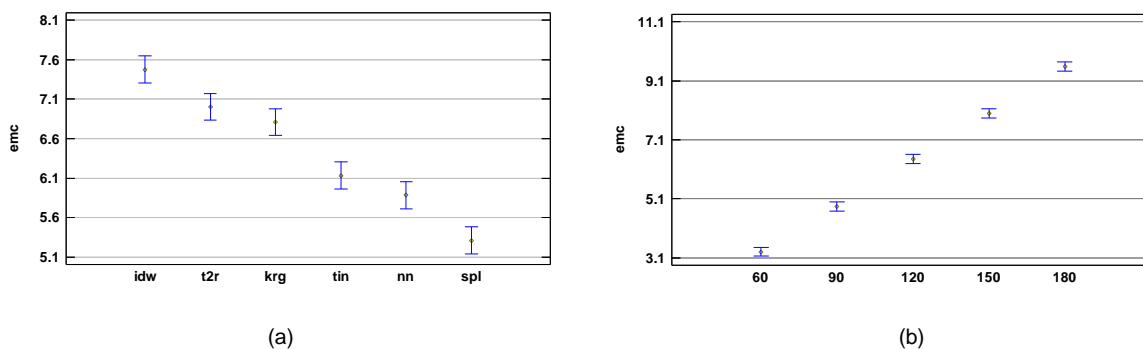


Figura 21 – Gráfico de medias, 95% de Fisher LSD. (a) EMC en función del algoritmo de interpolación. (b) EMC en función del tamaño de re-muestreo.

La tabla 3 muestra la media del *Error Medio Cuadrático* para cada uno de los niveles en los distintos factores (variables independientes). También muestra los errores estándar de cada media, los cuales se consideran una medida de la variabilidad en su muestreo. Las dos columnas de la derecha indican los intervalos de confianza del 95,0% para cada una de las medias (figura 21 (b)).

Tabla 3. Medias por Mínimos Cuadrados para el EMC con intervalos de confianza del 95.0%

Nivel	Casos	Media	Error. est.	Lim. Inf.	Lim. Sup.
MEDIA GLOBAL	30	6.436			
<b>Re-muestreo</b>					
60	6	3.30833	0.104659	3.09002	3.52665
90	6	4.84	0.104659	4.62169	5.05831
120	6	6.45833	0.104659	6.24002	6.67665
150	6	7.99167	0.104659	7.77335	8.20998
180	6	9.58167	0.104659	9.36335	9.79998
<b>interpolador</b>					
idw	5	7.476	0.114648	7.23685	7.71515
krig	5	6.808	0.114648	6.56885	7.04715
nn	5	5.884	0.114648	5.64485	6.12315
spl	5	5.312	0.114648	5.07285	5.55115
t2r	5	7.002	0.114648	6.76285	7.24115
tin	5	6.134	0.114648	5.89485	6.37315

En la tabla 4 se aplica un procedimiento de comparación múltiple para determinar cuáles medias son significativamente diferentes de otras. La mitad inferior de tabla de resultados muestra las diferencias estimadas entre cada par de medias. El asterisco que se encuentra al lado de los 10 pares (en el caso de los re-muestreos) indica que estas parejas muestran diferencias estadísticamente significativas con un nivel del 95,0% de confianza. (Esto actúa como soporte cuantitativo de las gráficas (a) y (b) de la figura 21). Se han identificado 5 grupos homogéneos según la alineación de las X's en columnas. No existen diferencias estadísticamente significativas entre aquellos niveles que compartan una misma columna de X's. El método empleado actualmente para discriminar entre las medias es el procedimiento

de diferencia mínima significativa (LSD) de Fisher. Con este método hay un riesgo del 5.0% al decir que cada par de medias es significativamente diferente, cuando la diferencia real es igual a 0.

Tabla 4. Pruebas de Múltiple Rangos para el EMC por el tipo de re-muestreo

re-muestreo	Casos	Media LS	Sigma LS	Grupos Homogéneos
60	6	3.30833	0.104659	X
90	6	4.84	0.104659	X
120	6	6.45833	0.104659	X
150	6	7.99167	0.104659	X
180	6	9.58167	0.104659	X

Contraste	Sig.	Diferencia	+/- Límites
60 - 90	*	-1.53167	0.308744
60 - 120	*	-3.15	0.308744
60 - 150	*	-4.68333	0.308744
60 - 180	*	-6.27333	0.308744
90 - 120	*	-1.61833	0.308744
90 - 150	*	-3.15167	0.308744
90 - 180	*	-4.74167	0.308744
120 - 150	*	-1.53333	0.308744
120 - 180	*	-3.12333	0.308744
150 - 180	*	-1.59	0.308744

\* indica una diferencia significativa.

Tabla 5. Pruebas de Múltiple Rangos para el EMC por el interpolador

interpolador	Casos	Media LS	Sigma LS	Grupos Homogéneos
spl	5	5.312	0.114648	X
nn	5	5.884	0.114648	X
tin	5	6.134	0.114648	X
krig	5	6.808	0.114648	X
t2r	5	7.002	0.114648	X
idw	5	7.476	0.114648	X

Contraste	Sig.	Diferencia	+/- Límites
idw - krig	*	0.668	0.338212
idw - nn	*	1.592	0.338212
idw - spl	*	2.164	0.338212
idw - t2r	*	0.474	0.338212
idw - tin	*	1.342	0.338212
krig - nn	*	0.924	0.338212
krig - spl	*	1.496	0.338212
krig - t2r	*	-0.194	0.338212
krig - tin	*	0.674	0.338212
nn - spl	*	0.572	0.338212
nn - t2r	*	-1.118	0.338212
nn - tin	*	-0.25	0.338212
spl - t2r	*	-1.69	0.338212
spl - tin	*	-0.822	0.338212
t2r - tin	*	0.868	0.338212

\* indica una diferencia significativa.

La figura 22 se conoce como el gráfico de interacción entre factores, debido a su configuración se puede señalar que las variables independientes: re-muestreo (60, 90, 120, 150, 180) y método de interpolación (IDW, KRG, N.N, SP, T2R, TIN) no interactúan, pudiéndose deducir esto a partir del comportamiento lineal y casi paralelo entre los factores evaluados. No obstante, esta gráfica es gran utilidad para observar la tendencia del error en cada interpolador para los distintos niveles de re-muestreo, pudiéndose identificar claramente a *Spline* como el algoritmo de mejor comportamiento para todos los casos.

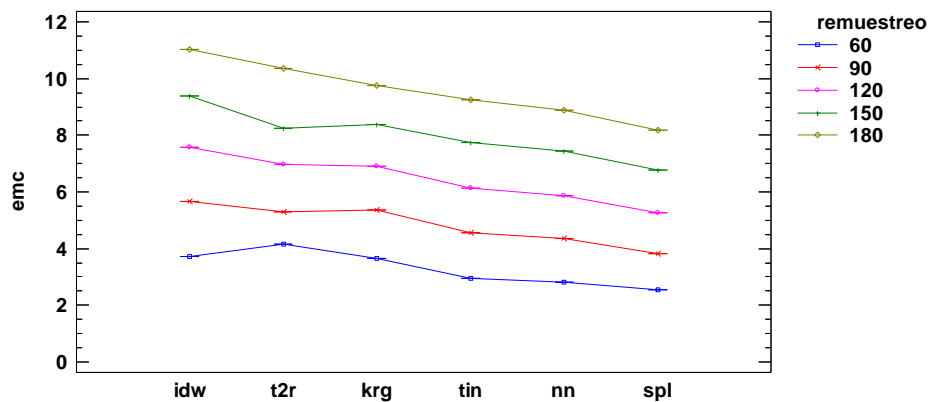


Figura 22 – interacción entre factores.

Analizando la figura 22 y de acuerdo con la prueba de múltiples rangos (tabla 5), el contraste entre los interpoladores indica que estadísticamente no existen diferencias consideradas significativas entre *KRG* y *T2R* puesto que su discrepancia es de -0,19, así mismo sucede con los interpoladores *N.N* y *TIN* ya que su diferencia es de -0,25; entre las demás posibles combinaciones de pares interpoladores si existen diferencias que deben ser tenidas en cuenta al momento de hacer la elección del algoritmo de interpolación.

## 4.1 Error en los valores de elevación

La cuantificación del error atributivo de la elevación suele hacerse mediante procesos comparativos de un conjunto de valores altimétricos de igual o mayor precisión, con sus correspondientes valores del modelo, estos valores se obtuvieron mediante el procedimiento de validación cruzada explicado en el numeral 3.3; haciendo uso de pruebas estadísticas convencionales, el error de elevación en un punto será igual a la diferencia existente entre el valor estimado o interpolado y su valor medido, las estadísticas del error “deben calcularse con un número suficiente de puntos homólogos convenientemente distribuidos” (Felicísimo, 1994). Para calcular la incertidumbre de los datos en un DEM se utilizan medidas descriptivas como:

**Error Medio Cuadrático:** es el descriptor tradicionalmente utilizado para cuantificar el error en un DEM, permite determinar la precisión en la estimación de la elevación en los puntos del muestreo.

**Media algebraica:** también conocida como error medio y corresponde a la media aritmética de los valores de error, e indica si la interpolación tiene una tendencia promedio a subestimar o sobreestimar en su predicción.

**Errores mínimo y máximo:** son indicadores de la magnitud que podrían llegar a tener los errores en casos extremos.



Tabla 6. Estadísticas del error de elevación para los diferentes interpoladores

re-muestreo	estadístico	idw	t2r	krig	tin	nn	spl
<b>180 x 180</b>	EMC	11.04	10.36	9.76	9.26	8.90	8.17
	media	0.03	-1.51	0.04	0.00	0.03	-0.01
	Min.	-67.07	-168.65	-58.51	-61.50	-57.50	-49.20
	Máx.	69.13	52.89	52.53	81.50	54.00	54.51
<b>150 x 150</b>	EMC	9.39	4.14	8.37	7.74	7.45	6.76
	media	0.02	-1.08	0.03	0.06	0.03	0.00
	Min.	-68.99	-154.97	-60.12	-55.00	-52.40	-48.16
	Máx.	51.17	43.58	42.27	55.40	43.32	41.41
<b>120 x 120</b>	EMC	7.56	6.98	6.92	6.15	5.88	5.26
	media	0.00	-0.94	0.01	0.02	0.00	-0.02
	Min.	-44.26	-148.15	-39.54	-39.67	-35.38	-38.65
	Máx.	42.86	34.85	38.87	38.33	37.63	39.19
<b>90 x 90</b>	EMC	5.66	5.29	5.35	4.56	4.36	3.82
	media	0.02	-0.67	0.03	0.00	0.02	0.01
	Min.	-36.70	-106.58	-34.35	-46.50	-32.89	-31.37
	Máx.	35.91	29.79	32.12	43.50	31.00	32.79
<b>60 x 60</b>	EMC	3.73	4.14	3.64	2.96	2.83	2.55
	media	0.01	-0.42	0.02	0.02	0.02	0.01
	Min.	-29.36	-138.67	-30.28	-30.50	-27.75	-32.71
	Máx.	25.46	29.98	26.45	29.00	29.25	38.10

Este trabajo considera el potencial de la validación cruzada como herramienta para estudiar el error de interpolación en los DEM , aunque esto es un tema tan complejo que muchos autores sugieren sea abordado mediante modelos estadísticos que se adecúen a las variables que se distribuyen espacialmente y que tratan de mostrar de alguna forma el patrón de distribución espacial de los errores de la elevación. Aún así este documento considera la naturaleza de los

errores derivados de los diferentes métodos de interpolación en los distintos niveles de muestreo.

## 4.2 El error y la curvatura

La curvatura se define como la tasa de cambio de la pendiente, por lo que depende de las derivadas de segundo grado de la altitud (Márquez, 2004). En otras palabras, la curvatura es la pendiente de la pendiente. Para calcular las derivadas de segundo grado no es posible usar la aproximación simple como en el caso de la pendiente, por lo cual se usa una superficie de ajuste de segundo grado:

$$Z = Ax^2y^2 + Bx^2y + Cxy^2 + Dx^2 + Ey^2 + Gx + Hy + I \quad (25)$$

La relación entre los coeficientes y los nueve valores circundantes para cada celda se enumeran de la siguiente forma:

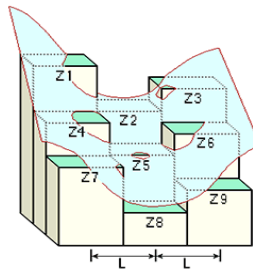


Figura 23 – Enumeración de celdas para cálculo de curvatura. Fuente: ArcGIS 10 Help.

$$A = [(Z_1 + Z_3 + Z_7 + Z_9)/4 - (Z_2 + Z_4 + Z_6 + Z_8)/2 + Z_5] / L^4$$

$$B = [(Z_1 + Z_3 - Z_7 - Z_9)/4 - (Z_2 - Z_8)/2] / L^3$$

$$C = [(-Z_1 + Z_3 - Z_7 + Z_9)/4 + (Z_4 - Z_6)/2] / L^3$$

$$D = [(Z_4 + Z_6)/2 - Z_5] / L^2$$

$$E = [(Z_2 + Z_8)/2 - Z_5] / L^2$$

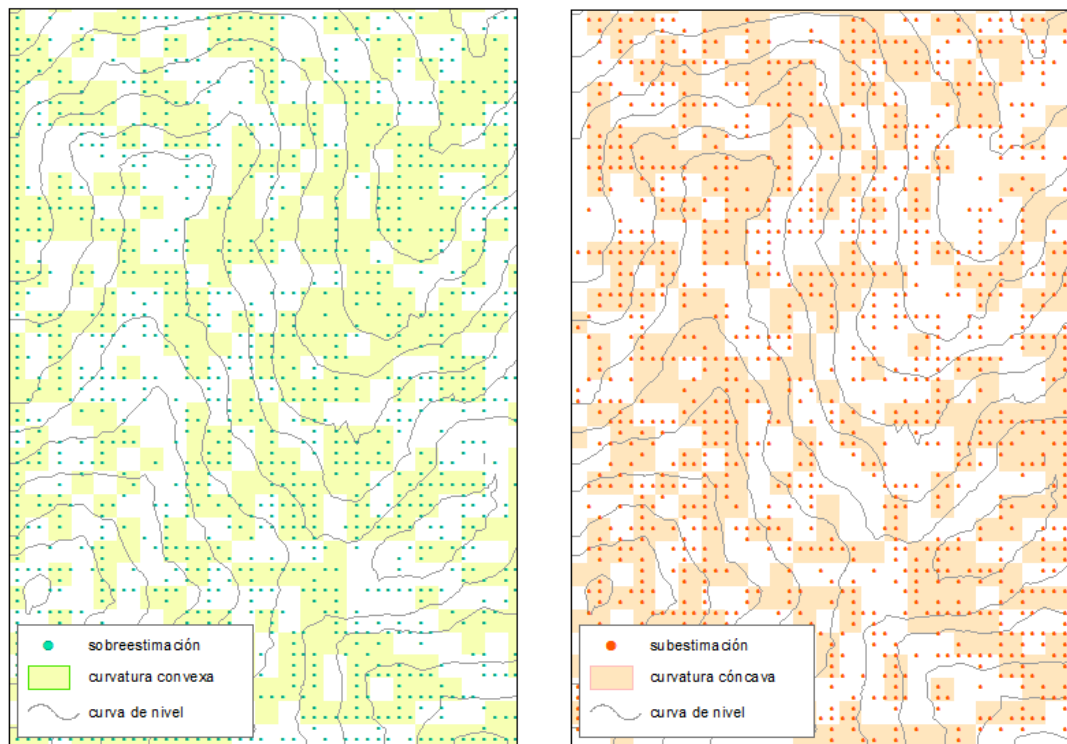
$$F = (-Z_1 + Z_3 + Z_7 + Z_9) / 4L^2$$

$$G = (-Z_4 + Z_6) / 2L$$

$$H = (Z_2 + Z_8) / 2L$$

$$I = Z_5$$

En todos los casos los algoritmos de interpolación presentan menor error a medida que la resolución del ráster se hace más fina. En términos generales la apariencia visual sigue el mismo patrón que el resultado del EMC, el patrón espacial del error es muy similar en todos los casos: la sobreestimación de elevación se produce cuando la geometría de la superficie es convexa y donde el cambio pendiente es muy marcado. Por lo tanto la gran mayoría de los errores se producen a lo largo de las líneas partidoras. La subestimación (o errores con signo negativo) se produce mayormente en áreas cóncavas y en especial hacia las zonas de vaguada.



