

Horacio, J. (2010): Procesamiento de las mallas de cotas altimétricas (MDT25) elaboradas por el Instituto Geográfico Nacional (IGN) para el análisis geoestadístico de los modelos digitales de elevaciones (MDE). En: Ojeda, J., Pita, M.F. y Vallejo, I. (Eds.), Tecnologías de la Información Geográfica: La Información Geográfica al servicio de los ciudadanos. Secretariado de Publicaciones de la Universidad de Sevilla. Sevilla. Pp. 172-178. ISBN: 978-84-472-1294-1

PROCESAMIENTO DE LAS MALLAS DE COTAS ALTIMÉTRICAS (MDT25) ELABORADAS POR EL INSTITUTO GEOGRÁFICO NACIONAL (IGN) PARA EL ANÁLISIS GEOESTADÍSTICO DE LOS MODELOS DIGITALES DE ELEVACIONES (MDE)

Horacio, Jesús

Dpto. de Geografía, Facultad de Geografía e Historia, Universidad de Santiago de Compostela. Plaza de la Universidad, 1, 15782 Santiago de Compostela (A Coruña). Correo electrónico: hhorahh@gmail.com

RESUMEN

Se estudia la información altimétrica desarrollada por el Instituto Geográfico Nacional (IGN) para la realización de modelos digitales de elevación. Dicha información, a escala 1:25.000, se puede solicitar en diversos tipos de formatos.

El artículo propone una metodología para implementar los datos de cotas en mallas regulares (MDT25) a los Sistemas de Información Geográfica (SIG) así como su adecuación para la aplicación de técnicas estadísticas, aspecto fundamental para poder determinar la calidad y precisión del modelo además de la propagación del error.

Palabras Clave: Modelo Digital de Elevaciones (MDE), Sistemas de Información Geográfica (SIG), error MDE, interpolación, Instituto Geográfico Nacional (IGN).

ABSTRACT

This paper study the altimetry information developed by the Geographic National Institute (Spain) for the accomplishment of Digital Models Elevation (DEM). This information, to scale 1:25:000, it´s posible to request in diverse formats.

The research proposes a methodology to implement information´s level in regular meshes (MDT25) to the Geographical Information System (GIS) as well as his adequacy for the application of statistical technologies, fundamental aspect to be able determine the quality and precisión to the model.

Key Words: Digital Model Elevation, (DEM), Geographic Information System (GIS), error DEM, Geographic National Institute (Spain).

INTRODUCCIÓN

El incremento del uso de los modelos digitales de elevación (MDE) en diferentes campos de aplicación, con atención especial a los procesos de simulación dinámica (hidrología, climatología, riesgos...), es paralela al fuerte auge experimentado por los Sistemas de Información Geográfica (SIG) en los últimos años. En esta misma línea, desde diferentes ámbitos administrativos se está haciendo un notable esfuerzo por mejorar el uso no comercial, libre y gratuito de información geográfica de índole variada. No obstante, es común que la estructura de los datos suministrados al usuario presente un formato 'propio' que dificulta su implementación en un alto número de programas SIG.

De este modo, el presente trabajo se centra en analizar la información vectorial a escala 1:25.000 proporcionada por el Instituto Geográfico Nacional (IGN) para la realización de MDEs (© Instituto Geográfico Nacional de España, 2008). A este respecto, se propone un método válido que adecue esta información para la legibilidad en todos los *software*, el análisis geoestadístico –técnicas básicas para determinar el error del MDE– y los estándares de precisión. Los datos de partida (denominados por el IGN como MDT25) se estructuran en mallas con cotas altimétricas equidistantes 25 m, con un rebase exterior de 500 m y organizados según las hojas de misma escala del Mapa Topográfico Nacional (MTN25). Actualmente, el producto MDT25 es distribuido, entre otros, como ficheros ASCII XYZ (sin formato) o bien como ASCII matriz de ESRI ('*agr', válido para el programa comercial ArcGIS ESRI®).

