UNIVERSIDAD POLITÉCNICA DE MADRID

DEPARTAMENTO DE INGENIERÍA TOPOGRÁFICA Y CARTOGRAFÍA

ESCUELA TÉCNICA SUPERIOR DE INGENIEROS EN TOPOGRAFÍA, GEODESIA
Y CARTOGRAFÍA

IMPORTANCIA DE LA CALIDAD DE LOS MODELOS DIGITALES DE ELEVACIÓN
(MDE) PARA LA TOMA DE DECISIONES TERRITORIALES. PROPUESTA DE UN
MÉTODO DE ESTIMACIÓN DE ERRORES Y COSTES DE PÉRDIDA

Autor

ELENA GABRIELA CHICAIZA MORA
Ingeniera Geógrafa y del Medio Ambiente

Directores

JOSÉ JUAN ARRANZ JUSTEL
Doctor en Ingeniería Geográfica

FRANCISCO JAVIER ARIZA LÓPEZ
Doctor Ingeniero Agrónomo

Mayo de 2017

Dedicado a mis seres amados e incondicionales

Agradecimientos

Deseo agradecer a todos quienes han sido parte del desarrollo de mi tesis, quienes han estado directamente relacionados, así como, a los que me han apoyado de lejos, ustedes han sido mi motivación en todo momento.

Al Dr. Francisco Javier Ariza por su sabiduría, paciencia y guía para la consecución de esta tesis doctoral.

Al Dr. José Juan Arranz por sus conocimientos y guía prestada en la elaboración de esta tesis doctoral.

Al Dr. Antonio Vázquez Hoehne por el apoyo y gestión en el Programa de Doctorado durante todo el proceso.

Al Dr. Xavier Buenaño por sus consejos y retroalimentación para la culminación de la presente tesis doctoral.

Al MSc. Marco Moreira por su guía y consejos en los temas relativos a los modelos hidráulicos.

A mis padres María Elena y Mario por haberme apoyado desde el primer día que decidí caminar hacia una meta que me ha traído grandes aprendizajes

A mis hijos Mateo y Nicolás por su amor y comprensión.

A Germán, Cristian, Henry, Azucena, Sandra, Ingrid y Victoria por su apoyo incondicional.

Resumen

La presente tesis doctoral tienen como objetivo proponer un método de estimación de errores de la componente vertical de los Modelos Digitales de Elevación, mediante simulaciones geoestadísticas, a fin de determinar su probabilidad de ocurrencia.

Las estimaciones de probabilidad de error vertical obtenidas son relacionadas con los resultados (calados de inundación) de un estudio de caso específico localizado en la microcuenca del Río Garrapata, Ecuador, en el que se ejecutó una simulación hidráulica de inundaciones con el uso de un MDE como insumo. El estudio de caso fue seleccionado usando criterios a partir de los resultados de la investigación de usos y usuarios de los Modelos Digitales de Elevación.

El enfoque usado en esta tesis constituye un punto de partida para la profundización de la determinación de los costes de la mala calidad de un grupo particular de datos geoespaciales como son los Modelos Digitales de Elevación.

Abstract

The current research project aims to propose an error estimation method of vertical component of Digital Elevation Models through geostatistical simulations in order to define its occurrence probability.

The vertical error probability estimations obtained are compared with the outputs (inundation heights) of a specific case study located in the Garrapata river watershed, Ecuador, where was carried out a flooding hydraulic simulation with the use of a DEM as input. The case study was chosen using experts's criteria from the results found of research of uses and users of Digital Elevation Models.

The approach used in this research project constitutes a starting point for the deepening of determination of poor quality of this particular geospatial data as Digital Elevation Models.

Índice general

Agradecimientos	III
Resumen	V
Abstract	VII
Índice figuras	XV
Índice de tablas	XIX
1 INTRODUCCIÓN	1
1.1 Hipótesis	3
1.2 Objetivo General	3
1.3 Objetivos Específicos	3
2 ANTECEDENTES	5
2.1 Información Geográfica y Modelos Digitales de Elevación	6
2.1.1 Información Geográfica	6
2.1.2 Modelo Digital de Elevación	7
2.1.3 Fuentes de obtención y modelos de datos de los MDE	9

2.1.4	Superficies derivadas del los MDE	11
2.1.5	Principales aplicaciones de los Modelo Digital de Elevaciones (MDE)	14
2.2	Teoría del error e incertidumbre en los MDE y herramientas	15
2.2.1	Definición de error e incertidumbre	15
2.2.2	Tipología de los errores en los MDE	16
2.3	Calidad de la Información Geográfica y de los Modelos Digitales de Elevación	18
2.3.1	Calidad de la Información Geográfica	18
2.3.2	Normas y Estándares de Calidad para la Información Geográfica	20
2.3.3	Calidad de los Modelos Digitales de Elevación	26
2.3.4	Evaluación de calidad en los MDE	27
2.4	Técnicas de generación de errores en los MDE	31
2.4.1	Simulación determinista de errores en los MDE	32
2.4.2	Simulación estocástica de errores en los Modelos Digitales de Elevación	33
2.5	Técnicas de propagación de errores	43
2.6	Valor de la Información Geográfica y los Modelos Digitales de Elevación	44
2.6.1	Valor Económico de la Información Geográfica	45
2.6.2	Costes de la calidad y la mala calidad	48
2.6.3	Función de costo de pérdida debido a la mala calidad	50
3	MATERIAL Y MÉTODOS	59
3.1	Fuentes de Información	59

3.1.1	Encuestas disponibles sobre los usuarios de MDE	59
3.1.2	Encuesta aplicada a usuarios especializados	60
3.1.3	Lista de términos específicos sobre campos de aplicación y uso de los MDE	60
3.1.4	Información cartográfica de zonas específicas en Ecuador . . .	60
3.2	Herramientas utilizadas	61
3.2.1	Entorno R	61
3.2.2	HEC-RAS y GeoHEC-RAS (ArcGis)	63
3.3	Selección de la aplicación y caso de uso	63
3.3.1	Análisis de usuarios y aplicaciones de los MDE en España . .	64
3.3.2	Encuesta a nivel mundial a usuarios especializados	66
3.3.3	Consulta a ScienceDirect	66
3.3.4	Aplicación seleccionada	69
3.4	Diseño de las simulaciones	69
3.4.1	Datos utilizados	70
3.4.2	Preparación de los datos	72
3.4.3	Cálculo de errores de la componente vertical	74
3.4.4	Análisis exploratorio de errores	74
3.4.5	Análisis de la estructura espacial	77
3.4.6	Simulación geoestadística	77
3.5	Caso: Inundación del Río Garrapata	78
3.5.1	Descripción del área de estudio	79
3.5.2	Herramientas para el modelado de inundaciones	81

3.5.3	Cálculo de errores de la componente vertical	89
3.5.4	Valoración del suelo y edificaciones en el área de estudio	91
3.5.5	Cualificación de pérdidas en términos monetarios	93
4	RESULTADOS Y DISCUSIÓN	97
4.1	Resultados de la selección de la aplicación y caso de estudio	97
4.1.1	Resultados de la encuesta sobre usuarios y usos de MDE en España	97
4.1.2	Resultados de la encuesta sobre usuarios y usos de MDE a nivel mundial	100
4.1.3	Resultado del análisis de las consultas a ScienceDirect	106
4.2	Ensayo: Obtención de simulaciones geoestadísticas de errores de la vertical en un MDE	109
4.2.1	Test estadístico de normalidad	112
4.2.2	Test de dependencia espacial de Moran	113
4.2.3	Análisis de la estructura espacial	113
4.3	Caso de estudio: Inundaciones del Río Garrapata	120
4.3.1	Determinación de áreas de inundación	120
4.3.2	Análisis exploratorio de errores	121
4.3.3	Variograma experimental	123
4.3.4	Ajuste del variograma experimental	124
4.3.5	Variabilidad de los errores de la vertical	125
4.3.6	Probabilidades de error de la vertical	126
4.3.7	Resultados de las pérdidas monetarias por errores en la vertical	131

5 CONCLUSIONES	133
5.1 Líneas futuras de investigación en torno a los costes de la mala calidad de los Modelos Digitales de Elevación	135
Bibliografía	137
Índice alfabético	149
Listado de Acrónimos	149
A Apéndices	151
A.1 Artículo: Spatial uncertainty of a geoid undulation model in 7 Guayaquil, Ecuador	151
A.2 Artículo: Modelling uncertainty using geostatistics, a case study in Ecuador	170
A.3 Artículo: Análisis de los usuarios y usos de los MDE en España . . .	177
A.4 Encuesta a usuarios especializados	205
A.5 Artículo: DEMs: An approach to users and uses from the quality perspective	210
A.6 Script de la simulación de errores	237
A.7 Informe de procesamiento del MDE Condado - Quito	244

Índice de figuras

2.1	Definición de pendiente. Tomado de El-Sheimy, Valeo y Habib (2005)	12
2.2	Relación entre los procesos de ISO 19157 y normas relativas a la calidad de la IG. Tomado de Ariza-López y Rodríguez-Pascual (2007)	25
2.3	Elementos importantes en el análisis variográfico.	36
2.4	Dos simulaciones condicionales en una dimensión. La simulación coincide con la realidad en los puntos condicionantes $(x_1, x_2, x_3, \dots x_N)$. Tomado de Alfaro (2008)	42
2.5	Flujo de valor de los beneficios de la IG. Tomado de Natural Resources Canada (2015)	46
2.6	Costes de la calidad y no calidad. Tomado de Yepes Piqueras, V. (2010)	49
2.7	Función de pérdida de la calidad. Modificado de Taguchi, Chowdhury y Wu (2005)	51
3.1	Librerías de R utilizadas comúnmente para manejo de datos espaciales	62
3.2	Estructura de la consulta a ScienceDirect y los resultados de la consulta	67
3.3	Ubicación del área de estudio en el sector El Condado, al nor-occidente del área urbana de la ciudad de Quito. Coordenadas en sistema TM-Quito.	71
3.4	MDE construido a partir de nube de puntos LAS. Coordenadas en el sistema de referencia espacial TM-Quito.	72

3.5	Discrepancia máxima D entre dos distribuciones. Tomado de Olea (2012)	75
3.6	Ubicación general del área de estudio caso Río Garrapata	80
3.7	Esquema del procedimiento para determinación de áreas de inundación. Tomado de Ackerman (2009)	83
3.8	Área de estudio del caso. Geometría del tramo del Río Garrapata analizado en el sector de San Andrés	85
3.9	Representación en HEC-RAS de la geometría del río con las secciones transversales y una los parámetros asociados a la sección transversal de la estación 1600	86
3.10	Procedimiento resumen de cálculo de errores de la componente vertical y análisis geoestadístico para la ejecución de simulaciones.	89
3.11	Algebra de mapas propuesto para la obtención de costes por la existencia de errores de la componente vertical en un MDE, caso de estudio Río Garrapata. Creado a través de la herramienta Model Builder de ArcGis.	94
4.1	Gráficos de redes de conexión entre las áreas temáticas.	98
4.2	Redes bipartita de los temas y tópicos relacionados a los MDE.	107
4.3	Mapa de errores obtenido de la sustracción entre MDE: evaluado y referencia. Coordenadas planas en sistema TM-Quito.	110
4.4	Mapa de errores usando una máscara. Coordenadas planas en sistema TM-Quito.	111
4.5	Gráficos exploratorios: histograma, gráfico cuantil-cuantil y diagrama caja-bijote (de izquierda a derecha)	112
4.6	Gráficas para evaluación de posible tendencia espacial	114
4.7	Variograma experimental de errores de la vertical	115
4.8	Variograma experimental de errores de la vertical	116

4.9	Variograma experimental de errores de la vertical	117
4.10	Variogramas experimentales direccionales (0, 45, 90 y 135 grados) de errores de la vertical	118
4.11	Simulaciones de errores de la vertical considerando un modelo gaussiano	119
4.12	Ejemplo del resultado de una sección transversal una vez ejecutada la modelización hidráulica	120
4.13	Áreas de inundación para los cinco períodos de retorno analizados: 5, 10, 25, 50 y 100 años. Sistema de referencia espacial WGS84, proyección UTM 17S.	122
4.14	Gráficas exploratorias de errores de la vertical en la microcuenca del Río Garrapata. De izquierda a derecha: histograma, diagrama cuantil-cuantil y caja-bigote.	123
4.15	Variograma experimental de errores de la vertical en zona de estudio: Río Garrapata.	124
4.16	Mapa o superficie variográfica de errores de la vertical en zona de estudio: Río Garrapata.	125
4.17	Variogramas direccionales en direcciones: 0, 45, 90 y 135 grados para error de la vertical para el caso de estudio Río Garrapata	126
4.18	Simulaciones de la estructura espacial de error de la vertical, caso de estudio: Río Garrapata	127
4.19	Ubicación de la recta que corta las simulaciones creadas, caso de estudio: Río Garrapata	128
4.20	Sección longitudinal trazada a través de simulaciones de error de la vertical, caso de estudio: Río Garrapata	129
4.21	Áreas inundables con probabilidades de error, subestimación de alturas con valores menores a -1,5 metros (subestimación de alturas)	130

4.22 Mapa de inundaciones y las pérdidas monetarias debido a la mala calidad del MDE relativo a la componente vertical; alturas con valores menores a -1,5 metros (subestimación)	132

Índice de tablas

2.1	Atributos primarios derivados de los MDE y sus aplicaciones. Tomado de El-Sheimy, Valeo y Habib (2005)	14
2.2	Número de normas desarrolladas relativas a la IG por cada institución. Tomado de Ariza-López (2011)	21
2.3	Resumen de los elementos de la calidad de la IG. Modificado de AENOR (2014)	22
2.4	Estándares de exactitud vertical establecidos por los diferentes organismos de cartografía	30
2.5	Descripción de 4 métodos para generar campos aleatorios conociendo únicamente RMSE. M representa la media, SD la desviación estándar, DxD la matriz donde se expresa la dependencia espacial. Tomado de Wechsler y Kroll (2006)	33
2.6	Contribución del sector y la industria de la IG en la economía nacional	46
2.7	Ejemplos de costos de la mala calidad en una empresa u organización. Tomado de Haug, Zachariassen y Van Liempd (2011)	49
3.1	Parámetros del sistema de referencia espacial empleado en el Distrito Metropolitano de Quito (DMQ).	73
3.2	Parámetros del sistema de referencia espacial empleado en el DMQ. .	77
3.3	Especificaciones del MDE que se utiliza como insumo para la estimación de áreas de inundación. Tomado de metadato MDE, de SIGTIERRAS	84

3.4	Caudales máximos en la zona estudiada. Tomado de (Gutiérrez, Góngora y Melo, 2008)	86
3.5	Caudales máximos en la zona estudiada. Tomado de (Moreira, 2016) .	87
3.6	Caudales máximos en la zona estudiada. Tomado de (Moreira, 2016) .	87
3.7	Intervalos de error.	90
3.8	Valor referencial para edificaciones. Tomado de SIGTIERRAS 2016 .	92
3.9	Valoración de la zonas al interior del territorio analizado. Los valores han sido establecidos tomando como referencia al informe de valoración masiva de predios rurales de un cantón cercano a la zona de estudio. Tomado Alcázar-Molina (2014), SIGTIERRAS (2016)	93
4.1	Frecuencias totales de áreas temáticas marcadas por usuarios encuestados.	99
4.2	Perfil de usuarios especializados de los MDE	101
4.3	Campos de aplicación de los MDE. Tomado de Ariza-López et al. (2017)	101
4.4	Frecuencias totales de los productos derivados de los MDE, marcadas por usuarios expertos encuestados	102
4.5	Frecuencias totales entre productos derivados MDE y exigencias de cálculo de variables y análisis	103
4.6	Índices empleados para expresar la calidad en los MDE	104
4.7	Valoración de la influencia de la mala calidad en los MDE, preguntas 17 y 18	105
4.8	Abreviaturas empleadas en el gráfico de red bipartita sobre temas y tópicos relativos a los MDE.	108
4.9	Parámetros de ajuste obtenidos mediante el método de mínimos cuadrados iterativos	124

4.10 Superficie (Ha) afectada por cada intervalo de pérdida monetaria (USD) para los períodos de retorno de 5, 10, 25 y 100 años. La pérdida monetaria está dada por celda.	129

Capítulo 1

INTRODUCCIÓN

Los costes de la mala calidad de la Información Geográfica (IG) en pocas ocasiones han sido abordados desde una perspectiva de cuantificación de las pérdidas incurridas en términos monetarios dependiendo de la aplicación en donde ésta sea usada.

Por otro lado, los Modelos Digitales de Elevación (MDE) se han constituido en uno de los productos que en la actualidad tienen mayor demanda debido a la incursión de nuevas geotecnologías como Vehículos Aéreos no tripulados (UAV), *Light Detection and Ranging* (LiDAR), entre otras, en un sinnúmero de campos de aplicación donde la variable altura es fundamental para el análisis y toma de decisiones territoriales.

Debido a que es muy complicado abarcar el análisis de varios elementos de la calidad de un producto cartográfico, de manera general, y en particular de los MDE, y que los tipos de error que pueden presentarse en estos son muy diversos, es necesario focalizar el estudio de un elemento en específico, el mismo que haya sido analizado y que sea crítico. En este contexto, la presente tesis centra su estudio en la exactitud de la componente vertical de los MDE.

Mediante la aplicación de una encuesta a los usuarios especializados de los MDE se pretende conocer, desde una perspectiva económica, el impacto de la mala calidad de este tipo de productos. De primera mano, se puede advertir que la mayoría de usuarios no son capaces de cuantificar el impacto económico de la mala calidad, aunque la mayoría de usuarios admiten que existe. La mayor proporción de entrevistados cuenta con un importante número de años de experiencia profesional.

Así, la presente tesis pretende determinar las probabilidades de error que podrían existir al momento de emplear un MDE y cuantificar, en función de ellas, las posibles pérdidas, en términos monetarios, debido a la mala calidad del MDE empleado en una aplicación específica.

El enfoque descrito permite adaptar el concepto empleado para la obtención de la función de pérdida de costos, en el que es necesario definir un atributo de calidad para determinar éstos.

Para ello, se propone la implementación de una metodología para estimar el patrón de distribución espacial de los errores en la componente vertical de un MDE mediante simulaciones, partiendo de la comparación del atributo de altura con un conjunto de datos de mayor exactitud. Esto se lo realiza mediante el uso de la técnica geoestadística, ejecutando inicialmente el análisis exploratorio y de estructura espacial, con el fin de obtener realizaciones de campos aleatorios de errores de la componente vertical de un MDE para una área de estudio en específico.

El proceso aquí reseñado es inicialmente realizado, a modo de ensayo o prueba, con un conjunto de datos del norte de la ciudad de Quito, Ecuador. El ensayo es ejecutado con la finalidad de determinar la factibilidad de aplicación de la metodología propuesta.

A continuación, la metodología es empleada en un estudio de caso y aplicación en específico, en la cual, la variable altura representada a través de un MDE juega un papel crítico en la obtención de un producto de análisis cartográfico; específicamente, se ha seleccionado a la simulación hidráulica para la determinación de áreas de inundación. La definición del sector de aplicación del MDE está dada mediante el análisis de los conocimientos e intereses de sus usuarios.

El conocimiento de las limitaciones que un producto cartográfico puede tener, en este caso, de un MDE es fundamental. La identificación de áreas donde existe una mayor probabilidad que ocurran o puedan presentarse errores en la componente vertical en un MDE con un determinado ámbito espacial constituye una importante herramienta en el momento de tomar decisiones o generar información geográfica en la cual el MDE sea un insumo crítico.

Por otro lado, la determinación de los costes o pérdidas que implica la mala calidad de un MDE en términos monetarios, puede ayudar para generar conciencia de la

importancia de la calidad de los MDE y además para crear nuevos estándares que comuniquen de mejor manera la calidad de los MDE y no únicamente con un solo valor que generalice la exactitud de la componente vertical de un MDE.

1.1 Hipótesis

La mala calidad de un MDE puede ser cuantificada en términos monetarios evaluando la exactitud vertical y la probabilidad de error dado un determinado umbral de acuerdo a un estudio de caso.

1.2 Objetivo General

Proponer un método de simulación de errores en la componente vertical en los Modelos Digitales de Elevación para determinar su probabilidad de ocurrencia y cuantificar las pérdidas económicas mediante un estudio de caso.

1.3 Objetivos Específicos

Los objetivos específicos de la presente tesis son los siguientes:

- Identificar los campos de aplicación de los MDE y los estándares utilizados para la evaluación de la calidad de los MDE.
- Recopilar información sobre un conjunto de casos reales de aplicaciones territoriales donde los MDE sean utilizados como una variable crítica.
- Seleccionar las técnicas para la simulación de errores de exactitud vertical en los MDE.
- Definir un conjunto de datos en el que se emplearán los métodos seleccionados para la simulación de errores de exactitud vertical en una aplicación específica de los MDE.

- Estimar los costes o pérdidas, en términos monetarios, originadas por la presencia de error en la componente vertical en un MDE, con la ayuda de variables como la probabilidad de error y la valoración de suelo, en un caso de estudio específico.

Capítulo 2

ANTECEDENTES

En este capítulo se realiza una descripción sobre los principales aspectos teóricos que van a ser abordados en la presente tesis, con el objetivo de establecer el marco teórico claro sobre el estado de la cuestión referente a la calidad de los MDE, y cómo su mala calidad afecta en los análisis geoespaciales y sus posteriores las aplicaciones territoriales para la toma de decisiones.

En este contexto, se realiza un breve repaso de las principales definiciones sobre la Información Geográfica (IG), la calidad de la IG, su importancia y valor; los MDE, aplicaciones, errores y medidas de error más comunes.

Por otro lado, se hace un resumen sobre las técnicas de simulación que han sido seleccionadas en la bibliografía como las mejores alternativas para simular errores de acuerdo a los casos de uso a ser utilizados.

Finalmente, se hace una breve revisión de los métodos de valoración económica que se han utilizado en el campo de la IG, así como los conceptos del costo de pérdida por mala calidad, para seleccionar un método que permita evaluar la importancia económica que tiene la (mala) calidad de los MDE en aplicaciones territoriales para la toma de decisiones.

2.1 Información Geográfica y Modelos Digitales de Elevación

2.1.1 Información Geográfica

A lo largo del tiempo las definiciones de la IG han ido evolucionado de acuerdo a las nuevas áreas de aplicación donde está siendo usada y al desarrollo de la tecnología asociada para su levantamiento, así, en la literatura se pueden encontrar diversidad de definiciones. Una de las definiciones más completa es la propuesta por Longhorn y Blakemore (2007), quienes realizan un repaso por las definiciones y términos más conocidos, para proponer una definición comprensiva para el siglo XXI: *“La Información Geográfica es una composición de atributos y datos espaciales que describen la localización y atributos de las cosas, incluyendo las formas y representaciones de éstas de acuerdo a un sistema de referencia 2D, 3D o 4D (x,y,z, tiempo), de tal manera que permita el análisis espacial de las relaciones entre las cosas así descritas, incluyendo sus diferentes atributos”*.

La IG ocupa un creciente lugar en las sociedades modernas (Genovese, Roche y Caron, 2009), el uso de la IG ha pasado de ser de uso exclusivo de especialistas (Longhorn y Blakemore, 2007) a convertirse en un insumo fundamental para la planificación y la toma de decisiones, la política pública, la gestión ambiental, la preparación para emergencias, la creación de productos de valor agregado (Donker, 2009) y un sinnúmero de aplicaciones por su característica ubicua.

Así, nos encontramos en un escenario donde la IG es accesible a todos los usuarios a través de las Infraestructuras de Datos Espaciales (IDE), las cuales han madurado desde su aparición y han surgido como un área de importante desarrollo en la IG, siendo un pilar de base para las estrategias e iniciativas gubernamentales (Nedovic-Budic, Cromptvoets y Georgiadou, 2012), facilitando el acceso a usuarios con diferentes perfiles de manera interoperable (INSPIRE, 2012), buscando ser más eficientes para servir sus necesidades y mejorar la capacidad en toma de decisiones. (Castelein, Bregt y Pluijmers, 2010; Natural Resources Canada, 2015; FGDC, 2016; Survey, 2015).

En el ámbito de la gobernanza territorial, la IG juega un papel importante, pues es la base para la generación de conocimiento, entendimiento y opinión sobre la realidad

del territorio, sus necesidades y posibilidades, no solo por parte de los tomadores de decisiones en materia de planificación territorial, sino de los actores del territorio (Dasí, 2011), de ahí que, la calidad las decisiones dependerá de la calidad de la IG.

Dentro del amplio abanico de productos de la IG, los MDE son un insumo fundamental en el análisis del territorio, por lo que la presente tesis centra su investigación en la calidad los MDE y las implicaciones de la mala calidad en casos de uso donde estos son considerados importantes para el análisis.

Actualmente los estándares ISO vinculados con la calidad de la información geográfica son muy conservadores, estáticos y genéricos en términos de muchas necesidades prácticas y en el entendimiento de la calidad de conjuntos de datos espaciales particulares como MDE (Podobnikar, 2016b) de ahí el interés en el estudio de la calidad de los MDE que permita plantear nuevas alternativas de evaluación de los mismos.

2.1.2 Modelo Digital de Elevación

Los datos de elevación son un elemento crucial en cualquier aplicación en las geociencias, por lo que, la elevación ha sido una variable considerada como una de las más necesarias y probablemente la más crítica en los Sistemas de Información Geográfica (SIG) (Atkinson, 2002), desde los fundamentales mapas geológicos a los más avanzados modelos tridimensionales (3D) de la superficie de la Tierra (Fleming, Giles y Marsh, 2010).

De acuerdo a Podobnikar (2016b) el MDE es un modelo superficial continuo usualmente en 2.5D, el cual consiste en valores de elevación que describen la superficie topográfica. Se puede describir de forma genérica del modo siguiente:

$$z = f(x, y)$$

Los MDE son una representación cuantitativa y matemática de la superficie de la Tierra (Mukherjee et al., 2013; Hyann et al., 2016), se han convertido en uno de los productos de datos espaciales más usados para el modelamiento de la superficie (Patel, Katiyar y Prasad, 2016), siendo la elevación uno de los descriptores más importantes de la morfología de la Tierra (INSPIRE, 2013).

A mediados de los años 50, un nuevo campo de investigación emergía en la fotogrametría, los Modelo Digital del Terreno (MDT) por sus siglas en inglés (Florinsky, 2012). Los investigadores Miller y Laflamme en 1958 definieron a los MDT como una representación estadística de la superficie continua del terreno, mediante un número elevado de puntos con coordenadas x,y,z conocidas, en un sistema de coordenadas arbitrario (Felicísimo, 1994).

Para algunos autores el término MDT es sinónimo de MDE, en este caso hay que hacer una distinción ya que, los modelos digitales de elevación hacen referencia al concepto de medición de altura con respecto a un datum y la generación por parte del modelo de valores absolutos de altura (Fallas, 2007). Los MDT por su parte, no solo contienen información de elevación, además incluye otros elementos geográficos y características naturales.

Así, Maune (2001) define al MDE como un término genérico para datos digitales topográficos y batimétricos en sus varias formas, éste implica elevaciones del terreno (suelo desnudo), vacío de vegetación y estructuras antrópicas. Un Modelo Digital de Superficie (MDS) a diferencia de un MDE, representa las elevaciones sobre las superficies de edificios, árboles, torres y otros elementos elevados sobre terreno (suelo desnudo).

Los MDE, han llegado a ser la principal fuente de información en topografía (Florinsky, 2012), lo que además les hace diferentes de otros conjuntos de datos geográficos por las siguientes razones (Fisher y Tate, 2006):

- Fueron una de las primeras formas de información digital geográfica que estuvieron disponibles.
- Son ampliamente usados.
- Están muy asociados con los conceptos matemáticos del modelamiento de superficies; y,
- Representan un fenómeno tangible y directamente observable, del cual todas las personas tienen experiencia directa (la superficie del terreno).

2.1.3 Fuentes de obtención y modelos de datos de los MDE

Los MDE pueden ser generados por medio de diversas tecnologías (topografía clásica, fotogrametría, LiDAR, sistemas RADAR, entre otras) y métodos (estereoscopia, estereo correlación, medición de rango, entre otras) y la modelación de superficies topográficas que pueden ser logradas por medio de diferentes opciones como: curvas de nivel, red de triángulos irregulares, superficies ráster, superficies paramétricas y más (Ariza et al., 2017).

Las fuentes más comunes de obtención de los MDE han sido clasificados por Nelson, Reuter y Gessler (2009) en tres grupos:

Fuentes primarias

- *Técnicas de levantamiento del terreno* que constituyen los levantamientos de alta precisión, obtenidos por métodos clásicos (nivelación, topografía, GNSS de alta precisión, entre otros), se caracterizan por su elevado costo ya que requieren de equipos de topografía dedicados, y éstos a menudo superan los recursos económicos de proyectos pequeños.
- *Sensores remotos* las plataformas de los sensores remotos pueden ser espaciales(satélites) o aerotransportadas (sistemas tipulados o UAV). Las imágenes resultantes pueden ser derivadas de tres tipos de fuentes: fotografía aérea, LiDAR y RADAR, la ventaja de éstas es que permite cubrir rápidamente grandes áreas.

Fuentes secundarias

- *Mapas topográficos existentes* información que proviene de las agencias cartográficas oficiales de la que se derivan curvas de nivel, ríos, puntos de elevación, entre otros. Por lo general, es información que puede estar en formato digital o es necesaria una digitalización aunque esta última opción ya es la menos usada.

- *Fotogrametría* uso de fotografías aéreas a través de la medición de pares estereoscópicos con métodos convencionales y digitales, para obtener ya sean curvas de nivel o puntos de elevación.

Modelos de datos de los MDE

De acuerdo con Felicísimo (1994), los conjuntos de datos de elevaciones que forman un MDE se pueden organizar dentro de dos modelos de datos geográficos básicos: vectorial y ráster

En los modelos vectoriales los atributos del terreno se representan mediante puntos, líneas y polígonos. Las estructuras de datos más comunes en este grupo son:

- *Curvas de nivel*, son probablemente la representación de la superficie del terreno más conocida, son líneas que unen puntos de elevación constante, uno de los inconvenientes de éstas es que solo indican los valores de la superficie a lo largo de las isolíneas y suelen utilizarse exclusivamente como paso intermedio entre la representación cartográfica y otras estructuras diferentes. La exactitud de las curvas de nivel depende de los datos de elevación y del método de interpolación.
- *La Red de Triángulos Irregular (TIN)*, por sus siglas en inglés, es la estructura de datos más utilizada en el campo de la hidrología. En un modelo TIN, los puntos de muestreo están conectados por arcos para formar triángulos y dentro de cada triángulo la superficie suele estar representada por un plano. El uso de triángulos permite que cada uno de ellos, que representa un pedazo de superficie, se ajuste a su vecino conformando así un mosaico de triángulos que cubren la superficie total, por lo que será continua, ya que la superficie de cada triángulo está definida por las elevaciones de los tres nodos de esquina. Los polígonos que son más complejos también pueden ser utilizados como mosaicos, siempre se puedan dividir en triángulos (El-Sheimy, Valeo y Habib, 2005). Los modelos TIN permiten la incorporación de datos adicionales que suponen una mejora del modelado es el caso de las líneas estructurales. Sin embargo, tienen el inconveniente del cálculo de las relaciones de vecindad que consumirá tiempo computacional (Arranz, 2013).

En los modelos ráster, los datos de elevación se interpretan como un mosaico formado por celdas donde a cada celda se le asigna el valor de la variable (elevación) en la de superficie que representa. La estructura ráster más común es la malla de celda cuadrada o matriz regular donde las celdas son cuadradas y de tamaño uniforme (Cuartero Sáenz, 2003). Es una estructura matricial donde se almacenan las alturas del terreno e implícitamente las relaciones topológicas entre los datos de puntos, por lo que, los algoritmos de modelado tienden a ser relativamente sencillos. Las mallas suelen ser generadas por interpolación, de ahí que para representar la complejidad del terreno, es necesario una gran cantidad de puntos que estén distribuidos por toda la superficie.

En la presente tesis nos centraremos en el estudio de los MDE de estructura de modelos ráster.

2.1.4 Superficies derivadas del los MDE

Las principales superficies derivadas de los MDE son:

1. Pendiente
2. Orientación
3. Curvatura
4. Visibilidad
5. Sombreado

Pendiente La pendiente es usualmente medida en grados o porcentaje y es definida como el grado de cambio en elevación δZ sobre un cambio en extensión lateral (El-Sheimy, Valeo y Habib, 2005).

El grado de cambio en elevación, tanto en la dirección x , como en y puede ser usado para identificar la dirección y magnitud del gradiente de mayor inclinación. En tal virtud, la magnitud de la pendiente se la calcula en base a la ecuación 2.1.

$$Slope = \frac{\delta Z}{\delta XY} = \sqrt{\left(\frac{\delta Z}{\delta X}\right)^2 + \left(\frac{\delta Z}{\delta Y}\right)^2} \quad (2.1)$$

El concepto de búsqueda de máxima inclinación, puede aplicarse de celda a celda en una ráster o entre triángulos en un TIN (Maune, 2001).

La elevación cambiará en las direcciones x e y , y por tanto la pendiente resultante de un cambio en extensión lateral es la suma vectorial de la pendiente en dichas direcciones, de acuerdo a lo mostrado en la Figura 2.1.

Figura 2.1: Definición de pendiente. Tomado de El-Sheimy, Valeo y Habib (2005)

Finalmente, nótese en la ecuación 2.2 que:

$$Slope = \sqrt{\left(\frac{\delta f}{\delta X}\right)^2 + \left(\frac{\delta f}{\delta Y}\right)^2} = \sqrt{\left(\frac{\Delta Z_x}{\Delta X}\right)^2 + \left(\frac{\Delta Z_y}{\Delta Y}\right)^2} \quad (2.2)$$

Con respecto a un conjunto de celdas, como en el caso del MDE, el valor de la pendiente resultante asignada a cada celda reflejará la pendiente total basada en la relación entre la celda y sus vecinos. Hay numerosas maneras en que la pendiente puede ser calculada cuando se trabajan con celdas, por ejemplo se pueden usar cuatro celdas vecinas en las orientaciones: norte, sur, este y oeste.

Orientación El valor asignado a cada celda en un mapa de orientación indica la dirección (norte, sur, etc.) en la cual esa celda está orientada. La ecuación 2.3 describe el cálculo de la dirección.

$$\tan\alpha = \frac{\delta f}{\delta X} / \frac{\delta f}{\delta Y} \quad (2.3)$$

De donde α es calculado como ángulo en el sentido de las manecillas del reloj, desde el norte (Véase Figura 2.1).

Curvatura La curvatura está definida por la segunda derivada de la superficie del terreno (o primera derivada de la pendiente) y describe la proporción del cambio de la pendiente (El-Sheimy, Valeo y Habib, 2005). Es usada para determinar las características físicas del terreno (morfología).

La ecuación 2.4 presenta la fórmula de cálculo de la curvatura.

$$Curvature = \sqrt{\left(\frac{\delta^2 f}{\delta X^2}\right)^2 + \left(\frac{\delta^2 f}{\delta Y^2}\right)^2} \quad (2.4)$$

Valores de curvatura negativa indican que el terreno es cóncavo hacia arriba. Valores de curvatura cero se deben a celdas planas y valores positivos indican que la superficie es convexa.

Visibilidad Un mapa de visibilidad presenta a las áreas visibles posibles dentro de la superficie cubierta por el MDE desde uno o más puntos de observación.

Sombreado Define la hipotética iluminación del terreno, ingresando parámetros para la caracterización de posición del sol, como son: el acimut y la altitud (Olaya, 1994), proveyendo un efecto visual de tres dimensiones sobre el área de estudio en una perspectiva de dos dimensiones.

2.1.5 Principales aplicaciones de los MDE

Existe una importante variedad de aplicaciones de los MDE, a partir de los atributos primarios derivados de los mismos, de las cuales se presentan algunas de ellas resumidas en la Tabla 2.1.

Atributo	Aplicación
Altura	Comportamiento de variables climáticas (Temperatura, presión) Vegetación y patrones de suelos Determinación de potencial energético
Pendiente	Flujo superficial y sub-superficial Erosión
Orientación	Irradiancia solar Evapotranspiración
Curvatura	Aceleración de flujo Patrones y tasas de erosión y deposición Índices de evaluación de suelo y tierra
Longitud, área y pendiente de cuencas de drenaje	Tiempo de concentración Volúmenes de contribución a escorrentía Tiempos y velocidades de atenuación de escorrentía
Longitudes y pendientes de drenajes	Tasas y velocidades de canales de flujo Tasas de erosión y producción de sedimentos

Tabla 2.1: Atributos primarios derivados de los MDE y sus aplicaciones. Tomado de El-Sheimy, Valeo y Habib (2005)

Las aplicaciones de los MDE son virtualmente ilimitadas (Maune, 2001). Cada decisión de aplicación debe ser hecha con datos de elevación apropiados. Esto dependerá de muchos factores incluyendo el costo, resolución y exactitud de los datos (Fleming, Giles y Marsh, 2010).