(a)

(b)

Figura 24 – Patrón espacial de los errores en elevación. (a) Sobreestimación. (b) Subestimación.

Para soportar estadísticamente lo anterior se aplicó un *análisis de varianza* de un solo factor (Tabla 7) el cual descompone la varianza del *error medio cuadrático* en dos componentes: un componente *entre-grupos* y otro componente *dentro-de-grupos*. La razón-F, que en este caso es igual a 108959, es el cociente entre el estimado entre-grupos y el estimado dentro-de-grupos. Puesto que el valor-P de la prueba-F es menor que 0.05, se puede decir que sí existe una diferencia estadísticamente significativa entre la media del *EMC* relacionándolo con cada tipo de curvatura (cóncava o convexa), en un nivel de confiabilidad del 95.0%.

Tabla 7. Análisis de varianza para EMC por tipo de curvatura (cóncava o convexa)

Fuente	Suma de Cuadrados	GI	Cuadrado Medio	Razón-F	Valor-P
<b>Entre grupos</b>	643159.	1	643159.	<b>108959.50</b>	<b>0.0000</b>
<b>Intra grupos</b>	1.23279E6	208850	5.90274		
<b>Total (Corr.)</b>	1.87595E6	208851			

Ahora bien, ya que se conoce de la existencia de una diferencia considerada estadísticamente significativa al relacionar los errores con el tipo de curvatura, es necesario acotar la magnitud de dicho error, esto se logra con una *prueba de múltiples rangos*, donde se indica cuáles medias (y en qué cantidad) son significativamente diferentes de otras. En la tabla 8 se ha colocado un asterisco junto al contraste *cóncavo-convexo*, indicando que este par muestra diferencias significativas con un nivel del 95.0% de confianza. El método empleado para discriminar entre las medias es el procedimiento de diferencia mínima significativa (LSD) de Fisher. Con este método hay un riesgo del 5.0% al decir que cada par de medias es diferente, cuando la diferencia real es igual a 0.

Tabla 8. Pruebas de Múltiple Rangos para el EMC por el tipo de curvatura (cóncava – convexa)

tipo	Casos	Media	Grupos Homogéneos
cóncavo	102416	-1.77003	X
convexo	106436	1.74032	X

Contraste	Sig.	Diferencia	+/- Límites
cóncavo - convexo	*	-3.51035	0.0208433

\* indica una diferencia significativa.

Mediante el gráfico de medias (figura 25 a) se corrobora que los errores con signo negativo (subestimaciones) se relacionan con las formas cóncavas (-1,77 m) y que las sobreestimaciones (errores signo positivo) se encuentran mayormente en formas topográficas de tipo convexo.

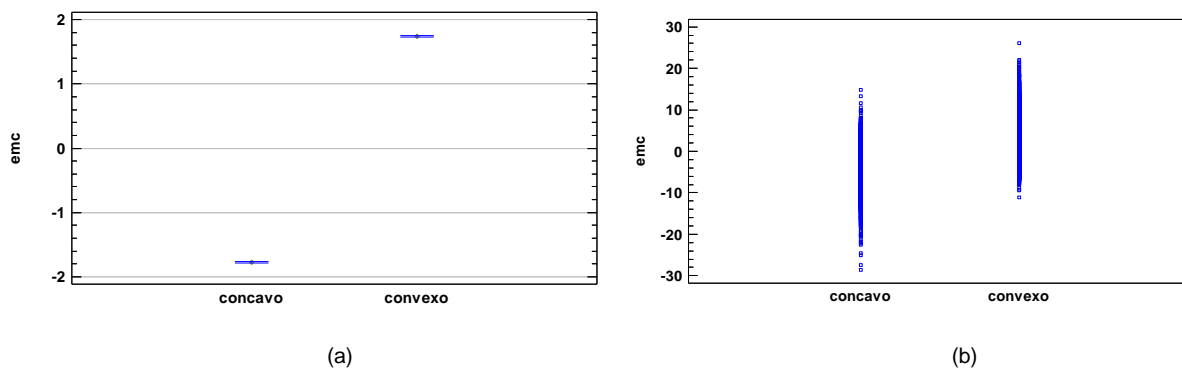


Figura 25 – (a) Gráfico de medias, 95% de Fisher LSD. EMC en función del tipo de curvatura. (b) Dispersión por código de nivel.

De la ilustración de la dispersión de datos (figura 23 b) se entiende que los valores del error para zonas cóncavas y convexas no son exclusivamente positivos o negativos para cada nivel, es por eso que en la figura 24 existen algunos puntos de sobreestimación en zonas cóncavas y subestimación en celdas convexas.

### 4.3 El error y la pendiente

La pendiente en un punto del terreno se define como el ángulo que existe entre el vector normal a la superficie en ese punto y la vertical (felicísimo, 1994) , esta ha sido valorada como uno de los más importantes descriptores del terreno y es ampliamente utilizada en la práctica normal de la topografía y cartografía (Balce, 1987; Li, 1993). Para usar la pendiente en términos de la descripción del terreno, existen dos parámetros relacionados con la estimación de sus valores: disponibilidad y variabilidad.

Por *disponibilidad* se supone que los valores de pendiente deben estar disponibles o ser estimados antes de que se haga el muestreo, a fin de ayudar en la determinación de sus intervalos. Si existe un DEM del área de trabajo, entonces los valores de pendiente para los puntos que lo conforman se pueden calcular y el promedio se puede usar como representativo (Zhu et al. 1999). Por otro lado, la pendiente podría estimarse desde un modelo estéreo, o por un par estereoscópico de fotografías aéreas con traslapo, o desde mapas con curvas de nivel.

Por *variabilidad* se entiende que los valores de pendiente podrían alterarse de un lugar a otro, así que la estimación de pendiente que es representativa para un área podría no adaptarse a otra. En este caso, los valores promedios podrían usarse tal como lo sugirió Ley (1986). Si la pendiente varía notablemente en un área, entonces ésta deberá dividirse en partes más pequeñas para su estimación.

Tabla 9. Rangos de Pendiente (Ley, 1986)

Tipo de terreno	Pendiente (°)	Pendiente (%)	Relieve (rango de alturas)
Plano	< 2	<4,4	< 80
Ondulado	2-6	4,4 – 13,3	80-300
Escarpado	6-25	13,3 – 55,6	300-600
Montañoso	> 25	>55,6	> 600

Fuente: Ley, R, 1986, Accuracy assessment of digital terrain models

El algoritmo usado para calcular este descriptor de la superficie en el proyecto es el usado por ArcGIS 10:

$$pend = \sqrt{s_1^2 + s_2^2} \quad (25)$$

$$s_1 = [(Z_3 + 2Z_6 + Z_9) - (Z_1 + 2Z_4 + Z_7)]/8d \quad (26)$$

$$s_2 = [(Z_1 + 2Z_2 + Z_3) - (Z_7 + 2Z_8 + Z_9)]/8d \quad (27)$$

$$pendiente \text{ en grados} = \tan^{-1}(\sqrt{s_1^2 + s_2^2}) \cdot 57,29578 \quad (28)$$

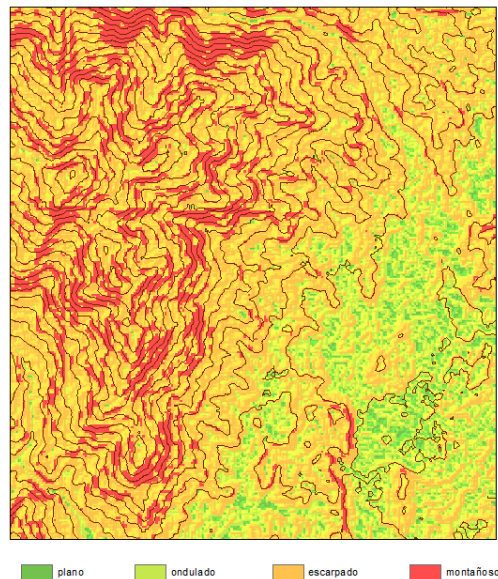


Figura 26 – Mapa de pendientes de acuerdo a la clasificación de Ley, R (1986).

Teniendo en cuenta la clasificación de la tabla 9 se elaboró el mapa de pendientes, en donde fácilmente se identifican las cuatro categorías: planas, onduladas, escarpadas y montañosas (figura 26). Para contrastar el error con la pendiente la figura 25 muestra la media del error absoluto para cada uno de los tipos de terreno, donde se indica un intervalo alrededor de cada media (los cuales son basados en el procedimiento de la diferencia mínima significativa LSD de Fisher), esta gráfica está construida de tal forma que si dos medias son iguales, sus intervalos se traslaparán un 95%. Cualquier par de intervalos que no se traslape horizontalmente (como es este caso) corresponden a pares de medias que tienen una diferencia estadísticamente significativa. Para todos los casos de interpolación evaluados es coincidente el comportamiento del incremento de *error absoluto* en la medida en que aumenta la *pendiente* del terreno. En el caso se la figura 26 se indican los resultados el para el mejor de los interpoladores a la menor resolución del re-muestreo: Spline.

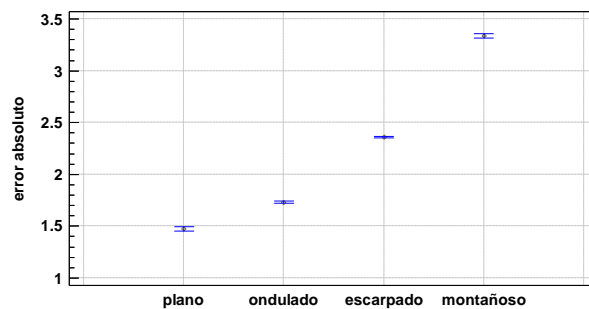


Figura 27 – Gráfico de medias, 95% de Fisher LSD. Error absoluto en función del tipo de terreno.

Se ha elaborado un mapa que contiene los errores absolutos clasificados en cuatro categorías donde se puede identificar que los errores menores a 1,28 m en su gran mayoría se encuentran en áreas clasificadas como *planas* y *onduladas*, así mismo los mayores errores (valores que fluctúan entre 1,28 y 3,48) se hallan en las zonas de pendientes *escarpadas* y *montañosas*.



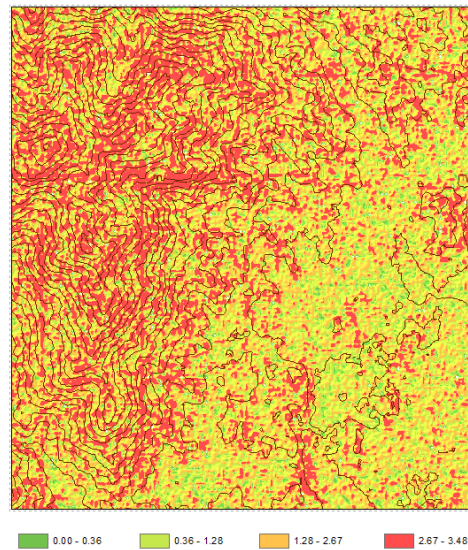


Figura 28 – Mapa de errores absolutos para el interpolador Spline a resolución de 30 m.

De acuerdo con la gráfica de dispersión de errores absolutos de la figura 29, se puede apreciar que los errores para superficies *planas* y *onduladas* se encuentran muy cercanos entre si, así mismo como sucede en las clasificaciones *escarpado* y *montañoso*, con lo cual se justifica que las dos clasificaciones de color verde en el mapa de errores de la figura 28 (rangos entre 0 – 0,36 y 0,36-1,28) se relacionen muy de cerca con las superficies que poseen pendientes hasta el 6%, es decir *planas* y *onduladas*, y las otras dos categorías (color naranja y rojo) se encuentren en las áreas de máxima pendiente, es decir *escarpadas* y *montañosas*.

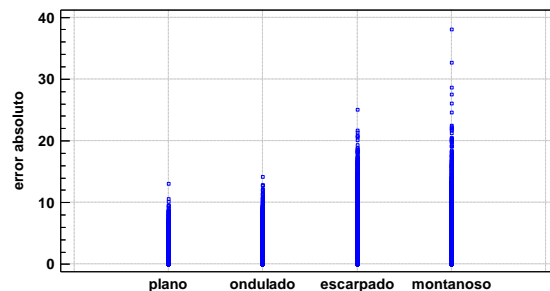


Figura 29 – Dispersión de errores absolutos por tipo de pendiente.

## 4.4 Características del error de elevación

Para intentar caracterizar el error, sus magnitudes se han modelado bajo una distribución Gaussiana  $N(0,1)$ , las pruebas de bondad del ajuste de normalidad *Kolmogorov-Smirnov* que se ejecutaron en todos los casos (todos los interpoladores a todos los niveles de re-muestreo), muestran que en su totalidad fueron significativamente diferentes de la distribución Gaussiana con valores de  $p$  para la prueba por debajo de 0,05 rechazando la idea de que estos errores provienen de una distribución normal con un 95% de confiabilidad (Tabla 12. Anexo D). Sin embargo, un estudio de las gráficas cuantil-cuantil (figura 31) y las estadísticas de distribución muestran resultados interesantes. La interpretación visual de los resultados, a primera vista sugiere una distribución Normal, pero si se mira en detalle se pueden apreciar colas muy largas y valores de curtosis muy altos, lo cual hace que los datos no se ajusten a una distribución Gaussiana sino una Logística.

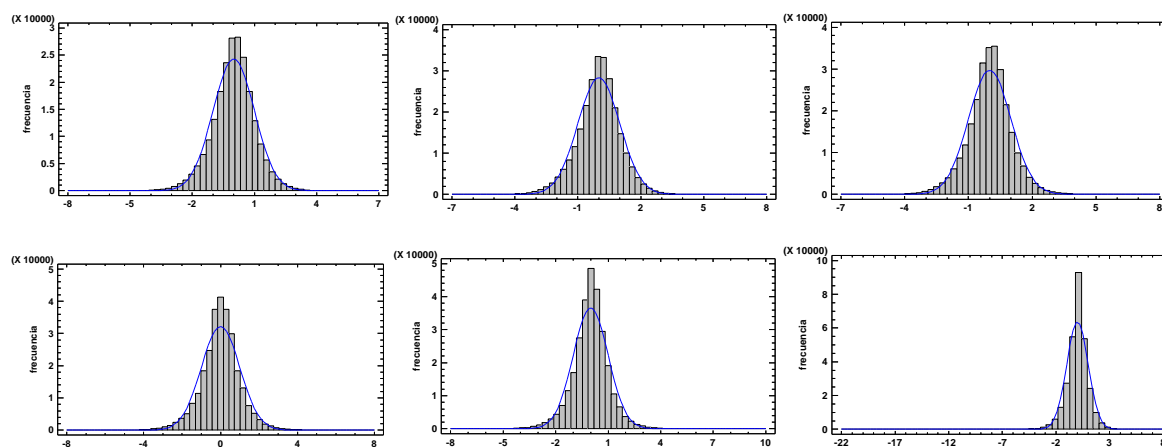


Figura 30 – Distribución de frecuencias del error del DEM. (a) para IDW con re-muestreo de 60x60. (b) para KRG re-muestreo de 90x90. (c) para NN re-muestreo 120x120. (d) para SPL re-muestreo 150x150. (e) para TIN re-muestreo 180x180. (f) distribución para T2R re-muestreo 90x90.