El estudio de la estructura de los datos (MDT25) ha servido para observar ciertos problemas, entre los que sobresalen: i) dificultades en la incorporación de los datos al programa; y ii) imposibilidad de determinar el error y precisión del modelo cuando el área de estudio precisa de más de una malla altimétrica, lo que genera zonas de solape y sus consecuentes complicaciones. Para solucionar éstos y otros aspectos se proponen una serie de medidas basadas en procedimientos SIG (análisis espacial con datos vectoriales), técnicas de interpolación o el apoyo en *software* auxiliares del tipo Access. El resultado último se sintetiza en un conjunto de etapas que, a juicio del autor, el usuario debe realizar o, cuando menos, considerar en el momento de ejecutar un MDE de garantía con algunos de los archivos proporcionados por el IGN.

ZONA DE ESTUDIO

El ensayo metodológico llevado a cabo ha sido realizado sobre la cuenca pontevedresa del río Lérez (453 km²) y la lucense del río Cabe de 733 km² (Figura 1). No obstante, el lugar de aplicación no deja de ser una mera anécdota en el sentido de que la problemática suscitada con la información del IGN aquí citada (© Instituto Geográfico Nacional de España, 2008) surge en el momento en que el área de estudio se ubica sobre los límites de dos o más MTN25.



Figura 1. Encuadre territorial de las zonas de estudio. Galicia (España)

Como ilustra la Figura 2, la cuenca del Lérez abarca al menos una parte de 12 MTN25 mientras que el Cabe lo hace sobre 15 MTN25. Aunque el grado de ocupación sobre cada MNT25 siempre difiere para una zona de estudio, todos ellos son necesarios para la realización del MDE, habida cuenta que en los procesos de interpolación sobre

cuencas es preciso extenderse más allá de los límites de ésta. Así, en la Figura 2 se muestra el límite de las cuencas trazadas según las curvas de nivel a escala 1:25000 y el área de la misma con un *buffer* de 1.000 m.

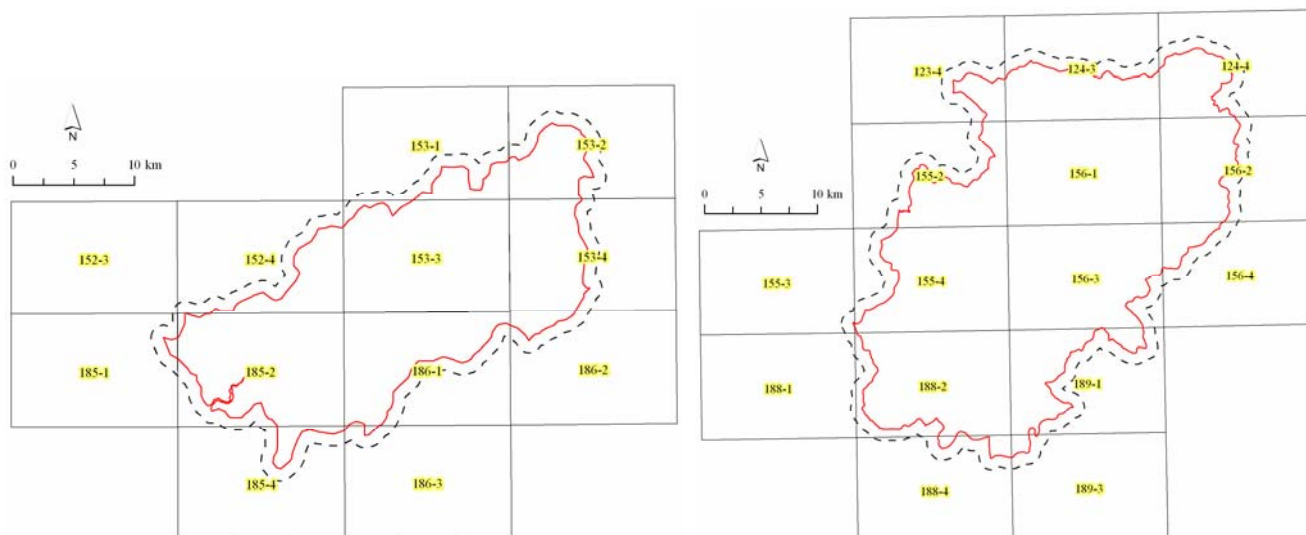


Figura 2. Distribución de los MTN25 en las cuencas del río Lerez (izquierda) y Cabe (derecha)

Topográficamente ambas cuencas presentan características contrastadas. En el caso del río Lerez predominan los relieves verticales con una red fluvial fuertemente encajada en numerosos sectores. Por el contrario, la cuenca del río Cabe se define por una amplia llanura central y por unos tramos altos y bajos con ratios de encajamiento A/P (anchura / profundidad) bajos.