En la actualidad, las técnicas y los métodos de obtención de los MDE se han ido mejorando y perfeccionando de tal forma que están al alcance de los usuarios a través de plataformas de visualización y descarga, dispuestas por los principales organismos proveedores de información espacial, como por ejemplo en las IDE, que juegan un rol importante como soporte del desarrollo sustentable de los gobiernos (Crompvoets et al., 2008). Sin embargo, en la mayoría de los casos éstos no proveen de estándares de exactitud vertical ajustados a las necesidades de los usuarios debido

a los diferentes criterios de exactitud de las diversas aplicaciones de los MDE (Liu, Hu y Hu, 2015).

En el caso de la presente tesis, para identificar las aplicaciones donde los MDE juegan un rol fundamental fue necesario conocer sobre los perfiles de usuarios y usos de los MDE, para lo cual, durante la investigación se realizó:

- Una caracterización de los usuarios y usos de los MDE a través de los resultados de la encuesta que se realiza en el formulario opcional de descargas del Centro Nacional de Información Geográfica (CNIG) del Instituto Geográfico Nacional de España, como resultado se obtuvo una publicación en la revista GeoFocus.
- Una encuesta a usuarios especializados en el uso de MDE, con el propósito de acercarse al tratamiento de la calidad de los MDE desde un enfoque de la aptitud para el uso y por ende una mejor comprensión de los usuarios. La encuesta en línea fue realizada en el año 2016, estuvo dirigida a profesionales relacionados con los campos de aplicación de los MDE, obteniendo respuestas de Europa (60,8 %), América (25,5 %) y Asia (13,7 %).

Los resultados son mostrados en la Sección 4.1. Dicha información fue necesaria para la selección de caso de uso que se emplea en la presente tesis.

2.2 Teoría del error e incertidumbre en los MDE y herramientas

2.2.1 Definición de error e incertidumbre

En la literatura técnica y científica, el término error se utiliza frecuentemente con dos significados bastante diferentes. En algunos casos se utiliza para cuantificar la diferencia entre el resultado de una medida y el considerado como valor de la misma (conocido como valor verdadero, valor real o estándar), mientras que en otras referencias se utiliza para denominar la incertidumbre del resultado de una medida, es decir, para cuantificar la imperfección del método e instrumento de medida empleado (Ruiz Armenteros, García Balboa y Mesa Mingorance, 2010).

El CEM (2008) define el término error de medida como la diferencia entre un valor medido de una magnitud y un valor de referencia (valor convencional o valor verdadero), mientras que define la incertidumbre de medida como un parámetro no negativo que caracteriza la dispersión de los valores atribuidos a un mensurando, a partir de la información que se utiliza.

El error puede ser definido como la diferencia entre la realidad y una representación de la realidad. En la práctica, los errores no son exactamente conocidos. En el mejor de los casos, los usuarios tienen alguna idea acerca de la distribución de valores que el error probablemente tome (Shekhar y Xiong, 2007).

Por otro lado, la incertidumbre se deriva de un conocimiento de la existencia de errores y su magnitud, por tanto, es proporcional a cuánto se conoce sobre ellos. Los términos error e incertidumbre no son sinónimos y no deberían usarse indistintamente, y si se lo hace advertir sobre esta situación.

2.2.2 Tipología de los errores en los MDE

Los errores en los MDE pueden ocurrir tanto en la elevación (Z) como en las coordenadas planimétricas (XY), pero está usualmente en el primero porque el error horizontal producirá error en la vertical (Fisher y Tate, 2006).

El error de elevación ϵ está definido por la ecuación 2.5.

$$\epsilon = |Z - z| \quad (2.5)$$

De donde:

Z es la altura estimada y z es la altura verdadera.

Típicamente los errores en los MDE son categorizados en tres grupos (Rieger, 1996; Cooper, 1998; Wise, 2000): errores groseros o equivocaciones (altitudes incorrectas en determinadas zonas), errores sistemáticos debido a sesgo determinista en la colección de datos o procesamiento, y errores aleatorios (Fisher y Tate, 2006).

La evidencia empírica demuestra que el error en un MDE es espacialmente variable, espacialmente correlacionado y heteroscedástico (Carlisle, 2005). Sin embargo varios

autores observan que los errores pueden ser: no aleatorios, a menudo no siguen una distribución normal (Maune, 2001), no son idénticos y no están distribuidos estacionariamente (Fisher, 1998; Liu, Hu y Hu, 2015).

La afirmación de que los errores de un MDE son espacialmente autocorrelados, quiere decir que los valores de error tienden a ser similares en ubicaciones cercanas y esta situación puede tener un considerable impacto en la salida de análisis espaciales que usan un MDE como entrada (Leon, Heuvelink y Phinn, 2014). Es conocido que la autocorrelación espacial¹ de los MDE de gran resolución/detalle, es el resultado de combinaciones complejas de componentes aleatorios y sistemáticos (Liu et al., 2012).

Debido a las características que tienen los errores en los MDE, éstos pueden presentarse como una combinación compleja de errores aleatorios y casi sistemáticos (Oksanen y Sarjakoski, 2006) o locamente sistemáticos, en consecuencia Podobnikar (2016a) propone la siguiente clasificación:

- **Errores aleatorios.**

No pueden ser eliminados, pueden ser errores verticales y planimétricos, son representados conceptualmente como variaciones al azar alrededor del valor real referido, pueden ser espacialmente correlados o no (Fisher y Tate, 2006).

- **Errores localmente sistemáticos.** Pueden ser parcialmente eliminados y únicamente inherentes a errores de la vertical, son muy complejos y los más típicos en los MDE, pero son generalmente considerados como aleatorios. Sin embargo son una combinación de sistemáticos y aleatorios. Se los conoce también como errores de mediana frecuencia, causados por errores en las mediciones fotogramétricas y procedimientos de interpolación (Florinsky, 2012). Son espacialmente correlados en el contexto de cierta área de vecindad. Son topográficos y semánticos con formas y patrones específicos, por lo que en este contexto el MDE debe ser considerado como una superficie y no como puntos independientes.

- **Errores sistemáticos.** Pueden ser eliminados, verticales y planimétricos. Son conocidos también como errores de baja frecuencia, resultado de un «sistema

---

¹En el caso de la presente tesis utilizaremos el término autocorrelación espacial, conocido también como dependencia espacial, asociación espacial o interdependencia espacial, por ser más usado.

determinista que, si se conoce, puede estar representado por alguna relación funcional», éstos incluyen errores debido a una pobre selección de puntos de control (López, 2002; Florinsky, 2012).

- **Errores groseros.** Generalmente removibles manual o automáticamente, pueden ser errores verticales y planimétricos, son fácilmente reconocibles, proceden de un filtrado inadecuado en el caso de una nube de puntos LiDAR o del método de interpolación, así como de una mala interpretación en el caso de estéreo-fotografía o digitalización manual.

2.3 Calidad de la Información Geográfica y de los Modelos Digitales de Elevación

En este apartado se hace un resumen de las principales definiciones, normas y estándares relativos a la calidad y usados en el contexto de la IG y los MDE.

2.3.1 Calidad de la Información Geográfica

La IG es la base para numerosos procesos de toma de decisiones sobre el territorio. Estas decisiones afectan a la vida, economía y derechos de los individuos y colectividades, y por ello se debe usar y exigir IG de calidad contrastada (Ariza-López, 2011).

La Norma ISO 9000:2015 define que la calidad *está determinada por la capacidad para satisfacer a los clientes, y por el impacto previsto y el no previsto sobre las partes interesadas pertinentes. La calidad de los productos y servicios incluye no sólo su función y desempeño previstos, sino también su valor percibido y el beneficio para el cliente.*

Es decir, en esta norma el cliente es parte de los grupos de interés (partes interesadas pertinentes) o “*stakeholders*”, por tanto ya no solo la calidad busca satisfacer los requerimientos del cliente, sino que los resultados de la calidad sean equilibrados en todos sus grupos de interés, para lo cual será necesario establecer claramente quiénes

son, qué requisitos, necesidades y expectativas tienen, así como el valor y beneficio para el cliente.

Así el enfoque de la aptitud para el uso (*fitness for use*) es vital ya que debería entenderse el propósito para la definición y diseño de las métricas empleadas en la calidad (Floridi e Illari, 2014) para los diferentes usos; ya que un mismo producto puede ser muy utilizado por un grupo de usuarios pero inadecuado para otros (Goodchild, Devillers y Jeansoulin, 2006). Dado que pueden existir múltiples usos de un mismo producto, existirán tantas adecuaciones como posibles usos, y tantas apreciaciones de la calidad (Ariza-López et al., 2013).

La visión de la calidad en los datos espaciales surgió de la academia durante los años ochenta y principios de los noventa. Su enfoque se ha diversificado a lo largo de los años (Devillers et al., 2010) desde un enfoque del control posicional únicamente orientado al producto, hasta la consideración de la metacalidad y la usabilidad como elementos de la calidad. En este contexto, la calidad de los datos espaciales comprende dos perspectivas (Mostafavi, Edwards y Jeansoulin, 2004; Stein, Shi y Bijker, 2009):

- **Calidad interna.** Perspectiva del productor donde la calidad de los datos espaciales está determinada por la exactitud posicional, temática y temporal, la genealogía, la coherencia lógica y la integridad de los datos.
- **Calidad externa.** Perspectiva del usuario cuya preocupación es la “aptitud para el uso” definido en términos de accesibilidad, relevancia, integridad, actualidad, interpretación, facilidad de comprensión y costo.

Para conseguir que la calidad de la IG sea la adecuada para su propósito hace falta una perspectiva general tanto sobre la calidad, como sobre el producto y las expectativas del usuario o cliente (Ariza-López et al., 2013). Por tanto, la calidad de la IG es un aspecto importante tanto para productores como para los usuarios, de cara a los retos de futuro como la evolución de la geomática, el internet de la cosas, el territorio inteligente, entre otros (Ariza-López, 2015).

En el ámbito de la investigación de la calidad de la IG en los últimos 30 años, uno de los logros más destacados es que la calidad de los datos espaciales se considera una subdisciplina de las geociencias, actualmente se cuenta con grupos especializados

en la temática que desarrollan eventos dedicados no solo a la investigación, sino a la aplicación de estándares en lo referente a la información geográfica. Además la calidad de datos espaciales cuenta con un enfoque internacional, pues se ha logrado establecer estándares dentro de la familia ISO 19100 (grupo de normas específicas para el sector de la IG), lo que ha permitido influir en los gobiernos y agencias cartográficas sobre la importancia de la calidad para los usuarios de la IG (Devillers et al., 2010).

2.3.2 Normas y Estándares de Calidad para la Información Geográfica

La tendencia de crecimiento continuo de generación y uso de la IG requiere especial atención en cuanto a la normalización y calidad, como claves de éxito (Ariza-López et al., 2013) al momento de la toma de decisiones basadas en esta.

Normalización y calidad van siempre de la mano (Ariza-López y Rodríguez-Pascual, 2007), la normalización juega un papel fundamental pues facilita la medición de la calidad, la armonización de los datos y es la base de la interoperabilidad. De acuerdo a la Directiva Europea INSPIRE (2007/02/CE), desde el punto de vista de una IDE, la interoperabilidad tiene como objetivo fijar un modelo para la aplicación de los elementos de la calidad como están definidos en los estándares (INSPIRE, 2012) y la pobre calidad puede comprometer la interoperabilidad.

Desde la perspectiva de la IG las primeras iniciativas de normalización provienen de aquellos países y sectores líderes en el uso de la IG (Ariza-López et al., 2013), las mismas que se han ido adaptando a las necesidades y al desarrollo de la geotecnologías. En la Tabla 2.2 se muestra el número de normas desarrolladas por las organizaciones más representativas en el ámbito de la IG, donde destaca el FGDC del Departamento de Defensa de los Estados Unidos de América, el cual ha sido el principal productor de normativas después de la ISO, lo que justifica su recorrido histórico durante varias décadas.

En la actualidad la principal iniciativa en normalización de la IG es desarrollada por el Comité Técnico (TC) 211 de la Organización Internacional de Normalización (ISO) y el *Open Geospatial Consortium* (OGC). Las normas ISO se formulan en colaboración con los organismos nacionales de normalización, mientras que los

Organización	Número de casos	%
FGDC	40	20.3
ISO	55	27.9
MIL	83	42.1
OGC	19	9.6
Total	197	100

Tabla 2.2: Número de normas desarrolladas relativas a la IG por cada institución. Tomado de Ariza-López (2011)

estándares OGC se crean con el apoyo de usuarios y proveedores de tecnología. Ambas organizaciones acumulan conocimientos sobre las mejores experiencias internacionales lo cual facilita su difusión (INSPIRE, 2012).

En el caso de la familia ISO 19100 la norma que aborda ésta temática de manera específica es ISO 19157:2013, la cual ha reemplazado las siguientes normas:

- ISO 19113 Principios de la calidad, establece los principios para describir la calidad de un conjunto de datos espaciales e informar sobre la misma.
- ISO 19114 Procedimientos de evaluación de la calidad, establece un marco para evaluar de la calidad y para informar sobre esa evaluación.
- ISO 19138 Medidas de la calidad, normaliza, identifica y define las medidas de la calidad a aplicar a la IG, es decir, utiliza un conjunto de medidas que siendo conocidas puedan ser bien y fácilmente interpretadas (e incluso pactadas) por las partes (productores y usuarios).

Las normas ISO 19113, 19114 y 19138 han sido adoptadas desde su aparición por muchos proveedores de datos geospaciales (Yang et al., 2013), en la actualidad la norma ISO 19157 integra éstas normas en un solo cuerpo normativo, que incluye elementos y conceptos que no fueron considerados como el caso de la usabilidad, la metacalidad, la unidad de calidad de datos (DQU), el ámbito de la calidad entre otros cambios.

Así, en la norma ISO 19157:2013 (AENOR, 2014) se:

- define las componentes para describir la calidad de datos;

- especifica las componentes y una estructura de contenidos para un registro de medidas de la calidad de datos;
- describe los procedimientos generales para evaluar la calidad de datos;
- establece principios para la presentación de informes de calidad de datos.

Los elementos de la calidad en la norma han sido clasificados en las siguientes categorías como se muestra en la Tabla 2.3.

Categorías	Elementos
Compleción Presencia o ausencia de objetos geográficos, sus atributos y relaciones	* Comisión * Omisión
Consistencia lógica Grado de adherencia a las reglas lógicas de la estructura de los datos, atributos y relaciones	* Conceptual * De dominio * De formato * Topológica
Exactitud posicional Exactitud de la posición de los objetos geográficos, en un determinado sistema de referencia espacial	* Exactitud absoluta o externa * Exactitud relativa o interna * Exactitud posicional de datos en malla
Exactitud temática Exactitud de atributos y clasificación	* Corrección de la clasificación * Corrección de atributos no cuantitativos * Exactitud de atributos cuantitativos
Calidad temporal Calidad de los atributos y de las relaciones temporales de los objetos geográficos	* Exactitud de una medida de tiempo * Consistencia temporal * Validez temporal
Usabilidad	
Se basa en los requerimientos de usuario (Todos los elementos de la calidad u otros pueden usarse)	

Tabla 2.3: Resumen de los elementos de la calidad de la IG. Modificado de AENOR (2014)

Otras normas ISO relacionadas a la calidad de la IG son:

Información Geográfica, familia ISO 19100

ISO 19115-1:2014 *Metadatos - Parte 1: Fundamentos* Proporciona un modelo y establece un conjunto común de terminologías, definiciones y procedimientos de aplicación para describir información o recursos que pueden tener una extensión geográfica.

ISO 19115-2:2009 *Metadatos - Parte 2: Extensión para imágenes y dato de malla* Permite la implementación de los metadatos de IG estructurada en forma de imágenes y mallas.

Las clases y atributos de ambas normativas se organizan en tres tipos: obligatorios, opcionales y condicionales. Los obligatorios deben aparecer siempre en todos los metadatos creados y en los perfiles y extensiones derivadas de estas normas. Los opcionales permiten una descripción normalizada más amplia de los recursos, si así se requiere. Los condicionales aparecerán según ciertos criterios indicados en la norma, convirtiéndose, de esta forma en obligatorios (Ureña Cámara, 2013).

Los metadatos en la IG son fundamentales para la toma de decisiones de la adquisición de los productos desde el punto de vista del usuario, ya que permiten seleccionar el producto que más se acomoda las necesidades de la aplicación donde serán usados. Por otra parte, desde el punto de vista del productor permiten advertir las limitaciones que podría tener el producto. Por tanto éstos también deben ser evaluados.

En el caso de la calidad de la IG, ésta se define en la Norma ISO 19157 y los metadatos de la calidad de los datos son comunicados mediante la ISO 19115-1 e ISO 19115-2.

ISO 19131:2007 *Especificaciones de producto de datos* El objetivo de la norma es establecer los requisitos que han de cumplir las especificaciones de producto en el caso de los datos geográficos. Así la norma describe los requisitos, contenido y estructura de como debería ser el producto. También proporciona ayuda en la creación de otras especificaciones de producto de datos, a fin de que puedan entenderse con facilidad y cumplan con el fin deseado.

Las especificaciones técnicas del producto son útiles para productores y usuarios, en el primer caso éstas pueden ser fijadas para describir los productos y en el caso de los usuarios para establecer los requisitos. Es importante tener clara la diferencia entre las especificaciones, que indican cómo debería ser el producto, y los metadatos, que reflejan cómo es realmente. Si un usuario quisiera evaluar si un producto es idóneo para un determinado fin, tendría que fijar su atención en los metadatos, antes que las especificaciones (García Balboa, 2013), ya que éstas indican los aspectos relevantes que deben evaluarse

para comprobar que se han alcanzado los niveles de calidad preestablecidos, mientras que los metadatos comunican los resultados de la evaluación de la calidad establecidos por la ISO 19157.

ISO/TS 19157-2:2016 *Calidad de los datos - Part 2: XML Esquema de Implementación de ISO 19157* Define la codificación de calidad de datos geográficos en XML. Es una implementación del esquema XML derivada de ISO 19157:2013 y los conceptos relacionados con la calidad de datos de ISO 19115-2. Este esquema puede ser utilizado para validar la conformidad de documentos XML con estos modelos conceptuales y realizar de forma estandarizada el informe de calidad independiente, pero hasta el momento no existe experiencia en su aplicación.

ISO/TS 19158:2012 *Aseguramiento de la calidad en el suministro de datos* Proporciona un marco de trabajo que permite a un cliente evaluar si sus proveedores, internos o externos, son capaces de proporcionarle información geográfica con la calidad que necesita. Por lo tanto, sirve para asegurar la calidad en la relación de producción entre el productor y el cliente.

También, identifica métodos de gestión de la calidad de la producción más eficientes y efectivos, permite la innovación y la mejora continua dentro del contexto de las normas existentes. La norma está basada en ISO 19157 e ISO 9000, no es aplicable en productos preexistentes donde no hay producción o actualización de datos que gestionar. Además, proporciona los principios y responsabilidades de la relación entre el cliente y el proveedor. La responsabilidad de la acreditación del procedimiento se comparte entre el cliente y el proveedor (IPGH, 2013).

Muestreo para la inspección

ISO 2859 *Procedimientos de muestreo para inspección de atributos*

ISO 3951 *Procedimientos de muestreo para inspección de variables*

En la norma ISO 19157 se establece que para los métodos de evaluación internos o externos pueden usarse métodos de inspección completa o muestreo, en este caso las normas ISO 2859 e ISO 3951, establecen los denominados planes de muestreo para la aceptación. Se trata de un conjunto de normas

de gran solidez, ampliamente utilizadas por la industria desde hace décadas y que permiten definir niveles de calidad y procesos de evaluación estándar. El concepto de Nivel de Calidad Aceptable desarrollado en estas normas es de plena aplicación al ámbito geomático así como los métodos de evaluación por técnicas de muestreo.

ISO 2859 hace referencia a los procedimientos de muestreo para la inspección por atributos, entendiéndose como tales a las características de interés de los elementos cuya consideración es desde una perspectiva cualitativa de que cumple o no cumple con una especificación, por ejemplo el cumplimiento de una regla topológica; mientras que ISO 3951 lo hace para la inspección por variables, considerando a éstas como características susceptibles de ser medidas sobre una escala continua, por ejemplo, un error cuya dimensión ha sido mensurada. (Ariza-López et al., 2013).

En la Figura 2.2 se muestra la relación entre los procesos de ISO 19157 y normas relativas a la calidad de la IG.

Figura 2.2: Relación entre los procesos de ISO 19157 y normas relativas a la calidad de la IG. Tomado de Ariza-López y Rodríguez-Pascual (2007)

2.3.3 Calidad de los Modelos Digitales de Elevación

Una importante cantidad de aplicaciones usan los MDE para inferir propiedades topográficas, por lo tanto la calidad de los MDE puede tener muchas implicaciones en las actividades desarrolladas en el territorio basadas en el análisis de estos (Liu et al., 2012). Sin embargo, muchos usuarios finales de MDE no son conscientes de los problemas relacionados con la calidad de los datos de elevación y su influencia en los cálculos de los subproductos derivados (Aguilar et al., 2010).

La calidad de los MDE puede ser abordada a partir de la evaluación de métodos cuantitativos y cualitativos, donde los primeros han sido los más clásicos y desarrollados por los principales organismos de estandarización cartográfica como: *Federal Geographic Data Committee* (FGDC), *Federal Emergency Management Agency* (FEMA), *National Digital Elevation Program* (NDEP) y *American Society for Photogrammetry and Remote Sensing* (ASPRS), que por lo general utilizan el Raíz del Error Medio Cuadrático (RMSE) (por sus siglas en inglés), como medida para reportar la calidad en los MDE.

El RMSE implica suposiciones sobre la naturaleza del error en las elevaciones, por lo que, la calidad no siempre está de acuerdo con las expectativas de los usuarios y sus aplicaciones. Sin embargo, los efectos del error no son frecuentemente evaluados por los usuarios y los métodos tampoco han sido integrados dentro los software SIG (Wechsler y Kroll, 2006).

Por otro lado, Podobnikar (2016b) considera que la cuantificación de los errores e incertidumbres necesita mejores soluciones que un simple RMSE para reportar la calidad de los MDE, pudiendo ser analizada desde enfoques complementarios: estadístico, empírico y visual.

- **Estadístico.** El análisis las características intrínsecas en un MDE; investigaciones han determinado que los errores de un MDE en el campo de la incertidumbre tienen un patrón específico en términos de geomorfología, densidad de muestreo, escala/resolución, método de generalización, interpolación, entre otros.
- **Empírico.** El análisis de otros MDE de mayor exactitud y la comparación con el MDE en estudio, usando puntos de referencia.

- **Visual.** Este método requiere un mayor nivel de adaptación a problemas particulares de la superficie del terreno que un solo estadístico. Depende de la experiencia del operador, para detectar valores extremos y el análisis puede estar acompañado de los criterios geomorfológicos/topográficos que consideran la forma y la semántica, por ejemplo considerando información contextual alrededor de cada punto del MDE.

La presente tesis se enfoca en los métodos cuantitativos, específicamente en los geoestadísticos para la definición y posterior alteración de los errores en los MDE, con el propósito de estimar los costos de la mala calidad derivada de los errores en la componente vertical en los MDE evaluados. Cabe indicar que no se evaluarán los errores producidos por los métodos de generación de los MDE, o errores debido a las fuentes de adquisición, sino que se considerarán los errores respecto a la comparación de productos de tres veces mayor exactitud como referencia de la realidad de la superficie analizada.

2.3.4 Evaluación de calidad en los MDE

La evaluación de calidad de los MDE, ha sido propuesta desde varios enfoques (Papasaika y Emmanuel, 2010), en función de las características de las fuentes de generación de esto y según el propósito de dicha evaluación ya sea controlar o estimar la calidad. La gran mayoría de estándares de calidad de los MDE están basados en el cálculo de la exactitud vertical (Maune, 2001), la cual ha sido un tópico ampliamente tratado desde la década de los 80 (Carlisle, 2005).

La medida de incertidumbre más empleada es el RMSE (Papasaika y Emmanuel, 2010; Liu et al., 2012; Podobnikar, 2016a) cuya fórmula se presenta en la ecuación 2.6, sin embargo ésta es una medida global simple, lo cual significa que el error que se asume es el mismo en cualquier parte del terreno cubierto por el MDE (Papasaika y Emmanuel, 2010), a menudo ésta estadística global no refleja la variabilidad del error de la superficie modelada (Coveney y Fotheringham, 2011).

El RMSE mide la dispersión de la distribución de desviaciones entre los datos de elevación tomados como referencia y los datos de elevación medidos en el producto evaluado. Es un índice sencillo y da un valor global de las desviaciones, pero no tiene en cuenta la variación espacial de los errores sobre la superficie (Weng, 2002) por

lo que, ésta medida simplemente resume los errores de elevación en un solo valor (Carlisle, 2005; Wechsler y Kroll, 2006).

$$RMSE_z = \sqrt{\frac{1}{n} \sum_{i=1}^n (Z_i - z_i)^2} \quad (2.6)$$

Donde:

Z_i es el i -ésimo valor de elevación medido (o evaluado) en el MDE,

z_i es la elevación de referencia (o punto de control) correspondiente; y,

n es el número de puntos de elevación revisados.

Razón por la cual, varios autores reconocen que el RMSE tiene sus limitaciones, ya que no logra captar las numerosas fuentes de error, no permite la identificación de áreas donde el error es mayor y áreas donde datos de fuentes adicionales podrían beneficiar la calidad del MDE, o predecir su efecto sobre el resultado de su uso (Wise, 2000; Carlisle, 2005; Devillers y Goodchild, 2010; Liu, Hu y Hu, 2015; Podobnikar, 2016b).

Por su parte, el percentil del 95 %, como una medida para expresar los resultados de la evaluación de la calidad, es recomendada en el caso de los MDE procedentes de datos LiDAR, debido a que la distribución estadística de los errores no es normal en zonas provistas de cobertura vegetal (Ariza-López et al., 2013; ASPRS, 2014; Heidemann, 2014). Los cálculos del percentil están basados en los valores absolutos del error, es decir en su magnitud; el signo no es tomado en cuenta.

Para la determinación del percentil del 95 % primero se calcula la posición del percentil 95 (n) usando la ecuación 2.7.

$$n = \left(\left(\left(\frac{P}{100} \right) * (N - 1) \right) + 1 \right) \quad (2.7)$$

Donde: n es la posición de la observación que contiene el percentil 95 (puede obtenerse un número decimal), P es la proporción (de 100) en el cual el percentil es deseado (95); y, N es el número de observaciones de la muestra.

Una vez que la posición de la observación 95 es determinada, el percentil (Q_p) puede ser interpolado desde las observaciones inmediatamente más alta y baja, usando la ecuación 2.8.

$$Q_p = (A[n_w] + (n_d * (A[n_w + 1] - A[n_w]))) \quad (2.8)$$

De donde: Q_p es el percentil 95, A es un vector de valores absolutos de la muestra, indexada en orden ascendente desde 1 a N , $A[i]$ es el valor del vector A en el índice i de la muestra, n es la posición de la observación que contiene el percentil 95, n_w es la parte entera de n ; y, n_d es la parte decimal de n (ASPRS, 2014).

En la Tabla 2.4 se resumen los estándares conocidos relacionados con la evaluación de la exactitud vertical en los MDE.

Un inconveniente con el uso del RMSE o del percentil 95, es que el productor por lo general provee un solo valor de exactitud de la evaluación global de la superficie, enmascarando la variación espacial de la calidad el MDE (Liu, Hu y Hu, 2015). Como se indicó anteriormente no siempre los errores cumplen con las condiciones para ser tratados con métodos estadísticos clásicos.

Además, la efectividad el RMSE y otros métodos estadísticos puede verse comprometidos (Liu, Hu y Hu, 2015), debido al número de puntos de control y al tipo de geoforma presente en el terreno (Hengl, Heuvelink y Loon, 2010).

En el caso de los estándares ISO, que son muy apreciados por la comunidad geoespacial, Podobnikar (2016a) considera que éstos son muy estáticos y genéricos en términos prácticos para las necesidades y entendimiento de los usuarios respecto a la calidad de los MDE.

Una descripción más detallada de la exactitud de los MDE permitiría un mejor entendimiento de su calidad y consecuente incertidumbre en las aplicaciones utilizadas (Carlisle, 2005).

Desde la perspectiva cuantitativa, en la literatura se puede encontrar trabajos donde se han explorado otros métodos como:

- **Quantiles.** Basado en métodos estadísticos robustos para evaluar el error vertical, donde la mediana, mediana normalizada, desviación absoluta y

Estándar	Organismo	Año	Alcance
National Map Accuracy Standard (NMAS)	Bureau of the Budget	1947	Reporta la exactitud vertical de curvas al 90% de intervalo de confianza. Más del 90% de los puntos de evaluación de elevación ("bien definidos") deben estar dentro de la precisión correspondiente a la mitad del intervalo de curva de nivel publicado.
National Standard for Spatial Data Accuracy (NSSDA)	Federal Geographic Data Committee (FGDC)	1998	Reporta la exactitud vertical de datos de elevación digital al 95% de intervalo de confianza. La exactitud vertical debería ser evaluada comparando elevaciones en un conjunto de datos con elevaciones de los mismos puntos de una fuente independiente de mayor precisión. El estadístico RMSE es solo válido cuando los errores siguen una distribución normal ($Z\text{-value} = 1.96$ en este caso). Se asume que los errores sistemáticos han sido eliminados.
Guidelines and Specifications for Flood Hazard Mapping Partners	Federal Emergency Management Agency (FEMA)	2003	La evaluación de la exactitud vertical debería ser hecha y reportada para cada categoría de cobertura vegetal en el área de inundación (20 puntos de evaluación por cada categoría (al menos tres categorías)). Las pendientes del área de análisis no deben exceder el 20%.
Guidelines for Digital Elevation Data	National Digital Elevation Program (NDEP)	2004	Los puntos de evaluación determinados por una fuente independiente de mayor exactitud deberían ser al menos tres veces más exactos que los datos que están siendo evaluados. El estándar dispone el uso del concepto de Exactitud Vertical Fundamental (EVF) en las áreas desprovistas de vegetación y opcionalmente el uso de la Exactitud Vertical Suplementaria (EVS) en otras categorías de cobertura vegetal individuales; y, el uso de la Exactitud Vertical Consolidada (EVC) en todas las categorías de cobertura vegetal combinadas. EVF es calculada al nivel del confianza del 95 % del RMSE, mientras que EVS y EVC son calculadas en el percentil 95 (El 95 % de los errores de elevación son iguales o menores que el valor del percentil 95).
Positional Accuracy Standards for Digital Geospatial Data	American Society for Photogrammetry and Remote Sensing (ASPRS)	2004	Alcance similar a Guidelines for Digital Elevation Data. Existen limitaciones cuando el número de puntos de evaluación es pequeño. Cuando procedimientos estadísticos más robustos estén disponibles se adjuntarán a los Anexos de este estándar.

Tabla 2.4: Estándares de exactitud vertical establecidos por los diferentes organismos de cartografía

quantiles de muestreo son utilizados para evaluar la influencia de la no-normalidad y los valores atípicos MDE en el caso de analizar la conformidad respecto a una especificación técnica (Höhle y Höhle, 2009).

- **Regresión Espacial.** Se basa en la hipótesis de que la distribución y la escala del error de elevación dentro de un MDE están al menos parcialmente relacionadas con las características morfométricas del terreno. La técnica consiste en generar un conjunto de parámetros de terreno para caracterizar la morfometría del terreno y desarrollar modelos de regresión para definir esta relación con el MDE, lo que mejorará la representación de la incertidumbre (Carlisle, 2005).
- **Geoestadístico.** Basados en técnicas estocásticas para la determinación de la incertidumbre, considerando la autocorrelación espacial de los errores de los MDE, mediante simulaciones para la propagación del error (Hengl, Heuvelink y Loon, 2010).
- **Teoría de la aproximación.** Ésta es capaz de estimar el error de propagación e interpolación de forma separada, ofreciendo una guía efectiva a la producción y control de los MDE, asumiendo que un usuario requiere que el error vertical en cualquier punto no exceda el valor de un umbral (Liu et al., 2012).

2.4 Técnicas de generación de errores en los MDE

En la presente tesis se busca estimar las pérdidas en términos monetarios por la mala calidad de los MDE, para lo cual, mediante la experimentación con conjuntos de datos espaciales seleccionados para los ensayos y la aplicación de procedimientos geoestadísticos para introducir errores aleatorios, se pretende demostrar de una manera simple la importancia de conocer la incertidumbre de los errores y los costos de la mala calidad en los MDE. Razón por la cual se hace una revisión en la literatura sobre las técnicas de generación de errores en los MDE.

La “escuela de la simulación” considera una representación espacial de errores como una distribución de posibles realizaciones dentro de las cuales se encuentra el verdadero valor (Wechsler, 2007).

Suponiendo la existencia de autocorrelación espacial en los errores de la componente vertical en un MDE, pueden ser generados y ajustados campos de error aleatorios mediante varios métodos (Wechsler y Kroll, 2006). Entiéndase como campo aleatorio a una superficie de valores al azar que estiman la magnitud, varianza y variabilidad espacial de la incertidumbre, donde cada valor representa el error potencial en una celda dada en el MDE (Wechsler, 2007).

El conocimiento previo de la estructura espacial del error en los MDE puede ser obtenido de una fuente de datos de mayor exactitud, tal como: alturas de puntos (puntos de control) o un MDE de mayor exactitud. Este es un enfoque empírico que asume que la estructura espacial de los errores del MDE está disponible (Wechsler, 2007).

En el caso de que no se disponga de información complementaria, es posible usar el RMSE provisto en el metadato del MDE y métodos para incorporar la autocorrelación espacial mediante la aplicación de una simulación determinista.

2.4.1 Simulación determinista de errores en los MDE

En el caso de que el usuario conozca únicamente el RMSE y no se disponga de información complementaria, Wechsler y Kroll (2006) sugieren la aplicación de los criterios establecidos en la Tabla 2.5 para la determinación de campos aleatorios deterministas², dependiendo del nivel de dependencia espacial (D) existente. El método genérico es conocido como medias móviles debido a que aplican un filtro de paso bajo al campo aleatorio para incrementar su dependencia espacial.

En el caso del método de dependencia espacial ponderada este tipo de filtro puede ser expresado a través de la ecuación 2.9.

$$Ponderación = \sum_{i=1}^{TL} \mu L_i * \left(\frac{TL - i + 1}{\sum_{i=1}^{TL} i} \right) \quad (2.9)$$

De donde:

---

²En el enfoque determinista las mismas entradas producirán invariablemente las mismas salidas, sin considerar el azar y la incertidumbre.

Método Campo Aleatorio	Descripción
No filtrado (U)	Campo aleatorio con $M = 0$, $SD = RMSE$
Auto-correlación con vecindarios (N)	Filtro de paso bajo Media 3×3 aplicado a (U), y re-escalado a $M = 0$, $SD = RMSE$
Dependencia espacial media (S)	Filtro de media $D \times D$ aplicado a (U); las celdas son re-emplazadas con la media de las celdas alrededor de $D \times D$ y re-escalados a $M = 0$, $SD = RMSE$
Dependencia espacial ponderada (W)	Filtro (U) con un kernel ponderado $D \times D$; las celdas más alejadas son asignadas con menor peso y re-escalados a $M = 0$, $SD = RMSE$

Tabla 2.5: Descripción de 4 métodos para generar campos aleatorios conociendo únicamente RMSE. M representa la media, SD la desviación estándar, $D \times D$ la matriz donde se expresa la dependencia espacial. Tomado de Wechsler y Kroll (2006)

μL_i es la media de los valores del anillo i ; y,
 TL es el número total de anillos.

2.4.2 Simulación estocástica de errores en los Modelos Digitales de Elevación

Una simulación de una función aleatoria es la obtención numérica de una o varias realizaciones de un modelo (simplificación de la realidad) (Alfaro, 2008).

El enfoque de simulación estocástica del modelamiento del error en un MDE requiere un número de realizaciones, bajo las cuales son calculados un conjunto de estadísticos. La incertidumbre es estimada al evaluar estos estadísticos.

La representación de estas distribuciones equi-probables de errores es referida como modelamiento estocástico o simulación de Monte Carlo, la cual asume que el MDE es únicamente una realización de una serie de potenciales realizaciones.

Por tanto, cada celda puede ser representada por una Función de Distribución de Probabilidad (PDF), razón por la cual, cada celda tendría una media y varianza conocida. Un valor es extraído desde la PDF y asignado para cada celda. Este proceso repetido muchas veces genera un conjunto de realizaciones, las cuales representan la realidad de la variable analizada (error de la altura).

La técnica no asegura que el error sea “real”, pero si provee un rango dentro del cual podemos afirmar que el verdadero error está incluido. Las técnicas de

simulación pueden, por tanto, ser usadas para representar la incertidumbre acerca de la verdadera elevación.

Métodos de simulación estocástica

En la literatura existe una amplia variedad de métodos de simulación estocástica entre los que destacan los geoestadísticos. Autores como Heuvelink (1998), Hengl, Heuvelink y Loon (2010) y Leon, Heuvelink y Phinn (2014) han mostrado las potencialidades y limitaciones del uso de la geoestadística en el campo de la propagación de errores y la cuantificación de las incertidumbres. La presente tesis se centra en este método para la simulación de errores, debido a que ésta permite conocer la estructura espacial de los errores.