La diferencias en las distribuciones por los distintos métodos interpolados son muy leves comparadas con las diferencias producidas por los diferentes niveles de re-muestreo como se

indica en la figura 30, donde la única diferencia visible entre las gráficas son las alturas de los picos de las curvas.

Tabla 10. Valores de la curtosis estandarizada para la distribución de los errores de elevación

re-muestreo	idw	t2r	krig	tin	nn	spl
<b>180 x 180</b>	194.23	1916.50	137.90	230.73	172.84	180.78
<b>150 x 150</b>	179.14	1067.19	125.31	218.86	155.22	146.94
<b>120 x 120</b>	145.19	2759.80	103.90	181.92	131.31	141.16
<b>90 x 90</b>	115.16	1460.47	91.43	157.14	122.11	159.38
<b>60 x 60</b>	99.037	8180.63	92.24	159.01	122.12	258.64

los valores de curtosis (Tabla 10) indican que las distribuciones se vuelven mucho menos leptocúrtica a medida que aumenta el nivel de re-muestreo, es decir, se intenta normalizar en la medida que la densidad de los datos de disminuye.

En muchos casos el valor de datos re-muestreados a 180x180 no se encuentra muy lejos del valor 0 de una distribución Gaussiana. En la figura 31 se muestran dos gráficas cuantil-cuantil dejando ver la normalidad no extrema cuando el re-muestreo es bajo y más cercana a la normal cuando el re-muestreo es alto. Cuando se modela el error de elevación siempre se asume que éste se encuentra espacialmente autocorrelacionado, esta es una presunción razonable dado que el terreno tiende a presentar variaciones leves en distancias cortas de tal modo que los pixeles vecinos probablemente tengan valores similares de altura y dado el hecho que la interpolación en si misma tiene la probabilidad de producir valores similares para los puntos más cercanos.

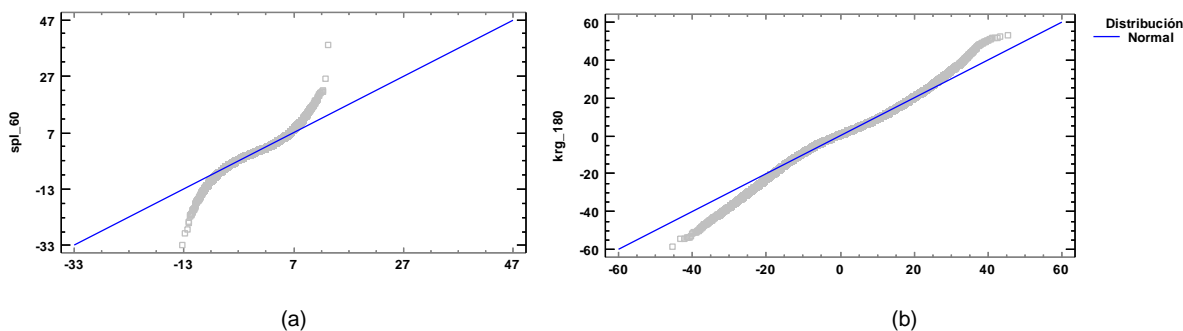


Figura 31 – Gráfica cuantil-cuantil para error producido usando (a) interpolador Spline con re-muestro de 60x60. (b) interpolador Kriging con re-muestro 180x180.

Con el fin de explorar la autocorrelación espacial del error se calculó el *índice de Moran* (*Moran's I*) a partir de los cuatro vecinos ortogonales de cada píxel (Tabla 11). Los valores para los niveles mayores del re-muestreo son muy altos, lo cual indica un fuerte grado de autocorrelación espacial en el patrón del error. Cuando el número de datos originales es pequeño habrá una fuerte correlación entre los valores de las elevaciones estimadas siempre que los valores vecinos estén siendo estimados a partir del mismo conjunto de puntos.

Tabla 11. Autocorrelación espacial del error en elevación usando *Moran's I*

re-muestreo	idw	t2r	krig	tin	nn	spl
<b>180 x 180</b>	0.97	0.97	0.97	0.96	0.97	0.96
<b>150 x 150</b>	0.95	0.95	0.96	0.97	0.96	0.95
<b>120 x 120</b>	0.91	0.91	0.95	0.92	0.94	0.94
<b>90 x 90</b>	0.72	0.74	0.69	0.71	0.65	0.67
<b>60 x 60</b>	0.60	0.39	0.48	0.39	0.41	0.37

Siempre un DEM tendrá gran cantidad de errores absolutos ya que es incapaz de representar totalmente la superficie del terreno en su verdadera forma, este modelo nunca podrá estar absolutamente definido pues para ello necesitaría contener tanta información como la geometría de la superficie que se pretende representar; algo inconcebible dado que se trata de

una superficie continua con infinitas variaciones; de cualquier forma, estos errores causados por la modelación también propenden a ser autocorrelacionados y la combinación de estos dos factores son conducentes a los valores del *índice de Moran* aproximándose a su máximo valor: uno (1). Sin embargo, en los niveles más bajos de re-muestreo, cuando el DEM está siendo generadas a partir de muchos puntos, la autocorrelación espacial en el error disminuye y se vuelve muy cambiante entre los métodos de interpolación. Esto sugiere que en algunos casos el simple cálculo de un solo error de autocorrelación realmente puede ser un indicador peligroso pues podría llegarse a juzgar que el error no presenta una autocorrelación significativa sobre la superficie estudiada.

Como caso especial dentro de este análisis el algoritmo T2R presenta datos atípicos con respecto a todos los otros algoritmos de interpolación. (figura 32).

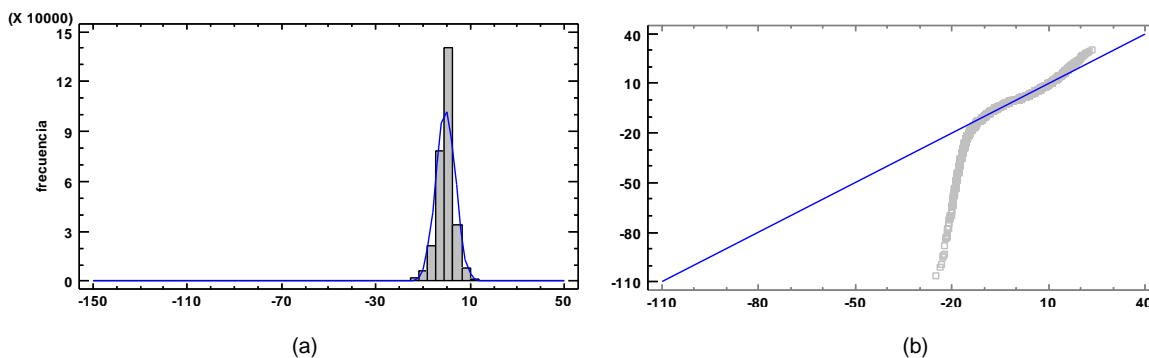


Figura 32 – Errores para algoritmo T2R (a) Distribución de frecuencias. (b) gráfica cuantil - cuantil.

Para todos niveles de re-muestreo este algoritmo en su distribución presenta una cola mucho más larga hacia el lado izquierdo que hacia el derecho, lo cual indica que las sub-estimaciones del error se alejan notoriamente de la media. Hacia el lado derecho, es decir los errores de sobre-estimación se comportan como datos muy cercanos a la *Distribución Normal*, esto se

puede ver claramente en el comportamiento de los puntos en la figura 32 (b). El comportamiento de las sub-estimaciones posiblemente se deba a que el algoritmo tiene como función principal la continuidad en la red hidrológica y es precisamente en las zonas de mayor depresión donde el error presenta sus máximas magnitudes.

## Capítulo 5: Conclusiones

Este trabajo ha introducido un método alternativo para estudiar algunas de las características del error en los DEM, mediante una técnica de re-muestreo a diferentes resoluciones para ser interpolado nuevamente a su resolución original. Cada momento surgen nuevas ideas por lo cual no se puede hablar de resultados definitivos, lo ideal es lograr acceder a más y mejor información, y mientras la ciencia avance y surjan nuevas técnicas, equipo, metodología y software se podrán obtener más y mejores resultados, sin embargo, asumiendo que los resultados de este trabajo puedan ser generalizados, se considera el EMC como una medida del control de calidad en un DEM, que sirve como una poderosa herramienta ya que puede utilizarse como rango de variación en técnicas de simulación. Se espera que el verdadero valor del error exista dentro del rango provisto por el EMC. Enmarcado en estos parámetros y usando la metodología propuesta en el capítulo 3, se han hallado conclusiones derivadas de este estudio entre las que se pueden destacar las siguientes:

- i.) Se aplicó el procedimiento de ANOVA (análisis de varianza) multifactorial para contrastar la hipótesis de que las medias de los errores de interpolación no dependen del *nivel de re-muestreo* ni del *método de interpolación*, dando como resultado a un nivel de confiabilidad del 95% que el EMC sí se encuentra altamente relacionado y dependiente de estas dos variables.
- ii.) Debido a la configuración del gráfico de interacción entre factores (figura 22), se puede cuantificar la magnitud del error en cada interpolador evaluado para cada uno de

- iii.) los niveles re-muestreo, pudiéndose identificar claramente a *Spline* como el algoritmo de menor error altimétrico y un patrón de variación más regular en todos los niveles de re-muestreo del DEM. Este algoritmo retiene características a pequeña escala que ninguno de los otros interpoladores evaluados pudo mantener.
  
- iv.) Gran cantidad de autores han modelado la propagación del error en los DEM mediante la aplicación del método de *Simulación de Monte Carlo* para evaluar la posible propagación del error, esos estudios usualmente asumen que el error posee una *Distribución Gaussiana* con un alto nivel de autocorrelación espacial. (Hunter y Goodchild, 1997; Fisher, 1992), los resultados de este trabajo sugieren que la situación podría ser un poco más compleja, en todos los métodos de interpolación y a todos los niveles de re-muestreo el error en elevación presenta una distribución que es estadísticamente diferentes de la *Gaussiana* con una agrupación muy fuerte de los valores alrededor de la media y con colas mucho más largas que la *Distribución Normal*. De todas formas cuando la fuente de los datos es relativamente escasa la diferencia con la *Distribución Normal* es poca, sugerido esto por las gráficas cuantil - cuantil que indican que la *Distribución Gaussiana* es una aproximación razonable en este caso.
  
- v.) Hay una relación similar entre la densidad de los datos iniciales y el nivel de autocorrelación espacial. Los altos valores de autocorrelación usualmente asumidos por los errores del modelado se encuentran en vigor cuando los datos de origen desde el cual se creó el DEM son escasos. En este caso la cercanía de los puntos están siendo interpolados a partir de los mismos puntos originales y por lo tanto están altamente



correlacionados. De todas formas, cuando hay más datos de entrada, los valores de autocorrelación espacial disminuyen y también se vuelven más dependiente del método de interpolación. Como era de esperarse métodos como *Spline* que ajustan una superficie a un número de puntos muestran una mayor autocorrelación espacial mayor que aquellos que están basados en la estimación a partir de la cercanía de sus puntos vecinos.

- vi.) El trabajo también confirma la importancia del fuerte vínculo que existe entre el error y la naturaleza del terreno que se manifiesta en sí mismo de dos maneras: primera: el patrón de los errores de elevación muestra un claro y enérgico enlace entre la geometría del terreno y su curvatura, mostrando subestimación para zonas cóncavas y sobre-estimación para zonas convexas, y segunda: las características generales del error del DEM muestran que son gravemente afectados por el grado de variabilidad de la pendiente del terreno que incide en la fiabilidad con la que se puede estimar el EMC, así como la de todos los niveles de error en los diferentes niveles del re-muestreo. El grado en que aumenta el error absoluto de la elevación es directamente proporcional al aumento de la pendiente del terreno, siendo esto un indicador útil en la detección de los niveles de error para aquellos productos derivados de la pendiente usados en la modelación digital de elevaciones.
- vii.) Este trabajo sugiere un nuevo enfoque al problema de la obtención de un muestreo de valores de error espacialmente distribuidos para poder estudiar algunas problemas de larga data del error en los DEM. Aunque sólo es aplicable a algunos de los métodos de generación de DEM la metodología ha revelado algunas características interesantes del

error producido por una variedad de métodos de interpolación aplicados al DEM. Se ha dado así alguna luz sobre la distribución de frecuencias y propiedades espaciales de la autocorrelación del error en el DEM.

## **5.1 Respuesta a la pregunta de investigación**

*¿Cómo la elección del algoritmo de interpolación incide sobre la representación de la superficie terrestre en un Modelo Digital de Elevación (DEM)?*

Analizando el estado del arte, no parece haber unanimidad a cerca de cuál es el método de interpolación que produzca los mejores resultados en la creación de un DEM, existen una cantidad de problemas conceptuales que deben ser considerados al utilizar los DEM como modelos de superficie. Por ejemplo, la superficie terrestre es de tipo continuo y un DEM es un conjunto de medidas discretas. Se dice que la fidelidad con que un DEM modela la verdadera superficie depende de: la rugosidad, pendiente, y curvatura propias del terreno, además del error relativo de la fuente de datos del DEM, pues siempre habrá detalles que se revelen en una escala más fina que las medidas en la resolución del DEM. (Por ello sería oportuno también generar modelos con resoluciones más altas).

La elección del algoritmo de interpolación en la construcción de un DEM debe estar fuertemente vinculada con el empleo final al cual éste vaya destinado, así mismo esta selección exige el conocimiento del tipo de estructura que almacena la información altimétrica, la resolución espacial (en el caso de estructuras ráster), la densidad de los puntos (en caso de las estructuras vectoriales), los descriptores topográficos del área que se quiere

modelar, la tolerancia en la magnitud del error final de representación, y finalmente la forma de medir este error.

De acuerdo con los resultados obtenidos en esta investigación es necesario medir la magnitud de los errores producidos por los algoritmos interpolantes sobre las distintas geometrías que se presentan en la naturaleza de la superficie terrestre, esto con el fin de poder tener una luz en la propagación de su comportamiento, ya que se comprueba que existe un alto índice autocorrelación espacial del error, así como vínculos muy fuertes entre éste (el error) y: la resolución espacial del DEM, la pendiente y la curvatura de la superficie a través del procedimiento de modelación.

Respecto de los métodos de interpolación analizados, el trabajo concluye que:

- *Spline* alcanza el menor EMC y el patón más regular en todos los grados de variabilidad de la pendiente, buscando el mejor ajuste de los puntos muestrales a una superficie continua.
- *Natural Neighbor* a pesar de operar en forma muy similar a IDW, utiliza un ajuste local que reduce el efecto de círculos concéntricos, pues tiene un parámetro de suavizamiento, además de tener en cuenta la anisotropía otorgando diferentes pesos a lo largo de los ejes de búsqueda.
- *Triangular Irregular Network* produce una superficie no suavizada lo que causa pendientes discontinuas en los bordes de la triangulación, este algoritmo solo tiene en cuenta la distribución espacial de los puntos originales y no la forma de la superficie generada y da cabida a la creación de triángulos que alteran la geometría de la superficie esperada, esto debido a que es un interpolador lineal, y las formas de la superficie terrestre por lo general no corresponden a modelos lineales.