METODOLOGÍA

El entramado metodológico ha de iniciarse cuestionándonos, ¿cómo de buena es la información con la que estamos trabajando? Es presumible que la información suministrada por un organismo oficial, en este caso el IGN, sea de calidad, si bien, dentro de esta suposición es necesario cerciorarse y dar respuesta a la pregunta planteada a fin de conocer los umbrales de error en los que nos movemos. Además, un error es algo inherente al modelo (Burrough and McDonnell, 1998) por lo que es necesario dotar al MDE de información que determine su magnitud (metadato).

El análisis del error en el MDE, entendido como la desviación altimétrica de éste respecto a la superficie real (Márquez, 2004), se realiza principalmente en función de criterios estadísticos. El conjunto de técnicas de evaluación del error del MDE se pueden agrupar, *grosso modo*, en globales –aquellas que consideran el conjunto de los datos– y locales –aquellas que se basan en un conjunto de puntos control para contrastar el valor del modelo en ese mismo punto–. La detección del error a través de técnicas locales ahonda en la cuantificación del error atributivo (Felicísimo, 1998) y, por consiguiente, gana en veracidad y objetividad frente a las técnicas globales.

La aplicación de las técnicas locales pretende comparar datos del modelo con otros más precisos para generar una superficie de error, lo que supone elaborar distintos modelos de interpolación con los mismos datos fuente y establecer un contraste con puntos de control (‘verdad terreno’) que suponemos libres de error. Como no siempre es posible contar con datos tomados en campo, las técnicas geoestadísticas recurren a la generación de un subconjunto de datos (*subset*) a partir del conjunto de la muestra. En el proceso es necesario indicar qué porcentaje de cotas de la muestra se va a usar para desarrollar los MDEs (*training*) y qué parte se reserva para cotejar el modelo (*test*). Resulta una condición fundamental que el número de puntos usados para testear la muestra sean suficientes y estén bien distribuidos (Felicísimo, 1994), con lo que puntos *test* que no cumplan estas condiciones pueden dar errores más altos de los esperados; una posible alternativa para evitar de antemano este problema consiste en analizar la distribución de los puntos mediante el uso de histogramas, gráficos de normalidad QQPlot, análisis espaciales de tendencias, mapas de Voronoi, etc. y, de ser necesario, volver a generar otra capa de cotas *test-training* antes de adentrarse en el desarrollo del modelo. Una vez que los resultados de testeado se ubiquen dentro de los umbrales adecuados se puede generar nuevamente el modelo con el 100% de los datos.

Como se indicaba en el capítulo introductorio, la información MDT25 se suministra como fichero ASCII XYZ (sin formato) y/o como ASCII matriz de ESRI. La principal diferencia entre ambas está en que los archivos de ESRI ya

vienen preparados para su lectura directa en el programa ArcGIS mientras que, por el contrario, los ficheros ASCII sin formato precisan de una serie de adecuaciones que faciliten su incorporación a un sistema SIG. En el caso del tratamiento geoestadístico de las cotas altimétricas es necesario emprender ajustes en ambos tipos de archivos, de modo que la metodología desarrollada en el presente artículo se va a centrar en: i) preparar la información del IGN MDT25 para su interpretación estadística; y ii) para los archivos sin formato (*.hxx) establecer una serie de pautas que permitan su visualización en programas SIG.

Adecuación del MDT25 para su visualización en programas SIG

El modo de suministrar el IGN la información MDT25 requiere, en el caso de los archivos ASCII XYZ sin formato, un profundo trabajo de preparación antes de su incorporación al *software* usado (esta serie de etapas se omiten para el programa ArcGIS ESRI® con los archivos '*.agr'). Los pasos aquí propuestos se sintetizan en cuatro y pueden llevarse a cabo en diferentes programas y/o sistemas operativos (*software* libre del tipo LINUX pueden ayudar sustancialmente a reducir el número de pasos mediante la implantación de un *script* que automatice el proceso), además han de repetirse tantas veces como archivos se deseen transformar, a menos que se agrupen varios en uno para minimizar el proceso, aunque ello también demanda una capacidad computacional superior.