La geoestadística es una rama de las matemáticas aplicadas de rápida evolución, cuyo origen se encuentra en la industria minera en los años 50, con el objetivo de ayudar a mejorar el cálculo de reservas minerales. Originalmente desarrollada en el campo de la minería, en los años 70 se extendió a otras áreas de las ciencias de la Tierra con el advenimiento de computadores de alta velocidad.

Así, la geoestadística se define como el estudio de fenómenos regionalizados es decir, fenómenos que se extienden en un espacio (geográfico o eje temporal) y que presentan una organización o estructura. No obstante, el objeto sobre el cual trabaja la geoestadística no es el fenómeno regionalizado propiamente dicho, que es una realidad física, sino que es una descripción matemática de esta realidad, en otras palabras, una función numérica aleatoria denominada variable regionalizada, definida en el espacio geográfico o temporal, que se supone representa y mide correctamente dicho fenómeno (Lorenzo e Iribas, 2008).

Para efectuar la inferencia estadística de la variable regionalizada se necesitan un sin número de realizaciones, por ello para simplificar se asume una hipótesis de estacionaridad u homogeneidad espacial. La idea es reemplazar las repeticiones sobre las realizaciones de la función aleatoria por una sola repetición en el espacio geográfico, es decir, los valores que se observan en las distintas localizaciones del dominio objeto de estudio presentan las mismas características y pueden considerarse como realizaciones de la misma función aleatoria.

Se plantean entonces tres hipótesis de estacionariedad:

1. Estacionariedad de 1^{er} Orden o Invariante a traslación, es decir que las propiedades probabilísticas de un conjunto de observaciones no dependen de las localizaciones concretas en las que hayan sido medidas, sino tan solo de sus separaciones.
2. Estacionariedad de 2^{do} Orden, la Esperanza de la función aleatoria es constante y se supone que la covarianza de lo que ocurre entre puntos depende de la distancia, por tanto media y varianza son constantes y finitas
3. Estacionariedad Débil o Intrínseca, cuando la varianza no es finita, se trabaja solo con la hipótesis que pide que los incrementos sean estacionarios.

Además existe una dependencia espacial que se traduce a una estructura de la aleatoriedad que puede ser representada a través de la función semivariograma.

La función semivariograma se representa por medio de la ecuación 2.10, en el que se calcula un promedio de las diferencias de la variable analizada, agrupándolas en un determinado número de intervalos (lag bins). Se usa un promedio, debido al importante número de pares que pueden generarse en el análisis.

$$\gamma(h) = \frac{1}{2N(h)} \sum_{n=1}^{N(h)} [Z(x_i + h) - Z(x_i)]^2 \quad (2.10)$$

De donde:

h : distancia de análisis (lag)

$N(h)$: número de pares de puntos a una distancia dada

$Z(x_i)$: atributo de la variable de análisis en el punto x_i

$Z(x_i + h)$: atributo de la variable de análisis en el punto x_i más una distancia h dada.

Los parámetros de la función semivariograma son:

- La meseta o sill es el valor estable de la función semivariograma y coincide con la varianza de la variable. En ocasiones tiene dos componentes: a) el efecto pepita o discontinuidad en el origen; y b) la meseta parcial (partial sill) (Moreno Jiménez, 2006).

- El alcance o rango es la distancia máxima de separación entre puntos en donde se presenta autocorrelación espacial.

Otro modo de expresar la autocorrelación espacial es con el gráfico de la covarianza empírica. Esta se obtiene con la Ecuación 2.11.

$$cov = (z(S_i) - \hat{z}) * (z(S_j) - \hat{z}) \quad (2.11)$$

Por otro lado, para determinar si la autocorrelación espacial de la variable es distinta en una o otra dirección, es decir, si los datos presentan anisotropía, así como, determinar la dirección del eje de anisotropía (Moreno Jiménez, 2006) se hace uso del conocido como superficie variográfica. Esta constituye una malla o grilla regular, en la que cada celda tiene un valor de semivariograma que se calcula mediante el promedio de los pares que caen dentro de cada celda. El tamaño de esta coincide con la amplitud del intervalo o lag, por tanto el número de celdas viene dado por el número de intervalos.

Algunos parámetros como el ángulo de dirección, anchura de banda, intervalo (lag) son presentados en la Figura 2.3.

Figura 2.3: Elementos importantes en el análisis variográfico.

Actualmente, la geoestadística es muy usada en los campos de la ciencia y la industria, en donde existe una necesidad por evaluar datos con correlación espacial o temporal. El análisis de los fenómenos espaciales y temporales son discutidos

tomando en cuenta tres aspectos: descripción de los datos o análisis exploratorio, interpretación o análisis estructural y estimación o caracterización (Wackernagel, 2013).

1. **Análisis exploratorio de datos**, cuyo objetivo es conocer el comportamiento de la variable a través del análisis de la estadística descriptiva del conjunto de datos.
2. **Análisis estructural**, que tiene como objetivo determinar la dependencia espacial en el sentido de la variabilidad, es decir si la varianza es muy pequeña la dependencia es mucho mayor y si la varianza es grande hay menos dependencia. Incluye el análisis de:
  - Nube variográfica, consiste en la gratificación de los valores del atributo de todas las diferencias al cuadrado.
  - SemiVariograma experimental o empírico (omnidireccional), es un gráfico en el que se representa en el eje de las ordenadas la mitad del promedio de las diferencias al cuadrado del atributo entre cada par de ubicaciones en un intervalo dado; y en el eje de las ordenadas la distancia o intervalo que los separa.
  - Variograma experimental para diferentes direcciones (0, 45, 90 y 135 grados)
  - Anisotropía o variomapa

3. **Caracterización**, que permite determinar el modelo que se ajusta a la variable que se está estimando. Existen varios modelos de funciones matemáticas para el ajuste de un variograma. Los modelos teóricos de semivariograma más comunes se tratan a continuación:

El modelo esférico o de Matheron (Lloyd, 2007), se representa mediante la ecuación 2.12.

$$\gamma(h) = \begin{cases} c * \left[1,5 \frac{h}{a} - 0,5 \left(\frac{h}{a} \right)^3 \right] & \text{si } h \leq a \\ c & \text{si } h > a \end{cases} \quad (2.12)$$

De donde c es la meseta, a es el rango y h es el intervalo o lag.

El modelo exponencial, se caracteriza por crecer más lentamente que el esférico y está dado por la ecuación 2.13.

$$\gamma(h) = c * \left[1 - \exp\left(-\frac{h}{a}\right) \right] \quad (2.13)$$

El modelo exponencial alcanza la meseta asintóticamente. El rango práctico es la distancia en la cual la semivarianza alcanza el 95 % de la varianza de la meseta y es aproximadamente $3a$ (Lloyd, 2007).

El modelo gaussiano no alcanza una meseta en una distancia finita (Lloyd, 2007), y el rango práctico es $\sqrt{3} * a$. Su representación matemática está dada por la ecuación 2.14.

$$\gamma(h) = c * \left[1 - \exp\left(-\frac{h^2}{a^2}\right) \right] \quad (2.14)$$

Existen varios métodos de ajuste del variograma empírico, entre los que destacan el propuesto por Desassis y Renard (2013), fundamentado en el método de mínimos cuadrados iterativos. Este es basado en una minimización numérica (determinista) de una suma de cuadrados. Cada parámetro es restringido a una definición consistente con la teoría geoestadística, aunque su dominio puede ser definido por el usuario. Este puede ingresar su propio conjunto de estructuras básicas y el algoritmo intenta ajustar los datos dentro de las opciones de modelos seleccionados, a la vez que evita el sobre-ajuste.

El algoritmo numérico del método de ajuste es una variante del algoritmo de Gauss-Newton y ha sido denominado foxleg. El modelo es reducido aplicando el principio de parsimonia.

La geoestadística se puede ser tratada desde distintos enfoques:

1. Geoestadística estacionaria lineal;
2. Geoestadística no estacionaria lineal;
3. Geoestadística no lineal; y
4. Simulaciones geoestadísticas.

En el primero, se asume que los datos provienen de la realización de un conjunto de funciones aleatorias las cuales son estacionarias o, al menos, cuyos incrementos espaciales o temporales son estacionarios.

En el segundo, se considera la incorporación de funciones de tendencia (función determinista dada por otras variables), las cuales no son invariantes a traslación.

En el tercero no se busca una estimación como tal, sino más bien se persigue una clasificación de áreas en función de la probabilidad de que los valores reales superen un determinado umbral (Rivoirard, 1994).

Finalmente, en el cuarto enfoque, las simulaciones geoestadísticas comprenden varios procedimientos probabilísticos para generar realizaciones de funciones aleatorias dado un variograma $\gamma(h)$ y, estos procedimientos se pueden agrupar en:

- Método del análisis armónico.
- Método de las medias móviles.
- Método de las líneas rotantes.
- Método gaussiano secuencial.

Método del análisis armónico. Este método se basa en la representación espectral de un campo aleatorio estocástico. Donde la señal del espectro es estacionaria y está constituida por una mezcla de componentes sinusoidales estadísticamente independientes a diferentes frecuencias. Estos constituyentes armónicos básicos pueden ser identificados físicamente por medio de filtros que dejan pasar oscilaciones en un intervalo de una frecuencia dada. También se puede usar una transformación discreta de Fourier. En el caso de procesos espaciales, el típico significado de los componentes de la frecuencia es generalmente menos claro, pero la representación espectral se mantiene como una herramienta teóricamente útil, especialmente para simulaciones (Chiles y Delfiner, 2012).

Método de las medias móviles. Este método consiste en:

Generar al azar realizaciones independientes w_i de una variable aleatoria W_i , la cual es normal o gaussiana con esperanza matemática 0 y varianza 1, y se asignan estos

valores en los vértices de una malla regular (en una, dos o tres dimensiones). Sea f una cierta función, la simulación es entonces definida por:

$$Y_i = \sum_k f(kb)w_{i+k} \quad (2.15)$$

De donde:

La sumatoria puede ser: simple, doble o triple; y, b es el paso de malla. Por lo tanto, para cada función f se tiene un variograma diferente. En este método se tiene una malla muy densa con valores normales con media 0 y varianza 1 (llamada también normal reducida). A cada punto de la malla se le asocia la suma ponderada (W) siguiente, la cual se realiza sobre todos los puntos que caen dentro de una elipse E :

$$Z_{x,y} = \frac{\sum_{x_i, y_j \in E} W(x_i y_j)}{\sqrt{n}} \quad (2.16)$$

Siendo n el número de puntos que caen dentro de la elipse E , por lo que se puede probar que $Z(x, y)$ también sigue una ley normal reducida.

En el caso específico de simular un modelo esférico (en dos dimensiones), existe un método más eficiente conocido como *random token*, el cual consiste en implantar en el plano, círculos cuyos centros y diámetros son aleatorios y asignar en cada implantación un número (aleatorio) a los puntos interiores, sumando la contribución de las implantaciones (Alfaro, 2008).

Método de las líneas rotantes. Sean k rectas D_i al azar en un espacio, donde i pertenece al intervalo $[1, k]$. Cada recta se caracteriza por un vector unitario e_i . Se genera, en cada recta D_i , una realización $Y_i(x)$ de una función aleatoria de una dimensión (mediante el método de las medias móviles).

$$Y(x) = \frac{1}{\sqrt{k}} \sum_{i=1}^k Y_i(x * e_i) \quad (2.17)$$

Para obtener una buena convergencia, es necesario utilizar por lo menos unas 50 rectas aleatorias o rectas uniformemente distribuidas (Alfaro, 2008).

Simulación secuencial Gaussiana. Dentro de las simulaciones gaussianas, podemos distinguir dos principales: condicionada (condicional) y no condicionada, dependiendo si el soporte de datos experimentales (realidad) y simulación coinciden o no, respectivamente. Las simulaciones condicionadas son las más usadas, por lo que serán descritas en la sección siguiente.

Cada ubicación del dominio espacial (celdas) es visitada con una secuencia aleatoria.

Para la primera ubicación, son consideradas la distribución normal con media μ_ε y desviación estándar σ_ε predefinidas. Un valor es asignado a esa ubicación desde la distribución de probabilidad que es creada usando un generador de números aleatorios.

En una segunda ubicación, la probabilidad condicional es calculada condicionando la variable en esa ubicación al valor que fue muestreado en la primera ubicación. Esto es realizado con kriging simple. Este método además de realizar una estimación, provee la incertidumbre de dicha estimación, a través de la varianza de kriging.

El método de simulación secuencial puede ser usado en casos en donde la variable es conocida en algunas ubicaciones, visitando esas ubicaciones primero y añadiendo estos valores al conjunto de datos condicionales.

Simulación condicional

Si consideramos una variable regionalizada conocida en los puntos experimentales $x_1, x_2, \dots, x_N$, la simulación condicional consiste en construir una realización que posee el mismo histograma (igual media y varianza), y variograma de los datos disponibles; además de estar condicionada por estos datos experimentales $z(x_1), z(x_2), \dots, z(x_N)$. Es decir, donde datos, simulación y realidad coinciden (Lantuéjoul, 2013).

Este último condicionante es resuelto con la ecuación siguiente que resuelve el problema (Alfaro, 2008):

$$Z_1(x) = Z_{k1}(x) + [Z_S(x) - Z_{k2}(x)] \quad (2.18)$$

De donde:

$z_1(x)$ es la simulación condicional, $Z_{k1}(x)$ es el krigado en el punto x , utilizando los datos disponibles en los puntos $x_1, x_2, \dots, x_N$.

Sin embargo, para garantizar la reproducción del variograma, la interpolación debe corresponder al krigado simple. $Z_S(x)$ es la simulación no condicional (resultado del método de medias móviles, bandas rotantes o cualquier otro). $Z_{k2}(x)$ es el krigado en el punto x , utilizando los valores simulados $Z_S(x)$ en los puntos $x_1, x_2, \dots, x_N$.

Dado que el krigado es un interpolador exacto, entonces la simulación condicional coincide con los datos en los puntos $x_1, x_2, \dots, x_N$.

La Figura 2.4 muestra dos simulaciones condicionales en una dimensión para la variable ley de mineral (en porcentaje). Se observa que todo estimador, en particular el krigado, suaviza la realidad, mientras que la simulación respeta la variabilidad real del fenómeno analizado.

Figura 2.4: Dos simulaciones condicionales en una dimensión. La simulación coincide con la realidad en los puntos condicionantes ($x_1, x_2, x_3, \dots, x_N$). Tomado de Alfaro (2008)

Existe una clara diferencia entre estimación y simulación condicional, por lo que ésta no debe ser utilizada para estimar la variable regionalizada. Esto ha sido ratificado en la literatura por Lantuéjoul (2013) y Chiles y Delfiner (2012).

Los objetivos de usar simulaciones condicionales siempre están relacionados con su característica fundamental que es reproducir la variabilidad espacial real de la variable regionalizada. A continuación se presentan algunos de ellos:

- Optimizar planes de reconocimiento en muestreo, definiendo mallas de muestreo óptimo.
- Contribuir a la enseñanza del comportamiento de variables.
- Apoyar en el estudio de caracterización de recursos naturales.
- Estimar la incertidumbre en el comportamiento de una variable.
- Obtener una estimación de propagación de errores en un modelo aplicado.

El siguiente apartado se centrará en las técnicas usadas para la propagación de errores (que han sido generados previamente).

2.5 Técnicas de propagación de errores

Varios investigadores se han focalizado en el uso de técnicas de simulación para propagar errores y cuantificar la incertidumbre en datos espaciales (Heuvelink, Burrough y Stein, 2006; Wechsler, 2007; Leon, Heuvelink y Phinn, 2014).

Alternativas a la simulación Monte Carlo incluyen modelos analíticos de propagación de error basados en la extensión de las Series de Taylor (Albani et al., 2004). Sin embargo, la simulación de Monte Carlo es el enfoque comúnmente aplicado para evaluar la incertidumbre de los MDE (Heuvelink, 1998). Esto puede deberse a su relativa simplicidad de concepto, avances en potencia de los computadores que han facilitado las demandas de este enfoque y, por otro lado, porque los métodos analíticos se vuelven incómodos y necesitan hacer uso de aproximaciones simplificadoras (Heuvelink, Burrough y Stein, 2006).

Los campos aleatorios usados para representar los errores en los MDE constituyen superficies de valores aleatorios que estiman la magnitud, varianza y variabilidad espacial de la incertidumbre.

De acuerdo a lo que afirma Isaaks y Srivastava (1989):

“El modelo para estimar la incertidumbre es originado en el campo de la geoestadística y está basado en la hipótesis de estacionaridad que asume que las propiedades espaciales son independientes de la ubicación. El error es complejo y es probablemente no estacionario, y espacialmente auto-correlado. La suposición de estacionaridad, sin embargo, aplica a la búsqueda de vecinos, no al conjunto de datos entero y como tal es un supuesto viable incluso en conjuntos de datos para los cuales la estacionaridad global es claramente inadecuada”

Las suposiciones base del procedimiento de simulación de Monte Carlo, tal como se aplica al análisis de incertidumbre (Lindsay, 2006; Wechsler, 2007) son las siguientes:

- El error en el MDE existe y constituye incertidumbre que es propagada con la manipulación de los datos de elevación.
- La verdadera naturaleza y grado de los errores es desconocido.
- Los errores pueden ser representados por una distribución de realizaciones de un MDE.
- La verdadera elevación está en algún lugar dentro de esta distribución.

2.6 Valor de la Información Geográfica y los Modelos Digitales de Elevación

Profesionales del sector de la industria de la IG, señalan que la información por sí misma no tiene un valor intrínseco, pero su valor está directamente ligado a su uso y naturaleza, es decir el valor que ésta ha añadido al proceso de toma de decisiones. Por lo que, el valor de la IG es altamente subjetivo y dependiente del contexto. Una misma información geográfica puede tener diferentes valores, dependiendo de las diferentes formas de uso, perfiles de usuarios y formatos, sumado a la temporalidad

y versiones de la misma o cuando dicha información es usada para un diferente propósito para la que fue creada (Longhorn y Blakemore, 2007). Además dependerá del sector (gubernamental, industrial privado o comercial) en el que sea valorada.

Hay varias maneras para definir un valor, algunas son cuantificables (tangibles) de forma monetaria, mientras otras (intangibles) no lo son (Genovese, Roche y Caron, 2009). De acuerdo con Altan et al. (2013) los estudios que analizan el valor de la información generalmente están basados en:

1. La estimación econométrica de la productividad o las ganancias debido a la información.
2. Estudios de los precios hedónicos suponiendo el valor de los bienes y servicios de la información.
3. Valoración contingente, basado en la voluntad de pagar por un bien o servicio.
4. Costos evitados, basados en una estimación de la ganancia de la eficiencia o reducción de las pérdidas.

2.6.1 Valor Económico de la Información Geográfica

La mayoría de los métodos existentes en la literatura se enfocan en medidas tangibles como la estimación econométrica de los beneficios directos (Houghton, 2011; Castelein, Bregt y Pluijmers, 2010; Genovese, Roche y Caron, 2009), otras se enfocan en valores intangibles como la percepción de los beneficios socio-económicos en aplicaciones de la IG en los campos de la salud, conocimiento, social, ambiental, entre otros. Estos beneficios intangibles pueden ser igual o más importantes que los tangibles (Borzacchiello y Craglia, 2011; Wong, 2015).

La Figura 2.5 muestra un ejemplo de flujo de valor de los beneficios de la IG tomando en cuenta los componentes para una valoración económica (Natural Resources Canada, 2015).

Las metodologías que más se han usado para la valoración económica de la IG son:

Figura 2.5: Flujo de valor de los beneficios de la IG. Tomado de Natural Resources Canada (2015)

País	Valor económico en billones	Contribución PIB	Año
Canadá	2.3 USD	0.15%	2013
Holanda	1.4 EU	0.25%	2008
Australia	1.37 AUD	0.6%	2006-2007
Reino Unido*	0.65 GBP	0.06%	2007
Estados Unidos	1.5-2.7 USD		2013

* solo considera el sector privado y no incluye el capital humano y los clientes

Tabla 2.6: Contribución del sector y la industria de la IG en la economía nacional

Volumen de negocios y el tamaño del mercado. Miden el valor económico estimando la importancia que tiene el sector de la IG en la economía nacional a través del uso de encuestas y datos económicos como el Producto Interno Bruto (PIB) para definir el sector de la IG y sus principales actividades. A modo de ejemplo, en la Tabla 2.6 se muestra el valor económico de la contribución al PIB del sector y la industria de la IG para algunos países en los que se dispone de este análisis.

Costo-beneficio. Es un método clásico usado principalmente en los SIG, donde se han reportado ratios de beneficio/coste de 2:1 a 150:1 por usar IG) y en las IDE, uno los más destacados es el caso de la IDE de Cataluña. El enfoque adoptado es recopilar y analizar la mayor cantidad de información pertinente para estimar los impactos

económicos de los costos y beneficios tangibles e intangibles (Houghton, 2011). Sin embargo muchos estudios de costo-beneficio tienen dificultad para cuantificar los beneficios en comparación con los costes (Longhorn y Blakemore, 2007).

Cadena de valor. Se relaciona con el conjunto de operaciones (directas e indirectas) de valor añadido realizadas por uno o más productores para transformar los datos geospaciales integrándolos con otras fuentes de información, atributos, modelos y tecnología para crear nuevos productos. Esta es una de las metodologías más complejas debido al número de variables que intervienen en análisis (Genovese, Roche y Caron, 2009).

En el caso de los MDE, en la literatura, los estudios relacionados con valor económico son escasos. A continuación se listan los más representativos:

1. El proyecto “EcoGeo Cookbook” llevado a cabo en el período de 2008-2010, con el propósito de evaluar el valor económico del sector de la IG en Quebec, mostró que el enfoque de cadena de valor es uno de los más adecuados para evaluar el valor IG. Sin embargo, tienen una complejidad asociada al número de variables involucradas en el análisis dentro de las organizaciones y entre ellas (Genovese, Roche y Caron, 2009; Genovese et al., 2010)
2. Por su parte Castelein, Bregt y Pluijmers (2010) para estimar la importancia de la IG como sector económico, utiliza el enfoque de volumen de negocios y empleo en el sector de la IG para medir el valor económico en Holanda. El estudio fue realizado en el 2008 y muestra que el sector de la IG es significativo con un rápido desarrollo y alto potencial. Destaca la definición del mercado de la IG, pues toma en cuenta el alcance del sector de la IG y a partir de ello se estima el valor económico que se estima es del 0.25 % del PIB.
3. Estudio realizado por Wong (2015) sobre el valor económico de la información geográfica 3D desde el enfoque de aplicaciones en áreas urbanas y no urbanas como la gestión de los recursos naturales. El informe destaca la necesidad de crear una metodología para cuantificar beneficios intangibles mediante la identificación de las aplicaciones y los usuarios finales; definición completa de la cadena de valor de la IG 3D y comparación de técnicas de valoración.

4. EuroSDR en marzo de 2017, presentó los resultados del proyecto “Identificando el valor económico de la geoinformación en 3D para las Agencias Nacionales de Cartografía”, donde fueron mostrados estudios de caso de valor económico desde los enfoques de cadena de valor y el análisis de costo-beneficio, en aplicaciones como la gestión forestal, planeamiento urbano, catastro, gestión de inundaciones y gestión de bienes.

La IG tiene una gama de costes directos, incluyendo: recolección, control de calidad, procesamiento, almacenamiento, difusión, publicidad de su existencia, agregación de valor y uso. La siguiente sección se centrará en la revisión bibliográfica relacionada con los costes de la calidad y la mala calidad.

2.6.2 Costes de la calidad y la mala calidad

De acuerdo a lo planteado por Crosby (1978), cuantificar los costos de la calidad de un producto requiere de un análisis tanto de los costos del aseguramiento de la calidad como de los costes de la mala calidad, comúnmente el primer análisis es el más usado por los productores, ya que éste se basa en todos aquellos costes en los que se incurre para asegurar que el producto alcanza sus especificaciones para lo cual se consideran dos categorías:

- Costes de prevención: actividades preventivas que forman parte de la estructura de aseguramiento de la calidad intentando evitar las causas de aparición de defectos y evitar la no calidad del producto.
- Costes de evaluación: actividades de inspección y control con el objetivo de impedir la salida de productos defectuosos al mercado.

Por su parte, los costes de la mala calidad se corresponden con los que generan productos defectuosos es decir aquellos que no cumplen con las especificaciones técnicas establecidas, en este grupo se tienen dos categorías:

- Fallos Internos: el producto defectuoso se identifica antes de que el cliente acepte el mismo, este puede ser: desechado, reparado, reprocesado o reclasificado.

Figura 2.6: Costes de la calidad y no calidad. Tomado de Yepes Piqueras, V. (2010)

- Fallos externos: los productos se identifican después de la aceptación del cliente, tras su salida al mercado, por lo que los costes suelen ser más altos (pago de reclamaciones, garantías, envíos, entre otros)

La Tabla 2.7 muestra algunos ejemplos de los costos de la mala calidad en una empresa u organización.

Efectos de la mala calidad de los datos		
Costos	Tareas operacionales	Decisiones estratégicas
Ocultos	Largos plazos de entrega, datos registrados múltiples veces, insatisfacción de los empleados	Focalización en segmentos de cliente erróneos, pobre planificación global de producción, deficiente política de precios
Directos	Errores de manufactura, entregas erróneas, errores en pagos	Pocas ventas, baja eficiencia, dificultades en mantener tiempos de entrega

Tabla 2.7: Ejemplos de costos de la mala calidad en una empresa u organización. Tomado de Haug, Zachariassen y Van Liempd (2011)

La presente tesis se centra en los fallos externos, para el caso de los MDE, específicamente en la componente de exactitud vertical, es decir los fallos una vez que el producto se encuentra disponible al cliente, con el objetivo de cuantificar los costes de la mala calidad en dichas condiciones en concreto. Tómese en cuenta que los fallos detectados a tiempo cuestan menos al principio que si éstos llega al cliente.

2.6.3 Función de costo de pérdida debido a la mala calidad

Dado que la valoración de la calidad de la IG es una tarea muy amplia y típicamente se la ha analizado a nivel organizacional nacional o regional, es necesaria la búsqueda de una opción para determinar los costes de pérdida debido a la mala calidad en un ámbito más específico.

Desde el ámbito empresarial, se puede afirmar que el coste de la mala calidad se asocia a la repetición de determinados procesos debido a fallos causados por una mala gestión de la misma o dicho de otra forma, está definida como la sumatoria de los gastos que han generado productos erróneos (Font, 2016).

Con respecto a los datos (en general), en sí mismos, no pueden clasificarse dentro de una empresa u organización como: operacionales y estratégicos, pueden ser reconocidos dentro de una de las categorías enunciadas, porque la empresa los percibe así (Haug, Zachariassen y Van Liempd, 2011).

De acuerdo a Taguchi, Chowdhury y Wu (2005) incluso cuando un producto que cumple las especificaciones establecidas para el mismo (sin defecto) es entregado al usuario, es necesario estimar su nivel de calidad. Para cumplir con este objetivo, la definición de una función de pérdida es propuesta.

La función de pérdida está definida por la ecuación 2.19:

$$L = k(y - m)^2 \quad (2.19)$$

En donde L es la pérdida en unidades monetarias, y es la característica o atributo de calidad analizado, m el valor objetivo o especificación de y , y k una constante económica de conversión de unidades. No se pretende realizar un análisis profundo de esta función y de los pormenores de su cálculo, sino mas bien comprender la motivación por la que fue propuesta por Taguchi.

En la Figura 2.7, se muestra la representación cuadrática de la función de pérdida $L(y)$, la misma que tiene un valor mínimo en $y = m$, incrementa cuando y se desvía de m , y es expresada en unidades monetarias (Taguchi, Chowdhury y Wu, 2005).

Figura 2.7: Función de pérdida de la calidad. Modificado de Taguchi, Chowdhury y Wu (2005)

En conclusión, Taguchi propone una visión determinista (basada en una fórmula o ecuación matemática) sobre la calidad en la que relaciona ésta al coste de pérdida impartida (en unidades monetarias) por el producto cuando este se aleja de la especificación o valor objetivo, y está enfocada principalmente en el consumidor o usuario y en la sociedad como un todo, por lo que el productor también está incluido.

Por lo anterior, el objetivo de la función de pérdida de calidad es la evaluación cuantitativa de la pérdida causada por la variación funcional de un producto.

Referencias

- AENOR (2014). *UNE-EN ISO 19157:2014 Información geográfica - Calidad de datos*. AENOR.
- Aguilar, Fernando J et al. (2010). “Modelling vertical error in LiDAR-derived digital elevation models”. En: *ISPRS Journal of Photogrammetry and Remote Sensing* 65.1, págs. 103-110.
- Albani, M et al. (2004). “The choice of window size in approximating topographic surfaces from digital elevation models”. En: *International Journal of Geographical Information Science* 18.6, págs. 577-593.
- Alfaro, Marco Antonio (2008). *La Simulación Condicional en un Depósito Minero*.
- Altan, O et al. (2013). *The Value of Geoinformation for Disaster and Risk Management (VALID): Benefit Analysis and Stakeholder Assessment. Published by the Joint Board of Geospatial Information Societies (JB GIS), Copenhagen*.
- Ariza-López, FJ et al. (2013). *Fundamentos de evaluación de la calidad de la información geográfica*.
- Ariza-López, Francisco Javier (2011). *Apuntes de clase: Introducción a la Calidad de la IG*.
- (2015). “Calidad de la IG: Situación actual y perspectivas de futuro”. En: *Seminario SIGTIERRAS-Quito*.
- Ariza-López, Francisco Javier y Antonio Federico Rodríguez-Pascual (2007). “La familia ISO 19100”. En:
- Arranz, José Juan (2013). “Diseño, optimización y análisis de sistemas basados en técnicas láser, para el modelado geométrico, registro y documentación, aplicados a entidades de interés patrimonial”. Tesis doct. Escuela Técnica Superior de Ingenieros en Topografía, Geodesia y Cartografía. Universidad Politécnica de Madrid.
- ASPRS (2014). *Positional Accuracy Standards for Digital Geospatial Data*. Photogrammetric Engineering & Remote Sensing Volume 81, No. 3, 53 p. URL: <http://www.asprs.org/Standards-Activities.html>.
- Atkinson, Peter M (2002). “Surface modelling: What’s the point?” En: *Transactions in GIS* 6.1, págs. 1-4.
- Borzacchiello, MT y M Craglia (2011). “Socio-economic benefits from the use of earth observation”. En: *Report from the International Workshop held at Joint Research Centre. European Commission Joint Research Centre, EUR*. Vol. 2495.

- Carlisle, Bruce H (2005). “Modelling the spatial distribution of DEM error”. En: *Transactions in GIS* 9.4, págs. 521-540.
- Castelein, Watse, Arnold Bregt e Yvette Pluijmers (2010). “The economic value of the Dutch geo-information sector”. En: *International Journal of Spatial Data Infrastructures Research* 5, págs. 58-76.
- CEM (2008). *Vocabulario Internacional de Metrología–Conceptos fundamentales y generales, y términos asociados (VIM)*. Ed. por CEM. CEM.
- Chiles, Jean-Paul y Pierre Delfiner (2012). *Geostatistics: modeling spatial uncertainty*. Vol. 497. John Wiley & Sons.
- Cooper, MAR (1998). “Datums, coordinates and differences”. En: *Landform monitoring, modelling and analysis*, págs. 21-35.
- Coveney, Seamus y A Stewart Fotheringham (2011). “The impact of DEM data source on prediction of flooding and erosion risk due to sea-level rise”. En: *International Journal of Geographical Information Science* 25.7, págs. 1191-1211.
- Crompvoets, Joep et al. (2008). “A multi-view framework to assess SDIs”. En: *Space for Geo-Information (RGI), Wageningen University and Centre for SDIs and Land Administration, Department of Geomatics, The University of Melbourne*.
- Cuartero Sáenz, Aurora (2003). “Análisis de los Modelos Digitales de Elevaciones (MDE) generados con imágenes SPOT-HRV y TERRA-ASTER”. Tesis doct. Departamento de Ingeniería Cartográfica, Geodésica y Fotogrametría de la Universidad de Jaén.
- Dasí, Joaquín Farinos (2011). “Inteligencia Territorial para la planificación y la gobernanza democráticas: los observatorios de los territorios”. En:
- Desassis, Nicolas y Didier Renard (2013). “Automatic variogram modeling by iterative least squares: univariate and multivariate cases”. En: *Mathematical Geosciences* 45.4, págs. 453-470.
- Devillers, Rodolphe y Helen Goodchild (2010). *Spatial data quality: from process to decisions*. CRC.
- Devillers, Rodolphe et al. (2010). “Thirty years of research on spatial data quality: achievements, failures, and opportunities”. En: *Transactions in GIS* 14.4, págs. 387-400.
- Donker, Frederika Welle (2009). “Public sector Geo Web Services: which business model will pay for a free lunch?” En: *SDI convergence*, pág. 35.
- El-Sheimy, Naser, Caterina Valeo y Ayman Habib (2005). *Digital terrain modeling: acquisition, manipulation, and applications*. Artech House.

- Fallas, Jorge (2007). *Modelos digitales de elevación: Teoría, métodos de interpolación y aplicaciones*.
- Felicísimo, Angel M (1994). *Modelos digitales del terreno*. Pentalfa Oviedo.
- FGDC (2016). *2015 Annual report: Reston, Virginia, USA*. Inf. téc. Federal Geographic Data Committee.
- Fisher, Peter (1998). "Improved modeling of elevation error with geostatistics". En: *GeoInformatica* 2.3, págs. 215-233.
- Fisher, Peter F y Nicholas J Tate (2006). "Causes and consequences of error in digital elevation models". En: *Progress in Physical Geography* 30.4, págs. 467-489.
- Fleming, Cory, JRA Giles y SH Marsh (2010). "Elevation models for geoscience". En: *Geological Society of London*. Geological Society of London.
- Floridi, Luciano y Phyllis Illari (2014). *The philosophy of information quality*. Vol. 358. Springer.
- Florinsky, Igor V (2012). *Digital terrain analysis in soil science and geology*. Academic Press.
- Font, Rosa M. (2016). "Implantación del Sistema de Gestión de Calidad según UNE en ISO 9001: 2008. Fundación Luis Giménez Lorente, en el ámbito de la cartografía y la gestión de sus fondos. Comparativa con otras instituciones cartográficas a nivel nacional." Tesis doct.
- García Balboa, J.L. (2013). *Normas sobre la calidad de la Información Geográfica*. En Ariza-López, F.J. (Editor), Fundamentos de evaluación de la calidad de la Información Geográfica (pp. 219-274). Jaén: Servicio de Publicaciones de la Universidad de Jaén.
- Genovese, Elisabetta, Stéphane Roche y Claude Caron (2009). "The value chain approach to evaluate the economic impact of Geographic Information: towards a new visual tool". En: *SDI Convergence; Research, Emerging Trends, and Critical Assessment*, págs. 175-188.
- Genovese, Elisabetta et al. (2010). "The EcoGeo Cookbook for the assessment of Geographic Information value". En: *International Journal of Spatial Data Infrastructures Research* 5, págs. 120-144.
- Goodchild, Michael F, Rodolphe Devillers y Robert Jeansoulin (2006). *Fundamentals of spatial data quality*. Vol. 662. John Wiley & Sons.
- Haug, Anders, Frederik Zachariassen y Dennis Van Liempd (2011). "The costs of poor data quality". En: *Journal of Industrial Engineering and Management* 4.2, págs. 168-193.