- *Kriging*: es un método de muy buenos resultados cuando hay suficientes datos para estimar el semivariograma, debido a que tratan el ‘ruido’ como parte de la señal. Por la robustez del algoritmo su procesamiento es muy lento especialmente cuando el volumen de datos es grande. La precisión de la interpolación por este método aumenta con la densidad de los puntos muestrales, según un modelo potencialmente decreciente que es igual para todos los interpoladores.
- *Topo to Raster* se caracteriza por presentar un excesivo suavizamiento de la superficie así como la omisión de gran cantidad de detalles de la misma, lo cual hace que pierda precisión planimétrica en la disposición de las formas que no corresponden a redes de drenaje.
- *Inverse Distance Weigth* es un método que procesa rápidamente los datos (esto puede ser algo ventajoso cuando el volumen es muy grande), aunque al modelar la superficie presenta la tendencia de generar patrones concéntricos alrededor de los puntos originales, (lo cual lo convierte en el interpolador con mayor EMC del estudio) esto sucede por el peso que se le da a la variación particular del valor de un punto del muestreo sobre los que están a su alrededor.

Para redondear la idea de la afectación que se da la selección del algoritmo de interpolación en la representación de una superficie, es un hecho que no existe un método de interpolación ideal para todos los fines, pues cada uno fue creado para la estimar información que caracterice el fenómeno de la mejor forma limitado a las asunciones inherentes en: el diseño del algoritmo, el tipo de superficie, su estructura de representación, y la tolerancia de los errores esperados.

## 5.2 Recomendaciones para futuros estudios

La elección de la fuente de datos es crítica para la calidad del DEM resultante, los datos deben consistir en observaciones sobre la elevación del terreno y siempre que sea posible, en información adicional sobre fenómenos que influyan significativamente en la forma de la superficie del terreno (Márquez, 2004), sería ideal aplicar esta metodología a una fuente de datos que garantice un error posicional (en términos de elevación) inferior que el otorgado por los modelos ASTER GDEM ( $\pm 20$  m), pues esto hace que las correcciones radiométricas (bandeamientos), geométricas (deficiencias en la georreferenciación) y atmosféricas (distorsiones ocasionadas por los valores de radiancia) tengan magnitudes muy grandes que hacen que el valor verdadero de la elevación fluctúe dentro de una tolerancia muy grande, esto se puede mejorar con DEM generado a partir de interferometría sobre pares de imágenes radar tomadas de satélite.

Los resultados de los errores producidos por los distintos algoritmos de interpolación emitidos en este proyecto son limitados a los valores por defecto que sugiere el software ArcGIS 10 para todos los casos, sería de mucho interés evaluar el comportamiento de los algoritmos aquí evaluados de forma independiente, pues al interior de cada uno de ellos existen parámetros de valores alternativos (como los semivariogramas en los casos de KRG e IDW) que son susceptibles de ajuste en función de los datos de entrada, produciendo resultados cambiantes en los datos de salida.

El trabajo también sugiere un fuerte vínculo entre el error y la naturaleza de la superficie del terreno, lo cual hace que se quiera seguir indagando en este campo, buscando la relación del

error con otros descriptores de la superficie como la orientación, la rugosidad y la dimensión fractal del terreno; ofreciendo una mejor manera de estudiar el patrón de comportamiento de los errores en los DEM. Por último sería muy interesante ver esta metodología aplicada a una gama más amplia de DEM y métodos de interpolación.

## Bibliografía

BALCE, A.E. (1987). Determination of optimum sampling interval in grid digital Elevation models (DEM) data acquisition. *Photogrammetric Engineering and Remote Sensing*, 53: 323-330.

BAKOSE, G.; NUMAN, M. (1990). Photostratigraphy based on drainage morphometry and lithologic character of three formations in northern Iraq. *ITC-Journal 1990 – 1*: 51-57.

BOSQUE SENDRA, JOAQUÍN (1992). *Sistemas de información Geográfica*. Madrid, Ediciones Rialp S.A.

BURROUGHS, P.A. Y MCDONNELL, A.(1998). *Principles of geographical information systems*. Somerset, Oxford University Press.

BURROUGHS, P.A. (1986). *Principles of Geographical Information System for land resources assessment*. Monographs on Soil and Resources Survey. No 12. Oxford University Press.

CARTER, J.R. (1988). Digital representation of topographic surfaces. *Photogrammetric Engineering & Remote Sensing*, vol 54, nº 11, pp 77-80.

COLLINS, F. AND BOLSTAD, P.V. (1996). Comparison of Spatial Interpolation Techniques in Temperature Estimation. In *Proceedings, Third International Conference/Workshop on integrating GIS and Environmental Modeling*, Santa Fe, NM, January 21-26, 1996. Santa Barbara, CA: National Center for Geographic Information and Analysis. CD

EL-SHEIMY, N; VALEO, C; HABIB, A. (2005). *Digital Terrain Modeling: Acquisition, Manipulation And Applications*. Calgary: Artech House.

FELICÍSMO PÉREZ, A.M. (1994). Modelos digitales de terreno. Oviedo, Pentalfa Ediciones.

Fisher, P.F.(1992).1<sup>st</sup> experiments in Viewshed uncertainty—simulating fuzzy Viewsheds. Photogrammetric Engineering and Remote Sensing 58(3), 345–352.

FISHER, P.F. (1996). Extending the applicability of viewsheds in landscape planning. Phothogrammetry Engineering and Remote Sensing 62 (11): 1279-1302.

GAO, J. (1997). Resolution and accuracy of terrain representation by Grid DEMs at a Micro-scale. International Journal of Geographical Information Science, 11(2), pp.192-212.

GARZON, J; JIMENEZ, G; VILA, J. (2012). Introducción a los modelos digitales de elevación en topografía. Saabrucken, Alemania: Editorial Académica Española.

GOLDEN SOFTWARE. (2011). Surfer 10. User's Guide. Contouring and 3D Surface Mapping for Scientist and Engineers. Golden Software, Inc.Colorado, USA. 619p. [www.goldensoftware.com](http://www.goldensoftware.com)

HUNTER, G.J, GOODCHILD, M.F. (1997). Modeling the uncertainty of slope and aspect estimates derived from spatial databases. Geographical Analysis 29 (1), 35–49. Huss, R.E., Pumar, M.A., 1997. Effect of database errors on intervisibility estimation. Photogrammetric Engineering and Remote Sensing 63 (4), 415–424.

IMHOF, EDUARD (1982). Cartographic relief representation, (ed Steward, H.J), Berlin: Walter de Gruyter.

KUMLER M. P., (1994). An intensive comparison of triangulated irregular networks (TINs) and digital elevation models. Cartographica. (1), pp. 1-9.



- LEY, R. (1986). Accuracy assessment of digital terrain models. *Auto-Carto London*, 1: 455-464.
- LI, ZHILIN (1993). Theoretical models of the accuracy of digital terrain models: an evolution and some observations. *Photogrammetric record*, 14(82): 651-660
- LI, ZHILIN; ZHU, QING. (2000). *The Digital Elevation Model*, Florida: CRC Press.
- LI, ZHILIN; ZHU, QING; GOLD CHRISTOPHER.(2005). *Digital Terrain Modeling: Principles and Methodology*. Boca Raton, Florida: CRC Press.
- MARQUEZ, JOAQUIN.(2004). Modelos digitales de elevaciones, métodos de interpolación y parámetros morfométricos del relieve. Departamento de geografía física y análisis geográfico regional, Universidad de Sevilla. Sevilla, España.
- MCCULLAGH, M.J. (1990). Digital terrain modeling and visualization. In Petri, G and Kennie, T.J.M (Ed.). *Terrain modeling in surveying and civil engineering*. Whittles Publishing, pp.128-151.
- MCCULLAGH, M.J. (1988). Terrain and surface modeling system: theory and practice. *Photogrammetric record* 12(72): 747-779.
- REES W. G. (2000). The accuracy of digital elevation interpolated to higher resolution. *International Journal of Remote Sensing*, (1), pp. 7-20.
- PEREZ , A; FRANCOIS J.(2009). Evaluación de los errores de modelos digitales de elevación obtenidos por cuatro métodos de interpolación. *investigaciones Geográficas: Boletín del Instituto de Geografía UNAM*, (69), pp. 53-67.

- PECKHAM, ROBERT; JORDAN GYOZO. (Ed.). (2007). Digital Terrain Modelling: Development and Applications in a Policy Support Environment. Berlin: Springer.
- PODOBNIKAR, T., (2005). Production of integrated digital terrain model from multiple datasets of different quality. *International Journal of Geographical Information Science* 19 (1), pp. 69–89.
- SHARIF, M., MAKAROVIC, B., (1989). Optimizing progressive and composite sampling for DTMs. *ITC Journal* 1989 (2), pp. 104–111.
- WEIBEL, R; HELLER, M.( 1991). Digital terrain modeling. In Maguire, D.J.; Goodchild, M.F. y Rhind, D.W. (Eds.). *Geographical Information Systems*. Longman, Scientific & Technical.(1), pp.269-297.
- WILSON, JOHN; GALLANT, JOHN. (Ed.). (2000). *Terrain Analysis: Principles and Applications*. New York: John Wiley & Sons.
- WISE, S. (2011). Cross validation as a means of investigating DEM interpolation error. *Computers and Geosciences*, 37(8), 978-991. Recuperado de <http://www.sciencedirect.com/science/article/pii/S0098300411000252>.
- ZHU, G., WANG, J.; JIANG, W. (1999). *Digital map analysis*. Publishing House of Wuhan Technical University of Surveying and Mapping, Wuhan, China.

## Anexo A. Esquema de geoprocesamiento

Este gráfico indica el geoprocesamiento programado a través de la herramienta *model builder*.

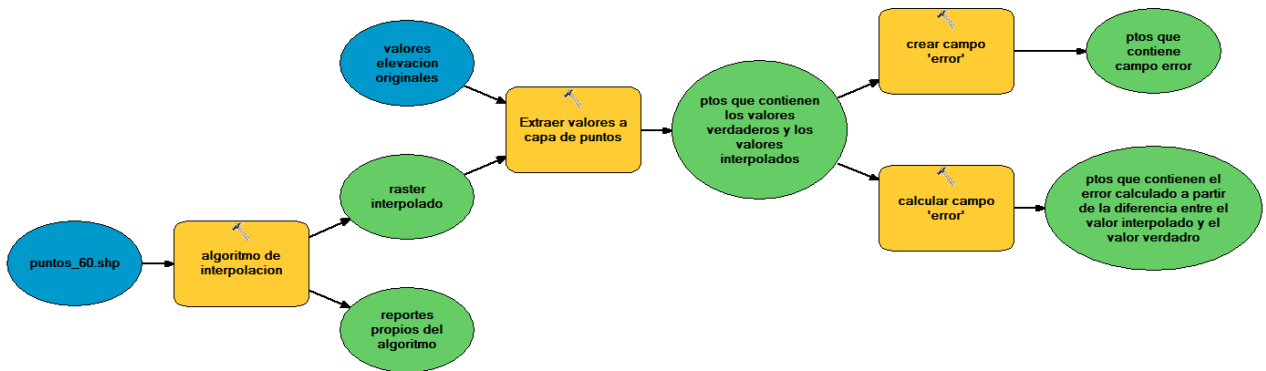


Figura 32– esquema de geoprocesamiento para interpolación y extracción de puntos.

Este mismo modelo fue usado para los seis algoritmos en evaluación en los 5 distintos niveles de re-muestreo.

## Anexo B. Raster de la zona de zona de estudio a distintos niveles de re-muestreo e interpolados por los algoritmos evaluados

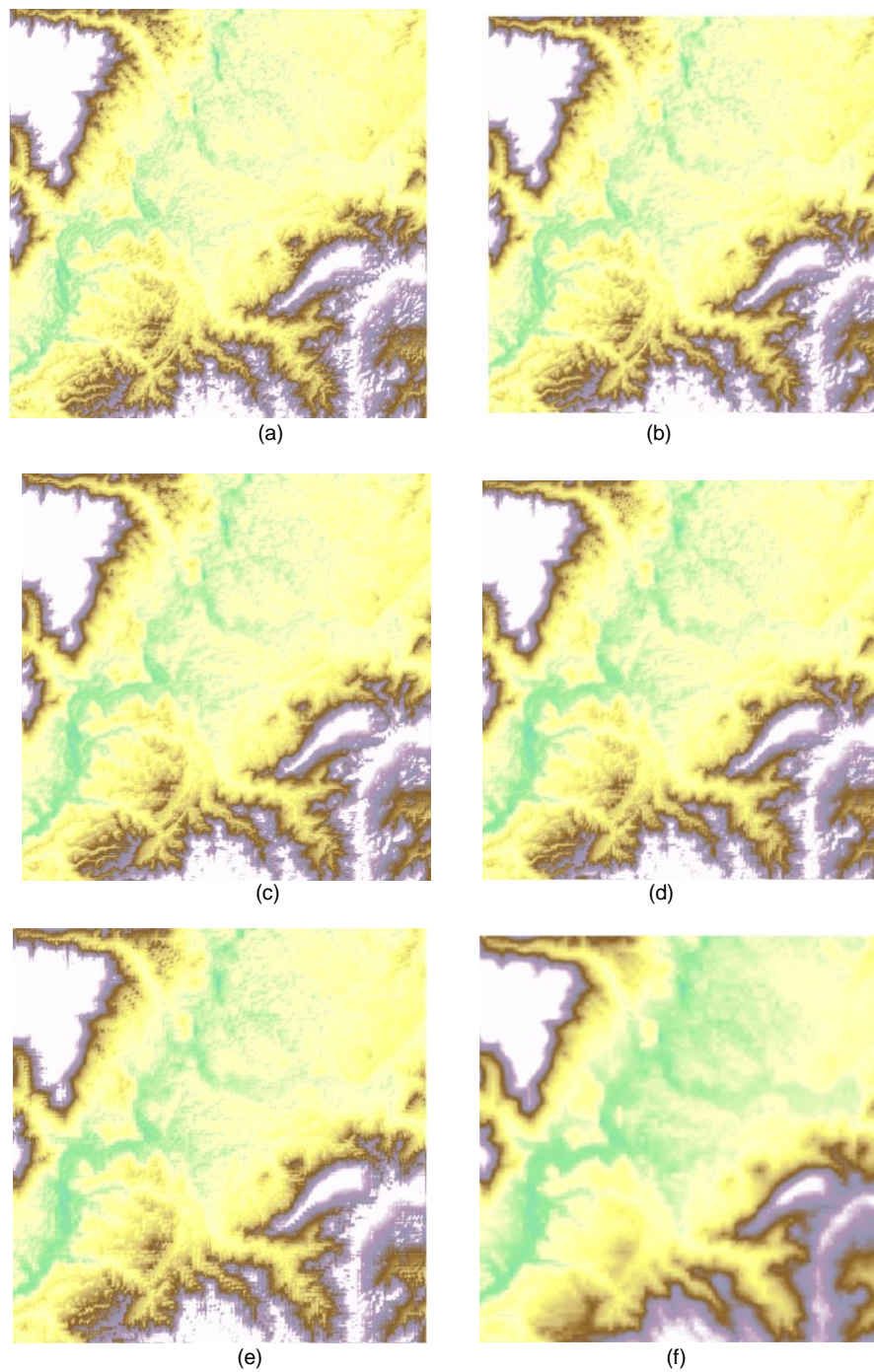


Figura 33– IDW: método de ponderación inverso a la distancia (Inverse Distance Weighted). (a) 30x30 (b) 60x60 (c) 90x90 (d) 120x120 (e) 150 x150 (f) 180x180