Las mallas altimétricas ASCII XYZ sin formato pueden abrirse para visualizar su contenido con un archivo de texto. La información ahí contenida consiste en una primera fila de datos con información genérica: fuente, escala, número de hoja, proyección y coordenadas (UTM) con inicio en la esquina inferior izquierda del cuadrante. Esta información, irrelevante para el levantamiento del modelo, puede ser eliminada, al igual que el espacio generado. Una vez efectuados estos cambios es recomendable convertir la extensión del archivo original (*.hxx) a formato texto (*.txt) y agruparlos todos en uno único para agilizar el desarrollo del ejercicio.

El resto de información del archivo consiste en tres columnas que almacenan, de izquierda a derecha, la coordenada UTM 'x', la coordenada UTM 'y' y la altura. Estas columnas con información es preciso individualizarlas con ayuda de algún programa (véase © *Microsoft Access*) y asignarles una cabecera, por ejemplo "x, y, z" para referirnos respectivamente a las coordenadas UTM y la altura. El nuevo archivo generado se recomienda guardarlo con extensión (*.dbf) pues es de sencilla lectura por la mayoría de instrumentos SIG.

Adecuación del MDT25 para el tratamiento geoestadístico

El tratamiento geoestadístico de los modelos digitales requiere de una muestra de puntos lo más amplia posible, bien distribuida y donde cada valor sea único en localización. Dicho esto, y como previamente se ha indicado, cada malla de cotas presenta un rebase exterior de al menos 500 m, lo que implica la duplicación de puntos entre mallas colindantes una vez que todos los archivos de la zona de estudio se unifican. Estas zonas de solape motivadas por el rebase exterior más allá de los límites de cada cuadrante 1:25.000 suponen la aparición de información, como mínimo, duplicada (ver el ejemplo gráfico de la Figura 3). Este hecho debe ser subsanado mediante la disolución de aquellos puntos que ocupen una misma posición geográfica (dato redundante), sin embargo, se ha observado también que en ocasiones se almacenan datos con la misma localización pero una altitud diferente (los contrastes no son muy marcados con una oscilación aproximada entre uno y tres metros).

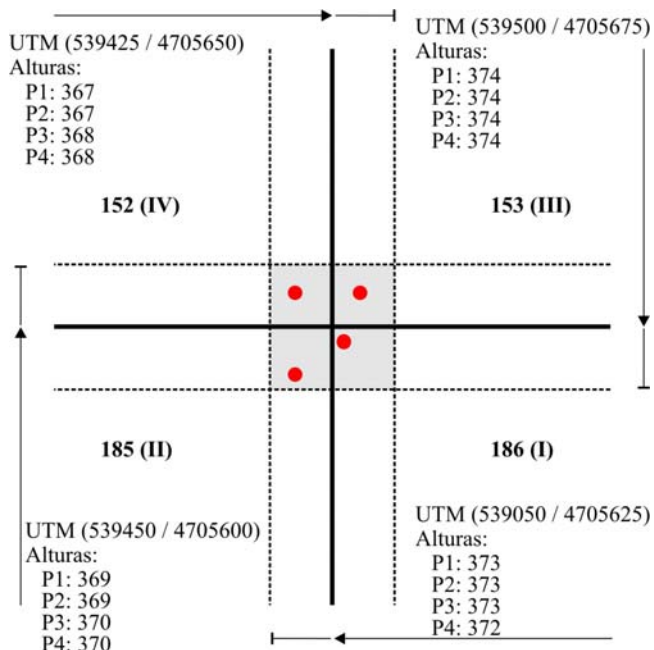


Figura 3. Ejemplo de 4 puntos en la zona de solape (gris) de 4 MTN25 con un rebase exterior (línea punteada)

Estos inconvenientes suponen un problema notable ya que para el testeado del modelo se corre el riesgo de que se consideren cotas con la misma localización geográfica y altitud con lo que se estaría falseando la veracidad de los resultados de calidad.

El procedimiento seguido para solucionar este doble problema (dato redundante e información altimétrica en ocasiones diferente en puntos con misma localización) ha consistido en la aplicación de una fusión espacial tomando como criterio de disolución las coordenadas puesto que es el elemento común que tienen todos los puntos repetidos (no así la altura).