- Heidemann, Hans Karl (2014). “Lidar Base Specification”. En: *Book 11: Collection and Delineation of Spatial*. Ed. por USGS. Vol. Section B, U.S. Geological Survey Standards. Cap. 4, págs. 1-78. URL: <https://pubs.usgs.gov/tm/11b4/pdf/tm11-B4.pdf>.
- Hengl, T, GBM Heuvelink y EEv Loon (2010). “On the uncertainty of stream networks derived from elevation data: the error propagation approach”. En: *Hydrology and Earth System Sciences* 14.7, págs. 1153-1165.
- Heuvelink, Gerard BM (1998). *Error propagation in environmental modelling with GIS*. CRC Press.
- Heuvelink, Gerard BM, Peter A Burrough y Alfred Stein (2006). “Developments in analysis of spatial uncertainty since 1989”. En: *Classics from IJGIS* 20, págs. 91-95.
- Höhle, Joachim y Michael Höhle (2009). “Accuracy assessment of digital elevation models by means of robust statistical methods”. En: *ISPRS Journal of Photogrammetry and Remote Sensing* 64.4, págs. 398-406.
- Houghton, John W (2011). *Costs and benefits of data provision*. Australian National Data Service.
- Hyann, Pinatibi et al. (2016). “Characterization of error on a digital elevation model (DEM) Based on morphological zones: Case of the Denguélé District (North-west of Côte d’Ivoire)”. En: *International Journal of Innovation and Applied Studies* 16.1, pág. 84.
- INSPIRE (2012). *A Conceptual Model for Developing Interoperability Specifications in Spatial Data Infrastructures*. Inf. téc. INSPIRE.
- (2013). *D2.8.II.1 INSPIRE Data Specification on Elevation, Technical Guidelines*. Inf. téc. European Commision Joint Research Centre, Thematic Working Group Elevation.
- IPGH, Instituto Panamericano de Geografía e Historia (2013). *Guía de Normas: Segunda edición en español*. Inf. téc. Comité ISO/TC 211 Información Geográfica / Geomática.
- Isaaks, Edward H y R Mohan Srivastava (1989). “Applied geostatistics”. En: Lantuéjoul, Christian (2013). *Geostatistical simulation: models and algorithms*. Springer Science & Business Media.
- Leon, Javier X., Gerard B. M. Heuvelink y Stuart R. Phinn (2014). “Incorporating DEM Uncertainty in Coastal Inundation Mapping”. En: *PLoS ONE* 9.9,

- págs. 1-12. DOI: [10.1371/journal.pone.0108727](https://doi.org/10.1371/journal.pone.0108727). URL: <http://dx.doi.org/10.1371%2Fjournal.pone.0108727>.
- Lindsay, John B (2006). "Sensitivity of channel mapping techniques to uncertainty in digital elevation data". En: *International Journal of Geographical Information Science* 20.6, págs. 669-692.
- Liu, XiaoHang, Hai Hu y Peng Hu (2015). "Accuracy assessment of lidar-derived digital elevation models based on approximation theory". En: *Remote Sensing* 7.6, págs. 7062-7079.
- Liu, XiaoHang et al. (2012). "Approximation theory applied to DEM vertical accuracy assessment". En: *Transactions in GIS* 16.3, págs. 397-410.
- Lloyd, Christopher D (2007). *Local models for spatial analysis*. CRC/Taylor & Francis London.
- Longhorn, Roger A y Michael Blakemore (2007). *Geographic information: value, pricing, production, and consumption*. CRC Press.
- López, Carlos (2002). "An experiment on the elevation accuracy improvement of photogrammetrically derived DEM". En: *International Journal of Geographical Information Science* 16.4, págs. 361-375.
- Lorenzo, Jose Maria Montero y Beatriz Larraz Iribas (2008). *Introducción a la Geoestadística Lineal*. Netbiblo.
- Maune, David Francis (2001). *Digital elevation model technologies and applications: the DEM users manual*. Asprs Publications.
- Moreno Jiménez, Antonio (2006). *Sistemas y análisis de la información geográfica: manual de autoaprendizaje con ArcGis*.
- Mostafavi, Mir-Abolfazl, Geoffrey Edwards y Robert Jeansoulin (2004). "An ontology-based method for quality assessment of spatial data bases". En: *Third International Symposium on Spatial Data Quality*. Vol. 1. 28a, págs. 49-66.
- Mukherjee, Sandip et al. (2013). "Evaluation of vertical accuracy of open source Digital Elevation Model (DEM)". En: *International Journal of Applied Earth Observation and Geoinformation* 21, págs. 205 -217. ISSN: 0303-2434. DOI: <http://dx.doi.org/10.1016/j.jag.2012.09.004>. URL: <http://www.sciencedirect.com/science/article/pii/S030324341200195X>.
- Natural Resources Canada, Minister of (2015). *Canadian geomatics environmental scan and value study*. Inf. téc. Minister of Natural Resources Canada.

- Nedovic-Budic, Zorica, Joep Crompvoets y Yola Georgiadou (2012). *Spatial data infrastructures in context: north and south*. Ed. por Zorica Nedovic-Budic, Joep Crompvoets y Yola Georgiadou. ISBN 9781439828021. Taylor & Francis.
- Nelson, A, HI Reuter y P Gessler (2009). “DEM production methods and sources”. En: *Developments in soil science* 33, págs. 65-85.
- Oksanen, Juha y Tapani Sarjakoski (2006). “Uncovering the statistical and spatial characteristics of fine toposcale DEM error”. En: *International Journal of Geographical Information Science* 20.4, págs. 345-369.
- Olaya, Victor (1994). *Hidrología computacional y modelos digitales del terreno*.
- Papasaika, Haris y Baltsavias Emmanuel (2010). “Quality evaluation of DEMs”. En: *Proceedings of the Ninth International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences 20-23rd July 2010, Leicester*, págs. 249-252.
- Patel, Arun, S.K. Katiyar y Vishnu Prasad (2016). “Performances evaluation of different open source {DEM} using Differential Global Positioning System (DGPS)”. En: *The Egyptian Journal of Remote Sensing and Space Science* 19.1, págs. 7 -16. ISSN: 1110-9823. DOI: <http://dx.doi.org/10.1016/j.ejrs.2015.12.004>. URL: <http://www.sciencedirect.com/science/article/pii/S1110982315000745>.
- Podobnikar, Tomaž (2016a). “Combination of error characterization and spatial error model to improve quality of digital elevation models”. En: *Proceedings of Spatial Accuracy 2016, July*.
- (2016b). “Relevant Quality of Digital Elevation Models in Earth and Environmental Studies ?” En: *International Journal of Earth & Environmental Sciences*. DOI: <https://doi.org/10.15344/2456-351X/2016/106>.
- Rieger, Wolfgang (1996). “Accuracy of slope information derived from DEM-data”. En: *International Archives of Photogrammetry and Remote Sensing* 31, págs. 690-695.
- Rivoirard, Jacques (1994). *Introduction to disjunctive kriging and non-linear geostatistics*. Clarendon Press.
- Ruiz Armenteros, Antonio, José García Balboa y José Mesa Mingorance (2010). “Error, incertidumbre, precisión y exactitud, términos asociados a la calidad espacial del dato geográfico”. En: *Catastro: formación, investigación y empresa: Selección de ponencias del I Congreso Internacional sobre catastro unificado y multipropósito*, págs. 95-102.

- Shekhar, Shashi y Hui Xiong (2007). *Encyclopedia of GIS*. Springer Science & Business Media.
- Stein, Alfred, Wenzhong Shi y Wietske Bijker (2009). *Quality aspects in spatial data mining*. CRC Press.
- Survey, Ordnance (2015). *Annual Report and Accounts 2014 2015*. Inf. téc. Ordnance Survey.
- Taguchi, G, S Chowdhury e Y Wu (2005). *Taguchi's quality engineering handbook*. Wiley.
- Ureña Cámara, M.A. (2013). *Normas sobre metadatos y evaluación de metadatos*. En Ariza-López, F.J. (Editor), Fundamentos de evaluación de la calidad de la Información Geográfica (pp. 329-387). Jaén: Servicio de Publicaciones de la Universidad de Jaén.
- Wackernagel, Hans (2013). *Multivariate geostatistics: an introduction with applications*. Springer Science & Business Media.
- Wechsler, SP (2007). "Uncertainties associated with digital elevation models for hydrologic applications: a review". En: *Hydrology and Earth System Sciences* 11.4, págs. 1481-1500. URL: <https://hal.archives-ouvertes.fr/hal-00305087/document>.
- Wechsler, Suzanne y Charles Kroll (2006). "Quantifying DEM uncertainty and its effect on topographic parameters". En: *Photogrammetric Engineering & Remote Sensing* 72.9, págs. 1081-1090.
- Weng, Qihao (2002). "Quantifying uncertainty of digital elevation models derived from topographic maps". En: *Advances in spatial data handling*. Springer, págs. 403-418.
- Wise, Stephen (2000). "Assessing the quality for hydrological applications of digital elevation models derived from contours". En: *Hydrological processes* 14.11-12, págs. 1909-1929.
- Wong, Kelvin Ka Yin (2015). "Economic Value of 3D Geographic Information". En: *EuroSDR and the Department of Computer Science, University College London: London, UK*.
- Yang, X et al. (2013). "An integrated view of data quality in Earth observation". En: *Phil. Trans. R. Soc. A* 371.1983, pág. 20120072.

Capítulo 3

MATERIAL Y MÉTODOS

En este capítulo se describen las fuentes de información, procedimientos y herramientas utilizadas para: la selección del caso de estudio, los ensayos de la simulación de errores en la componente vertical de los MDE (tomando como ejemplo un sector de El Condado, Quito) y su posterior aplicación en el caso de estudio el cual corresponde a un tramo del Río Garrapata, Manabí - Ecuador.

3.1 Fuentes de Información

Para la presente tesis se ha tomando en cuenta los materiales que a continuación se describen de manera general, los mismos que serán detallados en cada una de las secciones según sea el caso.

3.1.1 Encuestas disponibles sobre los usuarios de MDE

Los datos utilizados para el análisis, tanto de los usuarios como de los usos, proceden del servicio de descarga del CNIG (<http://centrodedescargas.cnig.es/CentroDescargas/index.jsp>). Este servicio de descarga incluye un sistema de alta de usuarios y una encuesta que se debe rellenar para cada descarga, los resultados de la encuesta utilizada corresponde al año 2014.

3.1.2 Encuesta aplicada a usuarios especializados

La encuesta fue realizada en colaboración del Grupo de Investigación en Ingeniería Cartográfica (GiiC) de la Universidad de Jaén, quienes también se encuentran investigando en temas relacionado con la calidad de los MDE.

3.1.3 Lista de términos específicos sobre campos de aplicación y uso de los MDE

Resultado de una consulta avanzada a ScienceDirect, que ha sido una fuente para identificar los campos de aplicación y casos de uso de los MDE, pues ofrece una base de datos muy grande de contenido que puede ser representativo, ofrece un sistema que permite búsquedas de expertos y como resultado presenta una lista de tópicos que pueden considerarse relacionados con casos de uso.

3.1.4 Información cartográfica de zonas específicas en Ecuador

Para la aplicación de la metodología de simulación de errores ya sea para el ensayo como para el caso de estudio, fue necesario contar con información cartográfica donde existan dos fuentes de datos independientes y una exactitud 3 veces mayor para los datos de referencia.

Una de las principales fuentes fue la información obtenida del Sistema Nacional de Información y Gestión de Tierras Rurales e Infraestructura Tecnológica SIGTIERRAS, el cual es ejecutado por el Ministerio de Agricultura, Ganadería, Acuacultura y Pesca – MAGAP, que busca proveer información para la planificación y ordenamiento territorial del área rural, con el objetivo de contribuir a la eficiente gestión y administración territorial en el Ecuador, mediante la gestión de ortofotografía basada en fotografía aérea, cartografía temática a nivel nacional y catastro e información predial de 57 cantones¹.

---

¹<http://www.sigtierras.gob.ec/descripcion-del-programa/>

Los productos que ha generado SIGTIERRAS como: la toma de fotografía aérea y generación de ortofotografía que cubre el 88 % de todo el territorio y el MDE a escala 1:5000 que son la base para el catastro rural, se encuentran disponibles a través de su portal web donde es posible descargar los productos cartográficos generados y además disponen servicio web interoperables.

Otra fuente importante para la aplicación de la metodología fue el Estudio de Impacto Ambiental del Proyecto Propósito Múltiple Chone, específicamente de apartado del Estudio Hidrológico de las Cuencas Garrapata y Mosquito, disponible en la página web La Empresa Pública ESPE-INNOVATIVA EP.

3.2 Herramientas utilizadas

3.2.1 Entorno R

R constituye un entorno y lenguaje de programación para el cálculo estadístico y gráfico, donde es posible realizar una amplia variedad de análisis estadísticos, que por supuesto incluye la generación de reportes y gráficos (Elosua, 2011). El software es fácilmente extensible gracias al uso y creación de funciones y librerías que se encuentran disponibles en repositorios del proyecto (CRAN) o en otros sitios como github. En el caso de CRAN, en el año 2016 se registraron aproximadamente 6.000 librerías, lo cual es una muestra de la demanda y desarrollo que en estos momentos tiene R.

En el ámbito de la Información Geográfica existen varias librerías que pueden ser usadas en las diferentes aplicaciones de la geociencias y que cuentan con un relevante número de recursos y material de apoyo para los usuarios con ejemplos de casos de uso. En la Figura 3.1 se muestra un esquema de las librerías empleadas para el análisis de la presente tesis.

Las principales librerías que se han utilizado se describen a continuación:

- Raster: permite la manipulación de objetos tipo ráster.
- Rgdal: establece enlaces con librerías para la lectura de datos espaciales, tanto en formato raster como en formato vectorial.

Figura 3.1: Librerías de R utilizadas comúnmente para manejo de datos espaciales

- Gstat: permite el modelamiento, estimación y simulación geoestadística de datos espaciales y temporales. Esta librería hace uso de objetos con una estructura especial denominada: *sp*, la misma que contiene una matriz con los datos de la variable analizada y otra matriz con las coordenadas para cada punto medido.
- GeoR: provee herramientas complementarias a la librería Gstat para análisis geoestadístico. Hace uso de objetos del tipo *geodata*, que es una lista, la cual tiene dos componentes obligatorios: una matriz con las coordenadas en dos dimensiones (*coords*) del diseño de muestreo y un vector que contiene el correspondiente valor medido en cada una de las ubicaciones del diseño (*data*).
- RGeostats: desarrollada por Mine Paris Tech, ofrece varias opciones para análisis geoestadístico, pero principalmente ofrece algoritmos de alto nivel para ajuste variográfico.
- Knitr: es una herramienta de propósito general para la generación de reportes dinámicos en R, tanto en formato *html* o *pdf*.

- RMarkdown: permite la generación de reportes dinámicos con la inclusión de lenguajes de programación como: R, Python y SQL (Structured Query Language). Los reportes pueden ser generados en *html* o *pdf* complementando su uso con la librería *knitr* y la instalación de $\text{\LaTeX}^2$.

3.2.2 HEC-RAS y GeoHEC-RAS (ArcGis)

Para la determinación de áreas de inundación relativas a un tramo del cauce del Río Garrapata, se han empleado los programas HEC-RAS 4.1 y el paquete HEC-GeoRAS 10.2 que funciona como una extensión en ArcGIS 10.2. HEC-RAS y HEC-GeoRAS son de libre distribución; creados por el Centro de Ingeniería Hidrológica (Hydrologic Engineering Center) del Cuerpo de Ingenieros de la Armada de los EE.UU. *United States Army Corps of Engineers* (USACE). HEC-RAS permite el modelado hidráulico unidimensional de cauces, canales artificiales e inundaciones, mientras que HEC-GeoRAS es una herramienta para el intercambio de datos espaciales entre los software ArcGIS y HEC-RAS.

3.3 Selección de la aplicación y caso de uso

Para seleccionar el estudio de caso donde se aplicará la metodología para cuantificar los costes de la mala calidad de los MDE, se ha realizado un análisis de las aplicaciones y usos más frecuentes de los MDE, así como para obtener información sobre los usuarios de los MDE.

La primera opción fue buscar cuestionarios web existentes vinculados a las Infraestructuras de Datos Espaciales (IDE) a nivel nacional, para luego solicitar los resultados de los cuestionarios. Así, se consultaron las IDE de: Argentina, Bolivia, Brasil, Canadá, Chile, China, Colombia, Costa Rica, Cuba, Ecuador, Francia, Alemania, Guatemala, Honduras, Italia, México, Nicaragua, Panamá, Paraguay, Perú, República Dominicana, España, Reino Unido, Uruguay, Estados Unidos, Venezuela, de las cuales solo se encontraron dos países con cuestionarios a los usuarios: España y Estados Unidos.

---

²Es un sistema de preparación de documentos de alta calidad diseñado para la producción de documentación técnica y científica y está disponible como software libre.

La *United States Geological Survey* (USGS) de Estados Unidos tiene un formulario para el registro de las descargas pero es muy general, el cual está asociado a un perfil "demográfico", este es para todas las descargas, no hay un cuestionario específico para cada descarga de un MDE, por lo que se descartó este registro para el análisis. En España se tiene un cuestionario vinculado a cada descarga y tipo de producto cartográfico que es descargado, como en el caso específico de los MDE.

Además se realizó una encuesta orientada a conocer los perfiles de usuario, campos de aplicación y requerimientos de calidad de los usuarios de los MDE.

Por otro lado, se analizó el resultado de una consulta a la base de datos científica ScienceDirect, orientado a una búsqueda de las publicaciones científicas y técnicas relacionadas a los campos de aplicación y términos asociados al los MDE.

3.3.1 Análisis de usuarios y aplicaciones de los MDE en España

Como se indicó anteriormente, los datos que se utilizaron para este análisis son las respuestas de la encuesta a los usuarios de los MDE desde el Centro de Descargas del CNIG <http://centrodedescargas.cnig.es/CentroDescargas/> en el año 2014. Se trata de una encuesta voluntaria que se ofrece para ser rellenada en cada sesión de descarga mediante un formulario disponible en línea, que permite al usuario seleccionar opciones en listas de una sola respuesta y de selección múltiple, así como campos de texto libre para descripción detallada del uso al que va destinado el MDE, observaciones y comentarios.

El objetivo de la encuesta es tener una realimentación por parte de los usuarios respecto a los productos que son descargados desde el CNIG y en el caso específico de los MDE conocer cuáles son las demandas en cuanto a los productos que son ofrecidos como:

- MDT05/MDT05. Modelo con paso de malla de 5x5 m, con la misma distribución de hojas que el MTN25. Formato de archivo ASCII matriz ESRI (asc). Sistema geodésico de referencia ETRS89 y proyección UTM en el huso correspondiente a cada hoja. Según la hoja de que se trate, el MDT05 se ha obtenido de una de las dos siguientes formas: por estereocorrelación automática

de vuelos fotogramétricos del Plan Nacional de Ortofotografía Aérea (PNOA) con resolución de 25 a 50 cm/píxel, revisada e interpolada con líneas de ruptura donde haya sido viable, o bien por interpolación a partir de la clase terreno de vuelos LIDAR del PNOA

- MDT25. Modelo con paso de malla de 25x25 m, con las mismas características de distribución y de sistema de referencia de coordenadas que el MDT05. El MDT25 se ha obtenido por interpolación de modelos digitales del terreno de 5x5 m de paso de malla procedentes del PNOA.
- MDT200. Modelo con paso de malla de 200x200 m, con distribución por provincias (rectángulo envolvente de cada provincia). Formato de archivo ASCII matriz ESRI (asc). Sistema geodésico de referencia ETRS89 y proyección UTM en el huso correspondiente a cada provincia. El MDT200 se ha obtenido por interpolación de modelos digitales del terreno de 5x5 m de paso de malla procedentes del PNOA.

Los resultados analizados fueron 12 493, que contienen información sobre la geografía de las descargas, el perfil de los usuarios, el uso de los productos y su valoración. Para la caracterización de los usos de los productos se utilizó el campo “Área Temática”, que en el formulario permite al usuario marcar una o más áreas temáticas. En el anejo de la Sección A.3 se muestra el formulario con el CNIG obtiene la información de los usuarios.

El análisis estadístico que se ha realizado se basa en conteos y obtención de porcentajes para variables, cruces de variables y atributos. Además se realizó un proceso de estandarización para los campos de texto libre, con son: “Observaciones a los sectores”, “Descripción de uso al que va destinado” y “Observaciones a la utilidad”, el propósito de recoger las ideas expresadas sin pérdidas de generalidad al momento de usar un sentido u otro de los términos utilizados, el trabajo fue realizado de forma manual y de forma iterativa hasta conseguir los términos que más se acercan al sentido de las respuestas dadas por los usuarios, para lo cual se consiguieron un total de 44 términos estandarizados. Con el resultado del análisis se realizó un artículo titulado: “Análisis de los usuarios y usos de los MDE en España” para la revista GeoFocus publicada en la edición 17 del año 2016.

3.3.2 Encuesta a nivel mundial a usuarios especializados

Una encuesta sobre el uso y aplicaciones a usuarios especializados, fue aplicada a 10 usuarios de 27 países donde el 60,8 % es de Europa, 25,5 % de América y 13,7 % de Asia, el formato de la encuesta es mostrado en el anejo de la Sección A.1. Para conseguir la participación de los usuarios especializado se recopiló más de 2500 direcciones de correo electrónico de diferentes fuentes: autores de artículos científicos, proceedings, documentos y páginas web, así como de las asociaciones profesionales en el campo de la investigación de aplicaciones con MDE (ASPRS, FIG, ICA, entre otras.)

El cuestionario estuvo disponible en línea desde noviembre de 2015 a diciembre de 2016, incluyó un total de 18 preguntas, vinculadas con las características del usuario (ocupación, institución, años de experiencia en el uso de MDE), la caracterización de las aplicaciones/usos, requerimientos o indicadores de calidad; y, experiencias vinculadas con la calidad de los MDE.

En base a los primeros resultados obtenidos mediante la aplicación de la encuesta, se procedió a homogeneizar las respuestas abiertas. Para el análisis de los resultados de la encuesta se utilizaron herramientas de estadística descriptiva, incluyendo el uso de gráficas de representación de variables de tipo categórico.

3.3.3 Consulta a ScienceDirect

Para complementar la comprensión sobre los campos de aplicación y usos de los MDE, se consultó además una base de datos científica en Inglés como es ScienceDirect^{TM}, que alberga más de 2500 revistas y más de 33000 libros Elsevier. La base de datos fue elegida porque: a) ofrece una base de datos muy grande que podemos considerar es representativa; b) ofrece un sistema que permite búsquedas expertas (por ejemplo, se puede usar conectores y operadores de proximidad, caracteres comodín, frases, buscar campos específicos de la base de datos), y c) ofrece como resultado una lista de temas que pueden considerarse relacionados con casos de uso.

Para las consultas a la base de datos se utilizó la opción “búsqueda experta”, en esta por medio de la función “tak( )”, que en sus siglas en inglés significa (Title-Abstr-

Key): título, resumen y palabras clave del autor o publicación, y tomando en cuenta un período comprendido entre los años 2006 y 2016, se obtuvieron como resultado el listado de artículos y las palabras clave asociadas “topic”.

En la Figura 3.2 se muestra la estructura de una consulta y los resultados obtenidos, donde se puede ver que la estructura de las búsquedas tuvieron dos partes principales. La primera para evitar ambigüedades con el término “modelo digital de elevación”, porque estas palabras pueden ser usadas en otros campos científicos (por ejemplo, en física de partículas, transferencia de calor y transferencia de masa) y una segunda con el fin de establecer un campo de aplicación específico (por ejemplo, hidrología, teledetección, agricultura, entre otros.). Los términos utilizados en la segunda parte de la consulta proceden de la lista de campos de aplicación obtenidos de las dos encuestas anteriores, para lo cual se utilizaron un total 43 términos. Así, la estructura de una consulta fue:

tak ((({DEM} or {elevation model} or {digital elevation})) AND "hydrology")

The screenshot displays the ScienceDirect search interface. At the top, the search query is entered: `tak ((({DEM} or {elevation model} or {digital elevation})) AND "hydrology")`. Below the search bar, there are filters for Year (2016, 2015, 2014, 2013, 2012), Publication title (Journal of Hydrology, Computers & Geosciences, Remote Sensing of Environment, Environmental Modelling & Software, Geomorphology), and Topic (soil, elevation model, flow, algorithm, catch, model, cell, drainage network, grid, hydrologic, lake, soil moisture, aster, basin, digital elevation, drainage, flood, flow direction, grid cell). The search results are listed on the right, showing the title, authors, and abstract of several articles. The search interface also includes a search bar, filters, and a list of search results.

Figura 3.2: Estructura de la consulta a ScienceDirect y los resultados de la consulta

Las palabras clave obtenidas en la búsqueda, se clasificaron en dos grupos: “temas” y “tópicos”, donde los “temas” corresponden a las palabras usadas en la segunda

parte de las consultas, siendo cuarenta y tres los temas utilizados; y los “tópicos” corresponden a las palabras clave de los resultados de cada consulta, cuyos valores indican la frecuencia con la que éstos parecen en cada tema.

Fue necesario un proceso de normalización de los términos, ya que inicialmente se contó con un total de 243 tópicos que debieron ser clasificados. Además, para mostrar los términos más relevantes, en caso de los temas se seleccionaron a los que tienen la sumatoria de pesos superior a 100 y para los tópicos se han eliminado los que tienen una suma de pesos sea menor a 10. Con los resultados obtenidos se realizó un gráfico de redes bipartitas que permiten visualizar las relaciones que existen entre los ellos.

Análisis de relaciones de términos relativos a los Modelos Digitales de Elevación

En términos generales una red es la forma más simple para representar las relaciones o conexiones entre puntos unidos en pares por líneas (Kolaczyk y Csárdi, 2014; Newman, 2010). En el análisis de redes los puntos son conocidos como nodos o vértices y las líneas como enlaces. El foco principal del análisis de redes son los gráficos de red que se han convertido en una herramienta útil para representar patrones de conexiones, interacciones entre elementos, ya que permiten ver instantáneamente características estructurales importantes de una red pues de otro modo sería difícil realizar un análisis de los datos en bruto (Newman, 2010).

En este contexto, para el análisis de los temas y tópicos se empleó el enfoque de redes bipartitas o de modo-2, donde se trabaja con dos conjuntos disjuntos, en este gráfico de red hay dos tipos de vértices, uno representa los vértices originales y el otro a los grupos a los cuales están relacionados.

Se construyó una red bipartita con la herramienta de “force-directed graph” a través del entorno R, haciendo uso de los paquetes igraph y statnet, se generó el gráfico considerando que el nivel más alto es el “tema” y el nivel inferior pertenece a los “tópicos” que son relacionados a cada uno de los temas. El algoritmo utilizado para la visualización de la red es Large Graph Layout, que está recomendado para redes de gran tamaño (Luke, 2016).

3.3.4 Aplicación seleccionada

En base a lo expuesto en las secciones anteriores y con los primeros resultados del análisis, se procedió a seleccionar una aplicación donde los MDE son cruciales para la determinación zonas de inundación, ya que de acuerdo a los primeros resultados las aplicaciones más representativas se encuentran en los capos de la hidrología superficial y geomorfología.

En el modelado hidrológico e hidráulico, los MDE son uno de los insumos de entrada más importantes, constituyen la clave para estimar las características hidrológicas y topográficas en un cuenca hidrográfica (Callow, 2013), por tanto su calidad afecta a los cálculos derivados como redes de canales de una cuenca, gradientes de pendiente, dirección del flujo y acumulaciones, entre otras variables. El uso de los MDE en estudios hidrológicos es ubicuo, sin embargo la incertidumbre en la representación de las alturas y los parámetros derivados de los MDE, raramente es considerada por los usuarios de los MDE en este campo de aplicación (Wechsler, 2007).

La calidad de los MDE tiene una gran influencia en la determinación de zonas de inundación (Leon, Heuvelink y Phinn, 2014). Ya que, la exactitud vertical y el tamaño de la celda son importantes para que los modelos de inundación puedan representar de mejor forma la realidad de las avenidas de inundación.

Cabe mencionar, que para la aplicación seleccionada se asume la consideración de que el MDE será el insumo que contiene errores en la componente vertical, para lo cual se generan los errores mediante un proceso de simulación geoestadística, por tanto, se asume que el resto de variables que se utilicen en los modelos están libres de error, así como los métodos empleados, pues no se cuestionan los mismos.

3.4 Diseño de las simulaciones

Como paso previo a los ensayos de simulación de errores de la componente vertical de los MDE, se realizó un análisis geoestadístico completo de la variable ondulación geoidal ³ con varias herramientas, en una zona de estudio localizada en

---

³Se define como la diferencia entre el elipsoide global de referencia y el geoide, que puede considerarse coincidente con la superficie de referencia materializada por el nivel medio del mar, o en su defecto la diferencia entre la altura elipsoidal y la nivelada sobre el nivel medio del mar

la región nor-occidental de la ciudad de Guayaquil, Ecuador. Lo anterior sirvió para familiarizarse con la aplicación de las herramientas geoestadísticas. Como producto de esa investigación, se publicó un artículo científico (Anejo en la Sección A.1).

Otra publicación relacionada con la geoestadística se presentó en la Conferencia International Spatial Accuracy (2016): Modelling Uncertainty using Geostatistics, a Case Study in Ecuador (Anejo en la Sección A.2).

Para emplear la metodología de simulación de errores de la componente vertical en los MDE que serán usados en el caso de estudio, fue necesario realizar consideraciones previas en torno a la estacionaridad de la variable error, principalmente sobre la presencia de autocorrelación espacial entre los errores en un determinado MDE evaluado.

Por tal motivo, para el ensayo de la metodología para la simulación de errores, se ha seleccionado un conjunto de datos de altura de un sector de Quito-Ecuador, donde el MDE de referencia fue obtenido mediante el uso de un sistema UAV cuyas características son detalladas en el anejo de la Sección A.7 y el MDE evaluado corresponde a datos obtenidos de la centro de descarga la IDE del DMQ.

3.4.1 Datos utilizados

Como se mencionó al inicio del presente capítulo, para la realización de ensayos en el cálculo de simulaciones de errores en la componente vertical, se usaron los siguientes datos:

- MDE del Distrito Metropolitano de Quito, con resolución espacial de 3 metros para el sector analizado.
- MDE generado a partir de nube de puntos en formato LAS obtenido en el trabajo de campo mediante un sistema UAV, con una densidad aproximada de 58 puntos por m^2 , el MDE fue generado usando el método de interpolación Ponderado Inverso a la Distancia (IDW), ya que al tener una alta densidad de puntos el método no es relevante.

La ubicación general del área de estudio se presenta en la Figura 3.3.

Figura 3.3: Ubicación del área de estudio en el sector El Condado, al nor-occidente del área urbana de la ciudad de Quito. Coordenadas en sistema TM-Quito.

El MDE generado a partir de nube de puntos en formato LAS se presenta en la Figura 3.4.

Para realizar operaciones entre ráster o vector dentro del entorno R, es necesario que el objeto tenga definido un sistema de referencia espacial. Para ello, y para las siguientes asignaciones de referencia espacial, se crea un objeto con la información de referencia espacial empleado, en este caso el usado dentro del DMQ.

Los parámetros del sistema de referencia empleado en el DMQ se presentan en la Tabla 3.1.

Otro punto importante para la ejecución de operaciones ráster es la coincidencia espacial de los objetos empleados. Para ello es necesario la aplicación de un proceso de remuestreo. Esto se realiza con la función *resample* y el método bilinear. Se usa como referencia o base el MDE construido a partir de nube de puntos en formato LAS.

Figura 3.4: MDE construido a partir de nube de puntos LAS. Coordenadas en el sistema de referencia espacial TM-Quito.

Para iniciar con el proceso de generación de simulaciones, es necesario realizar la preparación de los datos, que se describe a continuación.

3.4.2 Preparación de los datos

La fase de preparación de datos incluye la ejecución de procedimientos tendientes a obtener errores de las alturas en el área de estudio.

El cálculo de errores se lo realiza a través de una resta, celda a celda, entre el MDE evaluado y el MDE referencia. Como es lógico, los dos MDE deben tener

Parámetro	Valor
Datum	WGS84
Elipsoide	WGS84
Semieje mayor	6378137 m
Achatamiento	1/298.257223563
Semieje menor	6356752.314 m
Proyección cartográfica	Transversa de Mercator Modificada
Meridiano central	W 78° 30' 00''
Origen de latitudes	N 00° 00' 00''
Factor de escala central	1.0004584
Falso Este	500000 m
Falso Norte	10000000 m
Zona	17 Sur Modificada (W 77° - W80°)

Tabla 3.1: Parámetros del sistema de referencia espacial empleado en el DMQ.

la misma resolución y sistema de referencia espacial. Por lo anterior, es necesario realizar un procedimiento de remuestreo del MDE evaluado. Varios métodos pueden ser empleados para este propósito, siendo el método bilineal el más usado por el costo computacional que demanda y los resultados que ofrece. En este método, cada nuevo valor en la celda del MDE evaluado es una media ponderada por la distancia de los cuatro vecinos más próximos de la celda del MDE referencia.

Este proceso y los subsiguientes se han llevado a cabo en el entorno y software estadístico R. Detalles sobre éste se presentan a continuación.

El procedimiento metodológico propuesto, una vez que los datos han sido preparados, se presenta a continuación:

1. Cálculo de errores de las alturas entre el MDE evaluado y un MDE de referencia o un conjunto de puntos de control de mayor precisión.
2. Análisis exploratorio de errores.
3. Análisis de la estructura espacial de los errores (análisis variográfico).
4. Simulación geoestadística.

3.4.3 Cálculo de errores de la componente vertical

El cálculo de errores propiamente dicho se lo realiza a través de una sustracción entre el MDE analizado y el MDE de referencia. A fin de evitar cálculos erróneos por presencia de valores relacionados con extrapolación al momento de generar el MDE, se usa una máscara en el análisis.

Debido a que el cálculo de errores se realiza celda a celda es necesario realizar un muestreo pseudo-aleatorio simple (Nótese que es necesario definir un valor de partida o semilla para que el análisis sea reproducible, mediante la función `set.seed(1234)`). Con este procedimiento, se obtienen un conjunto de celdas con sus respectivos atributos de altura distribuidos en el área de estudio, además se consigue un número razonable de muestras, en términos de cálculo computacional, que representan los errores de las alturas del área de estudio con la que ensayará la simulación. Esta actividad se la realiza a través de la función `runif`.

3.4.4 Análisis exploratorio de errores

Con la finalidad de realizar el análisis estadístico de los errores de la componente vertical de forma más práctica, se procede a transformar los valores de las celdas de error muestreadas anteriormente a objetos espaciales tipo puntos y finalmente se convierten estos a un *dataframe* estándar (semejante a una hoja de datos electrónica con los campos de localización geográfica y altura).

El análisis exploratorio comprende un conjunto de herramientas estadísticas, complementadas con gráficas a fin de aceptar o rechazar la hipótesis planteada. Los gráficos estadísticos se dibujan en el entorno R a través del uso de las librerías base (instaladas por defecto). Adicionalmente, para mejorar la representación gráfica se hace uso de la librería *ggplot2*.

Análisis de la normalidad de los datos

Uno de los puntos más importantes del análisis exploratorio constituye el análisis de la normalidad de los datos. El estadístico D, presentado en la ecuación 3.1 es usado para medir el ajuste global entre dos distribuciones acumuladas.

$$D = \max_{-\alpha < z < \alpha} |F(z) - F_n(z)| \quad (3.1)$$

Siendo $F(z)$ la distribución acumulada univariante teórica para el atributo Z y $F_n(z)$ la función escalonada observada para una muestra de tamaño n . El estadístico D es ilustrado en la Figura 3.5.

Figura 3.5: Discrepancia máxima D entre dos distribuciones. Tomado de Olea (2012)

El estadístico D tiene el atractivo de no ser sensible a la transformación lineal de Z . En estadística clásica, asumiendo un muestreo aleatorio, es posible calcular la distribución de D en el caso de la también llamada hipótesis nula de dos colas. Por tanto, es posible investigar la significación de D para cualquier valor diferente de cero. Este es el test estadístico de Kolmogorov-Smirnov, que está basado en el cálculo de las máximas diferencias entre una función de distribución empírica y un valor de probabilidad puntual p_i (Thode, 2002).

Este test se lo ejecuta en R, mediante la función *lillie.test* de la librería *nortest*. En geoestadística, debido a la presencia de dependencia espacial entre las muestras, no se pueden calcular niveles de significación. La magnitud del estadístico D es lo que permite aceptar o rechazar la presunción de similaridad de dos distribuciones. La regla empírica es considerar que éstas son similares cuando $D \leq 0,1$. Aunque esto

no es totalmente satisfactorio, es mejor que la alternativa de realizar una afirmación subjetiva basada en la apariencia gráfica (Olea, 2012).

Test de dependencia espacial

Para confirmar la existencia de dependencia espacial de los errores, el coeficiente de autocorrelación I de Moran es calculado usando el método descrito por Gittleman y Kot (1990). Para el llenado de la matriz de ponderaciones se usan distancias como ponderaciones; y, posteriormente, el coeficiente es calculado usando la ecuación 3.2 (Lloyd, 2007).

$$I = \frac{n * \sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{(\sum_{i=1}^n (y_i - \bar{y})^2)(\sum_{i=1}^n \sum_{j=1}^n w_{ij})} \quad (3.2)$$

De donde: w_{ij} es un elemento de la matriz de proximidad espacial (distancia ponderada), y_i es el valor del atributo analizado con media $\hat{y}$, n es el número de elementos analizados u observaciones; y, $\sum_{i=1}^n \sum_{j=1}^n w_{ij}$ es la matriz de distancias. El cálculo del índice de Moran se lo realiza mediante la función *Moran.I* de la librería *ape*.

El p -valor obtenido en el test es analizado y si este es menor a 0,05 (considerando que la significación es del 5 %), no se puede aceptar la hipótesis nula del test que afirma que no existe autocorrelación espacial, por tanto los datos exhibirían autocorrelación espacial.

Adicionalmente, para evaluar la presencia de deriva (ausencia de estacionaridad) se han emplean gráficas de dispersión direccionales. En estas gráficas se confronta el valor de los errores de la componente vertical con respecto a las coordenadas Este y Norte, de forma independiente. Estas gráficas pueden ser obtenidas directamente a través de la función *plot* de la librería *geoR*⁴.

Para detectar visualmente la presencia de sesgo y valores atípicos se pueden usar otras gráficas, tales como: histograma de frecuencias, gráficos de probabilidad normal Q-Q, diagrama de caja y bigote. A continuación se presenta una tabla resumen de las gráficas, funciones y librerías con las que son generadas.

---

⁴No se debe confundir esta función con la homónima de la instalación base de *R*.

Gráfica	Función	Librería
Histograma	hist	base
Probabilidad normal Q-Q	qqnorm	base
Diagrama de caja y bigote	boxplot	base

Tabla 3.2: Parámetros del sistema de referencia espacial empleado en el DMQ.