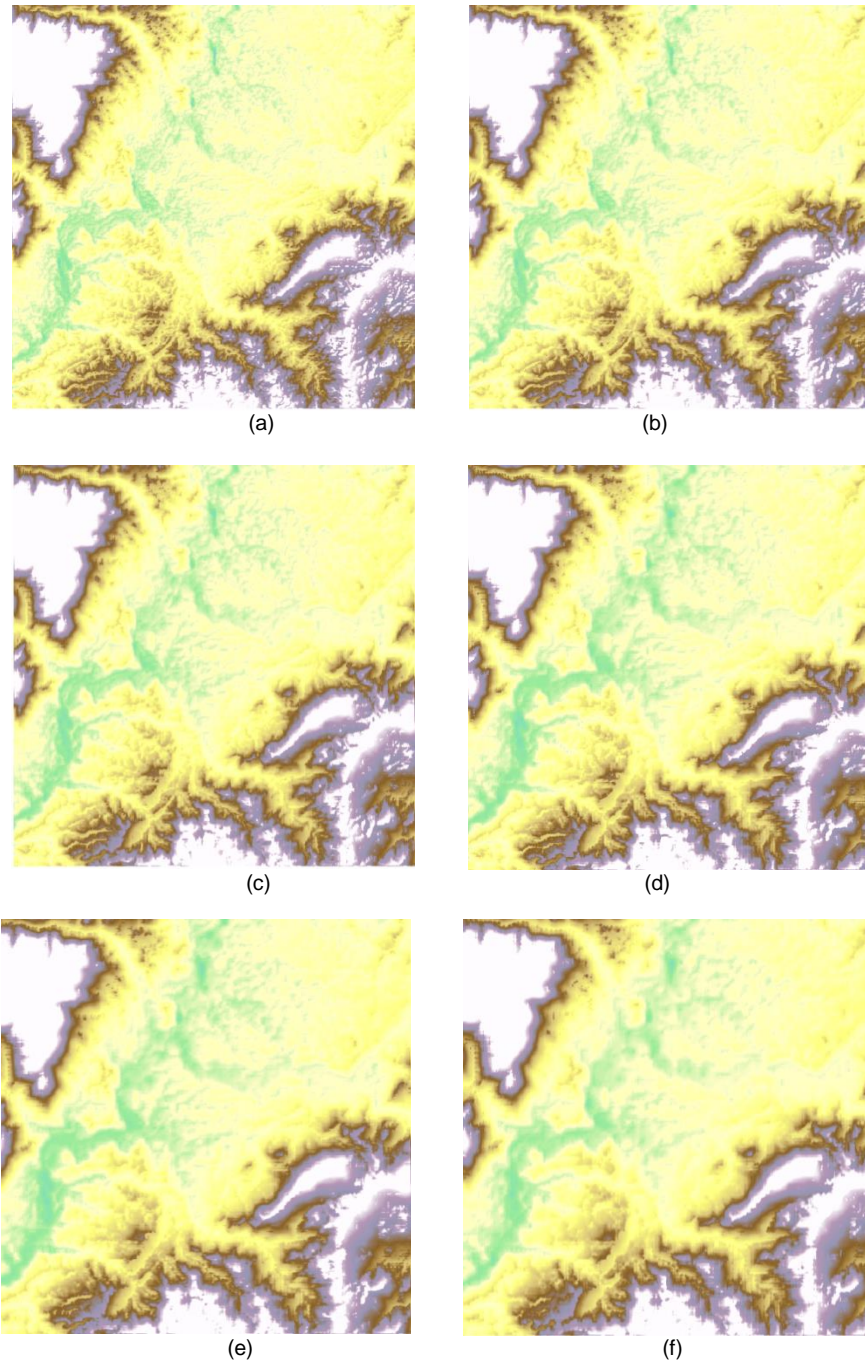


Figura 34– KRG: método de interpolación Kriging. (a) 30x30 (b) 60x60 (c) 90x90 (d) 120x120 (e) 150 x150 (f) 180x180

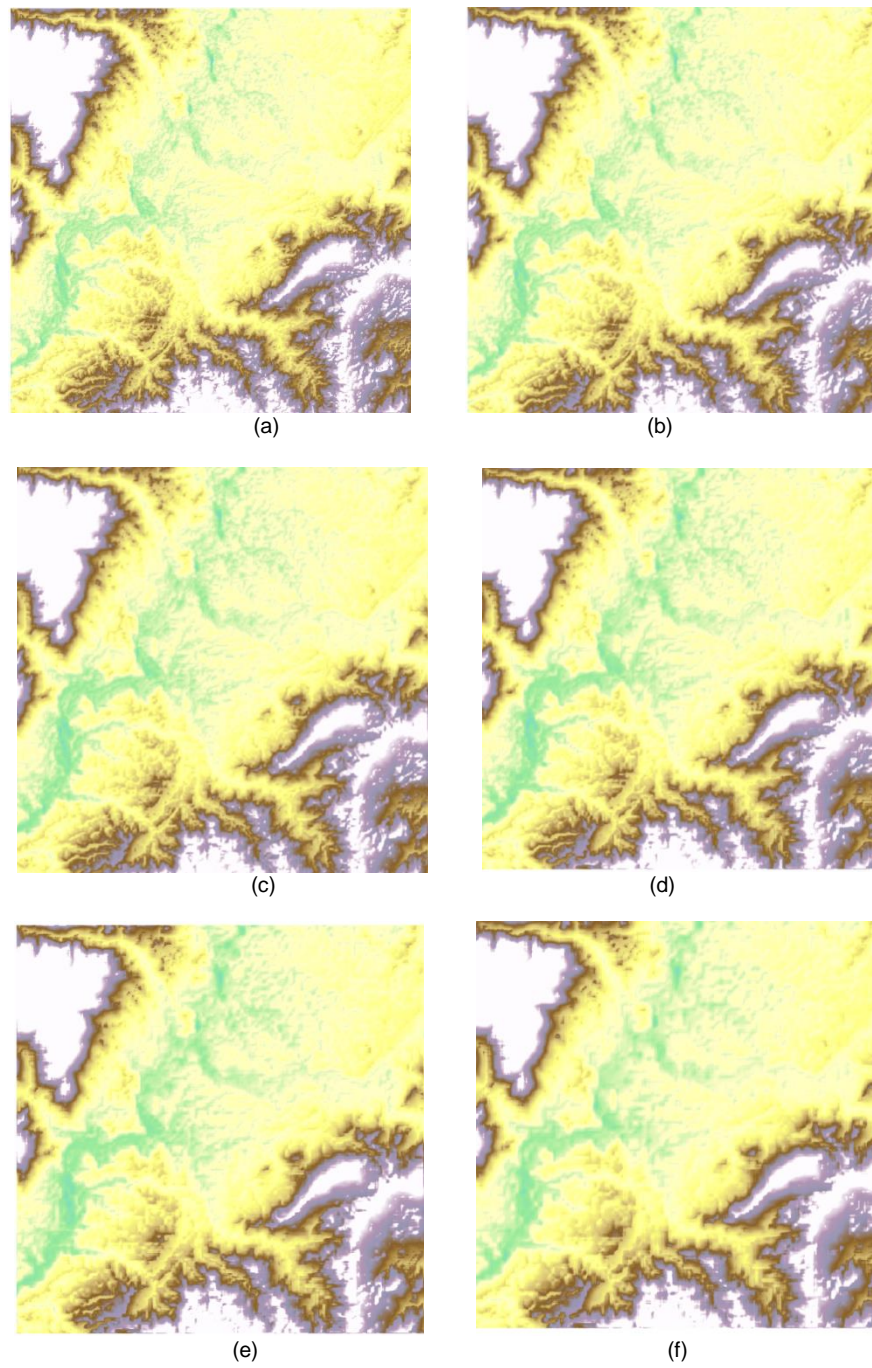


Figura 35– N.N: método de interpolación Natural Neighbor. (a) 30x30 (b) 60x60 (c) 90x90 (d) 120x120 (e) 150 x150 (f) 180x180

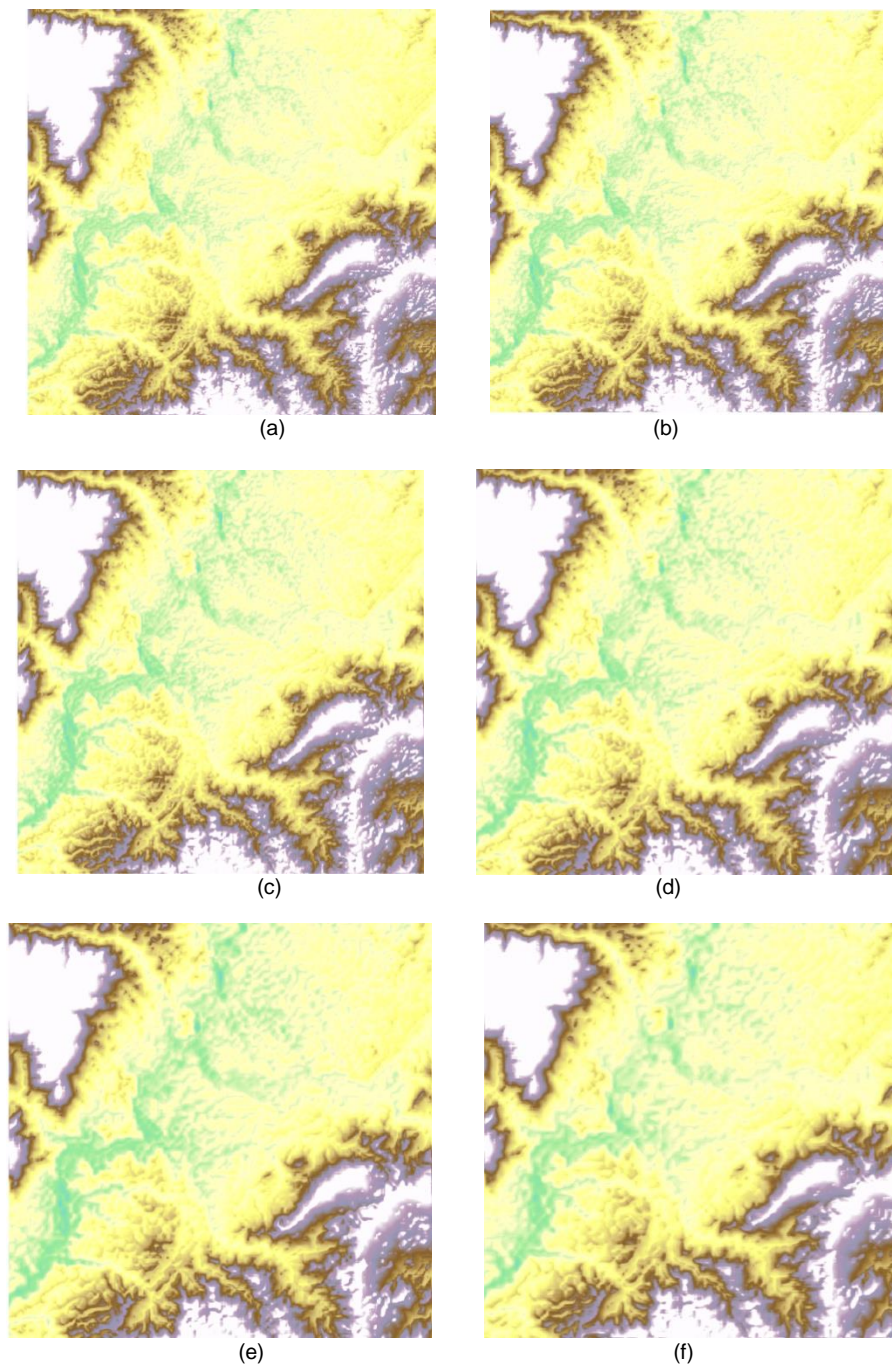


Figura 36– SPL: método de interpolación Spline. (a) 30x30 (b) 60x60 (c) 90x90 (d) 120x120 (e) 150 x150 (f) 180x180

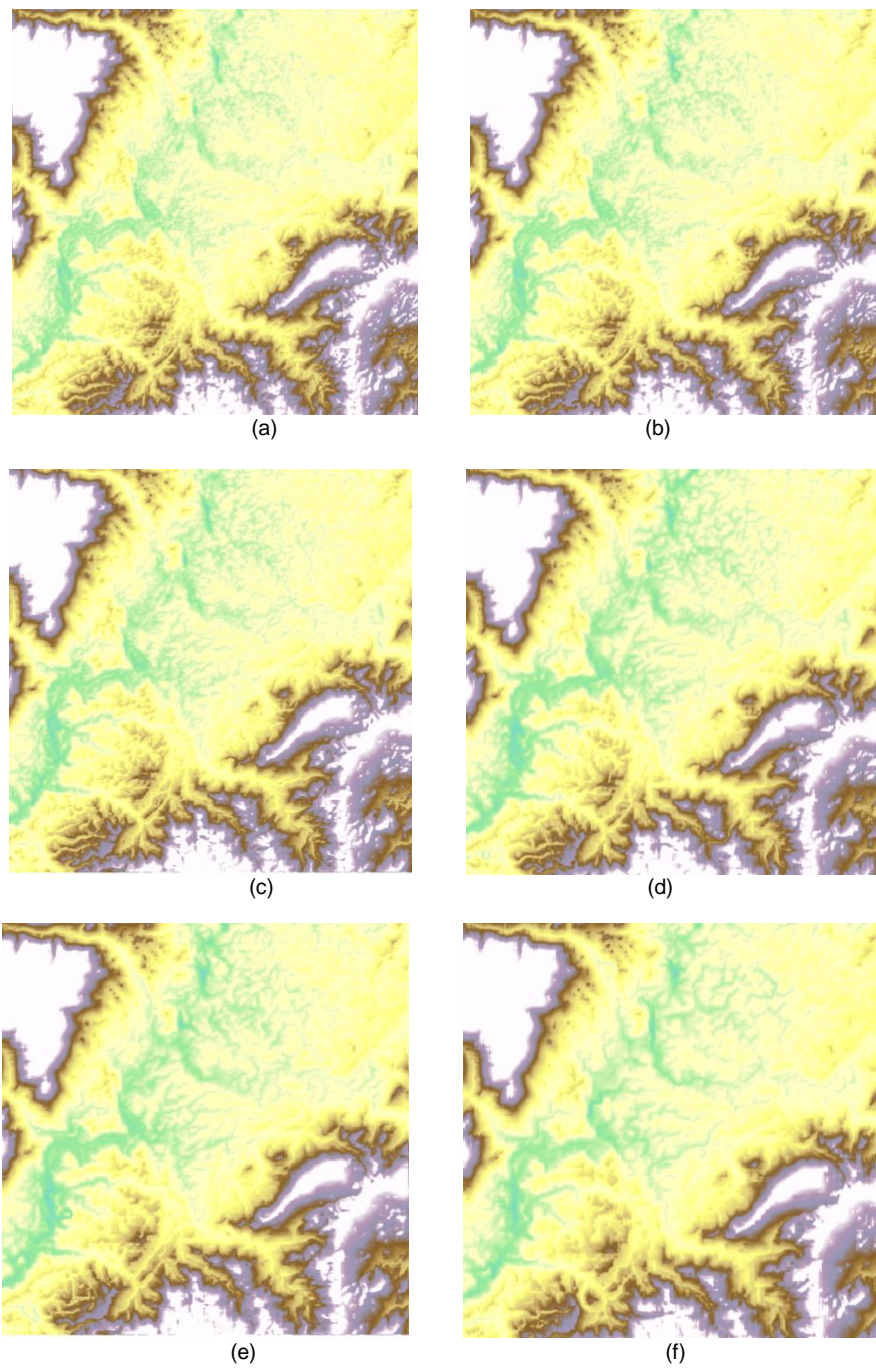


Figura 37– T2R: método de interpolación Topo to Raster. (a) 30x30 (b) 60x60 (c) 90x90 (d) 120x120 (e) 150 x150 (f) 180x180