De forma paralela se desarrolló un modelo digital de elevaciones con el archivo de datos redundantes mediante un interpolador exacto (*Spline, Inverse Distance Weighted...*). Esta condición de 'exacto' es fundamental para asegurarse que el valor predicho sea el mismo al que adopta realmente la variable altitud en el punto de muestreo. Ahora bien, ¿qué sucede en aquellos casos de puntos duplicados con altura diferente? Para estas situaciones, en las que se pueden juntar hasta un máximo de 4 valores de altura diferentes, uno por cada cuadrante MDT25, será el usuario quién determine qué desea hacer. La mayoría de los programas proponen diversas soluciones: valor máximo, mínimo, medio, anular todos, etc. Al ser diferencias altimétricas bajas, el valor promedio puede ser una buena alternativa.

En último término, se han extraído los valores de altitud del MDE efectuado con un interpolador exacto (un valor por cada localización x,y) y se han incorporado a la capa fusionada, es decir, aquella que únicamente almacena la localización UTM de todas los valores de cota.

Estadísticos de validación del modelo

La validación de los datos se hizo según 4 MDEs ráster: TIN (*Triangulated Irregular Network*) ráster, IDW (*Inverse Distance Weighted*), *spline* y *kriging*. Para conocer la incertidumbre de los mismos se han manejado las siguientes técnicas estadísticas (Li, 1991; Wood, 1996): error medio (MAE) para discernir entre errores aleatorios y/o sistemáticos; desviación estándar de la media algebraica para averiguar si el MDE está dentro del umbral de error máximo permitido (± 3); y error cuadrático medio (ECM) válido también para situar el MDE dentro de los estándares de precisión pero ignorando el signo, al contrario de lo que sucede con el error medio (Shearer, 1990), lo que convierte a esta fórmula en la más empleada para validar los MDEs (Burrough y McDonnell, 1998).

Por otro lado, la idea de precisión podemos definirla como la mayor o menor proximidad al estado de exactitud (diferencia entre el cálculo y el valor verdadero). Los métodos seguidos para conocer la precisión fueron el método de Ley (1986) y el método LMAS (*Linear Map Accuracy Standards*) del Servicio Cartográfico Militar Británico.

RESULTADOS

En la Tabla 1 se muestran los valores *test-training* adoptados para el análisis geoestadístico. No existe unanimidad científica en el establecimiento de estos porcentajes (Márquez, 2004), asentándose la mayoría de trabajos entre el 10-20% para el testeo, de modo que, en nuestro caso, se ha optado por una opción intermedia del 15% para el test y el resto como *training* (85%).

Tabla 1. Número de cotas por cuenca manejadas en el estudio y porcentajes de variación de la muestra total (M_1) a la muestra tratada (M_2), donde, M_1 : muestra total; M_2 : muestra fusionada; T_{R1} : *training* M_1 ; T_{S1} : *test* M_1 ; T_{R2} : *training* M_2 ; T_{S2} : *test* M_2 ; $Var.M$: variación muestras; $Var.TR$: variación *training*; $Var.TS$: variación *test*.

Río	M_1	M_2	T_{R1}	T_{S1}	T_{R2}	T_{S2}	$Var.M$	$Var.TR$	$Var.TS$
Lérez	954.651	919.911	811.453	143.198	781.924	137.987	-3,64%	-3,09%	-0,55%
Cabe	1.512.719	1.422.905	1.285.811	226.908	1.209.469	213.436	-5,94%	-5,05%	-0,89%

En la Tabla 2 se recogen las valoraciones de las técnicas estadísticas aplicadas sobre cada MDE y cuenca.

Tabla 2. Resultados de la validación estadística de los modelos digitales de elevación.

	TINgrid	IDW	Spline	Kriging	
Lérez	0,01	-0,19	-0,17	-0,54	MAE
	0,76	3,58	3,05	3,10	Desv. estándar
	0,62	3,49	2,98	3,82	ECM
Cabe	-0,50	-0,03	-0,02	-0,53	MAE
	-0,84	4,10	3,10	3,49	Desv. estándar
	0,59	3,86	2,95	3,11	ECM

De la Tabla 2 también se puede deducir el nivel de precisión de los modelos manejados. En el método de Ley (1986) el error estándar de las alturas debe ser 1/3 la resolución, es decir, ~8,33. En los estándares de precisión según el método de LMAS, los resultados son para el Lérez (%): 99,98; 99,91; 99,99; 99,75 (TINgrid, IDW, *spline* y *Kriging* respectivamente). En el caso del Cabe (%), y siguiendo el mismo orden de interpolador: 100; 99,10; 99,80; 99,55.