3.4.5 Análisis de la estructura espacial

Uno de los principales objetivos de la variografía es explorar y cuantificar la dependencia espacial de una variable, es decir, su autocorrelación espacial (Moreno Jiménez, 2006). El análisis variográfico contempla los siguientes análisis de:

1. Nube variográfica
2. Variograma experimental o empírico (omnidireccional).
3. Variograma experimental para diferentes direcciones (0, 45, 90 y 135 grados), que permite además determinar la presencia de anisotropía, a través de la comparación entre estos.
4. Anisotropía o variomapa, la posible existencia de anisotropía ha sido también analizada.

Caracterización: Ajuste del variograma empírico

Una vez que se ha realizado la revisión de la función variograma para la variable analizada, en el presente caso los errores de la componente vertical del MDE evaluado, se procede a realizar el ajuste del variograma a un modelo teórico como por ejemplo: esférico, exponencial y gaussiano entre los más comunes.

3.4.6 Simulación geoestadística

Los enfoques estadísticos basados en aleatorización y remuestreo pueden ser usados en casos en los que los datos son muestreados de distribuciones desconocidas o mixtas, donde los tamaños de muestra son pequeños, donde los valores atípicos son

un problema, o donde el diseño de un test apropiado basado en una distribución teórica es demasiado complejo y matemáticamente no manejable (Kabacoff, 2015). Es precisamente, en los datos espaciales en donde se presentan una o varias de estas condiciones.

Una vez que la estructura espacial de los errores de la componente vertical ha sido determinada, las simulaciones geoestadísticas pueden generarse a través de la función *krige* de la librería *gstat*. Esta función requiere el ingreso del número de simulaciones deseadas, el conjunto de datos, el modelo de estructura espacial, el número de observaciones más cercanas que serán consideradas; y, el mallado espacial en donde se llevará a cabo la simulación.

Las simulaciones realizadas se almacenan dentro de un objeto de R. A fin de llevar a cabo la revisión de los resultados de la simulación, se realiza la representación gráfica de una sección transversal. Lo anterior permite tener una idea de la variabilidad de los errores verticales en las simulaciones desarrolladas.

El procedimiento descrito para el diseño de la simulación será usado, de manera específico, en el caso de estudio que se describe a continuación.

3.5 Caso: Inundación del Río Garrapata

La zona seleccionada se caracteriza por frecuentemente sufrir problemas de sequías e inundaciones, lo cual produce pérdidas de los recursos hídricos y agrícolas, actualmente se lleva a cabo el proyecto Propósito Múltiple Chone, cuyas obras pretenden regular el caudal de Río Grande, abastecimiento de agua potable para Chone, riego y encauzamiento del río para la protección, contra las inundaciones periódicas que afectan en la zona.

El área de estudio abarca una franja comprendida entre las cotas 40.25 msnm en la parte alta y la cota 17.16 msnm en la parte baja en donde se asientan la población de San Andrés, la cual se tomará de referencia para valorar las pérdidas de la mala calidad debido a los errores en el MDE en el tramo seleccionado.

El modelo hidráulico unidimensional que se ha llevado a cabo, utiliza información del estudio hidrológico de la subcuenca del Río Garrapata, en el que se han determinado

los caudales de retorno de 5, 10, 25, 50 y 100 años, aplicando el Método del Número de Curva, para la transformación precipitación - escorrentía. Información generada en el proyecto de Sistema de Alerta Temprana de Control de Inundaciones en la cuenca del Río Chone. Por su parte, la información cartográfica es a partir del MDE y ortofotografías 1:5.000 obtenida del programa SIGTIERRAS.

3.5.1 Descripción del área de estudio

El área de estudio comprende una extensión de 1166,78 hectáreas, ubicada en la población de San Andrés, al norte de la ciudad de Chone, provincia de Manabí, región Costa del Ecuador. Es una zona caracterizada por sus condiciones climáticas extremas, con precipitaciones anuales que eventualmente superan los 3.500 mm en condiciones como las del fenómeno El Niño y períodos de lluvias extraordinarias como lo ocurrido en marzo del 2008, y lluvias menores a 500 mm en años con características de baja humedad. La variación estacional normal en la zona determina un período de 5 meses (enero a mayo) donde se concentra en promedio, aproximadamente el 85 % de la precipitación anual, mientras en el período junio a diciembre se observan 7 meses de bajas precipitaciones, incluso se registran meses con valores de 0 mm (Moreira, 2016).

Debido a estas características existe una frecuente escasez de agua y problemas de inundaciones. De acuerdo al estudio elaborado por la Banco de Desarrollo de América Latina (CAF) en el 2000, durante el desarrollo del Niño en los años 97-98, se produjeron precipitaciones excepcionales que correspondieron a períodos de retorno que superaron significativamente los 100 años.

En la Figura 3.6 se muestra el territorio que comprende la Cuenca del Río Chone donde se encuentra el área de estudio del caso.

El Río Garrapata junto con los ríos Mosquito y Grande, forman parte del sistema hidrológico que confluye en el Río Chone, en el transcurso se une con otros afluentes como el río Carrizal para desembocar en Océano Pacífico. El paisaje está constituido por materiales sedimentarios terciarios, que han dado lugar a relieves colinados medios a altos, sobre los que se han desarrollado suelos de texturas arcillosas y, valles y llanuras aluviales, de topografía plana donde los suelos de texturas limosas y arcillosas, dominan el medio. El uso de las tierras está dominado por la presencia

Figura 3.6: Ubicación general del área de estudio caso Río Garrapata

de pastos plantados y naturales, cultivos de ciclo corto, arboricultura tropical, plantaciones permanentes y pequeñas manchas de vegetación arbórea y arbustiva muy intervenidas.

Los peligros de inundaciones son altos e inciden en las poblaciones que se encuentran asentadas cerca de las márgenes de los ríos Garrapata, Mosquito y Grande que evidencian la falta planificación del territorio, ya que un alto porcentaje de pobladores carecen de los servicios básicos y sus viviendas han sido construidas en zonas altamente vulnerables a las inundaciones durante el período lluvioso (Gutiérrez, Góngora y Melo, 2008).

3.5.2 Herramientas para el modelado de inundaciones

Se han elegido HEC-RAS y HEC-GeoRAS para la determinación de las áreas de inundación debido a su uso extendido en el campo del modelamiento hidráulico. HEC-RAS trabaja con un modelo unidimensional el cual asume que la longitud del cause es la dimensión que prevalece, siendo x la distancia desde el principio del cause de la sección transversal e y es la elevación por encima de una cota de referencia. Así la información hidráulica y topografica es asignada por secciones.

HEC-RAS, permite realizar cuatro tipos de análisis en ríos:

- Modelización de flujo en régimen permanente
- Modelización de flujo en régimen no permanente
- Modelización del transporte de sedimentos
- Análisis de calidad de aguas

Para el caso del tramo del Río Garrapata analizado, se realizó una modelización de flujo en régimen permanente debido a que esta opción ofrece buenos resultados por ejemplo en inundaciones donde los cambios en la profundidad y la velocidad en un cierto punto dado ocurre de forma muy lenta (Haestad Methods Water Solutions, 2007).

El procedimiento para la determinación de áreas de inundación con HEC-RAS, HEC-GeoRAS y ArcGis consta de los siguientes pasos:

1. **Preproceso** Consiste en la preparación de los datos con el uso de HEC-GeoRAS y ArcGis para generar la información geométrica del cause y de las secciones transversales, que luego son importadas HEC-RAS en un solo archivo que contiene toda la información a ser modelada.
2. **Modelización** Trabajo con HEC-RAS para la simulación y generación de un archivo de exportación para ArcGis.
3. **Postproceso** Con los resultados de la simulación del modelo hidráulico, en ArcGis y HEC-GeoRAS es posible abrir los archivos generados por HEC-RAS de manera que se puede visualizar los resultados de las superficies de inundación en formato ráster y vector.

La Figura 3.7 muestra el esquema del procedimiento propuesto en el manual de HEC-GeoRAS para la determinación de la áreas inundadas.

Pre proceso o Preparación de los datos

En esta fase se realiza una revisión de estudios hidrológicos existentes en la zona, principalmente se usan los estudios de (Gutiérrez, Góngora y Melo, 2008) y (Moreira, 2016), a fin de obtener los parámetros para el cálculo del modelo hidráulico.

En cuanto la geometría del canal, ésta fue creada siguiendo el procedimiento propuesto en la Figura 3.7 a partir de cartografía existente. Para el presente caso los insumos utilizados fueron: Ortofotografía y MDE del área de estudio, información obtenida del centro de descarga del SIGTIERRAS.

Siendo el MDE, uno de los insumos fundamentales para la determinación de áreas de inundación, pues a partir de esta información de las alturas se derivan los datos para establecer los niveles de la inundación. La Tabla 3.3 que muestra las siguientes especificaciones.

Figura 3.7: Esquema del procedimiento para determinación de áreas de inundación. Tomado de Ackerman (2009)

GSD Ortofoto	Exactitud altimétrica del MDE	Paso de malla	Tipo de Representación	Sistema de Referencia de Coordenadas	Formato
40 cm	1,5 m	4mx4m	grid	EPSG:32717	tiff

Tabla 3.3: Especificaciones del MDE que se utiliza como insumo para la estimación de áreas de inundación. Tomado de metadato MDE, de SIGTIERRAS

HEC-GeoRAS permite trabajar en dos formatos: ráster y vector (TIN). Para el caso de la TIN, ésta se construyó a partir del MDE con el uso de la herramienta *ArcToolbox* de ArcGis

El siguiente paso es la creación de la geometría del cauce en HEC-GeoRAS que posteriormente será exportada a HEC-RAS, en un archivo que incluya los datos de geometría del cauce con los atributos obtenidos del MDE. En la Figura 3.8 se muestra el tramo del Río Garrapata que es analizado para determinar las áreas de inundación.

Modelización

Para la determinación de los niveles de crecida de agua en el tramo del Río Garrapata de las secciones de interés, el modelo hidráulico fue realizado en HEC-RAS, para lo cual se creo un proyecto que contenía:

1. La geometría del río: datos importados de HEC-GeoRAS.
2. Parámetros hidráulicos: condiciones de contorno.
3. Simulación del flujo

Geometría. El coeficiente de de rugosidad de Manning considerado para el cauce principal del río Garrapata es $n = 0,030$ (Cauce de planicie, limpio y tramo recto) y para la llanura de inundación $n = 0,045$ (Cauce de llanura de inundación, arbustos, pastos y cultivo)

Figura 3.8: Área de estudio del caso. Geometría del tramo del Río Garrapata analizado en el sector de San Andrés

La Figura 3.9 muestra la geometría del tramo del río seleccionado, junto con las secciones transversales y sus datos asociados.

Figura 3.9: Representación en HEC-RAS de la geometría del río con las secciones transversales y una los parámetros asociados a la sección transversal de la estación 1600

Parámetros hidráulicos. En la Tabla 3.4 se muestran los caudales máximos de crecida para periodos de retorno de 5, 10, 25, 50 y 100 años.

Periodo de retorno (T)	5 años	10 años	25 años	50 años	100 años
Caudal (m3/s)	199	248	311	358	405

Tabla 3.4: Caudales máximos en la zona estudiada. Tomado de (Gutiérrez, Góngora y Melo, 2008)

La capacidad de flujo para condiciones sin desborde y el nivel de crecida para el caudal máximo estimado con período de retorno de 100 años, es mostrada en Tabla 3.5, información que es importante al momento de calibrar el modelo

hidráulico, de acuerdo al Estudio Hidrológico de las cuencas Garrapata y Mosquito (Moreira, 2016).

Estación	Caudal sin desborde (m³/s)	Nivel sin desborde (m.s.n.m)	H (m)	V med (m/s)	Froude
H230 Garrapata AJ Chone	59,35	21,7	4,59	1,02	0,21

Tabla 3.5: Caudales máximos en la zona estudiada. Tomado de (Moreira, 2016)

Como se muestra en la Tabla 3.6 la capacidad hidráulica del río Garrapata es limitada por las características físicas del cauce como es la pendiente, ancho y profundidad. Así, con caudales de crecida de períodos de retorno menores a 5 años provocarán el desbordamiento del río, razón por la cual es una zona que periódicamente sufre desbordamientos e inundaciones.

Estación	Caudal máximo (m³/s)	Nivel máximo (m.s.n.m)	H (m)	V med (m/s)	Froude
H230 Garrapata AJ Chone	405	24,37	7,26	1,1	0,22

Tabla 3.6: Caudales máximos en la zona estudiada. Tomado de (Moreira, 2016)

En el caso de la simulación de flujo permanente, se procede mediante la opción *Steady Flow Data Editor*. Los datos necesarios son los siguientes:

1. Número de perfiles que serán calculados en este caso los períodos de retorno (T)
2. Datos del flujo, caudal de los períodos de retorno (T), que se muestran en la Tabla 3.4
3. Condiciones de contorno, mediante la opción *Reach Boundary Conditions*, se escogen los tipos: a) Nivel de agua conocido (indicar calado más la cota del fondo del cauce); b) Calado crítico (no requiere datos adicionales); Calado

uniforme (indicar la pendiente del tramo en cuestión); Curva de Gasto (útil si existe alguna sección de control con una determinada relación entre calado y caudal).

La pendiente (S) del cauce principal en el tramo de la modelación se ha considerado igual a 0.000415 m/m.

Simulación del flujo. Con la información de geometría y flujo del cauce, se procede a realizar el cálculo que para el caso corresponde a un análisis de flujo permanente con la opción *Run/Steady Flow Analysis*, se generará un plan de simulación que incorpore los archivos de datos de la geometría y de los datos hidráulicos.

Post-proceso

Una vez que se han exportado los resultados de la simulación del modelo hidráulico en HEC-RAS, el proceso de visualización y tratamiento de los resultados de las áreas de inundación se lo realiza en HEC-GeoRAS. La opción *RAS Mapping* permite configurar e importar resultados de la modelación.

En esta fase se sigue el procedimiento referido en la Figura 3.7:

1. Se configura el archivo de salida, así como el MDE (TIN o GRID) que se empleará como base para la obtención de los resultados en el SIG.
2. Lectura y procesamiento de los datos que han sido importados desde HEC-RAS.
3. Se seleccionan los perfiles correspondientes a los caudales que se haya establecido en las condiciones de flujo en los períodos de retorno que se quiere analizar el área de inundación y se genera un TIN para cada perfil seleccionado.
4. Con la opción *RAS Mapping / Inundation Mapping / Floodplain Delineation*, se crean las zonas de inundación para cada perfil.
5. Se visualizan de resultados con las herramientas disponibles en ArcGis para 3D.

3.5.3 Cálculo de errores de la componente vertical

El procedimiento para el cálculo de errores de la componente vertical y el análisis geoestadístico es similar al detallado en el ensayo; el mismo que es resumido en la Figura 3.10.

Figura 3.10: Procedimiento resumen de cálculo de errores de la componente vertical y análisis geoestadístico para la ejecución de simulaciones.

Cálculo de probabilidad de errores

Una vez que las simulaciones han sido generadas, es posible obtener funciones de distribución de probabilidad de los errores de la componente vertical. El enfoque de uso de funciones de distribución de probabilidad de los errores de la componente vertical, dado un intervalo crítico de errores es el propuesto en la presente tesis.

En la Tabla 3.7 se muestran los intervalos de error en las alturas considerados.

Intervalo de error (m)	Tipología de errores
Mayores a 1,5	Exceso o sobreestimación
Menores a -1,5	Defecto o subestimación
Mayores a 1,5 y menores a -1,5	Sobreestimación y subestimación
Mayores a 2	Exceso o sobreestimación
Mayores a -2	Defecto o subestimación
Mayores a 2 y menores a -2	Sobreestimación y subestimación
Mayores a 3	Exceso o sobreestimación
Menores a -3	Defecto o subestimación
Mayores a 3 y menores a -3	Sobreestimación y subestimación

Tabla 3.7: Intervalos de error.

En todos los casos se realiza una clasificación de áreas en donde se espera obtener valores de error de la altura en el intervalo requerido. Esto se lleva a cabo mediante el uso de la función *ifelse* seguido de la condición (intervalo) que se desea evaluar. Luego, se calcula el valor medio de cada celda del conjunto de simulaciones creadas mediante la función *apply* y el parámetro $MARGIN = 1$, el cual permite realizar el cálculo de la media por filas, registros (celdas).

Cabe mencionar que los valores de error de la altura se encuentran almacenados en un objeto con estructura matricial en la que los errores para cada simulación se almacenan en columnas, por lo que las filas representan las celdas del ráster del área estudiada.

A modo de ejemplo se incluye la sintaxis empleada para la acción descrita.

```
apply(ifelse((errores >= 1,5), 1, 0), 1, mean)
```

En donde, *errores* constituye el objeto de *R* en el cual se encuentra almacenada la información correspondiente a las simulaciones de error de la componente vertical de los MDE.

Una vez que los objetos son generados, son exportados a formato GeoTiff, mediante la función *writeRaster*, para su posterior uso en el software ArcGis.

Con los mapas de probabilidad de error de la componente vertical, inundaciones (con calados) con un período de retorno dado; y, de valoración de suelo, se procede

a realizar álgebra de mapas mediante la herramienta Model Builder de ArcGis, de la manera que a continuación se describe.

En síntesis, se realiza una multiplicación entre el mapa de inundaciones (el cual es previamente reclasificado, a través de la herramienta *Reclassify* en dos clases: 1) calados mayores o iguales a 0,5 metros (valor de 1); y, 2) menores a este umbral (valor de 0)) con el mapa de probabilidad de error de exactitud vertical.

3.5.4 Valoración del suelo y edificaciones en el área de estudio

Con el resultado anterior, nuevamente se realiza una multiplicación con el mapa de valoración del suelo, en el cual cada polígono o zona posee el valor monetario del suelo (dólares/hectárea).

Para fines de comparación, se usan mapas de inundaciones con períodos de retorno de 5, 10, 25, 50 y 100 años.

Las zonas del mapa de valoración han sido delimitadas y caracterizadas apoyándose en los conceptos de zonas agroeconómicas homogéneas (ZAH). No se considera los valores a nivel predial provenientes del catastro, debido a que en la zona no se cuenta con dicha información.

La valoración tomó en cuenta lo estipulado en el artículo 495 del Código Orgánico de Organización Territorial, Autonomía y Descentralización (COOTAD) de Ecuador, en el que se reglamenta que: *“El valor de la propiedad mediante la suma del valor del suelo y, de haberlas, el de las construcciones que se hayan edificado sobre el mismo”*

Los valores de las zonas han sido calculados en referencia a la propuesta de normativa técnica de valoración catastral predial de suelo rural planteada por Alcázar-Molina (2014) para el Programa SIGTIERRAS en cantones cercanos a la zona de estudio.

Donde las consideraciones fueron las siguientes: El valor del suelo es determinado por un proceso de comparación con precios unitarios de venta de inmuebles de condiciones similares u homogéneas del mismo sector multiplicado por la superficie del inmueble (Alcázar-Molina, 2014). Para el caso del análisis se consideraron las

zonas en su conjunto conociendo el valor por hectárea de zonas con características homogéneas.

Mientras que para estimar el valor de las construcciones se utilizó el método del coste, el valor de las edificaciones, que es el precio de las construcciones que se hayan desarrollado de carácter permanente sobre un inmueble, calculado sobre el método de reposición (Alcázar-Molina, 2014).

Dentro del ámbito rural lo anterior se puede resumir en la ecuación 3.3.

$$V_p = V_{suelo} + V_{const} \quad (3.3)$$

:

donde: V_{suelo} es el valor del suelo

V_{const} es el valor de la construcción

Así, la valoración de edificaciones existentes en cada una de las zonas, fueron calculadas en referencia a los valores del mismo informe de valoración masiva de predios que se muestra en la Tabla 3.8. Las tipologías más frecuentes en las zonas corresponden a la primera clase.

Tipologías Edificaciones (unidades)	Valor por m2
1 Madera +Madera + Zinc	\$ 48,56
2 Hormigón Armado +ladrillo o bloque + Zinc	\$ 114,69
3 Hormigón Armado +ladrillo o bloque + Asbesto+cemento	\$ 124,20
4 Hormigón Armado +ladrillo o bloque +Losa de	\$ 178,35
5 Paredes Soportantes +Ladrillo o Bloque + Zinc	\$ 80,20
6 Otras tipologías	\$ 67,92

Tabla 3.8: Valor referencial para edificaciones. Tomado de SIGTIERRAS 2016

La descripción y valoración económica de las zonas del mapa de valoración se presentan en la Tabla 3.9.

Cod	Zona	Descripción	Valor referencial por ha
0	ZR	Río Garrapata	
1	Z1	Centro poblado, caracterizado por construcciones de diversos materiales	\$71.497,57
2	Z2	Zona definida por tener un uso principalmente para el cultivo de arroz, zona propensa a inundaciones debido a su geoforma. La viabilidad del sector es buena a pesar de que muchas veces no son transitables en épocas lluviosas.	\$6.001,52
3	Z3	Zona de uso agrícola donde existen cultivos de ciclo corto	\$3.000,97
4	Z4	Zona de uso principalmente agrícola, existiendo mayor diversificación de cultivos entre ellos limón, plátano, cacao, caña de azúcar, banano, mango y de ciclo corto (maíz y yuca)	\$4.003,00
5	Z5	Zona de uso agrícola donde predomina el cultivo de cacao y mandarina	\$5.002,16
6	Z6	Zona de uso principalmente agrícola, cultivos de ciclo corto como el maíz y la yuca, huertos y pastos para uso pecuario. Existen contrucciones.	\$6.004,09
7	Z7	Zona de uso habitacional, en estas extensiones si bien puede existir algún tipo de cultivo este es principalmente utilizado para autoconsumo familiar y no representa algún tipo de ingreso económico.	\$25.603,49
8	Z8	Zona caracterizada por matorrales, bambú y vegetación herbácea	\$700,69
9	Z9	Zona dedicada a la cria de aves de corral, donde existen infraestructuras dedicadas a la actividad avícola.	\$12.202,29

Tabla 3.9: Valoración de la zonas al interior del territorio analizado. Los valores han sido establecidos tomando como referencia al informe de valoración masiva de predios rurales de un cantón cercano a la zona de estudio. Tomado Alcázar-Molina (2014), SIGTIERRAS (2016)

3.5.5 Cualificación de pérdidas en términos monetarios

A fin de realizar el cálculo, se procede a ejecutar una rasterización del mapa de valoración de suelo, que se encuentra en formato vectorial. El valor asignado a cada

celda corresponde al valor unitario multiplicado por un factor 16/10000 (dado que la celda tiene una dimensión de 4 x 4 metros).

El mapa de valoración de suelo, permite tener una idea de las posibles pérdidas ocasionadas por un evento de inundación del río Garrapata, mediante una ponderación, considerando las diferentes probabilidades de presencia de errores de la componente vertical. De esta manera, no se obtiene un límite único para áreas con un determinado error, sino mas bien escenarios probabilísticos, que incluso para efectos de planificación y sociabilización, permiten obtener mayor credibilidad.

La Figura 3.11 presenta el algebra de mapas descrito.

Figura 3.11: Algebra de mapas propuesto para la obtención de costes por la existencia de errores de la componente vertical en un MDE, caso de estudio Río Garrapata. Creado a través de la herramienta Model Builder de ArcGis.

Referencias

- Ackerman, C (2009). *HEC-GeoRAS User's Manual*.
- Alcázar-Molina, Manuel (2014). "Propuesta de normativa técnica de valoración catastral predial del suelo rural". Bloque temático No.2.
- Elosua, Paula (2011). "Introducción al entorno R". En: *Bilbao: Euskal Herriko Unibertsitateko Argitalpen Zerbitzua/Servicio Editorial de la Universidad del País Vasco*.
- Gittleman, John L y Mark Kot (1990). "Adaptation: statistics and a null model for estimating phylogenetic effects". En: *Systematic Biology* 39.3, págs. 227-241.
- Gutiérrez, C, E Góngora y P Melo (2008). *Estudio Hidrológico de Inundaciones en la Cuenca Alta del Río Chone (Garrapata, Mosquito y Grande)*. Inf. téc. INAMHI, CLIRSEN y FAO.
- Haestad Methods Water Solutions David Klotx, Adam Strafaci Annaleis Hogan Kristen Dietrich (2007). *Floodplain modeling using HEC-RAS*. First edition. Bentley Institute Press. ISBN: 1934493023,978-1-934493-02-1,9781628709483,1628709480. URL: <http://gen.lib.rus.ec/book/index.php?md5=a9a6224964ac28ff889177f7bd4bb53e>.
- Kabacoff, Robert (2015). *R in action: data analysis and graphics with R*. Manning Publications Co.
- Kolaczyk, Eric D y Gábor Csárdi (2014). *Statistical analysis of network data with R*. Springer.
- Leon, Javier X., Gerard B. M. Heuvelink y Stuart R. Phinn (2014). "Incorporating DEM Uncertainty in Coastal Inundation Mapping". En: *PLoS ONE* 9.9, págs. 1-12. DOI: [10.1371/journal.pone.0108727](https://doi.org/10.1371/journal.pone.0108727). URL: <http://dx.doi.org/10.1371%2Fjournal.pone.0108727>.
- Lloyd, Christopher D (2007). *Local models for spatial analysis*. CRC/Taylor & Francis London.
- Luke, Douglas A (2016). *A user's guide to network analysis in R*. Springer.
- Moreira, Marco (2016). *Estudio de Impacto Ambiental del Proyecto Propósito Múltiple Chone: Estudio Hidrológico de las Cuencas Garrapata y Mosquito*. Inf. téc. Empresa Pública del Agua, ESPE-Innovativa.
- Moreno Jiménez, Antonio (2006). *Sistemas y análisis de la información geográfica: manual de autoaprendizaje con ArcGis*.

- Newman, Mark (2010). “Networks: an introduction”. En: *United States: Oxford University Press Inc., New York*, págs. 1-2.
- Olea, Ricardo (2012). *Geostatistics for engineers and earth scientists*. Springer Science & Business Media.
- Thode, Henry C (2002). *Testing for normality*. Vol. 164. CRC press.
- Wechsler, SP (2007). “Uncertainties associated with digital elevation models for hydrologic applications: a review”. En: *Hydrology and Earth System Sciences* 11.4, págs. 1481-1500. URL: <https://hal.archives-ouvertes.fr/hal-00305087/document>.

Capítulo 4

RESULTADOS Y DISCUSIÓN

En este capítulo se presentan los resultados obtenidos en la presente investigación.

4.1 Resultados de la selección de la aplicación y caso de estudio

Como se indicó en el capítulo anterior, se han utilizado diversas fuentes de información para conocer qué campos de aplicación de los MDE son los más usuales y para, a partir de ahí, seleccionar el caso de uso sobre el que trabajar en esta investigación.

4.1.1 Resultados de la encuesta sobre usuarios y usos de MDE en España

Como resultado de la caracterización de los usuarios y usos de los MDE que se descargan del CNIG de España se encontró que el 70.74 % de los usuarios tienen un perfil profesional y además el 82 % de los usuarios españoles clasificaron la utilidad de los MDE como alta o muy alta; destacan las aplicaciones relacionadas con la cartografía y medio ambiente con un 60,57 % y un segundo grupo compuesto por aquellas que están en el entorno al 42,62 %, que son: hidrología, educación y

Figura 4.1: Gráficos de redes de conexión entre las áreas temáticas.

En la Tabla 4.1 se presentan las frecuencias totales obtenidas para cada área temática y porcentajes calculados respecto a los 12.493 registros, tomando en cuenta que el usuario que realiza una descarga puede marcar más una categoría, no obstante se muestra el porcentaje respecto a las descargas totales.

ID	Áreas temáticas	%
CART	Cartografía	35.25
MA	Medioambiental	25.32
EF	Educación y Formación	9.83
F/B	Forestal/Biodiversidad	9.80
OTL	Ocio y Tiempo Libre	7.76
HY	Hidrología	7.75
OTU	Ordenación del Territorio y Urbanismo	7.48
IIC	Infraestructuras e Ingeniería Civil	6.15
GEO	Geología	5.76
AG	Agricultura	4.95
AET	Arqueología y Estudios Temporales	4.58
I+D	Investigación, Ciencia e Innovación	4.57
T	Turismo	3.44
O	Otros	3.11
CA	Catastro	2.72
EDA	Suelos (Edafología)	2.35
ERM	Energía y Recursos Minerales	1.63
FAU	Fauna, Entomología/Biodiversidad	1.51
ZR/PC	Zonas de Riesgo/Protección Civil	1.47
GEOF	Geofísica	1.41
AC	Artística y Cultural	1.38
CC	Cambio Climático	1.38
NL	Navegación y Localización	1.29
OC	Oceanografía/Costas	1.11
V	Vivienda	1.03
TL	Transporte y Logística	1.02
TC	Telecomunicaciones	0.93
DS	Defensa y Seguridad	0.87
DEU	Demografía, Expansión Urbana	0.77
S	Sociología	0.55
JD	Jurídico	0.32
SN	Sanidad	0.22
MK	Marketing (Mercadotecnia)	0.17

Tabla 4.1: Frecuencias totales de áreas temáticas marcadas por usuarios encuestados.

Cabe indicar que, aproximadamente el 73 % de los usuarios marcan una única área temática, casi el 97.7 % de los usuarios está incluido en el grupo formado por las cinco primeras categorías. Hay casos de usuarios que llegan a marcar un número muy alto de las opciones éstas han sido incluidas en un solo grupo que corresponde al 2.3 %.

4.1.2 Resultados de la encuesta sobre usuarios y usos de MDE a nivel mundial

Los resultados de la encuesta a nivel mundial de los usuarios especializados en los MDE, aportan importante información que ha servido para la selección del caso de aplicación, así como para conocer las demandas de calidad de los MDE en un grupo especializado de usuarios.

Perfil de usuarios

En cuanto a la información sobre el perfil de usuario recogida en la esta encuesta, se ha encontrado que de alguna forma coincide con los resultados del perfil de usuarios de España y los campos de aplicación encontrados en el registro de descargas del CNIG.

En esta encuesta se pudo conocer los perfiles de usuarios avanzados, los cuales tienen un carácter más científico y demandas más específicas en el caso de los temas relacionados con la calidad de los MDE, lo cual se evidencia en que el 63.8 % corresponde a los perfiles relacionados con la enseñanza e investigación (suma de perfil profesor, investigador y estudiante), en la Tabla 4.2 se muestran los resultados de las preguntas 1, 2 y 3 relacionadas con el perfil de usuario.

Perfil de usuario	%	Perfil de la Organización	%	Años de experiencia	%
Profesor de universidad (profesor, emerito, etc)	27,5	Universidad	47	Más de 4	47
Investigador	2,5	Agencias Nacionales de Cartografía	23,5	Entre 1 y 4	23,5
Profesional (Geomatica)	24,4	Instituciones de investigación	9,8	Menos de 1	9,8
Profesional (Otros campos)	10,8	Agencias ambientales	9,8		
Estudiantes (PhD, master, etc.)	8,8	Compañías	5,9		
Otros	1	Profesionales	2		
P1		Otros	2		
		P2		P3	

% del total de encuestados, N=102

Tabla 4.2: Perfil de usuarios especializados de los MDE

Campos de aplicación y productos derivados

La Tabla 4.3 muestra los resultados de la encuesta respecto a los campos de aplicación del uso de los MDE, los cuales son clasificados en 4 grupos, donde destacan las aplicaciones relacionas con estudios hídricos y redes de drenaje.

Propósito	%
Estudios hídricos y redes de drenaje	19,51
Fotogrametría, LiDAR, Láser escáner, UAV, Teledetección, etc.	18,7
Estudios medioambientales	14,63
Producción de Modelos Digitales de Elevación	8,13
Geología	6,5
Evaluación y estudio de cultivos agrícolas	4,88
Evaluación de desastres, protección civil, etc.	4,07
Ciencia de suelos y botánica	4,07
Ingeniería Civil	2,44
Glaciología	2,44
Vulcanología	2,44
Impresión 3D	1,63
Arquitectura	1,63
Geodesia	1,63
Otros	7,31

Número de encuetados = 102,

respuestas contestadas = 98 , malas respuestas = 3

Tabla 4.3: Campos de aplicación de los MDE. Tomado de Ariza-López et al. (2017)

Para conocer sobre los productos derivados que más son utilizados por los usuarios especializados, se analizaron los resultados de las preguntas 5 y 6 relacionadas con los productos derivados de los MDE y los modelos de datos del MDE.

En la Tabla 4.4 se muestran en resumen los productos derivados del MDE, en orden de importancia, correspondiente a la pregunta 5. Los productos clásicos superan el 30 %.

Productos derivados del MDE	%
Elevación	75,5
Pendiente	71,6
Orientación	52,9
Delineación de la red de drenaje	51,0
Curvas de nivel	48,0
Acumulación de flujo	46,1
Análisis de cuencas visuales	44,1
Área de capatación	36,3
Detección de canales	33,3
Sombras	31,4
Curvatura	29,4
Índice de humedad topográfica	28,4
Rugosidad	27,5
Concavidad y convexidad	26,5
Detección de divisorias	26,5
Generación de parámetros del modelo	23,5
Otros	11,0
Análisis de costo	10,8

Número de encuestados = 102,
Respuestas con contestación = 102

Tabla 4.4: Frecuencias totales de los productos derivados de los MDE, marcadas por usuarios expertos encuestados

El 91,2 % de usuarios indica que el modelo de datos más usado es el raster, un modelo predominante frente a la TIN con el 30,4 %, las curvas de nivel con el 26,5 % y los perfiles con el 12,7 %, esto puede ser debido al que la mayoría de los MDE están disponibles en el modelo raster.

Aspectos de la calidad

Las preguntas 10 a la 18 estaban orientadas a conocer aspectos relacionados con la calidad

En la Tabla 4.5 se muestran en porcentaje la principal exigencia del cálculo de las aplicaciones en función de los productos derivados (pregunta 10) de los cuales se han seleccionado los 9 primeros en grado de importancia. Las dos variables guardan relación y los resultados son muy similares. Siendo la altura y la pendiente las principales exigencias en el cálculo.

Producto derivado / Exigencia de cálculo (%)	Z	Pendiente	Orientación	Volumen	Curso de agua	Insolación	Cuenca visual	Cuenca hidrográfica
1 Elevación	72,7	45,5	23,4	22,1	35,1	20,8	26,0	31,2
2 Pendiente	72,6	52,1	27,4	21,9	39,7	23,3	27,4	35,6
3 Orientación	74,1	50,0	33,3	27,8	35,2	20,4	25,9	33,3
4 Delineación Red de Drenaje	59,6	42,3	15,4	19,2	46,2	32,7	40,4	53,8
5 Curvas de nivel	79,6	55,1	28,6	20,4	38,8	14,3	20,4	30,6
6 Acumulación de Flujo	63,8	44,7	17,0	27,7	48,9	29,8	36,2	46,8
7 Análisis de cuencas visuales	57,8	46,7	15,6	22,2	51,1	28,9	35,6	55,6
8 Área de captación	56,8	51,4	21,6	29,7	56,8	32,4	40,5	48,6
9 Detección de canales	61,8	50,0	20,6	32,4	55,9	26,5	32,4	41,2

Tabla 4.5: Frecuencias totales entre productos derivados MDE y exigencias de cálculo de variables y análisis

La preguntas 11, 12, 13 y 14 aportan información valiosa para conocer sobre la manera de evaluar, los índices empleados para expresar la calidad, la satisfacción con los mismos y cómo se podrían mejorar.

Los resultados de cómo los usuarios están acostumbrados a expresar la exactitud muestran que el 47 % usa un enfoque estadístico, frente al 26,5 % de los usuarios que usan juicios subjetivos (por ejemplo, inspección visual). Por otra parte el 11 % de los casos, los usuarios de los MDE no tienen preocupación por la calidad de los datos.

El uso de índices para expresar la calidad es un elemento clave para entender las necesidades reales de los usuarios. Los resultados muestran que el 28,43 % de los encuestados utiliza índices relacionados con la exactitud posicional siendo estos los más usados en la evaluación de la calidad de los MDE y el 29,41 % indica que no usa índices para expresar la calidad. Sin embargo, el 42,16 % muestra una tendencia a utilizar índices con una mejor descripción de la calidad que las medidas tradicionales basadas sólo en la exactitud posicional. En la Tabla 4.6 se muestran los resultados.

Opción: tipos de índices para expresar calidad	%
Sólo usas una medida de exactitud posicional	28,43
Ninguna	20,59
Exactitud posicional más otra medida de manera conjunta pero independientemente	17,65
Consideras medidas relacionadas con los resultados	16,67
No responde	8,8
Dependiendo del caso	3,92
Construyes un índice compuesto con ambas	3,92

% del total de encuestados, N=102

Tabla 4.6: Índices empleados para expresar la calidad en los MDE

Respecto a los índices 49 % de los encuestados indican que están satisfechos. En la pregunta sobre algo que mejorar, el 31,37 % los encuestados no respondió a la cuestión y el 14,71 % de los encuestados dio respuestas no directamente relacionadas con esta pregunta. El 3.92 % de las respuestas indican que no se necesita mejorar.

Además, el 2,94 % de los encuestados indicó que no tienen criterio para responder a esta pregunta. En relación a aquellas respuestas directamente relacionadas con esta pregunta, es interesante notar que la mayoría de los encuestados están exigiendo más información, no mejores medidas o índices. La demanda de una mejor información sobre exactitud vertical (Z) con un 17,65 y planimétrica (XY) con el 12,75 son las dos respuestas con los mayores porcentajes.