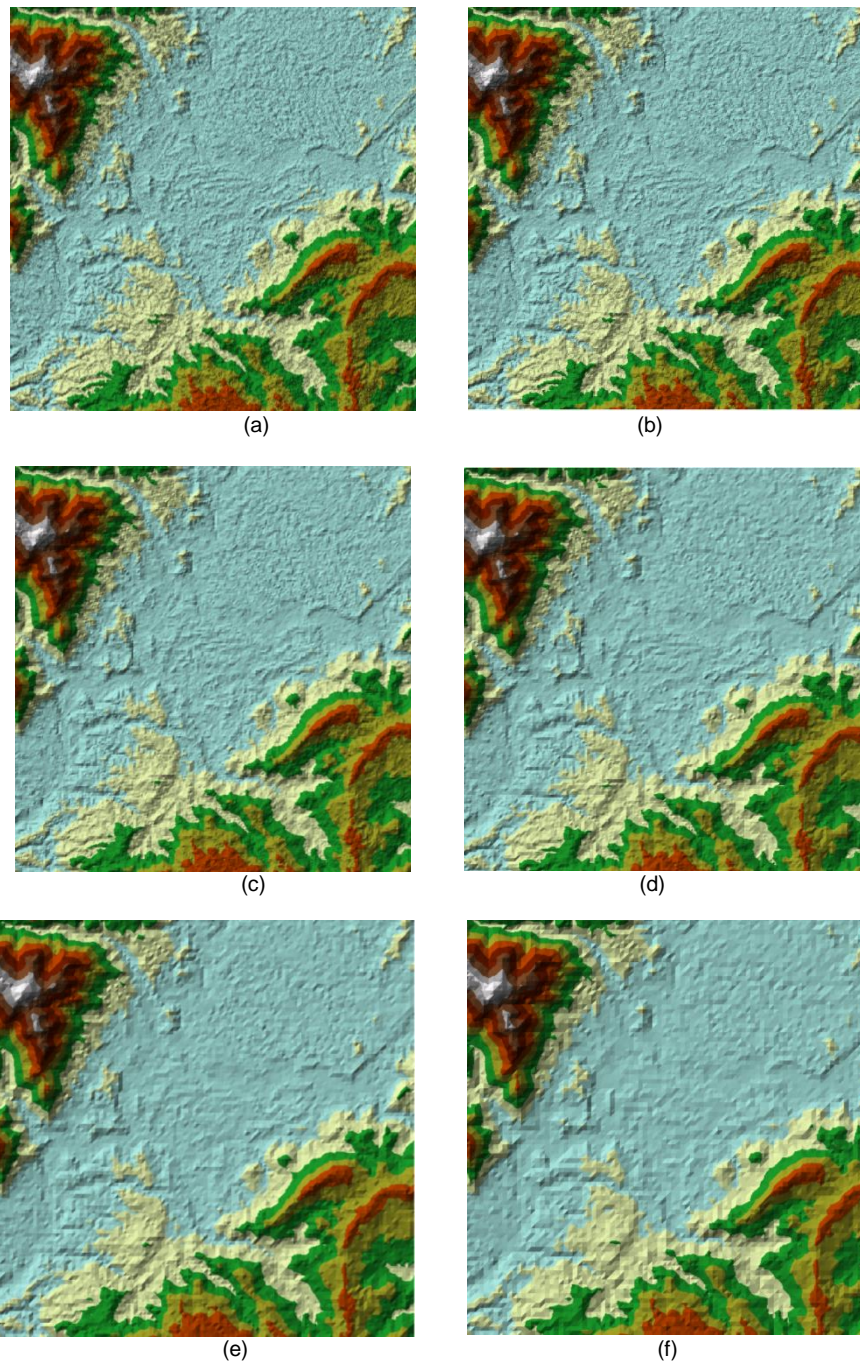


Figura 38– TIN: método de interpolación Triangular Irregular Network. (a) 30x30 (b) 60x60 (c) 90x90 (d) 120x120 (e) 150x150 (f) 180x180

## Anexo C. Modelación de error en los distintos algoritmos evaluados

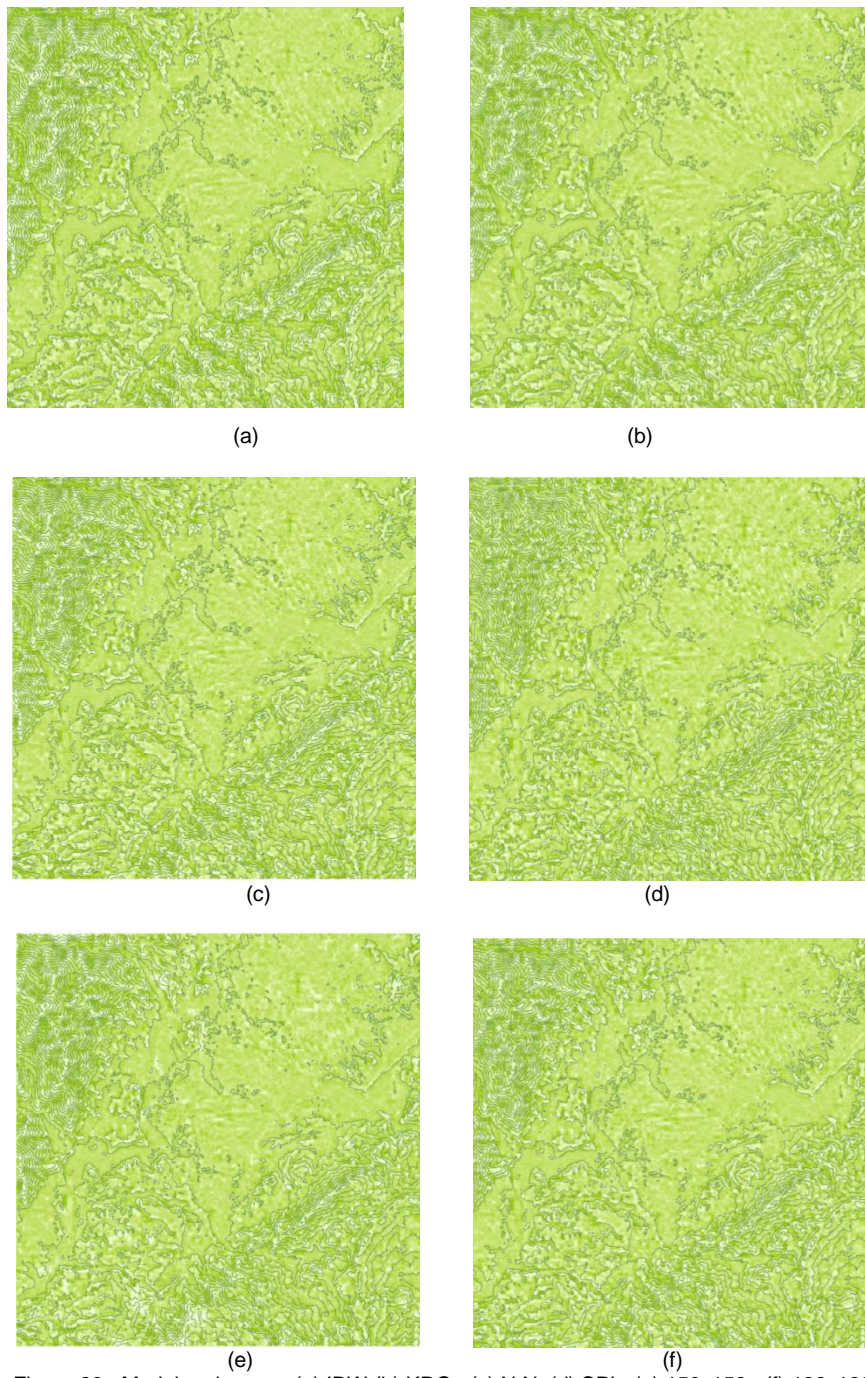


Figura 39– Modelos de error (a) IDW (b) KRG (c) N.N (d) SPL (e) 150x150 (f) 180x180

## Anexo D. Pruebas de Bondad de ajuste de Kolmogorov-Smirnov

Tabla 12. Estadístico de pruebas de bondad de ajuste a todos los interpoladores en los diferentes niveles de re-muestreo

re-muestreo	estadístico	idw	t2r	krig	tin	nn	spl
<b>180 x 180</b>	DMAS	0.044783	0.0528014	0.0327893	0.0468692	0.0371132	0.035852
	DMENOS	0.059606	0.0786322	0.0484206	0.055252	0.0500015	0.0402468
	DN	0.059606	0.0786322	0.0484206	0.055252	0.0500015	0.0402468
	Valor-P	0.0	0.0	0.0	0.0	0.0	0.0
<b>150 x 150</b>	DMAS	0.0419258	0.0436014	0.0318104	0.0451104	0.03622	0.0337781
	DMENOS	0.0558202	0.0652176	0.0460806	0.0507581	0.0460193	0.0351512
	DN	0.0558202	0.0652176	0.0460806	0.0507581	0.0460193	0.0351512
	Valor-P	0.0	0.0	0.0	0.0	0.0	0.0
<b>120 x 120</b>	DMAS	0.0355864	0.0523575	0.0266368	0.0429243	0.0308743	0.02903
	DMENOS	0.0495309	0.0743123	0.0416444	0.0487984	0.0411557	0.0310799
	DN	0.0495309	0.0743123	0.0416444	0.0487984	0.0411557	0.0310799
	Valor-P	0.0	0.0	0.0	0.0	0.0	0.0
<b>90 x 90</b>	DMAS	0.0284099	0.0508376	0.023892	0.0414203	0.0282579	0.028983
	DMENOS	0.0412397	0.0705956	0.0364723	0.0449353	0.0349871	0.0290864
	DN	0.0412397	0.0705956	0.0364723	0.0449353	0.0349871	0.0290864
	Valor-P	0.0	0.0	0.0	0.0	0.0	0.0
<b>60 x 60</b>	DMAS	0.0250595	0.0780522	0.0233607	0.0527844	0.0282579	0.0374025
	DMENOS	0.0344029	0.0979281	0.0320139	0.0630603	0.0349871	0.0380912
	DN	0.0344029	0.0979281	0.0320139	0.0630603	0.0349871	0.0380912
	Valor-P	0.0	0.0	0.0	0.0	0.0	0.0

## Anexo E. Distribución de frecuencias de errores y gráficas cuantil - cuantil

### - cuantil

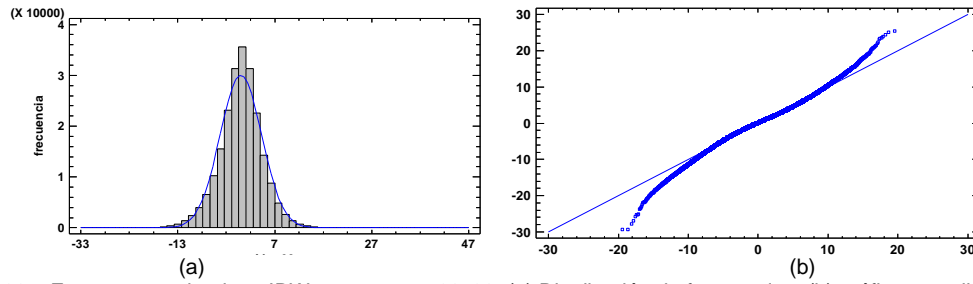


Figura 33 – Errores para algoritmo IDW re-muestreo 60x60 (a) Distribución de frecuencias. (b) gráfica cuantil - cuantil.

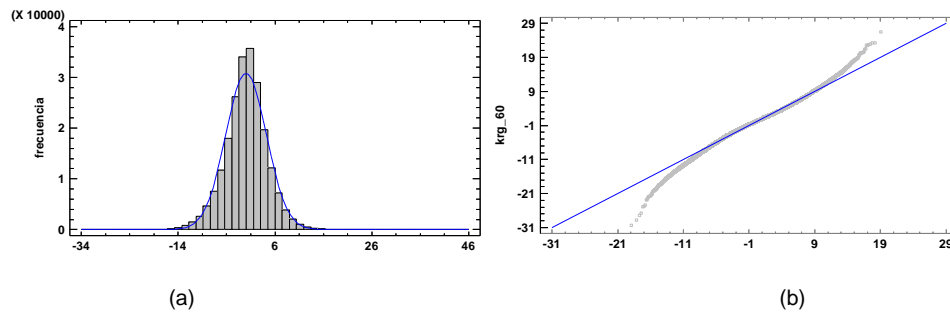


Figura 341 – Errores para algoritmo KRG re-muestreo 60x60 (a) Distribución de frecuencias. (b) gráfica cuantil - cuantil.

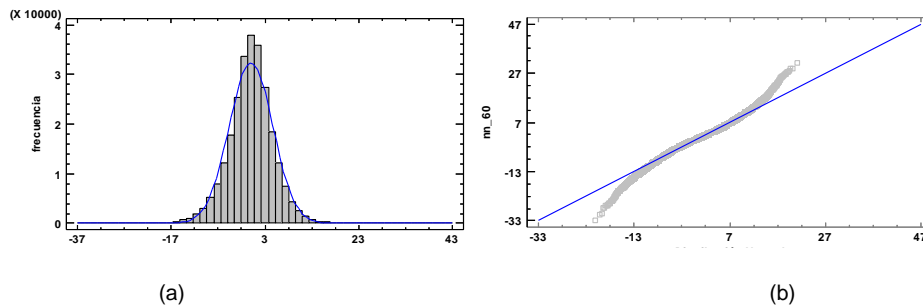


Figura 352 – Errores para algoritmo N.N re-muestreo 60x60 (a) Distribución de frecuencias. (b) gráfica cuantil - cuantil.

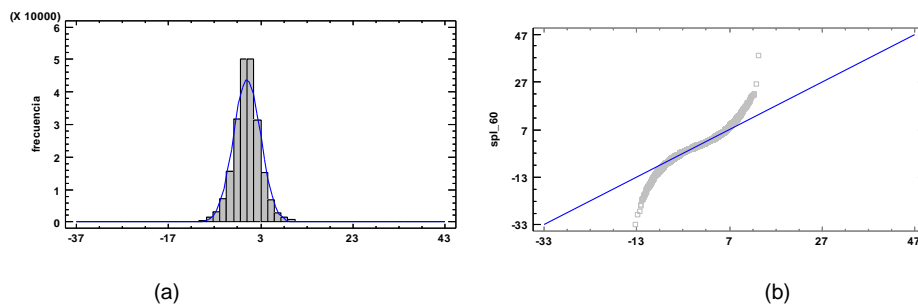


Figura 363 – Errores para algoritmo SPL re-muestreo 60x60 (a) Distribución de frecuencias. (b) gráfica cuantil - cuantil.