CONCLUSIONES

Con el ejercicio llevado a cabo es posible no solo contrastar la presumible calidad de los datos sino cuantificar el error. Igualmente, se deja el material preparado para aplicaciones estadísticas varias en lo referente a la propagación del error, consideración menos tratada en la literatura científica pero de importancia notable (Burrough, 1986; Márquez, 2004).

Los resultados de la aplicación estadística, con especial atención al ECM, ponen de relieve diferentes niveles de error para distintos interpoladores, aspecto muy a considerar según sean los fines de la modelización, principalmente en lo que atañe a procesos de simulación dinámica donde el nivel de error puede influir ostensiblemente en el resultado final. En las dos cuencas trabajadas el TINgrid se erige como el modelo más robusto estadísticamente, con una desviación del ECM inferior al metro en ambos casos. La situación contraria aparece con el IDW y el *kriging*, modelos éstos que rebasan los umbrales de idoneidad. El *spline* está en el límite de la permisividad estadística con desviaciones del ECM próximas a los 3 m pero rebasadas por la desviación estándar.

Los estándares de precisión atestiguan valoraciones diferentes según el método empleado, sin embargo, en todos los casos y para los dos métodos manejados se presentan valores de calidad muy elevada.

El resultado final de calidad almacenado como metadato debe ser tratado como una combinación entre el dato, el objetivo de estudio, el interpolador y la topografía del terreno, de modo que unos errores serán más asumibles que otros, sin olvidar el hecho de que toda estructura matricial implica *per se* una pérdida de fidelidad ya que se pasa de un punto con precisión espacial a una cuadrícula que adopta el valor de ese punto para un área.

En suma, los datos del IGN (© Instituto Geográfico Nacional de España, 2008) se pueden calificar con un alto grado de calidad, si bien, no están exentos de variaciones según la zona de estudio y su cotejo con diferentes métodos de interpolación.

En último lugar es conveniente mencionar el mal uso del término MDT (Modelo Digital del Terreno) en vez de MDE (Modelo Digital de Elevaciones) ya que el primero, según Felicísimo (1994), se refiere a una estructura numérica de datos que representa la distribución espacial de una propiedad de la superficie de terreno (de una variable cuantitativa y continua), mientras que el MDE es una estructura de datos numérica que representa la distribución espacial de la altitud de la superficie del terreno.

AGRADECIMIENTOS

A Ángel Manuel Felicísimo y Juan Manuel Rodríguez Borreguero por su amabilidad y valiosas sugerencias para el desarrollo del artículo. A la Consellería de Educación e Ordenación Universitaria da Xunta de Galicia por la bolsa concedida para la realización de estudios de Tercer Ciclo.

REFERENCIAS BIBLIOGRÁFICAS

Burrough, P. A. (1986): *Principles of Geographical Information Systems for land resources assessment*. Clarendon, Oxford, 203 pp.

Burrough, P. A. y McDonnell, A. (1998): *Principles of Geographical Information Systems for land resources assessment*. Somerset, Oxford University Press, 333 pp.

Felicísimo, A. M. (1994): *Modelos digitales del terreno*. Pentalfa Ediciones, Oviedo, 220 pp.

Felicísimo, A. M. (1998): *Modelos digitales del terreno*. <http://207.248.141.2:80/SIR/CursoMDT/cursoMDT.html>

Ley, R. (1986): "Accuracy assessment of digital terrain models. Proceedings". *AutoCarto*, (1): 455-464

Li, Z. (1991): "Effects of check points on the reliability of DTM accuracy estimates obtained from experimental tests". *Photogrametric Engineering & Remote Sensing*, 47(10): 1.333-1.340

Márquez, J. (2004): *Modelos digitales de elevaciones, métodos de interpolación y parámetros morfométricos del relieve*. Tesis doctoral, Universidad de Sevilla, 594 pp.

Shearer, J. W. (1990): "The accuracy of digital terrain models". En Petrie, G. & Kennie, T. J. (Ed.): *Terrain modelling in surveying and civil engineering*. Whittles Publishing Services in association with Thomas Telford, London, pp. 315-336

Wood, J. (1996): *The geomorphological characterisation of digital elevation models*. Tesis doctoral, Universidad de Leicester, 109 pp.