Además existe una preocupación (7.84 %) sobre el control del terreno (número de puntos de control, tamaño de la muestra, distribución, entre otros). Por otra parte, en otras respuestas obtenidas de esta pregunta se relacionan con metadatos.

Aspectos de valoración de la mala calidad

En la pregunta 16 se abordó la influencia de la mala calidad en los MDE, donde el 30.39 % de los encuestados no respondió y el 13.73 % no están directamente relacionadas con esta pregunta. Las respuestas con mayor porcentaje indican problemas de estimación con el 24,51 %, independientemente del parámetro (por ejemplo, pendiente, orientación, volumen, etc.). El 22,55 % de las respuestas indican problemas causados por los datos (por ejemplo, reelaboración, confusión, ubicación, etc.) y el 15,68 % de las respuestas indican la influencia de la mala calidad sobre las decisiones, los costos y los peligros. Finalmente, el 5,88 % de las respuestas indican que la mala calidad no es relevante para los resultados.

Desde una perspectiva cualitativa y cuantitativa la valoración de la influencia de la mala calidad de los MDE fue abordada en las preguntas 17 y 18. Los resultados se muestran en la Tabla 4.7.

Influencia de la mala calidad (P17)	%	Monetarización de la consecuencias (P18)	%
Alto	56,9	No responden	45,63%
Medio	30,4	Ninguna	23,30%
Bajo	7,8	No es posible cuantificar	23,30%
Ninguna	2	x1.000.000 USD	2,91%
No responde	2,9	x100.000 USD	0,97%
		x1.000 USD	0,0097
		x100 USD	0,0097
		x10 USD	0,0097
		≈0 USD Insignificante	0,0097

(1) N=102; (2) N = 103 (Múltiples respuestas son permitidas)

Tabla 4.7: Valoración de la influencia de la mala calidad en los MDE, preguntas 17 y 18

Desde una perspectiva económica, las preguntas 17 y 18 proponen el mismo tema. En la pregunta 18 la mayoría de los encuestados, el 45,63 %, no responde, esta pregunta en relación a las otras de la encuesta es la menos respondida. Esto significa que los usuarios realmente no son capaces de dar una respuesta. No tienen una idea cuantitativa sobre el impacto económico de la mala calidad. Usan rutinariamente los conjuntos de datos del MDE, pero no conocen las consecuencias económicas de

los datos deficientes. Esto está de acuerdo con los resultados obtenidos para en la pregunta 16 (opción sin respuesta) pero en este caso el porcentaje es más alto porque no se exige una explicación sino una cifra. Además, el 23,30 % de los encuestados indican que el uso de los MDE de mala calidad no tiene consecuencias económicas. Esta respuesta es muy contradictoria con respecto a pregunta 17.

La única justificación que se encuentra es que la palabra “influencia” de la pregunta 17 se entiende como “influencia en los resultados” que los resultados de mala calidad no tienen consecuencias económicas para estos usuarios. Un porcentaje similar el 23.30 % de los encuestados no cuantifica, pero explica las consecuencias. Esto significa el reconocimiento de la influencia de mala calidad pero también la incapacidad de traducir problemas de mala calidad a dinero. Sólo el 7 % de los encuestados son capaces de monetizar las consecuencias de la mala calidad dada un orden de magnitud o rango del impacto económico.

En este contexto, con los resultados obtenidos en la encuesta internacional se puede concluir que, a los encuestados en su mayoría les cuesta cuantificar en términos monetarios la mala calidad en los MDE que son empleados en las diversas aplicaciones. Por tal motivo, esta investigación pretende ser un primer acercamiento para relacionar la mala calidad con las pérdidas en términos monetarios en un caso concreto del uso de los MDE en para determinar áreas de inundaciones.

4.1.3 Resultado del análisis de las consultas a ScienceDirect

Como resultado de las consultas realizada a ScienceDirect^{TM} y el proceso de normalización, se obtuvieron 60 nodos (21 temas y 39 tópicos) y 222 aristas. Tanto los temas y tópicos están en Inglés, ya que la base de datos se encuentra en ese idioma y no se ha realizado una traducción con el fin de evitar ambigüedades en los términos que fueron consultados.

El gráfico de análisis de la red bipartida se muestra en la Figura 4.2, donde además se calcularon los parámetros de la red con el fin de analizar con detalles las relaciones. En la Tabla 4.8 se muestran las abreviaturas de términos empleados para mejor comprensión del gráfico de red bipartita.

Figura 4.2: Redes bipartita de los temas y tópicos relacionados a los MDE.

ABREVIATURAS	
C_C	Climate Change
S_R	solar radiation
B_W	body water
C_D	channel drainage
E_M	elevation model
E_S	erosion sediment
G_I	Geographic Information
H_A	historical age
I_M	image_map
L_A_S	Landsat_Aster_Spot
L_R_S	Lidar_Radar_SAR
U_C_B	urban_city_build
R_S	river_stream

Tabla 4.8: Abreviaturas empleadas en el gráfico de red bipartita sobre temas y tópicos relativos a los MDE.

La densidad es de 0,125, lo que significa que la red tiene los enlaces suficientes para permitir una visualización correcta. Se midieron la centralidad de proximidad y la centralidad de intermediación. Estas centralidades se basan en la perspectiva de que la importancia se refiere a la localización de un vértice con respecto a las direcciones del gráfico de red.

El tema “Soild” es 0,562 de centralidad de proximidad, considerando un rango de 0,302 a 0,562. Y la centralidad de intermediación, para el tema “Soild” es 220 y el tema “Medioambiente” es 204. Estos resultados muestran la fuerte relación que existe entre ambos respecto a la consulta sobre modelos de elevaciones digitales.

Los puntos de corte identificados son: Biodiversidad, Medioambiente y Radiación Solar(S_R). Si estos nodos fueran separados de la red, podrían generar nuevos componentes.

En los resultados obtenidos de los materiales utilizados para el análisis se ha podido observar que existe una gran tendencia del uso de los MDE en campos afines al Medioambiente y las geociencias, sobre todo en aquellas donde la variable altura es fundamental, para realizar cálculos derivados como en el caso de las pendientes en la temática suelos que de cierta forma era de esperarse por la larga trayectoria del uso de los MDE en estos campos.

Respecto a las encuestas hay que tomar en cuenta que, las dos estaban orientadas a conocer el propósito en el que son utilizados los MDE y a través de éstas conocer el perfil de usuarios y sus demandas respecto a la calidad de los MDE. Sin embargo, este tipo de encuestas no son muy comunes en los centros de descarga y las que existen son opcionales, por lo que sería una buena opción contar con este tipo de herramientas que nos permitan conocer sobre el propósito del uso que los usuarios dan a los MDE, así como las necesidades de calidad para dichos propósitos.

Los primeros resultados de las encuestas fueron tomados en cuenta para la selección del campo de aplicación y el área de estudio donde se empleará la metodología para la simulación de errores y determinar las implicaciones de la mala calidad en los MDE respecto a su componente vertical.

Finalmente, con los resultados en el análisis del registro de usuarios, la encuesta en línea y la consulta a Science Direct se realizó una publicación que fue enviada a una revista científica y se encuentra en proceso de revisión. El publicación se encuentra en el anejo de la Sección A.6

4.2 Ensayo: Obtención de simulaciones geoestadísticas de errores de la vertical en un MDE

Como se indicó en el capítulo anterior se realizó una prueba para determinar la validez de la metodología propuesta con un MDE de referencia y uno evaluado del sector conocido como El Condado en la ciudad de Quito.

El mapa de errores obtenido entre la diferencia de los MDE (evaluado y referencia) se aprecia en la Figura 4.3.

En primera instancia, se aprecian valores muy altos de error (alrededor de ± 40 metros). Esta situación se debe a que dentro del área de estudio (por defecto un área rectangular), existen sectores en donde no se realizaron mediciones (puntos topográficos) y por ende al momento de crear el MDE, se produce una extrapolación en dichos sectores. Por lo anterior, es necesario usar una máscara para poder realizar una análisis de errores más consistente al considerar sectores en donde efectivamente

Figura 4.3: Mapa de errores obtenido de la sustracción entre MDE: evaluado y referencia. Coordenadas planas en sistema TM-Quito.

se realizaron mediciones. El mapa de errores enmascarado se puede visualizar en la Figura 4.4.

Figura 4.4: Mapa de errores usando una máscara. Coordenadas planas en sistema TM-Quito.

A continuación se realiza un muestreo aleatorio de celdas, a fin de obtener un número de datos factible para ejecutar el análisis geoestadístico. La función usada es `errores_sample[runif(312285) >= 0,005] < -NA`. El valor 312285 corresponde al número total de celdas del MDE evaluado. La función `runif` realiza un muestreo aleatorio simple de celdas.

Dado que se trabaja inicialmente con estructura tipo ráster, es necesario realizar una conversión a estructura vectorial de las celdas muestreadas a puntos para facilitar el análisis estadístico y finalmente a un *data frame* estándar (a través de la función `as.data.frame`).

Por otro lado, cabe mencionar que se analiza una área relativamente homogénea con respecto a pendiente y geoforma, con lo que se está asegurando las condiciones de

estacionaridad. También se excluyen errores de la vertical atípicos para el tipo de MDE analizado. Los errores excluidos son mayores a 2 m o menores a -2 m.

En la Figura 4.5 se aprecian los gráficos exploratorios: histograma, gráfico cuantil-cuantil y diagrama caja-bijote.

Figura 4.5: Gráficos exploratorios: histograma, gráfico cuantil-cuantil y diagrama caja-bijote (de izquierda a derecha)

Aparentemente, de acuerdo a las gráficas anteriores, los errores de la componente vertical del MDE evaluado tienen una distribución aproximadamente normal. Sin embargo, es necesario realizar un test estadístico de normalidad para confirmar este supuesto.

4.2.1 Test estadístico de normalidad

El estadístico D del test de Kolmogorov-Smirnov obtenido es 0,06768, con un tamaño de muestra de 497. De acuerdo a Olea (2012), cuando usamos la técnica geoestadística, la normalidad es confirmada únicamente realizando la comparación

de la magnitud del estadístico D con un valor $D \leq 0,1$. Cuando se cumple esta condición, se asevera que los datos son normales. Como se indicó en el capítulo anterior, no es posible considerar o analizar los niveles de significación (p -valor) debido a la presencia de dependencias espaciales en las muestras.

4.2.2 Test de dependencia espacial de Moran

Para definir si existe o no dependencia o autocorrelación espacial entre los errores la vertical, es necesario analizar el p -valor obtenido en el test de Moran. En este caso, se obtuvo un p -valor de 0 y como este es menor a 0,05 (considerando un nivel de significación del 5 %), no se puede aceptar la hipótesis nula del test que afirma que no existe dependencia espacial por tanto los datos exhiben autocorrelación espacial.

A continuación se realiza la revisión de posible existencia de tendencia espacial.

Como se puede apreciar en la Figura 4.6, concretamente, en las gráficas superior derecha e inferior izquierda, la presencia de tendencia no es evidente. Con respecto a la gráfica inferior derecha, se puede advertir la presencia de un ligero sesgo a la izquierda; lo que se confirma con el cálculo del coeficiente de sesgo con un valor de -0,585. Oliver y Webster (2014) sugiere la ejecución de transformaciones de escala cuando el coeficiente de sesgo supera un valor de 0,5 para el caso de variables ambientales (concentración de elementos químicos en suelos). Con respecto a la magnitud del sesgo de los errores de la componente vertical de un MDE y la necesidad de realizar transformaciones no existen antecedentes en la revisión bibliográfica realizada. Por lo anterior, en el presente caso, no se ha considerado necesaria una transformación.

4.2.3 Análisis de la estructura espacial

El variograma experimental obtenido a partir del conjunto de datos se presenta en la Figura 4.7.

En la Figura 4.7 se puede apreciar un variograma acotado (que alcanza una meseta), a un valor de semivarianza de 0,25 aproximadamente.

Figura 4.6: Gráficas para evaluación de posible tendencia espacial

Adicionalmente, en la Figura 4.8 se puede apreciar que existen pocos pares de puntos en donde las diferencias de errores de la vertical entre ellos, son más grandes que el promedio.

Con la ayuda de un mapa o superficie variográfica (Figura 4.9) podemos detectar la presencia de anisotropía. En este caso, se puede apreciar que existe una dirección de mayor semejanza entre los errores verticales, y está dada a los 135 grados aproximadamente.

De forma complementaria, con la interpretación de los variogramas direccionales calculados a 0, 45, 90 y 135 grados (Figura 4.10) se confirma o no la presencia de

Figura 4.7: Variograma experimental de errores de la vertical

Figura 4.8: Variograma experimental de errores de la vertical

anisotropía. El variograma de 135 grados es el que tiene la menor semivarianza, por lo que se confirma la existencia de anisotropía zonal en esta dirección, ya que las mesetas son diferentes en las direcciones analizadas.

Para modelar este tipo de anisotropía (zonal), se puede usar un modelo anisotrópico geométrico (las mesetas son similares en las direcciones evaluadas) definiendo una relación de semieje mayor/semieje menor muy grande. En el presente caso, dada que la extensión del área de análisis es pequeña y por tanto, se presenta un rango corto; los variogramas direccionales son similares a cortas distancias, por lo que no se considera la anisotropía.

Figura 4.9: Variograma experimental de errores de la vertical

A continuación, se procede a realizar el ajuste del variograma experimental a uno teórico.

En el presente caso, visualizando el variograma experimental, es posible afirmar que el modelo de variograma teórico que mejor se ajusta a los datos es el Gaussiano.

En el siguiente caso de estudio se realizará un análisis más detallado con respecto al ajuste del variograma experimental.

Finalmente, se crean 100 simulaciones de errores de la vertical que obedecen a la estructura espacial previamente definida. La Figura 4.11 presenta, a modo de ejemplo, cuatro simulaciones ejecutadas.

Nótese el efecto granulado presente en las simulaciones, el mismo que está dado por la definición de efecto pepita en el modelo de dependencia espacial empleado.

Figura 4.10: Variogramas experimentales direccionales (0, 45, 90 y 135 grados) de errores de la vertical

Figura 4.11: Simulaciones de errores de la vertical considerando un modelo gaussiano

4.3 Caso de estudio: Inundaciones del Río Garrapata

4.3.1 Determinación de áreas de inundación

Resultado de la modelización hidráulica en HEC-RAS en cada uno de los perfiles transversales se obtuvo el calado y velocidad como se muestra en la Figura 4.12 del tramo del río analizado.

Figura 4.12: Ejemplo del resultado de una sección transversal una vez ejecutada la modelización hidráulica

La Figura 4.13 muestra las áreas de inundación resultado de la modelación hidráulica para los cinco períodos de retorno. Las diferencias de áreas no es muy representativa ya que como se dijo anteriormente por la características climáticas y geomorfológicas de la zona como son las altas precipitaciones en la época invernal y la bajas pendientes, así como la poca profundidad y ancho de la sección del río hacen que éste se desborde con caudales de crecida en períodos de retorno menores a 5 años.

Sin embargo a pesar de ser una zona de constantes inundaciones, existen poblaciones asentadas al margen del Río Garrapata, aumentando aún más la vulnerabilidad de los sitios con peligro de inundaciones. Por tal motivo, para una correcta planificación territorial y de la gestión del riesgo de inundaciones se requiere de información de

calidad que se encuentre enmarcada en estándares. Además que permita conocer los posibles errores que podrían contener los productos utilizados para las simulación de inundaciones.

Resultados de simulación de errores A continuación se presentan los resultados de la simulación de errores de la componente vertical del MDE, que se usa como entrada para la modelación hidráulica.

4.3.2 Análisis exploratorio de errores

En la Figura 4.14 se presentan algunos gráficos exploratorios de los errores de la vertical en el área de estudio de la microcuenca del Río Garrapata.

Al parecer, existe una subestimación de alturas (errores por defecto) en el área de estudio, ya que existe un mayor número de puntos con errores menores a cero.

Con respecto al test de normalidad, el valor del estadístico D obtenido es ligeramente mayor (0,10184) al valor crítico de 0,1, por lo que se rechaza la hipótesis nula de normalidad, es decir los errores de la vertical no son normales. Esta situación se hace evidente en la gráfica cuantil-cuantil, en la que los extremos de la distribución presentan distanciamiento de la línea cuantil-cuantil teórica.

La media de la distribución (-0,82) es mayor que la mediana (-0,87) por lo que la distribución estadística está positivamente sesgada. La cola está alargada a la derecha. Existe una mayor cantidad de puntos con error de la vertical negativos.

Por otro lado, no es posible considerar los niveles de significación por la presencia de dependencia espacial de las muestras (Olea, 2012).

Evalando la dependencia espacial entre puntos muestreados, se puede apreciar que el p -valor obtenido es cero y por tanto menor a 0,05; por lo que no se puede aceptar la hipótesis nula del test que afirma que no existe autocorrelación espacial, por tanto los datos exhiben dependencia espacial.

Figura 4.13: Áreas de inundación para los cinco períodos de retorno analizados: 5, 10, 25, 50 y 100 años. Sistema de referencia espacial WGS84, proyección UTM 17S.

Figura 4.14: Gráficas exploratorias de errores de la vertical en la microcuenca del Río Garrapata. De izquierda a derecha: histograma, diagrama cuantil-cuantil y caja-bigote.

4.3.3 Variograma experimental

En este sentido, se realiza el cálculo del variograma experimental (Figura 4.15).

El variograma alcanza una meseta, lo cual demuestra las condiciones de estacionaridad de la variable analizada.

La nube variográfica calculada para el conjunto de datos analizado no muestra diferencias de errores de la vertical importantes entre los puntos analizados.

El mapa variográfico para el conjunto de datos se presenta en la Figura 4.16. Como se puede apreciar, no existe la evidencia de presencia de anisotropía de manera extendida en el área de estudio. A nivel local (distancias menores de 100 metros) se puede ver algún efecto anisotrópico en la dirección norte.

En la Figura 4.17, los variogramas direccionales, de forma general, muestran una relativa coincidencia, lo cual confirma la ausencia de anisotropía regional.

Figura 4.15: Variograma experimental de errores de la vertical en zona de estudio: Río Garrapata.

4.3.4 Ajuste del variograma experimental

Los resultados del ajuste del variograma experimental por el método de mínimos cuadrados iterativos se presentan en la Tabla 4.9.

Modelo	Rango	Rango teórico	Meseta
Exponencial	442.284	147.638	0.332
Esférico	83.975		0.196
Gaussiano	634.837	366.785	0.040

Tabla 4.9: Parámetros de ajuste obtenidos mediante el método de mínimos cuadrados iterativos

Figura 4.16: Mapa o superficie variográfica de errores de la vertical en zona de estudio: Río Garrapata.

Como se puede evidenciar, una combinación de tres modelos (modelo anidado) ha sido obtenida a partir del ajuste realizado. Los modelos teóricos considerados por este son: exponencial, esférico y gaussiano.

Una vez que el modelo de estructura espacial ha sido definido, los resultados de la simulación se presentan a continuación. En la Figura 4.18 se muestran 4 realizaciones, a modo de ejemplo de las simulaciones creadas.

4.3.5 Variabilidad de los errores de la vertical

Para analizar la variabilidad de los errores de la vertical en las simulaciones generadas, se realiza el cálculo de celdas de cada una de las simulaciones que se

Figura 4.17: Variogramas direccionales en direcciones: 0, 45, 90 y 135 grados para error de la vertical para el caso de estudio Río Garrapata

cortan con una recta definida por $y = 9928,514$ (Véase Figura 4.19). Los resultados son mostrados en la Figura 4.20.

4.3.6 Probabilidades de error de la vertical

Una de las novedades de la metodología propuesta es la visualización de diferentes probabilidades de error de la vertical, lo que permite que los actores y tomadores de decisión puedan mejorar la comunicación acerca de la confiabilidad de determinación espacial de los errores la componente vertical de un MDE.

En la Figura 4.21 se muestran los resultados obtenidos con la modelación de inundación considerando un período de retorno de 100 años, donde se estima la

Figura 4.18: Simulaciones de la estructura espacial de error de la vertical, caso de estudio: Río Garrapata

probabilidad de la ocurrencia de errores de la vertical utilizando un umbral de calado de inundación.

Las áreas en rojo corresponden a zonas con mayores probabilidades de presencia de error en la vertical. Por el contrario, las zonas con menor probabilidad de error corresponden a las áreas de color celeste. Zonas con un valor intermedio de

Figura 4.19: Ubicación de la recta que corta las simulaciones creadas, caso de estudio: Río Garrapata

probabilidad de error de la vertical (áreas con color amarillo) se evidencian al sur-occidente del área de estudio.

En la Tabla 4.10, se presenta un resumen de la superficie inundada por intervalo de pérdida monetaria debido a la presencia de errores de la vertical para períodos de retorno de 5, 10, 20, 50 y 100 años.

Figura 4.20: Sección longitudinal trazada a través de simulaciones de error de la vertical, caso de estudio: Río Garrapata

Intervalo de pérdida (USD)	Período de Retorno (años)/Superficie (Ha)				
	5	10	25	50	100
0 - 0.5	48.42	50.55	52.26	52.74	52.94
0.5 - 5	15.08	16.52	18.97	20.64	21.92
5.0 - 25.0	2.72	3.34	4.13	5.03	5.60
25.0 - 50.0	0.93	1.14	1.32	1.53	1.68
50.0 - 114.0	0.11	0.15	0.20	0.23	0.25

Tabla 4.10: Superficie (Ha) afectada por cada intervalo de pérdida monetaria (USD) para los períodos de retorno de 5, 10, 25 y 100 años. La pérdida monetaria está dada por celda.

Figura 4.21: Áreas inundables con probabilidades de error, subestimación de alturas con valores menores a -1,5 metros (subestimación de alturas)

La superficie media de inundación en el área de estudio tomando en cuenta todos los períodos de retorno modelados es de aproximadamente 75,7 Ha.

El 68 % del área de estudio se encuentra dentro del intervalo de pérdida menor a 0,5 USD por m², mientras que el 2 % del área total analizada tiene una pérdida monetaria mayor a 25 USD.

Tomando como referencia la Tabla 4.10, se puede observar que para cualquiera de los intervalos de pérdida monetaria analizados, la superficie (Ha) de afectación se va incrementando gradualmente desde el período de retorno de inundaciones de 5 años hasta 100 años. Esta situación permite concluir que el modelo hidráulico usado brinda resultados consistentes. La mayor superficie de afectación para todos los períodos de retorno analizados se encuentra dentro del intervalo de pérdida entre 0 y 0,5 USD por m², con una superficie de 51,38 Has. Por otro lado, la menor superficie de afectación para los períodos de retorno evaluados se encuentra dentro del intervalo 50,0 a 114,0 USD por m², con una superficie de 0,19 Has.

4.3.7 Resultados de las pérdidas monetarias por errores en la vertical

En la Figura 4.22 se presenta el mapa de resultados de pérdidas monetarias debido a la presencia de errores de la vertical para un período de retorno de 100 años (que constituye el peor escenario).

En la parte sur del área de estudio se encuentra localizado el poblado de San Andrés, que coincide espacialmente con la categoría de mayor pérdida monetaria (50,0 a 114,0 USD por m²). Adicionalmente, en el nor-orienté de la zona de estudio (zona Z1) se puede ver otra zona con valores altos de pérdidas monetarias (intervalos entre 5-25, 25-50 y 50-114 USD por m²).

Figura 4.22: Mapa de inundaciones y las pérdidas monetarias debido a la mala calidad del MDE relativo a la componente vertical; alturas con valores menores a -1,5 metros (subestimación)

Capítulo 5

CONCLUSIONES

Con los resultados del análisis de las encuestas y la consulta a ScienceDirect, se evidencia que la calidad de los MDE todavía requiere de gran atención, especialmente en los siguientes aspectos:

- Evaluación de la exactitud de los MDE.
- Uso de los índices de calidad que expresen mejor los requisitos de los usuarios.
- Influencia de la mala calidad de los resultados.
- Monetización de la influencia de la mala calidad en los resultados.

La gran demanda que tienen los MDE en los diversos campos de aplicación está relacionada con el sector del Medioambiente.

En el caso de España, se ha constatado que es importante contar con una encuesta para los usuarios de los MDE lo cual, ha permitido caracterizar el perfil general de los usuarios en España, así se pudo conocer que el 70,74 % de los usuarios son particulares y que más del 82 % de los usuarios españoles clasificaron la utilidad de los MDE como alta o muy alta, lo cual representa una oportunidad para a través de las repuestas de la encuesta indagar en aspectos relacionados con la calidad. Sin embargo este tipo de encuestas no son muy comunes en los centros de descarga de las Agencias Cartográficas Nacionales, como se pudo constatar en la primera fase de la investigación.

Por otra parte, los resultados de la encuesta de los usuarios y usos de CNIG-IGN ha proporcionado una gran lista de dominios de aplicación y su importancia correspondiente que fueron el hilo conductor para la selección del caso de uso junto con las encuesta en línea y la lista de temas y tópicos de ScienceDirect.

En la encuesta en línea, cabe destacar los resultados que muestran que los usuarios de MDE expresan su calidad principalmente a través de evaluaciones estadísticas, aunque la evaluación subjetiva es de gran importancia. Además, se han puesto de manifiesto que las cuestiones analizadas de calidad (expresión de precisión, tipo de índices, satisfacción, posibles mejoras, influencia de mala calidad y monetización de las consecuencias) requieren atención.

Se ha podido constatar la importancia de conocer a los usuarios de los MDE, sus necesidades y conocimiento en lo que refiere a la calidad, por lo cual es necesario plantear alternativas desde la perspectiva de las pérdidas en términos monetarios de la mala calidad de los MDE, pues los resultados muestran una dificultad al momento de traducir los problemas de mala calidad a dinero. Sólo el 7 % de los encuestados son capaces de monetizar las consecuencias de la mala calidad dada un orden de magnitud o rango del impacto económico, esto puede deberse a que el 85,4 % de los encuestados son usuarios avanzados con más de 4 años de experiencia en el campo de los MDE.

En referencia a la metodología para la simulación de errores, la principal limitación relacionada con la aplicación de la técnica geoestadística radica en la existencia de estacionaridad de la variable error de la componente vertical del MDE. En los ensayos realizados y en el caso de estudio, la estacionaridad de la variable se cumple para un territorio pequeño. Adicionalmente, para obtener un modelo de dependencia espacial consistente con la realidad es necesario contar con un cantidad suficientemente grande de puntos de mayor exactitud para poder ejecutar el análisis geoestadístico.

El uso de un número considerable de realizaciones de error de la vertical (al menos 100) permite conformar escenarios probabilistas de su comportamiento espacial. Una vez que el modelo probabilista es obtenido, es posible plantear escenarios de probabilidad de error de la vertical dado un determinado umbral crítico. Con el incremento de la potencia de los ordenadores, la creación de un mayor número de campos aleatorios es factible.

Para el caso de estudio seleccionado (modelación hidráulica de inundaciones) se considera que el MDE es una variable crítica susceptible de error y que el resto de parámetros empleados en el modelo de inundaciones son correctos.

La mayor superficie de afectación para cada uno de los períodos de retorno analizados (5, 10, 25, 50 y 100 años), respectivamente, presenta la menor pérdida económica, encontrándose dentro del intervalo de pérdida entre 0 y 0,5 USD por m², con una superficie de afectación promedio de 51,38 Ha. La menor superficie de afectación para cada uno de los períodos de retorno evaluados (5, 10, 25, 50 y 100 años), respectivamente, presenta la mayor pérdida económica, se encuentra dentro del intervalo de 50,0 a 114,0 USD por m², con una superficie de afectación promedio de 0,19 Ha.

La presente tesis no adopta una metodología económica, en particular, para determinar el valor económico de los MDE, sino mas bien obtiene y muestra una relación entre el coste de la mala calidad de los MDE y la probabilidad de existencia de errores de la componente vertical de éstos (enfoque probabilista). El planteamiento propuesto usó conceptos aplicados en la definición de la función de pérdida de la calidad de la teoría de Taguchi que relaciona los costes de pérdida asociados con el atributo de un producto que no cumple el nivel de conformidad, una vez que ha finalizado el proceso de producción del mismo, en el caso de este estudio un MDE.

El enfoque usado en esta tesis constituye un punto de partida para la profundización de la determinación de los costes de la mala calidad de este grupo particular de datos geoespaciales como son los MDE.

5.1 Líneas futuras de investigación en torno a los costes de la mala calidad de los Modelos Digitales de Elevación

A continuación se presenta un listado de las principales líneas de investigación en torno a la presente tesis y que podrían llevarse a cabo en el futuro:

- En relación con los usuarios de los MDE, sería interesante realizar encuestas y entrevistas a usuarios que sí puedan valorar económicamente la mala calidad y en función de ello establecer líneas generales sobre las pérdidas.
- En cuanto a las técnicas de simulación, se podrían utilizar otras variantes de Kriging como Co-Kriging, a fin de incorporar una variable en el análisis con la cual se pueda tener más control y sirva incluso como indicador de calidad.
- Investigaciones relacionadas con el coste de la mala calidad de los MDE desde una perspectiva económica, tomando en cuenta la definición de la cadena de valor donde se comparen los costes de la buena calidad y la mala calidad.
- Respecto a la calidad en términos monetarios se podrían realizar investigaciones centradas en casos específicos de uso y en comunidades específicas de usuarios de los MDE.
- Investigaciones relacionadas con la generación de herramientas que permitan al usuario evaluar los MDE desde una perspectiva integral e incluya la valoración económica las consecuencias de un MDE con posibles fallos.

Bibliografía

Ackerman, C (2009). *HEC-GeoRAS User 's Manual*.

AENOR (2014). *UNE-EN ISO 19157:2014 Información geográfica - Calidad de datos*. AENOR.

Aguilar, Fernando J et al. (2010). “Modelling vertical error in LiDAR-derived digital elevation models”. En: *ISPRS Journal of Photogrammetry and Remote Sensing* 65.1, págs. 103-110.

Albani, M et al. (2004). “The choice of window size in approximating topographic surfaces from digital elevation models”. En: *International Journal of Geographical Information Science* 18.6, págs. 577-593.

Alcázar-Molina, Manuel (2014). “Propuesta de normativa técnica de valoración catastral predial del suelo rural”. Bloque temático No.2.

Alfaro, Marco Antonio (2008). *La Simulación Condicional en un Depósito Minero*.

Altan, O et al. (2013). *The Value of Geoinformation for Disaster and Risk Management (VALID): Benefit Analysis and Stakeholder Assessment. Published by the Joint Board of Geospatial Information Societies (JB GIS), Copenhagen*.

Ariza-López, FJ et al. (2013). *Fundamentos de evaluación de la calidad de la información geográfica*.

Ariza-López, Francisco Javier (2011). *Apuntes de clase: Introducción a la Calidad de la IG*.

- Ariza-López, Francisco Javier (2015). “Calidad de la IG: Situación actual y perspectivas de futuro”. En: *Seminario SIGTIERRAS-Quito*.
- Ariza-López, Francisco Javier y Antonio Federico Rodríguez-Pascual (2007). “La familia ISO 19100”. En:
- Ariza-López, Francisco Javier et al. (2017). “An approach to users and uses of DEM from the quality perspective”. En:
- Arranz, José Juan (2013). “Diseño, optimización y análisis de sistemas basados en técnicas láser, para el modelado geométrico, registro y documentación, aplicados a entidades de interés patrimonial”. Tesis doct. Escuela Técnica Superior de Ingenieros en Topografía, Geodesia y Cartografía. Universidad Politécnica de Madrid.
- ASPRS (2014). *Positional Accuracy Standards for Digital Geospatial Data*. Photogrammetric Engineering & Remote Sensing Volume 81, No. 3, 53 p. URL: <http://www.asprs.org/Standards-Activities.html>.
- Atkinson, Peter M (2002). “Surface modelling: What’s the point?” En: *Transactions in GIS* 6.1, págs. 1-4.
- Borzacchiello, MT y M Craglia (2011). “Socio-economic benefits from the use of earth observation”. En: *Report from the International Workshop held at Joint Research Centre. European Commission Joint Research Centre, EUR*. Vol. 2495.
- Carlisle, Bruce H (2005). “Modelling the spatial distribution of DEM error”. En: *Transactions in GIS* 9.4, págs. 521-540.
- Castelein, Watse, Arnold Bregt e Yvette Pluijmers (2010). “The economic value of the Dutch geo-information sector”. En: *International Journal of Spatial Data Infrastructures Research* 5, págs. 58-76.
- CEM (2008). *Vocabulario Internacional de Metrología–Conceptos fundamentales y generales, y términos asociados (VIM)*. Ed. por CEM. CEM.

- Chiles, Jean-Paul y Pierre Delfiner (2012). *Geostatistics: modeling spatial uncertainty*. Vol. 497. John Wiley & Sons.
- Cooper, MAR (1998). "Datums, coordinates and differences". En: *Landform monitoring, modelling and analysis*, págs. 21-35.
- Coveney, Seamus y A Stewart Fotheringham (2011). "The impact of DEM data source on prediction of flooding and erosion risk due to sea-level rise". En: *International Journal of Geographical Information Science* 25.7, págs. 1191-1211.
- Crompvoets, Joep et al. (2008). "A multi-view framework to assess SDIs". En: *Space for Geo-Information (RGI), Wageningen University and Centre for SDIs and Land Administration, Department of Geomatics, The University of Melbourne*.
- Cuartero Sáenz, Aurora (2003). "Análisis de los Modelos Digitales de Elevaciones (MDE) generados con imágenes SPOT-HRV y TERRA-ASTER". Tesis doct. Departamento de Ingeniería Cartográfica, Geodésica y Fotogrametría de la Universidad de Jaén.
- Dasí, Joaquín Farinos (2011). "Inteligencia Territorial para la planificación y la gobernanza democráticas: los observatorios de los territorios". En:
- Desassis, Nicolas y Didier Renard (2013). "Automatic variogram modeling by iterative least squares: univariate and multivariate cases". En: *Mathematical Geosciences* 45.4, págs. 453-470.
- Devillers, Rodolphe y Helen Goodchild (2010). *Spatial data quality: from process to decisions*. CRC.
- Devillers, Rodolphe et al. (2010). "Thirty years of research on spatial data quality: achievements, failures, and opportunities". En: *Transactions in GIS* 14.4, págs. 387-400.
- Donker, Frederika Welle (2009). "Public sector Geo Web Services: which business model will pay for a free lunch?" En: *SDI convergence*, pág. 35.

- El-Sheimy, Naser, Caterina Valeo y Ayman Habib (2005). *Digital terrain modeling: acquisition, manipulation, and applications*. Artech House.
- Elosua, Paula (2011). “Introducción al entorno R”. En: *Bilbao: Euskal Herriko Unibertsitateko Argitalpen Zerbitzua/Servicio Editorial de la Universidad del País Vasco*.
- Fallas, Jorge (2007). *Modelos digitales de elevación: Teoría, métodos de interpolación y aplicaciones*.
- Felicísimo, Angel M (1994). *Modelos digitales del terreno*. Pentalfa Oviedo.
- FGDC (2016). *2015 Annual report: Reston, Virginia, USA*. Inf. téc. Federal Geographic Data Committee.
- Fisher, Peter (1998). “Improved modeling of elevation error with geostatistics”. En: *GeoInformatica* 2.3, págs. 215-233.
- Fisher, Peter F y Nicholas J Tate (2006). “Causes and consequences of error in digital elevation models”. En: *Progress in Physical Geography* 30.4, págs. 467-489.
- Fleming, Cory, JRA Giles y SH Marsh (2010). “Elevation models for geoscience”. En: *Geological Society of London*. Geological Society of London.
- Floridi, Luciano y Phyllis Illari (2014). *The philosophy of information quality*. Vol. 358. Springer.
- Florinsky, Igor V (2012). *Digital terrain analysis in soil science and geology*. Academic Press.
- Font, Rosa M. (2016). “Implantación del Sistema de Gestión de Calidad según UNE en ISO 9001: 2008. Fundación Luis Giménez Lorente, en el ámbito de la cartografía y la gestión de sus fondos. Comparativa con otras instituciones cartográficas a nivel nacional.” Tesis doct.
- García Balboa, J.L. (2013). *Normas sobre la calidad de la Información Geográfica*. En Ariza-López, F.J. (Editor), Fundamentos de evaluación de la calidad de

- la Información Geográfica (pp. 219-274). Jaén: Servicio de Publicaciones de la Universidad de Jaén.
- Genovese, Elisabetta, Stéphane Roche y Claude Caron (2009). "The value chain approach to evaluate the economic impact of Geographic Information: towards a new visual tool". En: *SDI Convergence; Research, Emerging Trends, and Critical Assessment*, págs. 175-188.
- Genovese, Elisabetta et al. (2010). "The EcoGeo Cookbook for the assessment of Geographic Information value". En: *International Journal of Spatial Data Infrastructures Research* 5, págs. 120-144.
- Gittleman, John L y Mark Kot (1990). "Adaptation: statistics and a null model for estimating phylogenetic effects". En: *Systematic Biology* 39.3, págs. 227-241.
- Goodchild, Michael F, Rodolphe Devillers y Robert Jeansoulin (2006). *Fundamentals of spatial data quality*. Vol. 662. John Wiley & Sons.
- Gutiérrez, C, E Góngora y P Melo (2008). *Estudio Hidrológico de Inundaciones en la Cuenca Alta del Río Chone (Garrapata, Mosquito y Grande)*. Inf. téc. INAMHI, CLIRSEN y FAO.
- Haestad Methods Water Solutions David Klotx, Adam Strafaci Annaleis Hogan Kristen Dietrich (2007). *Floodplain modeling using HEC-RAS*. First edition. Bentley Institute Press. ISBN: 1934493023,978-1-934493-02-1,9781628709483,1628709480. URL: <http://gen.lib.rus.ec/book/index.php?md5=a9a6224964ac28ff889177f7bd4bb53e>.
- Haug, Anders, Frederik Zachariassen y Dennis Van Liempd (2011). "The costs of poor data quality". En: *Journal of Industrial Engineering and Management* 4.2, págs. 168-193.
- Heidemann, Hans Karl (2014). "Lidar Base Specification". En: *Book 11: Collection and Delineation of Spatial*. Ed. por USGS. Vol. Section B, U.S. Geological Survey Standards. Cap. 4, págs. 1-78. URL: <https://pubs.usgs.gov/tm/11b4/pdf/tm11-B4.pdf>.