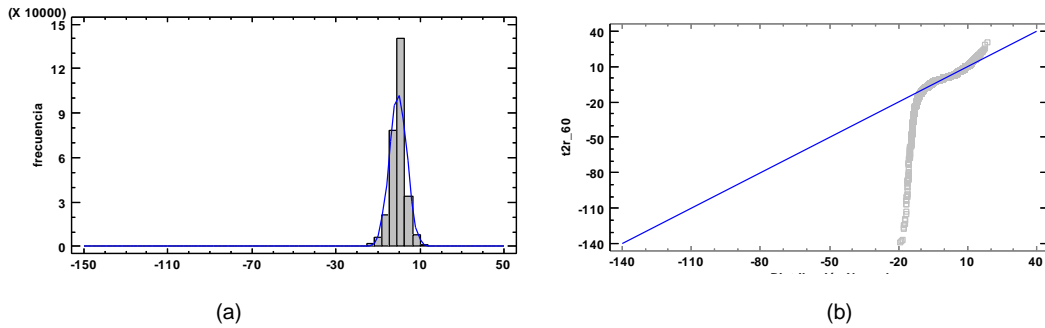


Figura 374 – Errores para algoritmo T2R re-muestreo 60x60 (a) Distribución de frecuencias. (b) gráfica cuantil - cuantil.

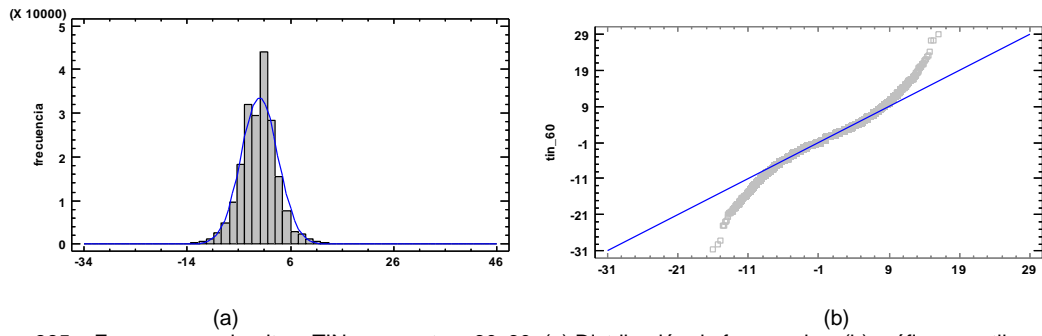


Figura 385 – Errores para algoritmo TIN re-muestreo 60x60 (a) Distribución de frecuencias. (b) gráfica cuantil - cuantil.

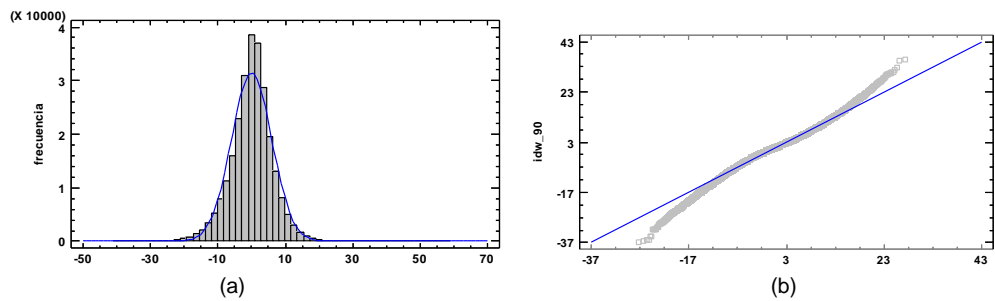


Figura 396 – Errores para algoritmo IDW re-muestreo 90x90 (a) Distribución de frecuencias. (b) gráfica cuantil - cuantil.

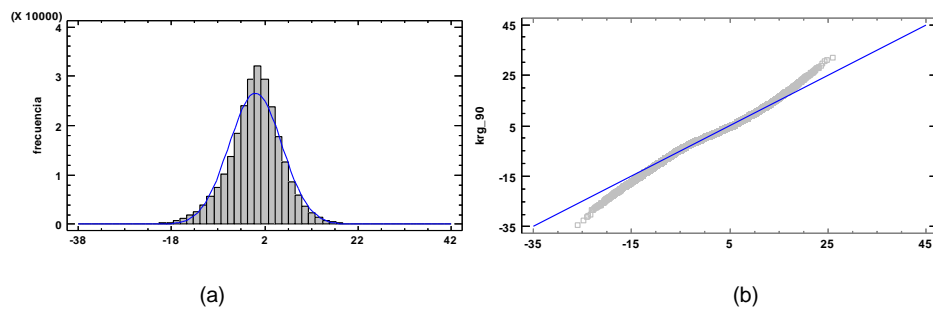


Figura 407 – Errores para algoritmo KRG re-muestreo 90x90 (a) Distribución de frecuencias. (b) gráfica cuantil - cuantil.

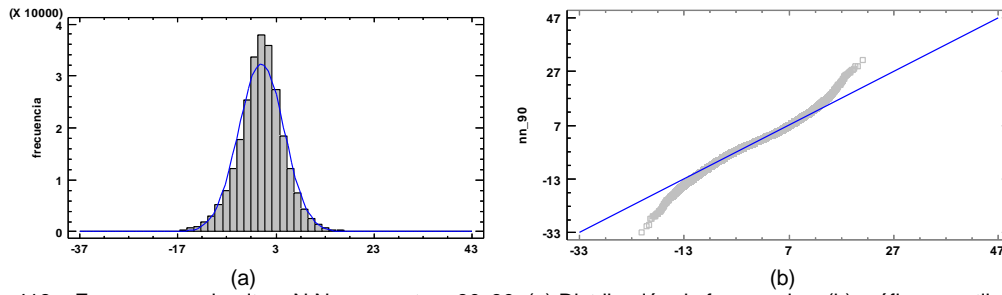


Figura 418 – Errores para algoritmo N.N re-muestreo 90x90 (a) Distribución de frecuencias. (b) gráfica cuantil - cuantil.

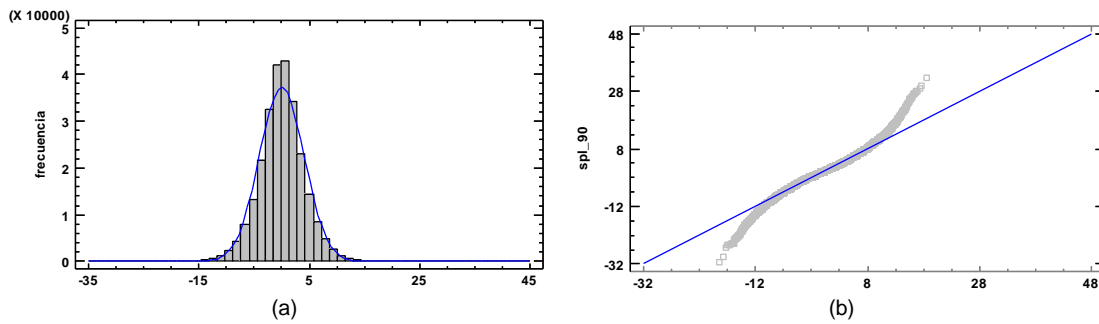


Figura 429 – Errores para algoritmo SPL re-muestreo 90x90 (a) Distribución de frecuencias. (b) gráfica cuantil - cuantil.

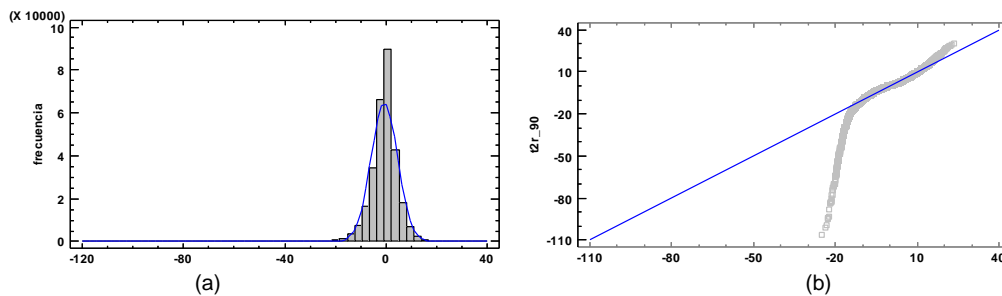


Figura 50 – Errores para algoritmo T2R re-muestreo 90x90 (a) Distribución de frecuencias. (b) gráfica cuantil - cuantil.

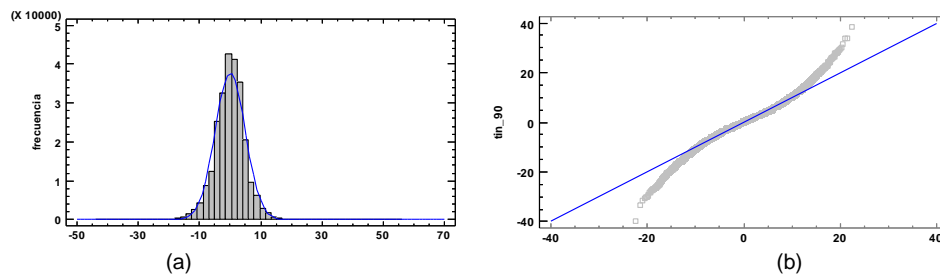


Figura 51 – Errores para algoritmo TIN re-muestreo 90x90 (a) Distribución de frecuencias. (b) gráfica cuantil - cuantil.

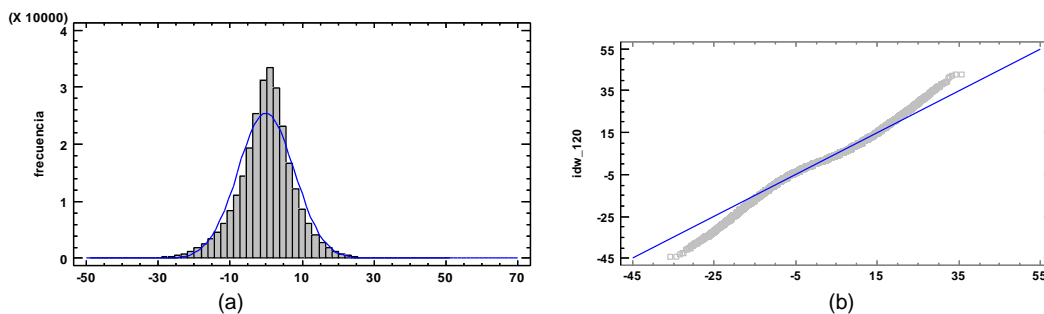


Figura 52 – Errores para algoritmo IDW re-muestreo 120x120 (a) Distribución de frecuencias. (b) gráfica cuantil - cuantil.

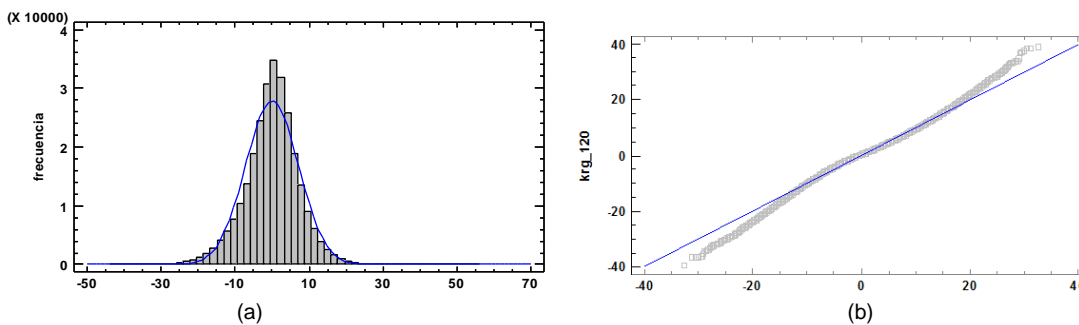


Figura 53 – Errores para algoritmo KRG re-muestreo 120x120 (a) Distribución de frecuencias. (b) gráfica cuantil - cuantil.

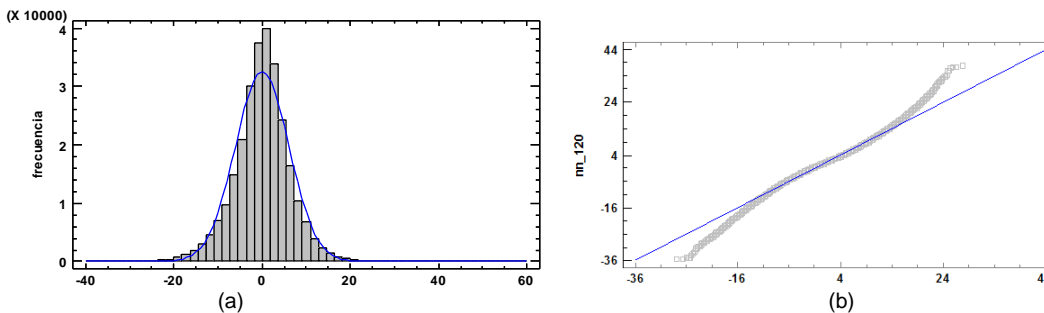


Figura 54 – Errores para algoritmo N.N re-muestreo 120x120 (a) Distribución de frecuencias. (b) gráfica cuantil - cuantil.

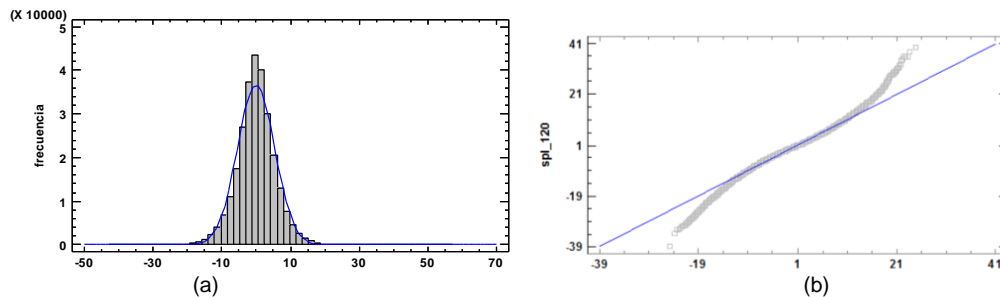


Figura 55 – Errores para algoritmo SPL re-muestreo 120x120 (a) Distribución de frecuencias. (b) gráfica cuantil - cuantil

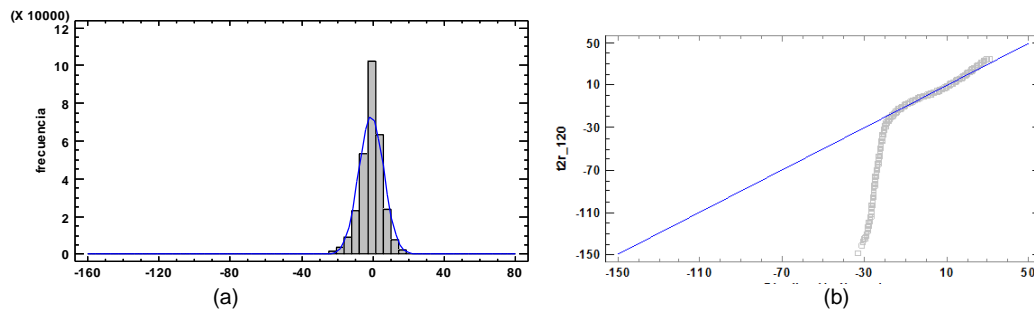


Figura 56 – Errores para algoritmo T2R re-muestreo 120x120 (a) Distribución de frecuencias. (b) gráfica cuantil - cuantil

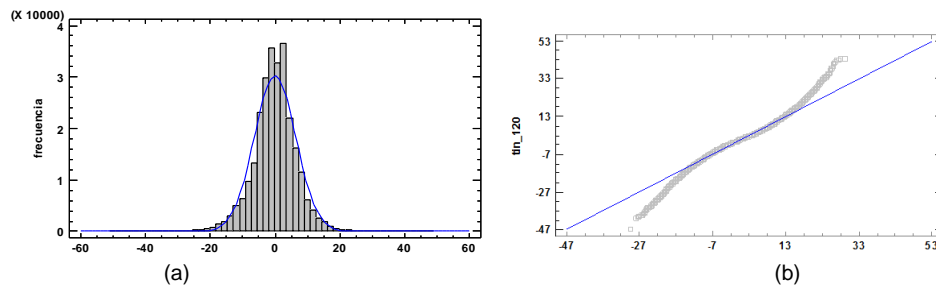


Figura 57 – Errores para algoritmo TIN re-muestreo 120x120 (a) Distribución de frecuencias. (b) gráfica cuantil - cuantil