- Hengl, T, GBM Heuvelink y EEv Loon (2010). “On the uncertainty of stream networks derived from elevation data: the error propagation approach”. En: *Hydrology and Earth System Sciences* 14.7, págs. 1153-1165.
- Heuvelink, Gerard BM (1998). *Error propagation in environmental modelling with GIS*. CRC Press.
- Heuvelink, Gerard BM, Peter A Burrough y Alfred Stein (2006). “Developments in analysis of spatial uncertainty since 1989”. En: *Classics from IJGIS* 20, págs. 91-95.
- Höhle, Joachim y Michael Höhle (2009). “Accuracy assessment of digital elevation models by means of robust statistical methods”. En: *ISPRS Journal of Photogrammetry and Remote Sensing* 64.4, págs. 398-406.
- Houghton, John W (2011). *Costs and benefits of data provision*. Australian National Data Service.
- Hyann, Pinatibi et al. (2016). “Characterization of error on a digital elevation model (DEM) Based on morphological zones: Case of the Denguélé District (North-west of Côte d’Ivoire)”. En: *International Journal of Innovation and Applied Studies* 16.1, págs. 84.
- INSPIRE (2012). *A Conceptual Model for Developing Interoperability Specifications in Spatial Data Infrastructures*. Inf. téc. INSPIRE.
- (2013). *D2.8.II.1 INSPIRE Data Specification on Elevation, Technical Guidelines*. Inf. téc. European Commision Joint Research Centre, Thematic Working Group Elevation.
- IPGH, Instituto Panamericano de Geografía e Historia (2013). *Guía de Normas: Segunda edición en español*. Inf. téc. Comité ISO/TC 211 Información Geográfica / Geomática.
- Isaaks, Edward H y R Mohan Srivastava (1989). “Applied geostatistics”. En:

- Kabacoff, Robert (2015). *R in action: data analysis and graphics with R*. Manning Publications Co.
- Kolaczyk, Eric D y Gábor Csárdi (2014). *Statistical analysis of network data with R*. Springer.
- Lantuéjoul, Christian (2013). *Geostatistical simulation: models and algorithms*. Springer Science & Business Media.
- Leon, Javier X., Gerard B. M. Heuvelink y Stuart R. Phinn (2014). “Incorporating DEM Uncertainty in Coastal Inundation Mapping”. En: *PLoS ONE* 9.9, págs. 1-12. DOI: [10.1371/journal.pone.0108727](https://doi.org/10.1371/journal.pone.0108727). URL: <http://dx.doi.org/10.1371%2Fjournal.pone.0108727>.
- Lindsay, John B (2006). “Sensitivity of channel mapping techniques to uncertainty in digital elevation data”. En: *International Journal of Geographical Information Science* 20.6, págs. 669-692.
- Liu, XiaoHang, Hai Hu y Peng Hu (2015). “Accuracy assessment of lidar-derived digital elevation models based on approximation theory”. En: *Remote Sensing* 7.6, págs. 7062-7079.
- Liu, XiaoHang et al. (2012). “Approximation theory applied to DEM vertical accuracy assessment”. En: *Transactions in GIS* 16.3, págs. 397-410.
- Lloyd, Christopher D (2007). *Local models for spatial analysis*. CRC/Taylor & Francis London.
- Longhorn, Roger A y Michael Blakemore (2007). *Geographic information: value, pricing, production, and consumption*. CRC Press.
- López, Carlos (2002). “An experiment on the elevation accuracy improvement of photogrammetrically derived DEM”. En: *International Journal of Geographical Information Science* 16.4, págs. 361-375.
- Lorenzo, Jose Maria Montero y Beatriz Larraz Iribas (2008). *Introducción a la Geoestadística Lineal*. Netbiblo.

- Luke, Douglas A (2016). *A user's guide to network analysis in R*. Springer.
- Maune, David Francis (2001). *Digital elevation model technologies and applications: the DEM users manual*. Asprs Publications.
- Moreira, Marco (2016). *Estudio de Impacto Ambiental del Proyecto Propósito Múltiple Chone: Estudio Hidrológico de las Cuencas Garrapata y Mosquito*. Inf. téc. Empresa Pública del Agua, ESPE-Innovativa.
- Moreno Jiménez, Antonio (2006). *Sistemas y análisis de la información geográfica: manual de autoaprendizaje con ArcGis*.
- Mostafavi, Mir-Abolfazl, Geoffrey Edwards y Robert Jeansoulin (2004). "An ontology-based method for quality assessment of spatial data bases". En: *Third International Symposium on Spatial Data Quality*. Vol. 1. 28a, págs. 49-66.
- Mukherjee, Sandip et al. (2013). "Evaluation of vertical accuracy of open source Digital Elevation Model (DEM)". En: *International Journal of Applied Earth Observation and Geoinformation* 21, págs. 205 -217. ISSN: 0303-2434. DOI: <http://dx.doi.org/10.1016/j.jag.2012.09.004>. URL: <http://www.sciencedirect.com/science/article/pii/S030324341200195X>.
- Natural Resources Canada, Minister of (2015). *Canadian geomatics environmental scan and value study*. Inf. téc. Minister of Natural Resources Canada.
- Nedovic-Budic, Zorica, Joep Crompvoets y Yola Georgiadou (2012). *Spatial data infrastructures in context: north and south*. Ed. por Zorica Nedovic-Budic, Joep Crompvoets y Yola Georgiadou. ISBN 9781439828021. Taylor & Francis.
- Nelson, A, HI Reuter y P Gessler (2009). "DEM production methods and sources". En: *Developments in soil science* 33, págs. 65-85.
- Newman, Mark (2010). "Networks: an introduction". En: *United States: Oxford University Press Inc., New York*, págs. 1-2.

- Oksanen, Juha y Tapani Sarjakoski (2006). "Uncovering the statistical and spatial characteristics of fine toposcale DEM error". En: *International Journal of Geographical Information Science* 20.4, págs. 345-369.
- Olaya, Victor (1994). *Hidrología computacional y modelos digitales del terreno*.
- Olea, Ricardo (2012). *Geostatistics for engineers and earth scientists*. Springer Science & Business Media.
- Oliver, MA y R Webster (2014). "A tutorial guide to geostatistics: Computing and modelling variograms and kriging". En: *Catena* 113, págs. 56-69.
- Papasaika, Haris y Baltsavias Emmanuel (2010). "Quality evaluation of DEMs". En: *Proceedings of the Ninth International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences 20-23rd July 2010, Leicester*, págs. 249-252.
- Patel, Arun, S.K. Katiyar y Vishnu Prasad (2016). "Performances evaluation of different open source {DEM} using Differential Global Positioning System (DGPS)". En: *The Egyptian Journal of Remote Sensing and Space Science* 19.1, págs. 7 -16. ISSN: 1110-9823. DOI: <http://dx.doi.org/10.1016/j.ejrs.2015.12.004>. URL: <http://www.sciencedirect.com/science/article/pii/S1110982315000745>.
- Podobnikar, Tomaž (2016a). "Combination of error characterization and spatial error model to improve quality of digital elevation models". En: *Proceedings of Spatial Accuracy 2016, July*.
- (2016b). "Relevant Quality of Digital Elevation Models in Earth and Environmental Studies ?" En: *International Journal of Earth & Environmental Sciences*. DOI: <https://doi.org/10.15344/2456-351X/2016/106>.
- Rieger, Wolfgang (1996). "Accuracy of slope information derived from DEM-data". En: *International Archives of Photogrammetry and Remote Sensing* 31, págs. 690-695.

- Rivoirard, Jacques (1994). *Introduction to disjunctive kriging and non-linear geostatistics*. Clarendon Press.
- Ruiz Armenteros, Antonio, José García Balboa y José Mesa Mingorance (2010). “Error, incertidumbre, precisión y exactitud, términos asociados a la calidad espacial del dato geográfico”. En: *Catastro: formación, investigación y empresa: Selección de ponencias del I Congreso Internacional sobre catastro unificado y multipropósito*, págs. 95-102.
- Shekhar, Shashi y Hui Xiong (2007). *Encyclopedia of GIS*. Springer Science & Business Media.
- Stein, Alfred, Wenzhong Shi y Wietske Bijker (2009). *Quality aspects in spatial data mining*. CRC Press.
- Survey, Ordnance (2015). *Annual Report and Accounts 2014 2015*. Inf. téc. Ordnance Survey.
- Taguchi, G, S Chowdhury e Y Wu (2005). *Taguchi's quality engineering handbook*. Wiley.
- Thode, Henry C (2002). *Testing for normality*. Vol. 164. CRC press.
- Ureña Cámara, M.A. (2013). *Normas sobre metadatos y evaluación de metadatos*. En Ariza-López, F.J. (Editor), *Fundamentos de evaluación de la calidad de la Información Geográfica* (pp. 329-387). Jaén: Servicio de Publicaciones de la Universidad de Jaén.
- Wackernagel, Hans (2013). *Multivariate geostatistics: an introduction with applications*. Springer Science & Business Media.
- Wechsler, SP (2007). “Uncertainties associated with digital elevation models for hydrologic applications: a review”. En: *Hydrology and Earth System Sciences* 11.4, págs. 1481-1500. URL: <https://hal.archives-ouvertes.fr/hal-00305087/document>.

- Wechsler, Suzanne y Charles Kroll (2006). “Quantifying DEM uncertainty and its effect on topographic parameters”. En: *Photogrammetric Engineering & Remote Sensing* 72.9, págs. 1081-1090.
- Weng, Qihao (2002). “Quantifying uncertainty of digital elevation models derived from topographic maps”. En: *Advances in spatial data handling*. Springer, págs. 403-418.
- Wise, Stephen (2000). “Assessing the quality for hydrological applications of digital elevation models derived from contours”. En: *Hydrological processes* 14.11-12, págs. 1909-1929.
- Wong, Kelvin Ka Yin (2015). “Economic Value of 3D Geographic Information”. En: *EuroSDR and the Department of Computer Science, University College London: London, UK*.
- Yang, X et al. (2013). “An integrated view of data quality in Earth observation”. En: *Phil. Trans. R. Soc. A* 371.1983, pág. 20120072.

Listado de Acrónimos

ASPRS	<i>American Society for Photogrammetry and Remote Sensing.</i>
CAF	Banco de Desarrollo de América Latina.
CNIG	Centro Nacional de Información Geográfica.
COOTAD	Código Orgánico de Organización Territorial, Autonomía y Descentralización.
DMQ	Distrito Metropolitano de Quito.
FEMA	<i>Federal Emergency Managment Agency.</i>
FGDC	<i>Federal Geographic Data Committee.</i>
IDE	Infraestructuras de Datos Espaciales.
IG	Información Geográfica.
LiDAR	<i>Light Detection and Ranging.</i>
MDE	Modelo Digital de Elevaciones.
MDS	Modelo Digital de Superficie.
MDT	Modelo Digital del Terreno.
NDEP	<i>National Digital Elevation Program.</i>
PDF	Función de Distribución de Probabilidad.
PIB	Producto Interno Bruto.

RMSE	Raíz del Error Medio Cuadrático.
SIG	Sistemas de Información Geográfica.
TIN	Red de Triángulos Irregular.
UAV	Vehículos Aéreos no tripulados.
USACE	<i>United States Army Corps of Engineers.</i>
USGS	<i>United States Geological Survey.</i>

Apéndice A

Apéndices

A.1 Artículo: Spatial uncertainty of a geoid undulation model in 7 Guayaquil, Ecuador

Open Geosciences

Spatial uncertainty of a geoid undulation model in Guayaquil, Ecuador

--Manuscript Draft--

Manuscript Number:	OPENGEO-D-17-00024R3
Full Title:	Spatial uncertainty of a geoid undulation model in Guayaquil, Ecuador
Article Type:	Research Article
Keywords:	ellipsoidal height; stationarity; anisotropy; trend
Corresponding Author:	Elena Chicaiza, MSc. Universidad Politécnica de Madrid Madrid, Madrid SPAIN
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Universidad Politécnica de Madrid
Corresponding Author's Secondary Institution:	
First Author:	Elena Chicaiza, MSc.
First Author Secondary Information:	
Order of Authors:	Elena Chicaiza, MSc.
	César Leiva, MSc.
	José Arranz, Dr.
	Xavier Buenaño, Dr.
Order of Authors Secondary Information:	
Abstract:	Geostatistics is a discipline that deals with the statistical analysis of regionalized variables. In this case study, geostatistics is used to estimate geoid undulation in the rural area of Guayaquil town in Ecuador. The geostatistical approach was chosen because the estimation error of prediction map is getting. Open source statistical software R and mainly geoR, gstat and RGeostats libraries were used. Exploratory data analysis (EDA), trend and structural analysis were carried out. An automatic model fitting by Iterative Least Squares and other fitting procedures were employed to fit the variogram. Finally, Kriging using gravity anomaly of Bouguer as external drift and Universal Kriging were used to get a detailed map of geoid undulation. The estimation uncertainty was reached in the interval $[-0.5; +0.5]$ m for errors and a maximum estimation standard deviation of 2 mm in relation with the method of interpolation applied. The error distribution of the geoid undulation map obtained in this study provides a better result than Earth gravitational models publicly available for the study area according the comparison with independent validation points. The main goal of this paper is to confirm the feasibility to use geoid undulations from Global Navigation Satellite Systems and leveling field measurements and geostatistical techniques methods in order to use them in high-accuracy engineering projects.

Research Article

Open Access

E.G. Chicaiza*, C.A. Leiva, J.J. Arranz, and X.E. Buenaño

Spatial uncertainty of a geoid undulation model in Guayaquil, Ecuador

DOI: DOI, Received ..; revised ..; accepted ..

Abstract: Geostatistics is a discipline that deals with the statistical analysis of regionalized variables. In this case study, geostatistics is used to estimate geoid undulation in the rural area of Guayaquil town in Ecuador. The geostatistical approach was chosen because the estimation error of prediction map is getting. Open source statistical software R and mainly geoR, gstat and RGeostats libraries were used. Exploratory data analysis (EDA), trend and structural analysis were carried out. An automatic model fitting by Iterative Least Squares and other fitting procedures were employed to fit the variogram. Finally, Kriging using gravity anomaly of Bouguer as external drift and Universal Kriging were used to get a detailed map of geoid undulation. The estimation uncertainty was reached in the interval $[-0.5; +0.5]$ m for errors and a maximum estimation standard deviation of 2 mm in relation with the method of interpolation applied. The error distribution of the geoid undulation map obtained in this study provides a better result than Earth gravitational models publicly available for the study area according the comparison with independent validation points. The main goal of this paper is to confirm the feasibility to use geoid undulations from Global Navigation Satellite Systems and leveling field measurements and geostatistical techniques methods in order to use them in high-accuracy engineering projects.

Keywords: ellipsoidal height, stationarity, anisotropy, trend

1 Introduction

Data quality is a concept related with uncertainty [1] that needs to fit within the perspective of producer (internal) or users (external). Positional accuracy is an internal data quality element of importance due to errors could propagate to derived maps and that has been studied widely in the literature. On the other hand, the elevation attribute is a critical element in any geoscience application from geologic maps to 3D models [2], specially important in order to generate digital elevation models (DEMs) and sub-products of them. Several factors affects the overall vertical accuracy of DEMs [3], nevertheless one of the most important is the uncertainty introduced by geoid transformations added to the theoretical model [4] that produces differences between local and global DEM vertical accuracy (theoretically the same for a specific application).

The term geoid was employed by mathematician J.B. Listing [5] to define mathematical surface of equipotential surfaces of Earth's gravitational field. The geoid's calculation is usually carried out by combining a global geopotential model of gravitational field with terrestrial gravity anomalies measured locally and complemented with local or regional topographic data according the desired accuracy. Geoid height or geoid undulation is enunciated as height above the ellipsoid taken along the normal line to ellipsoid's surface [5].

On the other side, the vertical gradient of the gravity anomaly can be related to the horizontal derivatives of the deflection of the vertical and the geoid height (undulation) N and its first and second derivatives [6]. This first relation, however, may be less suited for practical application as inaccuracies of N are greatly amplified by forming the second derivatives.

Different estimation methods of geoid heights models such as least-squares collocation (LSC), artificial neural networks, fuzzy logic and radial basis functions (Multiquadric-MQ, Thin Plate Spline-TPS) have been widely advocated [7]. Some of these methods are analyzed in [8] and constitute deterministic methods. Geostatistical techniques have not been explored in this context.

*Corresponding Author: E.G. Chicaiza: Universidad Politécnica de Madrid, E-mail: gabriela.chicaizam@alumnos.upm.es

C.A. Leiva: Universidad de las Fuerzas Armadas ESPE, Campus Sangolquí, Ecuador, E-mail: caleiva@espe.edu.ec

J.J. Arranz: Universidad Politécnica de Madrid, E-mail: josejuan.arranz@upm.es

X.E. Buenaño: Universidad Politécnica de Madrid, E-mail: xgoodyear@gmail.com

Geostatistics is a discipline that deals with the statistical analysis of regionalized variables (spatially distributed). Therefore, applications of geostatistics are wide. Complementarily, in the Global Navigation Satellite System (GNSS)/levelling method, existence of uniformly and homogeneously scattered points on the study area generally increases the quality of the estimation [7]. In this study, geostatistics is applied in order to obtain geoid undulation estimation and error models in the rural area of Guayaquil town, Ecuador.

The stationarity is one of main hypothesis of geostatistics and can be reached with different assumptions. The first order or strict hypothesis that assumes a mean constant translation invariant; the second order which argues that the mean is constant and the variance is finite; and, the intrinsic or weak stationarity that deals with the increments in order to reach a finite variance of these one. In this context, the application of kriging interpolation method is feasible in order to get an error estimation that cannot be obtained with deterministic methods. When the presence of drift is determined, various alternatives can be implemented. One option is carried out the geostatistical analysis in smaller areas, subdividing the original data set [9]. Nevertheless, if the number of samples is little, geostatistical analysis is not feasible due to few pair of points would use to build the variogram function and conclusions from this would not be correct. Another approach is the definition of a function of drift through the deterministic definition of this, which can be explained by a spatial trend or another variable. Once this has been reached, universal kriging or kriging with external drift can be applied.

On the other hand, multiple simulations can be used to quantify heights uncertainty through the analysis and evaluation of statistics related with a distribution of spatial random fields [10]. Nevertheless, this option has not been chosen because of the typical smooth spatial change behavior of geoid undulation which can be better depicted by kriging estimation.

In the present case, study area is located in Ecuador, in the western of Andean chain as shown in Figure 1. This area is mainly buildup of accreted fragments of Late Cretaceous mafic oceanic plateau basement [11]. The lithology varies through the study area in relation with different processes of weathering.

Fig. 1. Study area (red circle) is located in the southern west area of Guayaquil city, in the coastal region of Ecuador.

The number of measurements is not enough in order to subdivide the analysis in smaller areas. For this reason, the analysis has been developed with the whole data set.

Another important task is determination of a possible trend that will be analyzed in the next section. After this task, computation of experimental variogram was carried out and anisotropy was also analyzed.

The structural analysis constitutes a critical step in geostatistical modeling. The computation of variogram

model tries to find the model that minimizes a cost function measuring the distance between model and experimental variograms [12]. In the univariate case, it is enough that the sills fulfill a positivity constraint and they can be fitted by ordinary least squares (OLS), weighted least squares (WLS) or generalized least squares (GLS) [13].

Another interesting approach is the estimation of the trend and the variogram of the random residuals from it simultaneously. The current best practice is restricted maximum likelihood (REML) technique [14].

Intrinsic random functions of order k (IRF- k), which are associated with generalized covariances, also provide a class of non stationary models that are useful to represent the non stationary solutions of stochastic partial differential equations such as found in hydrogeology [15]. It is possible to use this alternative that can be complemented with iterative least squared technique, which provides other interesting approach for variogram fitting. This last method is based on a numerical minimization (deterministic) of a sum of squares and uses specific techniques to face with numerical singularities and convergence problems [12]. The aim of this paper is firstly obtain an uncertainty error estimation and a detailed geoid undulation estimation map in order to avoid additional field work according the fitness for use (engineering design).

2 Materials and Methods

The study area is located in the southern west of Guayaquil city, in the coastal region of Ecuador, within of Guayas basin. The area is a floodplain with low slope. Guayaquil data set includes 299 measurements of heights and geoid undulation determined in the same locations with an average distance of 3 km. Geoid undulations were obtained by second order geometric leveling (tolerance $\pm 8.4 \text{ mm}\sqrt{k}$) or trigonometric levelling with reciprocal (tolerance $\pm 3 \text{ cm}\sqrt{k}$) to get mean sea level height (where k is the displacement distance) and cover an average area of near 2,400 km^2 that includes a variety of landforms.

Ellipsoidal heights were obtained by Global Position System (GPS) measurements from geodesic supplementary control points with differential static method using double frequency GPS receptors (relative accuracy $10 \text{ mm} \pm 2 \text{ ppm}$).

The geostatistical approach used in this article to get geoid undulation estimations is novel because typically this is developed with a deterministic approach, such as using polynomial surface fitting, least squares collocation (LSC), finite-element methods and artificial neural networks (with less frequency), that do not let to obtain an uncertainty level related with the estimation. A workflow of the methodology is shown in Figure 2.

The geostatistical analysis can be carried out in a lot of commercial software, but using open source software for research purposes is an interesting option. One of the available alternatives is R environment and statistical program [16]. Inside R, there are an important number of libraries related with the spatial data analysis. Nevertheless, a criterion to choose a specific library is the updating level and the quantity and quality of algorithms available. In this sense, interesting options are: a) RGeostats developed by the Geostatistical Team of the Geosciences Research Center of MINES ParisTech [17], b) geoR developed by Ribeiro Jr. and Diggle P. [18]; and c) gstat developed by Pebesma E. [19]. As a first step, exploratory data analysis (EDA) (see Section 2.1) was carried out as a mechanism for data quality assurance in spatial (coordinates) and thematic (geoid undulation) components. This step was critical, since any incorrect value alters any subsequent analysis. An important activity constitutes the review of statistical and spatial distribution.

Fig. 2. Overall workflow methodology that summarizes the tasks carried out in the current work (focused in geostatistical analysis).

2.1 Exploratory data analysis

The EDA includes basic descriptive statistics of geoid undulation.

To confirm the validity of experimental variogram and data intrinsically, an envelope (Figure 3) drawn based on simulations obtained by repeatedly simulating from the fitted model (parameters of spherical covariance) in the data locations [18]. This envelope shows the variability of the empirical variogram. On the other hand, sometimes a preliminary variography on raw data before minimization (drift removal) could be useful [20].

Fig. 3. Experimental variogram (circles) and envelopes (dashed lines) based in spherical covariance model building by simulations.

2.2 Trend analysis

In the current study, trend was identified through plots of spatial distribution in two dimensions (postplot) and directional clouds.

One very general form for a lineal model [21] is shown in Equation 1:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon \quad (1)$$

Where β_i , $i = 0, 1, 2, 3$ are unknown terms, β_0 is called the intercept term, ε is the fit error and X_n are explanatory variables. This approach can be extended to spatial coordinates with East and North as predictors and trend as explained variable.

Then the equation for linear or first order trend (Eq. 2) will be:

$$Z(s_i) = \beta_0 + \beta_1 East_i + \beta_2 North_i + \varepsilon \quad (2)$$

The predicted value ($Z(s_0)$) at location s_0 will be again a linear combination of the observed $Z(s_i)$, $i = 1, \dots, n$ values as shown in Equation 3:

$$\hat{Z}(s_0) = \omega_1 Z(s_1) + \omega_2 Z(s_2) + \dots + \omega_n Z(s_n) = \sum_{i=1}^n \omega_i Z(s_i) \quad (3)$$

Where $\omega_1 + \omega_2 + \dots + \omega_n = 1$

Suppose now that a trend of the linear form is present. Then the value $\hat{Z}(s_0)$ can be expressed as Equations 4 or 5:

$$\hat{Z}(s_0) = \sum_{i=1}^n \omega_i Z(s_i) = \sum_{i=1}^n \omega_i \beta_0 + \sum_{i=1}^n \omega_i \beta_1 X_i + \sum_{i=1}^n \omega_i \beta_2 Y_i + \sum_{i=1}^n \omega_i \delta(s_i) \quad (4)$$

$$\hat{Z}(s_0) = \sum_{i=1}^n \omega_i Z(s_i) = \beta_0 \sum_{i=1}^n \omega_i + \beta_1 \sum_{i=1}^n \omega_i X_i + \beta_2 \sum_{i=1}^n \omega_i Y_i + \sum_{i=1}^n \omega_i \delta(s_i) \quad (5)$$

The same approach is used to model a spatial second order trend and other covariates, the mathematical development for these cases is not detailed in this work. The trend detection and removal is a previous stage to continue with the next steps of prediction process.

In order to review an important set of spatial trend models, a regression graphical method for all subsets was employed. In this method (interactions of predictors), every possible model is inspected [22]. Additionally, the relative weights of predictors were analyzed according the heuristic method proposed by [23].

On the other hand, Bouguer gravity anomaly is an interesting factor which can be reviewed in the trend modeling. The anomalies map of the study area was obtained from International Gravimetric Bureau [24].

For an easy visualization, variograms with a set of different trend considerations have been computed as shown in Figure 4.

Fig. 4. Sample variograms for data. The distances are expressed in meters. The upper-left panel uses the unadjusted geoid undulation response, the upper-right panel uses residuals of a linear trend surface, the lower-left panel uses residuals of a quadratic trend surface, the lower-right panel uses Bouguer's gravity anomaly as covariate.

The upper-left panel shows a variogram calculated from the original data, instead the upper-right panel uses residuals from a linear model. The lower-left panel uses residuals from a quadratic model and the lower-right panel uses residuals from Bouguer gravity anomaly as covariate.

The differences amongst the resulting variograms illustrate the inter-play between the specification of a model for the mean response and the resulting estimated covariance structure.

In order to assess the normality of geoid undulation's residuals, a Lilliefors (Kolmogorov-Smirnov) test was carried out. The results in this case indicate that by incorporating gravity anomaly into a model for the geoid undulation appears to achieve approximate stationarity of the residuals, because the corresponding variogram reaches a plateau [25]. Another important observation is that the semi-variance decreases significantly between raw variogram and residual variogram of a model adjusting for gravity anomaly as covariate.

2.3 Anisotropy detection

Once the trend has been removed, it is fundamental to detect remainder anisotropy. For this purpose, graphical tools such as variogram surface and directional variograms are used [26]. The original data showed geometric anisotropy that could be solved with a linear transformation of the spatial coordinates that will make the variation isotropic. Nevertheless, analysis of anisotropy of model trend residuals is a tool to define if a linear transformation is necessary.

2.4 Model validation

Two methods has been carried out to validate a model: 1) leave-one out (LOO) re-estimation, whereby one sample at a time is removed from the data set and re-estimated from the remaining data [27] (typically used when is not available an independent data set). If enough points are available, the effect of the removed point on the model (which was estimated using that point) is minor [28]; and, 2) jack-knife, whereby an independent data set or a subset of the data is removed completely from the data set, and reestimated using the original or remaining data.

3 Results and discussion

In practical terms, the multinormality nature of the data is a requirement to fulfill and one that is not possible to verify, but if the univariate distribution approximates a normal distribution, the working hypothesis is that the data were generated by a multivariate normal distribution [29]. For this reason, it is important to review the data distribution. The Figure 5 shows the histograms of heights and geoid undulations data.

In the present case, the skewness coefficient is 0.52 for geoid undulations and 2.81 for heights. In some cases, for example in environmental variables such as metal concentrations in soils, if the skewness coefficient is between 0.5 and 1, transformation to square roots might normalize the distribution approximately [14].

The D statistic for Lilliefors (Kolmogorov-Smirnov) test is 0.046335. According [29] the normality is confirmed with $D \leq 0.1$. It is not possible to compute significance levels because of the presence of spatial dependences in the samples. As mentioned earlier, outliers cause serious distortions in geostatistics [14]. In the current case there are not outliers according the review of box-plot shown in Figure 6 for geoid undulation (A preliminary analysis also was carried out), but for heights, there are an important number of outliers. This last situation is common because there are little peaks in the study area.

Through the analysis of variogram cloud was possible to detect some points which produce relevant differences between point pairs for the calculation of experimental variogram.

Postplot is shown in Figure 7 and directional clouds taking account geoid undulation in Figure 8. According these graphics, apparently there is a systematic increment of geoid undulation from south-west to north-east, then the use of Kriging with external drift constitutes an interesting option to implement.

Fig. 5. Frequency histograms of: a) heights (m) and b) geoid undulation (m). The heights between 0 m to 100 m are more frequent in the study area. The geoid undulation histogram does not show a predominant range.

Fig. 6. Box-plots of: a) heights (m) and b) geoid undulation (m) measurements in sample points. Outliers in heights reflect the variety of geomorphology in the study area in contrast with the absence of outliers in geoid undulation determinations.

3.1 Detrending

The regression graphical method for all subsets is shown in Figure 9; The best model is which that includes the terms: $North$, $North^2$, $East^2$ and the intercept showing highest R^2 with minor numbers of variables.

Another criteria to choose the best covariates configuration was the most parsimonious equation as measured by the Akaike Information Criterion (AIC) [30].

The relative weights for predictors are 22.76% for $East^2$, 38.60% for $North$ and 38.64% for $North^2$.

The final trend equation (Eq. 6) obtained is:

$$Z(s_i) = 1.026932 \times 10^5 - 3.619433 \times 10^{-11} East^2 - 2.120668 \times 10^{-2} North - 1.094790 \times 10^{-9} North^2 \quad (6)$$

Fig. 7. Spatial plot by geoid undulation (m) attribute. Projected coordinates in Projection Universal Transversa Mercator (UTM), Zone 17 South, Datum WGS84.

Fig. 8. Plots of geoid undulation data set in East and North directions, respectively (Directional clouds). The plots show a clear trend.

The coefficients were computed with ordinary least squares due to the observation points are approximately regularly spaced. Otherwise, the proper approach is residual maximum likelihood (REML) [26]. It is often more important to use more useful and higher quality data than to use sophisticated statistical methods. In some situations; for example: if the points are extremely clustered, and/or if the sample is ≤ 100 , and/or if the measurements are noisy or obtained using non-standard techniques; however, the model needs to be fitted using the most sophisticated technique to avoid making biased predictions [31].

Fig. 9. A regression graphical method for modeling of spatial trend of geoid undulation for all subsets. Axis X shows predictors of the regression model considered in this study. Axis Y shows fitting coefficient (R^2) for each predictions interaction.

A.1. Artículo: Spatial uncertainty of a geoid undulation model in 7 Guayaquil, Ecuador

Regarding Bouguer gravity anomaly, the correlation coefficient between this and geoid undulation for the study area is 0.81. This value shows a strong positive linear relationship between gravity anomaly and geoid undulation, situation that makes physical sense. Incorporating gravity anomaly into a model for the geoid undulation seems to achieve approximate stationarity of the residuals, because the corresponding variogram reaches a plateau [25]. Another important observation is that the semi-variance decreases significantly between raw variogram and residual variogram of a model adjusting using gravity anomaly as covariate.

Once the trend was identified, it was possible to compute the residual or true variogram. In order to compare the trend regression models, some statistics were computed. A summary of these one are shown in Table 1.

Table 1. Statistics (ME: mean error, MSE: mean square error, MSNE: mean square normalized error and RMSE: root mean square error) for spatial trend model and Bouguer anomaly as external drift.

Trend definition	ME	MSE	MSNE	RMSE
Spatial trend	-0.00053	0.0058996	0.9966555	0.07681
Bouguer anomaly	-0.00057	0.0058973	0.9966555	0.07679

For both trend models, the mean error (ME) proves the lack of bias of the estimate, with a value close to 0 [32].

The experimental variogram was computed with the classical equation (Eq. 7), (where Nh is the number of pairs of points at distance h (lag)), because in this case there are not outliers according the analysis of geoid undulation carried out. When an important number of these is detected, the use of a robust variogram estimators, such as Cressie and Hawkins, Dowd and Genton [14] is suggested.

$$\gamma(h) = (1/2N_h) \sum [Z(x_i + h) - Z(x_i)]^2 \quad (7)$$

The spherical, exponential, gaussian and cubic covariance models were analyzed in order to carry out the variogram fitting. Conceptually, the gaussian model is suitable for slowly-varying variables such as elevations [27]. Nevertheless, in this study, the spherical model shows the smallest residual sum of squares (RSS) (See Table 2) and the sill (0.8) is reached in 12.1 km. The spherical model has been used in other geostatistical analysis to carry out the variogram modelling [33]. The REML estimation was carried out based in the parameters of this model. Figure 10 shows the REML and spherical OLS fittings. Both of them are very similar. The REML parameters are shown in Table 3. The AIC shows that the absence of spatial dependence in this study is not a valid hypothesis (high AIC in nonspatial model).

Table 2. Diagnostics of model fitting with spherical, exponential, gaussian and cubic covariance models. The lowest RSS is obtained for spherical covariance model.

Model	RSS
Spherical	0.24142
Exponential	0.31492
Gaussian	0.25672
Cubic	0.25415

Fig. 10. Variogram fitting with a spherical covariance model with two methods: REML (continuous line) and OLS (dashed line). The experimental variogram is represented by circles.

Table 3. REML parameters obtained with a spherical covariance model. The tables shows the Akaike Interval Confidence (AIC).

Parameter	Value
Sill	0.713
Range	11,174
AIC	
Maximized likelihood	357.2
Non spatial model	782.1

3.2 Anisotropy evaluation

The Figure 11 shows the variogram surface, computed in gstat package [19], up to 45 km and with a bin width of 2.5 km. These parameters were selected because the variogram is only valid for a distance of one half of the field size since for larger distances the variogram begins to leave data out of the calculations [27] and the bin width is related with the spatial distribution of samples. The anisotropy is evident, with maximum continuity in the direction 110° clockwise from north. To confirm this assumption, directional variograms are shown in Figure 12. Notice that the variograms are computed in 100°, 105°, 110° and 115°. The 110° variogram shows the least variation.

84

85

86

87

88

Fig. 11. Variogram surface of geoid undulation with a bin width of 2,500 m. Semi-variances for each pixel are shown and defined with directions: X axis (dx) and Y axis (dy).

89

90

91

92

93

94

As mentioned in Section 3.1, once that the trend was modeled, a variogram surface was drawn from the residuals of trend model. The Figure 13 shows anisotropy with few significant variances so this could be negligible in the kriging procedure.

95

96

97

98

99

00

01

02

Fig. 12. Directional variograms of geoid undulations in 100°, 105°, 110° and 115°.

Fig. 13. Variogram surface of geoid undulation's residuals. Semi-variances for each pixel are shown and defined in directions: X axis (dx) and Y axis (dy). The semi variance decreases significantly.

3.3 Validation

The result of LOO cross validation method is shown in Figure 14. The errors have a normal distribution, with the major proportion of errors in the range (-0.5; 0.5).

Fig. 14. Histogram (left) and plot of predicted vs. data (red circles) in relation with the data set.

3.4 Estimation

Initially, the direction of maximum continuity was identified in 110° and the distance parameter for this direction was 40,000 m, nevertheless the trend model identified incorporates this particularity as evidenced in the variogram surface of residuals.

The estimation phase is the last one in Kriging which is an interpolation method that provides the best unbiased linear estimates (minimum variance) of point values or block averages [34]. The spatial continuity aspect is included in Kriging by variogram model fitted as mentioned above. The error estimation map of geoid

Fig. 15. Map of error estimation of geoid undulation (m). A mask has been applied in order to avoid the visualization of extrapolated areas.

Finally, in order to compare the geoid undulation estimations carried out in this study with Earth Gravitational models (EGM) used in Ecuador, a set of 33 validation points (Figure 16) which were measured in an independent field trip with the same methodology that were obtained the sampling points were employed. It is important to notice that these points were not used in the computation of geoid undulation estimation map. The mean square error (MSE) ratios in relation with the estimation map are 300 and 225 for EGM96 (spherical harmonic coefficients complete to degree and order 360) and EGM08 (this model contains the spherical harmonic coefficients of the gravity field and their errors to degree and order 2,160 and partially to 2,190 [35], respectively.

Fig. 16. Localization of validation points (red circles) in relation with the data set.

The results of estimation in validation points are shown in Figure 17.