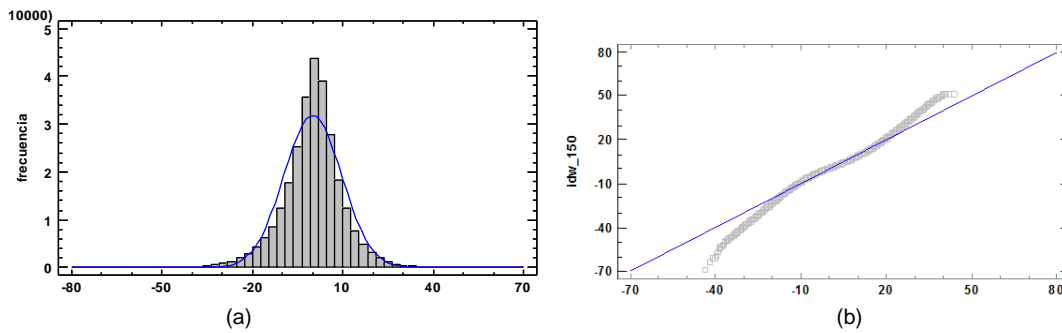


Figura 58 – Errores para algoritmo IDW re-muestreo 150x150 (a) Distribución de frecuencias. (b) gráfica cuantil - cuantil

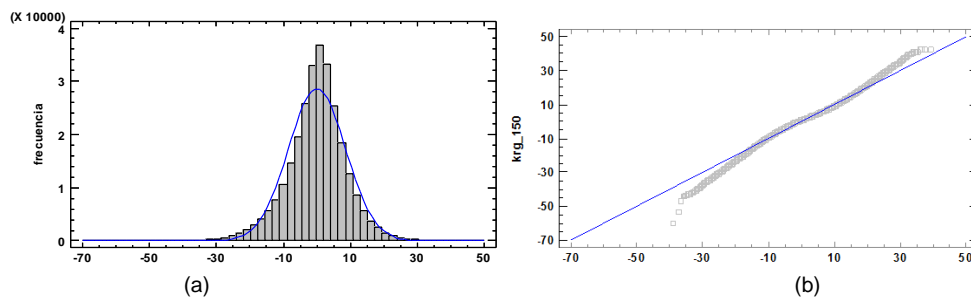


Figura 59 – Errores para algoritmo KRG re-muestreo 150x150 (a) Distribución de frecuencias. (b) gráfica cuantil - cuantil



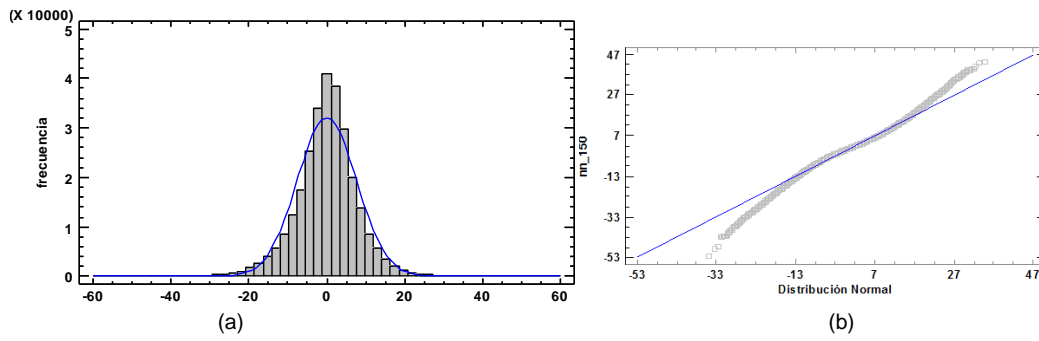


Figura 60 – Errores para algoritmo N.N re-muestreo 150x150 (a) Distribución de frecuencias. (b) gráfica cuantil - cuantil

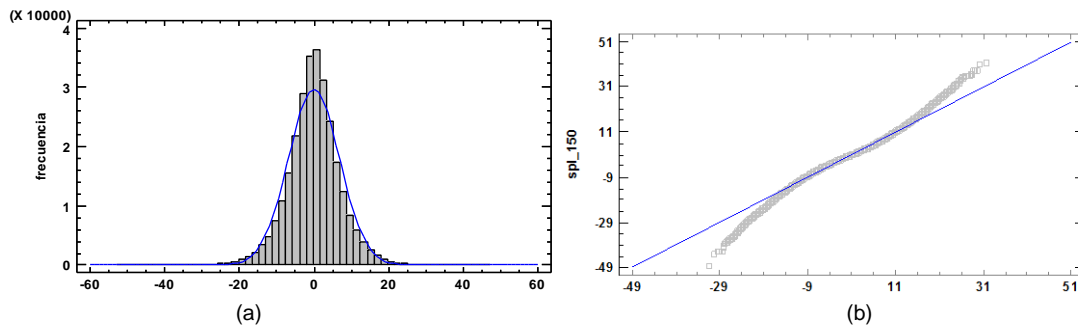


Figura 61 – Errores para algoritmo SPL re-muestreo 150x150 (a) Distribución de frecuencias. (b) gráfica cuantil - cuantil

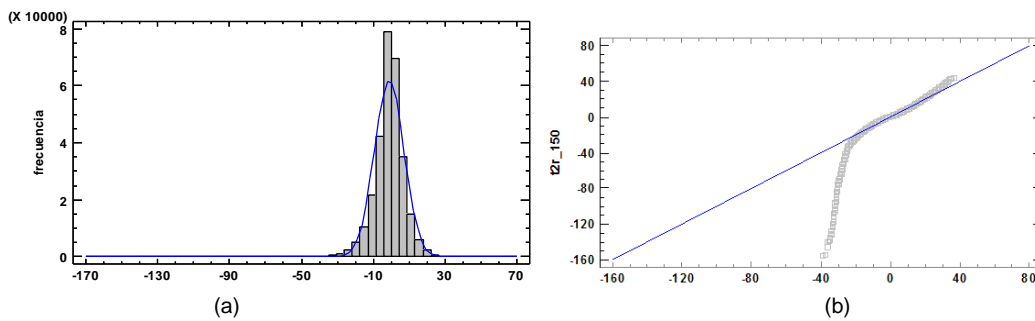


Figura 62 – Errores para algoritmo T2R re-muestreo 150x150 (a) Distribución de frecuencias. (b) gráfica cuantil - cuantil

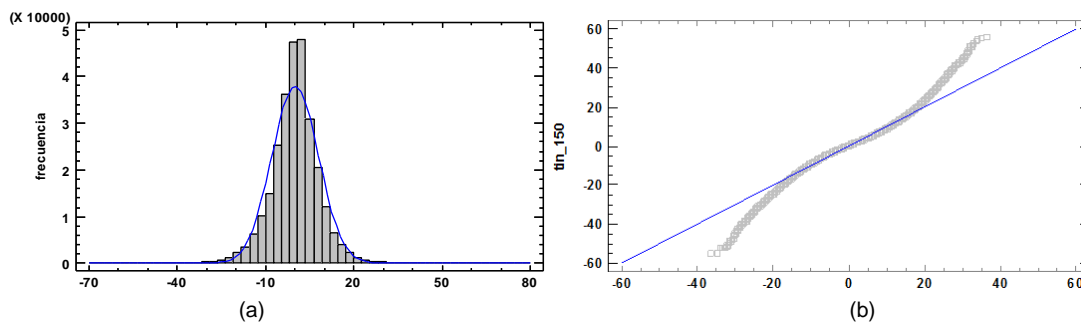


Figura 63 – Errores para algoritmo TIN re-muestreo 150x150 (a) Distribución de frecuencias. (b) gráfica cuantil - cuantil

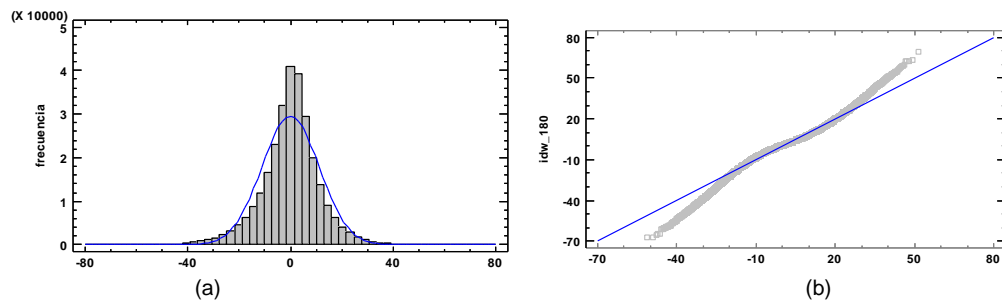


Figura 64 – Errores para algoritmo IDW re-muestreo 180x180 (a) Distribución de frecuencias. (b) gráfica cuantil - cuantil

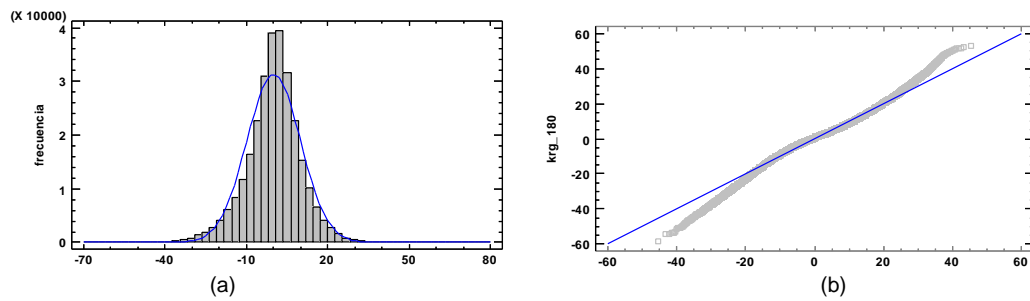


Figura 65 – Errores para algoritmo KRG re-muestreo 180x180 (a) Distribución de frecuencias. (b) gráfica cuantil - cuantil

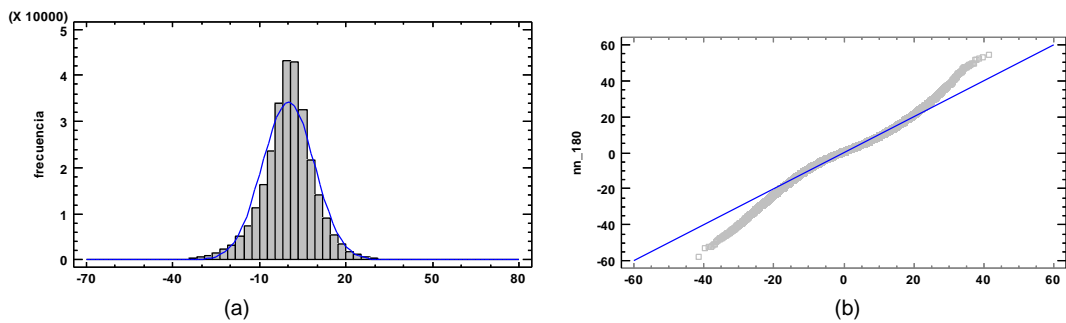


Figura 66 – Errores para algoritmo N.N re-muestreo 180x180 (a) Distribución de frecuencias. (b) gráfica cuantil - cuantil

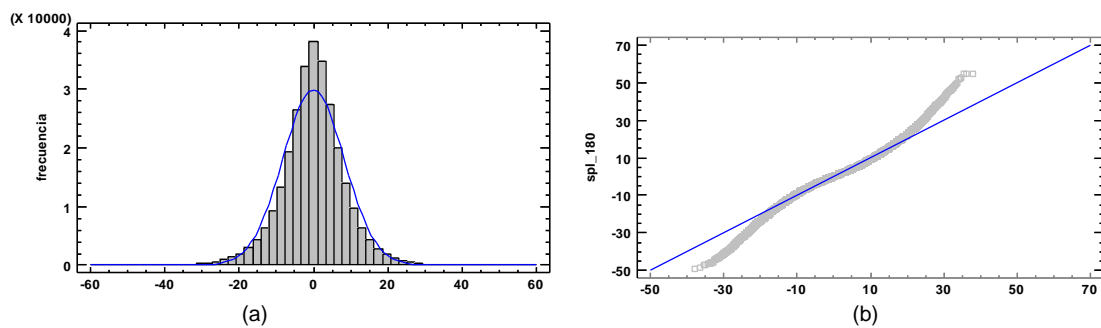


Figura 67 – Errores para algoritmo SPL re-muestreo 180x180 (a) Distribución de frecuencias. (b) gráfica cuantil - cuantil

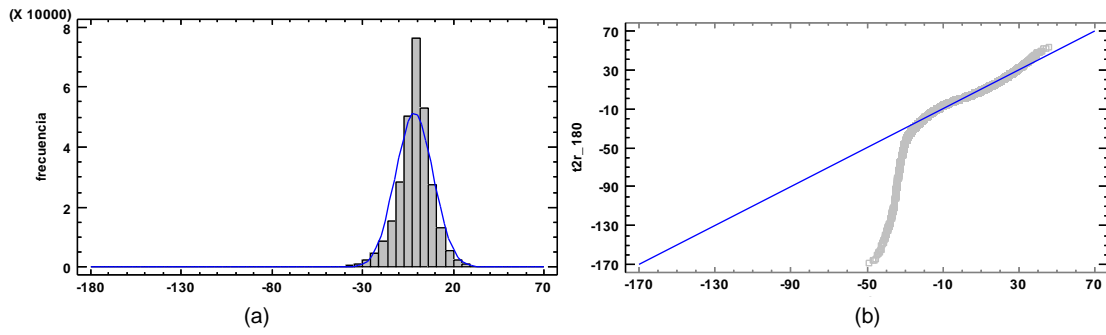


Figura 68 – Errores para algoritmo T2R re-muestreo 180x180 (a) Distribución de frecuencias. (b) gráfica cuantil – cuantil

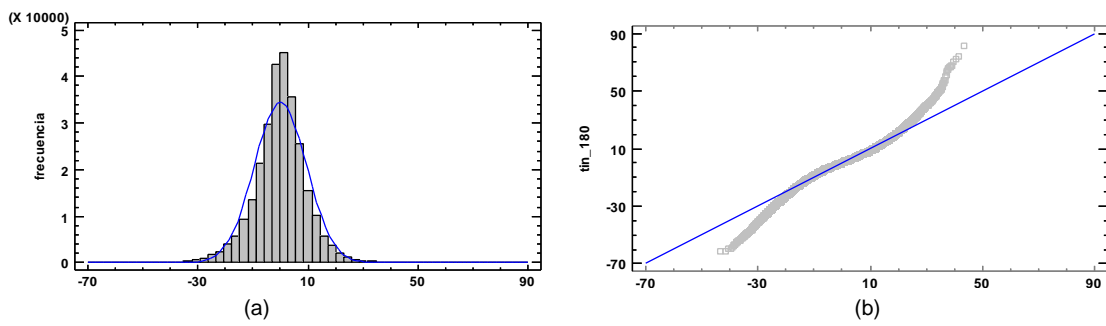


Figura 69 – Errores para algoritmo TIN re-muestreo 180x180 (a) Distribución de frecuencias. (b) gráfica cuantil – cuantil