Fig. 17. Results of errors modeled by EGM08, EGM96 and Kriging in validation points location.

4 Conclusions

The classical use of deterministic methods in order to get geoid undulations maps have not let to define the uncertainty of this variable in these estimations. For this reason, the present work uses a geostatistical approach. The advantage of kriging consists of involving all data set in the analysis. Weights applied do not depend only on the distance among the measured points and location of the predictor, but also on the spatial arrangement of measured points [36]. Since the geoid undulation is typically a slowly-varying variable, this study carries out the analysis of trend with emphasis.

The sill parameter of geoid undulation variogram decreases almost 100 times when gravity anomaly is considered as trend. For this reason, Bouguer anomaly is critical in the geoid undulation estimation process.

The statistics computed for the trend models analyzed in this study: a) spatial trend model and b) Bouguer anomaly as external drift are very similar. This situation is explained because the geoid undulation, typically, varies slowly in the geographical space and is related with physical processes such as tectonic structure dynamics [37].

A comparison of the estimation results with Earth gravitational models (typically used in Ecuador) was carried out in order to compare the performance of the proposal methodology. In comparison with Earth gravitational models: EGM96 and EGM08, the geoid undulation prediction map generated in this study shows an outstanding improvement analyzing the mean square error and the distribution of errors (Figure 17). According the characteristics of kriging method, the distribution of errors is unbiased. In relation with EGM96, the kriging's range of errors is almost the third part of this one. On the other hand, the EGM08 errors distribution is left-skewed. For this reason, the application of geostatistical techniques in this field represents a practical tool in order to use its products in detailed engineering projects where altimetric accuracy is fundamental.

Future research must be focused on optimal sampling design based in geoid undulations surveys carried out in Ecuador and the subsequent geostatistical analysis. Applications related with tectonic structure analysis must be addressed.

5 Acknowledgments

The authors wish to acknowledge to Instituto Geográfico Militar (IGM) of Ecuador for providing data collected in various field trips.

References

- [1] Goodchild, M. F., Devillers, R., and Jeansoulin, R., Fundamentals of spatial data quality, volume 662. John Wiley & Sons, 2006.
- [2] Fleming, C., Giles, J., and Marsh, S., Introducing elevation models for geoscience. Geological Society, 345:1–4, 2010.
- [3] Su, J. and Bork, E., Influence of vegetation, slope, and Lidar sampling angle on Dem accuracy. Photogrammetric Engineering & Remote Sensing, 72(11):1265–1274, 2006.
- [4] Katerji, W., Vario-Model for Estimating and Propagating DEM Vertical Accuracy: Case of Lebanon. PhD thesis, UPM, 2016.
- [5] Kopeikin, S. M., Mazurova, E. M., and Karpik, A. P., Towards an exact relativistic theory of earth's geoid undulation. Physics Letters A, 379(26):1555–1562, 2015.
- [6] Bouman, J., Relation between geoidal undulation, vertical deflection, vertical gravity gradient. Journal of Geodesy, 86(4):287–304, 2012.
- [7] Doganalp, S. and Selvi, H. Z., Local geoid determination in strip area projects by using polynomials, least-squares collocation and radial basis functions. Measurement, 73:429–438, 2015.
- [8] Sansò, F. and Sideris, M. G., Geoid determination: theory and methods. Springer Science & Business Media, 2013.
- [9] Isaaks, E. H. and Srivastava, R. M., An introduction to applied geostatistics. Oxford University Press, 1989.
- [10] Wechsler, S. P. and Kroll, C. N., Quantifying Dem uncertainty and its effect on topographic parameters. Photogrammetric Engineering & Remote Sensing, 72(9):1081–1090, 2006.
- [11] Machiels, L., La Roca Magica-Zeolite occurrence and genesis in the Late Cretaceous Cayo arc of Coastal Ecuador. PhD thesis, Department of Earth and Environmental Sciences. Division of Geology. KU LEUVEN, 2010.
- [12] Desassis, N. and Renard, D., Automatic variogram modeling by iterative least squares: Univariate and multivariate cases. Mathematical Geosciences, 45(4):453–470, 2013.
- [13] Cressie, N. A. and Cassie, N. A., Statistics for spatial data, volume 900. Wiley New York, 1993.
- [14] Oliver, M. and Webster, R., A tutorial guide to geostatistics: Computing and modelling variograms and kriging. Catena, 113:56–69, 2014.
- [15] Chiles, J.-P. and Delfiner, P., Geostatistics: modeling spatial uncertainty, volume 497. John Wiley & Sons, 2012.
- [16] R Development Core Team., R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2006. ISBN 3-900051-07-0.
- [17] Renard, D., Bez, N., Desassis, N., and Beucher, H., RGeostats: The Geostatistical package 10.0.8. MINES ParisTech. Free download from: <http://cg.ensmp.fr/rgeostats>, 2014.
- [18] Ribeiro Jr., P. and Diggle, P., geoR: a package for geostatistical analysis. R-NEWS, 1(2):15–18, 2001.
- [19] Pebesma, E., Multivariable geostatistics in S: the gstat package. Computers & Geosciences, 30:683–691, 2004.
- [20] Hatvani, I., Leuenberger, M., Kohán, B., and Kern, Z., Geostatistical analysis and isoscape of ice core derived water stable isotope records in an Antarctic macro region. Polar Science, 2017.
- [21] Faraway, J. J., Linear models with R. CRC Press, 2014.
- [22] Kabacoff, R., R in action: data analysis and graphics with R. Manning Publications Co., 2015.
- [23] Johnson, J. W., A heuristic method for estimating the relative weight of predictor variables in multiple regression. Multivariate behavioral research, 35(1):1–19, 2000.
- [24] Bonvalot, S., Balmino, G., Briais, A., Kuhn, M., Peyrefitte, A., Vales, N., et al., World gravity map, 636 1: 50,000,000 map, eds. Technical report, BGI-CGMW-CNES-IRD, Paris. 637, 2012.
- [25] Diggle, P. and Ribeiro, P. J., Model-based geostatistics. Springer Science & Business Media, 2007.
- [26] Rossiter, D. G., Geostatistics & open-source statistical computing. Exercise 5: Predicting from point samples, 2014.
- [27] Rossi, M. and Deutsch, C., Mineral resource estimation. Springer Science & Business Media, 2013.
- [28] Rossiter, D. G., Geostatistics & open-source statistical computing. Lecture 6 - Assessing the quality of spatial predictions, 2014.
- [29] Olea, R., Geostatistics for engineers and earth scientists. Springer Science & Business Media, 2012.
- [30] Akaike, H., Information theory and an extension of the maximum likelihood principle. In Selected Papers of Hirotugu Akaike, pages 199–213. Springer, 1998.
- [31] Hengl, T., A practical guide to geostatistical mapping. University of Amsterdam, 2009.
- [32] Guagliardi, I., Cicchella, D., De Rosa, R., and Buttafuoco, G., Assessment of lead pollution in topsoils of a southern Italy area: Analysis of urban and peri-urban environment. Journal of Environmental Sciences, 33:179–187, 2015.
- [33] Fabijńczyk, P., Zawadzki, J., and Wojtkowska, M., Geostatistical study of spatial correlations of lead and zinc concentration in urban reservoir, study case Czerniakowskie lake, Warsaw, Poland. Open Geosciences, 8(1):484–492, 2016.
- [34] Armstrong, M., Basic linear geostatistics. Springer Science & Business Media, 1998.
- [35] Eshagh, M. and Zoghi, S., Local error calibration of Egm08 geoid using Gnss/levelling data. Journal of Applied Geophysics, 130:209–217, 2016.
- [36] Zůvala, R., Fišerová, E., and Marek, L., Mathematical aspects of the kriging applied on landslide in Halenkovice (Czech Republic). Open Geosciences, 8(1):275–288, 2016.
- [37] Banerjee, P., Foulger, G., Dabral, C., et al. Geoid undulation modelling and interpretation at Ladak, nw Himalaya using Gps and levelling data. Journal of Geodesy, 73(2):79–86, 1999.

A.2 Artículo: Modelling uncertainty using geostatistics, a case study in Ecuador

Modelling uncertainty using geostatistics, a case study in Ecuador

Chicaiza Elena ^{*1} and Buenaño Xavier¹

¹Technical University of Madrid (UPM), Spain

*Corresponding author: gachisbar@gmail.com

Abstract

Spatial data includes positional and thematic components. In this study, the uncertainty analysis has been focused in thematic aspect. A total of 200 soil samples with metal determinations were analyzed. Graphical and numerical tools were employed to summarize and detect critical metals. Copper was identified as the cation of major concern in the study area. Kriging and Co-located Co-Kriging (with Zinc as secondary variable) were employed to carry out predictions. The leave-one-out cross validation method was employed to assess the variogram model used in estimation process, specifically mean square deviation ratio (MSDR) statistic was considered. Finally, unconditional simulations (500 realizations) were computed in order to get mean value and uncertainty of the random field of variable analyzed. The mean of variances obtained from simulations is 1.6 times higher than the mean of kriging variances.

Keywords

Uncertainty, simulation, validation.

I INTRODUCTION

According Griffith et al. (2015), spatial data comprise two components: attribute and location. In this work, we focused the analysis in attribute component.

Currently, the statement that descriptions of spatial phenomena are subject to uncertainty is now generally accepted as explained by Chiles and Delfiner (2009). In this context, Chiles and Delfiner (2009) argues Geostatistics can be defined as "the application of probabilistic methods to regionalized variables".

This study uses R Development Core Team (2013) open-source statistical software. In order to carry out the geostatistical analysis, two specific packages or libraries developed by Ribeiro Jr. and Diggle (2001) (geoR) and Pebesma (2004) (gstat) were employed.

The study area is located in the southeastern of Ecuador, in Zamora province. The area is inside a mining concession where informal mining activities are carried out. The accumulation of heavy metals in soils can bring human health problems, specially when these are exposed directly, for example in this kind of mining activities.

In this study, metals in 200 soil samples were analyzed. A multivariate data exploration analysis was carried out with some metals (Cu, Zn, Cd, and Pb). Kabacoff (2011) recommends the use of relatively new graphical statistical analysis tool named corrgram (see Figure 1) that summarize the relationships between variables very well.

The Figure 1 shows that these variables have a positive correlation (hatching in 45 degrees) with exception of Cd-Pb relation (hatching in -45 degrees); the color saturation in the figure repre-

Figure 1: A correlogram of heavy metals soil samples.

sents magnitude of correlation between variables. Notice the significant correlation between Cu and Zn.

II CALCULATIONS

Copper has been identified as an element of major concern. For this reason, the Cu variable was chosen for the spatial estimation by two methods: a) Ordinary Kriging and b) Co-located Co-Kriging. It's interesting to estimate this variable with a covariate, in this case the Zn metal, in order to define the uncertainty threshold between the two methods employed. The Figure 2 shows variograms of Cu and Zn and the cross-variogram.

Figure 2: Direct variograms of Cu, Zn, cross-variogram and fitted functions.

A initial exploration data analysis (EDA) of Cu variable was carried out in order to identify outliers and a possible presence of trend that could limit the application of geostatistical theory.

Through directional clouds plotting, spatial trend was discarded. Directional variograms (0, 45, 90, 135 degrees) confirm this assumption.

An important aspect to notice is that exists a nugget effect in the variogram model of both elements analyzed (Cu and Zn). In this case, the nugget effect is originated in the presence of micro-structure.

The same approach was carried out to analyze Zn variable, and the results were similar.

Some variogram fitting techniques can be employed, but according Oliver and Webster (2014) residual maximum likelihood (REML) method is the current best practice. This approach was carried out in the present study.

According Goovaerts (2001), the choice of an approach for uncertainty modeling should be guided by the answer to these questions: What type of uncertainty model is being sought?, What is the support for the uncertainty assessment?, Why do we model uncertainty?.

In this context, it is important to carry out a validation procedure. There are some ways to validate the model decisions, one of them is the cross validation technique.

Bivand et al. (2008) denotes that cross validation splits the data set into two sets: a modeling set and a validation set. When the number of splits is equal to the number of observations, the procedure is called leave-one-out cross validation.

For the cross-validation, leave-one-out technique was used. There are some statistics to analyze the cross-validation. The most interesting is the mean squared deviation ratio (MSDR), which is the mean of the squared errors divided by the corresponding kriging variances according Oliver and Webster (2014). The formula is shown in Equation 1.

$$MSDR = \frac{1}{N} \sum_{i=1}^N \frac{\{z(x_i) - \hat{Z}(x_i)\}^2}{\hat{\sigma}_K^2(x_i)} \quad (1)$$

In these equation $z(x_i)$ is the i th datum at x_i , $\hat{Z}(x_i)$ is the kriged prediction there, and $\hat{\sigma}_K^2(x_i)$ is the kriging variance.

Also mean error (ME) and root mean squared error (RMSE) were computed. The results are shown in Table 1.

	ME	RMSE	MSDR
Ordinary Kriging	0.104	45.318	1.017
Co-Kriging	-0.085	40.688	1.099

Table 1: Uncertainty results with two geostatistical methods.

Another important tool to define uncertainty is simulation. Bivand et al. (2008) defines geostatistical simulation as the simulation of possible realisations of a random field, given the specifications for that random field (e.g. mean structure, residual variogram, intrinsic stationarity) and possibly observation data.

In this context, the drawing of envelopes is an interesting technique. Envelopes were computed assuming a (transformed) Gaussian random field model. According Ribeiro Jr. and Diggle (2001), simulated values are generated at the data locations, given the model parameters obtained by REML. Empirical variogram is computed for each simulation using the same lags as for the original variogram of the data. The envelopes are computed by taking, at each lag, the

maximum and minimum values of the variograms for simulated data. In this case, 99 simulations were carried out. Figure 3 shows the envelopes based in this criteria. The figure shows the range in that we would expect to model the variogram of Copper variable.

Figure 3: Envelopes of simulations.

Also, unconditional simulations that ignore observations and only reproduce means and prescribed variability, have been carried out. A mean value of 71, corresponding to the mean of prediction map was considering to carried out the simulations.

The results of four (4) first realizations are shown in Figure 4.

Figure 4: Unconditional simulation of 4 realizations with REML variogram parameters

A total of five hundred (500) realizations were carried out. The mean of variances obtained in these was computed; this mean is 1.6 times higher than the mean of kriging variances. This is a reasonable value, considering that observations were not included in simulations. Nevertheless

the high values of kriging variances and simulations's mean variances could be explained for nugget effect inclusion in the models. The mean of random field generated by simulations was 70.38 and the mean of kriging predictions was 70.96. The proximity of these values reflects that the random field was modeled correctly.

III RESULTS

ME of two methods is almost equal. This situation demonstrates that kriging (in general) is an unbiased estimator. RMSE of ordinary kriging is higher than co-kriging due to the auxiliary variable provides more information to the model prediction.

On the other hand, MSDR of Ordinary Kriging is slightly lower than MSDR of co-kriging, probably due to the characteristics of this statist. Lloyd (2010) in his study concerned with mapping precipitation amount, found that OK outperformed CK (with elevation as the secondary variable) and this was due, at least in part, to the weak global relationship between the two variables.

The map of kriging prediction is shown in Figure 5. The gradient color represents the Copper variable estimation and the contour lines represent variance prediction. The highest variances are located in the corner of southwestern sector. This zone exhibits a lower sampling density in relation with the rest of study area.

Figure 5: Map of kriging prediction and variance.

References

- Bivand R. S., Pebesma E. J., Gomez-Rubio V. (2008). *Applied spatial data analysis with R*. Springer. Springer.
- Chiles J.-P., Delfiner P. (2009). *Geostatistics: modeling spatial uncertainty*, Volume 497. John Wiley & Sons.

A.2. Artículo: Modelling uncertainty using geostatistics, a case study in Ecuador

---

- Goovaerts P. (2001). Geostatistical modelling of uncertainty in soil science. *Geoderma* 103(1), 3–26.
- Griffith D., Wong D., Chun Y. (2015). Uncertainty-related research issues in spatial analysis. *Uncertainty Modelling and Quality Control for Spatial Data*, 3.
- Kabacoff R. (2011). *R in Action*. Manning Publications Co.
- Lloyd C. D. (2010). *geoENV VII – Geostatistics for Environmental Applications*, Chapter Multivariate Interpolation of Monthly Precipitation Amount in the United Kingdom, pp. 27–39. Dordrecht: Springer Netherlands.
- Oliver M., Webster R. (2014). A tutorial guide to geostatistics: Computing and modelling variograms and kriging. *Catena* 113, 56–69.
- Pebesma E. J. (2004). Multivariable geostatistics in s: the gstat package. *Computers & Geosciences* 30, 683–691.
- R Development Core Team (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <http://www.R-project.org>.
- Ribeiro Jr. P., Diggle P. (2001). geoR: a package for geostatistical analysis. *R-NEWS* 1(2), 15–18.

A.3 Artículo: Análisis de los usuarios y usos de los MDE en España

ANÁLISIS DE LOS USUARIOS Y USOS DE LOS MDE EN ESPAÑA

JOSÉ LUIS MESA-MINGORANCE¹, ELENA GABRIELA CHICAIZA², XAVIER BUENAÑO³,
CAI JIANHONG⁴, ANTONIO FEDERICO RODRÍGUEZ-PASCUAL⁵, FRANCISCO JAVIER
ARIZA-LÓPEZ¹

¹Universidad de Jaén, Dpto. de Ingeniería Cartográfica, Geodésica y Fotogrametría
Campus Las Lagunillas s/n. 23071, Jaén, España
jlmesa@ujaen.es

²Instituto Panamericano de Geografía e Historia (IPGH)

³Universidad Politécnica de Madrid

⁴Beijing University of Civil Engineering and Architecture, China

⁵Centro Nacional de Información Geográfica

RESUMEN

Se ha realizado una caracterización de los usuarios y usos de los MDE (MDE05, MDE25 y MDE200) que se descargan del centro de descargas del CNIG. Para realizar este trabajo se ha utilizado un conjunto de 12 493 registros de encuestas realizadas en línea y voluntariamente en el momento de descarga y correspondientes al año 2014. Con estos datos se ha determinado la geografía de las descargas, el perfil de los usuarios, el uso de los productos y su valoración. Se han identificado 6 087 usuarios distintos, y de esta población los usuarios mayoritarios tienen el perfil de profesionales particulares (71 %) y aquellos relacionados con actividades docentes (18 %); además, la mayoría de ellos solo realizan una descarga. Los usos predominantes son los relacionados con el hacer docente e investigador y los de carácter profesional. Destaca que los usos relacionados con el ocio, deporte y turismo alcanzan un 9.5 % de los casos. La valoración de la utilidad de los productos es muy alta, pero no tanto las necesidades de actualización.

Palabras clave: MDE, uso, usuario.

ABSTRACT

We have performed a characterization of users and uses of DEM (DEM05, DEM25, DEM200) are downloaded from the download center of CNIG. To make this work we used a set of 12.493 records online surveys and voluntarily at the time of discharge and for the year 2014. With this data we have determined the geography of downloads, the profile of users, the use of the

Recibido: 07/03/2016

Aceptada versión definitiva: 30/05/2016

[Attribution-NonCommercial-NoDerivatives 4.0 International \(CC BY-NC-ND 4.0\)](#)

© Los autores

www.geofocus.org

Mesa-Mingorance, J.L., Chicaiza, E.G., Buenaño, X., Jianhong, C., Rodríguez-Pascual, A.F., Ariza-López, F.J. (2016): "Análisis de los usuarios y usos de los MDE en España", *GeoFocus (Artículos)*, nº17, p. 133-160. ISSN: 1578-5157

products and their valuation. We have identified 6087 different users, and the majority of this population profile users have individual professionals (71%) and those related to educational activities (18%); Moreover, most of them only perform one download. The predominant uses are related to teaching and research and professional character. It's noteworthy that the uses related to leisure, sport and tourism reached 9,5% of cases. The assessment of the utility of the products is very high, but not so much needs updating.

Keywords: DEM, use, user.

1. Introducción

En España, la abundancia de productos de datos geoespaciales de carácter digital (p.e. ortofotos, modelos de elevaciones, callejeros, etc.) es una realidad destacada en comparación con otros países. Esta realidad se ha alcanzado con el esfuerzo de todos los españoles, vía presupuestos, y el buen hacer de las agencias y servicios cartográficos regionales (p.e. Instituto de Estadística y Cartografía de Andalucía, el Instituto Cartográfico Valenciano, el Instituto Cartográfico de Cataluña, etc.) y la coordinación nacional del Instituto Geográfico Nacional y el Centro Nacional de Información Geográfica.

En apenas algo más de una década hemos pasado de un sistema obscuro y de acceso caro a los datos geoespaciales, mediante pago (p.e. en 1996 el precio de una hoja del MTN50 con altimetría era de 50 000 pts, unos 300 €), a una situación de gran transparencia, disponibilidad y fácil acceso: hay metadatos, acceso sin coste, disponibilidad en línea, sistemas de descarga, etc. En todo este proceso se debe reconocer el empuje recibido por el marco regulatorio de la UE, principalmente por la Directiva Inspire (Directiva 2007/2/CE), y la alineación legislativa que se ha realizado en todo nuestro sistema de legislación y normativas cartográficas (p.e. LISIGE, Decretos de Andalucía, etc.).

En este macroproceso se ha prestado mucha atención a crear productos de datos (p.e. vuelos PNOA, CartoCiudad, etc.), a implantar servicios (p.e. WMS, WFS, etc.) y a crear los correspondientes metadatos y sus servicios de catálogo, e incluso a analizar las ventajas económicas que supone invertir en las IDE (EC, 2008). Sin embargo, bajo nuestro punto de vista, existe, y ha existido, una falta clara de atención al análisis de los usuarios de los productos de datos y de los usos que realizan esos usuarios. Toda producción industrial, ya sea de bienes o servicios generales, o de productos geomáticos, se justifica sólo y exclusivamente por el valor y utilidad que genera su uso, por ello en los sistemas de gestión de la calidad con una perspectiva de calidad total, y también en los sistemas de excelencia, se dedican esfuerzos considerables a disponer de realimentación de los usuarios. En algunos modelos de gestión de la calidad de datos geoespaciales, como el recogido en la Norma Técnica Cartográfica (IECA 2011), se incluye la necesidad de identificar a los usuarios y evaluar su satisfacción a lo largo del tiempo.

El objetivo de este trabajo es realizar una aproximación actual a los usuarios y usos de los Modelos Digitales de Elevaciones (MDE) en España. Nuestra intención es conocer mejor el perfil de los usuarios y saber en qué tipo de aplicaciones se utilizan los datos de MDE. Consideramos que este análisis es de interés para justificar mejor la inversión en esos datos ciertamente costosos, poder

Recibido: 07/03/2016

Aceptada versión definitiva: 30/05/2016

© Los autores

www.geofocus.org

Mesa-Mingorance, J.L., Chicaiza, E.G., Buenaño, X., Jianhong, C., Rodríguez-Pascual, A.F., Ariza-López, F.J. (2016): "Análisis de los usuarios y usos de los MDE en España", *GeoFocus (Artículos)*, nº17, p. 133-160. ISSN: 1578-5157

segmentar el mercado, determinar tendencias y posibles oportunidades de desarrollo de productos de valor añadido, etc.

Hemos decidido centrarnos en los MDE por ser un producto muy explotado, con muchas y diversas oportunidades de uso. Los MDE aportan la componente altimétrica al análisis del territorio dentro de los Sistema de Información Geográfica (SIG), o bien en los servicios de procesamiento en la web (WPS). Los MDE tienen aplicación directa en la descripción geomorfológica del terreno (Gomez, Hayakawa, Obanawa, 2015), son la superficie de referencia para todas las aplicaciones hidrológicas (ciclo del agua, erosión, inundaciones, etc.) (Saksena, Merwade, 2015), sirven de base para el desarrollo de modelos de tipo forestal (Juel, Groom, Svenning, Ejrnaes, 2015) y agrario (Rekha, Gangadharan, Ravichandran, Mahalakshmi, Panigrahi, Pillai, 2015) y tienen una utilidad inestimable en todo lo que es ingeniería civil (Stroeve, Li, Le, He, Stroeve, 2015).

Para desarrollar nuestro análisis, tanto de los usuarios como de los usos, utilizaremos datos procedentes del servicio de descarga del CNIG (<http://centrodedescargas.cnig.es/CentroDescargas/index.jsp>). Este servicio de descarga incluye una encuesta voluntaria que se ofrece para ser rellenada en cada descarga si el usuario lo desea. Este tipo de utilidad no está muy extendida en los centros de descarga de información geográfica, pero como se podrá comprobar, es de mucha utilidad para tener realimentación de los usuarios. El análisis estadístico realizado se basa en conteos y obtención de porcentajes para variables, y cruces de variables y atributos.

Este trabajo consta de tres apartados además de esta introducción. En el primero de ellos se presenta el material y el método aplicado. El segundo presenta los resultados y su análisis. Este apartado se centra en describir los hallazgos relativos a: las descargas (lugar, cantidad, etc.), el perfil de los usuarios, el uso de los productos y la valoración de los productos. El trabajo acaba con unas conclusiones que abarcan todos los aspectos presentados.

2. Material y método.

Como se ha indicado previamente, la información utilizada en este análisis procede de la encuesta voluntaria (véase la Figura 1, <http://centrodedescargas.cnig.es/CentroDescargas/index.jsp>) que, de manera opcional, deben rellenar los usuarios que deseen hacer descargas del Centro de Descargas del CNIG.

Mesa-Mingorance, J.L., Chicaiza, E.G., Buenaño, X., Jianhong, C., Rodríguez-Pascual, A.F., Ariza-López, F.J. (2016):
“Análisis de los usuarios y usos de los MDE en España”, *GeoFocus (Artículos)*, n°17, p. 133-160. ISSN: 1578-5157

Figura 1. Pantalla del centro de descargas del CNIG

Los MDE son conjuntos de datos que, según la Ley 14/2010, la llamada LISIGE (Ley de Infraestructuras y Servicios de Información Geográfica de España) forman parte del conjunto de datos de referencia y pertenecen a la lista del conjuntos de datos del Anexo II de Inspire, cuya implementación está prevista de manera más inmediata y en plazos más cortos. Según la legislación comunitaria estos datos deben estar accesibles para todos los ciudadanos y administraciones. Esto es

Recibido: 07/03/2016

Aceptada versión definitiva: 30/05/2016

© Los autores

www.geofocus.org

Mesa-Mingorance, J.L., Chicaiza, E.G., Buenaño, X., Jianhong, C., Rodríguez-Pascual, A.F., Ariza-López, F.J. (2016): "Análisis de los usuarios y usos de los MDE en España", *GeoFocus (Artículos)*, nº17, p. 133-160. ISSN: 1578-5157

posible por medio de los servicios de mapas (WMS, WMTS) y los servicios de descarga (WCS y Atom). Además, en España los datos MDE eran durante el año 2014 descargables y utilizables libremente siempre que su uso fuera comercial (<http://centrodedescargas.cnig.es/CentroDescargas/cambiarMenu.do?destino=infoUsoNoComercial>), lo cual facilita enormemente la extensión de su conocimiento y su utilización. En España, los productos que están disponibles para su descarga y sobre los cuáles se ha podido disponer de información facilitada por los usuarios son:

- MDT05/MDT05-Lidar. Modelo con paso de malla de 5x5 m, con la misma distribución de hojas que el MTN50. Formato de archivo ASCII matriz ESRI (asc). Sistema geodésico de referencia ETRS89 y proyección UTM en el huso correspondiente a cada hoja. Según la hoja de que se trate, el MDT05 se ha obtenido de una de las dos siguientes formas: por estereocorrelación automática de vuelos fotogramétricos del Plan Nacional de Ortofotografía Aérea (PNOA) con resolución de 25 a 50 cm/píxel, revisada e interpolada con líneas de ruptura donde haya sido viable, o bien por interpolación a partir de la clase terreno de vuelos LIDAR del PNOA.
- MDT25. Modelo con paso de malla de 25x25 m, con las mismas características de distribución y de sistema de referencia de coordenadas que el MDT05. El MDT25 se ha obtenido por interpolación de modelos digitales del terreno de 5x5 m de paso de malla procedentes del PNOA.
- MDT200. Modelo con paso de malla de 200x200 m, con distribución por provincias (rectángulo envolvente de cada provincia). Formato de archivo ASCII matriz ESRI (asc). Sistema geodésico de referencia ETRS89 y proyección UTM en el huso correspondiente a cada provincia. El MDT200 se ha obtenido por interpolación de modelos digitales del terreno de 5x5 m de paso de malla procedentes del PNOA.

En total se ha dispuesto de 12 493 registros correspondientes a todas las descargas realizadas a lo largo del año 2014 por todos los usuarios que rellenaron la encuesta y que descargaron datos de cualquiera de los tres productos MDT ya indicados. En cada encuesta registrada se dispone de los campos que rellena el usuario (véase la Figura 1), así como de otros campos no visibles en la encuesta (como la fecha). Para una contestación sencilla y normalizada, algunos campos permiten la selección de una única opción de una lista enumerada y otros una selección múltiple sobre una lista de chequeo. En los campos de observaciones se ofrece la opción de contestación en forma de texto libre. Los campos y sus tipologías son las que se presentan en la Tabla 1.

Tabla 1. Campos de la encuesta a usuarios del centro de descargas del CNIG

Campo	Valores que puede tomar
MDT	MDT-05
	MDT-25
	MDT-200
Fecha	Fecha de la descarga
Perfil usuario	Uso particular
	Microempresa (<10 empleados)
	PYME (<50 empleados)

Recibido: 07/03/2016

Aceptada versión definitiva: 30/05/2016

© Los autores

www.geofocus.org

Mesa-Mingorance, J.L., Chicaiza, E.G., Buenaño, X., Jianhong, C., Rodríguez-Pascual, A.F., Ariza-López, F.J. (2016):
“Análisis de los usuarios y usos de los MDE en España”, *GeoFocus (Artículos)*, nº17, p. 133-160. ISSN: 1578-5157

	Medina empresa (entre 51 y 250 empleados)
	Gran empresa (>250 empleados)
	Administración General del Estado
	Comunidad Autónoma
	Universidad
	Administración Local
	Otros
Utilidad	1 poco útil 2 3 4 5 mucha utilidad
Observaciones a la utilidad	Libre: Observaciones del usuario
Necesidad de actualización	1 poca necesidad 2 3 4 5 mucha necesidad
Observaciones a la necesidad de actualización	Libre
Provincia	Provincia
País	País
Descripción de uso al que va destinado	Libre: Descripción del usuario.
Área Temática (Sectores de empleo del MDT) (Se pueden marcar varios sectores)	Medioambiental
	Forestal / Biodiversidad
	Cambio Climático
	Oceanografía / Costas
	Catastro
	Infraestructuras e Ingeniería civil
	Geofísica
	Defensa y Seguridad
	Jurídico
	Telecomunicaciones
	Vivienda
	Cartografía
	Ordenación del Territorio y Urbanismo
	Investigación, Ciencia e Innovación
	Geología
	Ocio y Tiempo libre
	Energía y recursos minerales
	Zonas de Riesgo / Protección civil
	Arqueología y estudios temporales
	Sociología
	Turismo
	Educación y Formación
	Agricultura
	Demografía, expansión urbana

Recibido: 07/03/2016

Aceptada versión definitiva: 30/05/2016

© Los autores

www.geofocus.org

Mesa-Mingorance, J.L., Chicaiza, E.G., Buenaño, X., Jianhong, C., Rodríguez-Pascual, A.F., Ariza-López, F.J. (2016): "Análisis de los usuarios y usos de los MDE en España", *GeoFocus (Artículos)*, nº17, p. 133-160. ISSN: 1578-5157

	Hidrología
	Suelos (Edafología)
	Transporte y Logística
	Navegación y Localización
	Fauna, Entomología / Biodiversidad
	Marketing (Mercadotecnia)
	Sanidad
	Artística y Cultural
	Otros
Observaciones a los sectores	Libre: comentarios aclaratorios del usuario
Ámbito	Internacional
	Europeo
	Nacional
	CC. AA. / Regional
	Provincial
	Otro (local, comarcal, cuenca hidrográfica, etc.)

Una vez se dispuso de los 12 493 registros o casos (N), se procedió a analizarlos para definir un proceso que permitiera utilizarlos en un estudio estadístico. Como se ha indicado, la encuesta está diseñada para que muchos ítems se contesten a partir de selecciones, pero campos de posible gran interés, como son los de observaciones, se recogen en forma de texto libre, por lo que se ha hecho necesario un proceso de normalización. También hay campos (p.e. área temática), que presentan registros que son agregación de respuestas de selección múltiple que deben ser desagregados de manera previa a su análisis. A continuación se presentan estos procesos de preparación de los datos.

Se ha procedido a la normalización de todos los campos que permiten entradas de texto libre. Estos campos son: "Observaciones a los sectores", "Descripción de uso al que va destinado", "Observaciones a la utilidad". En primer lugar se debe indicar que no todos los registros incluyen entradas en estos campos. El porcentaje de entradas es 4.64 %, 99.95 %, 1.53 % para cada uno de los campos indicados, respecto al total de casos N. En todos ellos existe una gran variedad de respuestas, situación que dificulta cualquier análisis estadístico. El proceso de normalización ha sido manual, por lo que ha sido tedioso y costoso en tiempo. La normalización se ha realizado de manera conjunta para los 3 productos, al objeto de que sea homogénea. El proceso consiste en sustituir los contenidos de estos campos, ya sean palabra(s), frase(s) o textos más largos, por un único término o clave que clasifique la contestación. La consideración de cada uno de estos términos obedece a la idea de recoger, lo más fielmente posible, los casos existentes, sin perder generalidad. Estos términos no se pueden generar a la primera, se trata de un proceso necesariamente iterativo, de tal manera que, a medida que se van analizando los registros, se van aclarando el número de términos a emplear y su propio alcance.

En este trabajo tras el análisis de los 12 486 registros con aportaciones en las observaciones del campo "Descripción de uso al que va destinado," y de los otros dos campos de comentarios, se han propuesto los 44 términos para normalizar las contestaciones, éstos se presentan en la Tabla 2.

Tabla 2. Términos de usos normalizados

Recibido: 07/03/2016

Aceptada versión definitiva: 30/05/2016

© Los autores

www.geofocus.org

Mesa-Mingorance, J.L., Chicaiza, E.G., Buenaño, X., Jianhong, C., Rodríguez-Pascual, A.F., Ariza-López, F.J. (2016): "Análisis de los usuarios y usos de los MDE en España", *GeoFocus (Artículos)*, nº17, p. 133-160. ISSN: 1578-5157

1.- Acústico. Hay aplicaciones de los MDT en el campo del análisis de ruidos (p.e. cerca de aeropuertos, ciudades, etc.). Si bien este tipo de aplicación se podría haber considerado dentro del campo ambiental, se considera interesante destacar su presencia.
2.- Agricultura. Se engloban todas las actividades que se han identificado relacionadas con el agro.
3.- Análisis SIG. Se engloban procesos de derivación de pendientes, orientaciones, cuencas visuales, transformación de datos, etc.
4.- ArqueologíaPatrimonio. Se engloban todas las actividades que se han identificado relacionadas con la arqueología y patrimonio.
5.- Arquitectura. Se engloban todas las actividades que se han identificado relacionadas con la arquitectura. La ordenación del territorio queda bajo otro término.
6.- Aviación. Se engloban todas las actividades que se han identificado relacionadas con la aviación (vuelos, aproches, etc.).
7.- Biología. Se engloban todas las actividades que se han identificado relacionadas con las plantas, insectos, etc., que presentan un carácter más específico que las que aparecen bajo el término genérico de medioambiental.
8.- CatastroMercado. Se engloban todas las actividades que se han identificado relacionadas con el catastro, las parcelas y aspectos de valoración y mercado (p.e. <i>geomarketing</i>). La cantidad de registros en catastro supera a las de mercado.
9.- Comprobación. Numerosos registros indican "comprobación", "comprobación de valores", "comprobación de cotas", etc. Todos estos casos se incluyen bajo este término. Realmente son un caso sin especificar pero es interesante que se da el sentido de comprobación.
10.- Curiosidad. Numerosos registros indican que los usuarios se acercan a estos productos por mera curiosidad. Se entiende que no realizan aplicaciones específicas, simplemente toman contacto con el producto, por lo general en un entorno geográfico y temático que les resulte conocido.
11.- DefensaSeguridadPCivil. Se engloban todas las actividades que se han identificado relacionadas con la defensa, la seguridad o los servicios de protección civil, dejando a un lado los incendios por su especificidad.
12.- DocenciaEstudio. Bajo este término se engloban todas las actividades que se han identificado relacionadas con la docencia, en todos sus niveles, incluido el universitario y los cursos cortos de formación. Se indican como simple docencia, preparación de clases por parte del profesorado (teóricas o prácticas). También se incluyen los registros relacionados con el estudio, es decir, la actividad por parte del alumnado, en todos sus niveles.
13.- EnergíasRenovables. Se engloban todas las actividades que se han identificado relacionadas con la energía solar (térmica y fotovoltaica), biomasa, etc.
14.- Forestal. Se engloban todas las actividades que se han identificado relacionadas con los montes y las especies forestales.
15.- GeografíaTerritorio. Se engloban todas las actividades que se han identificado relacionadas con el conocimiento del medio geográfico y del territorio. La ordenación y actuación sobre el territorio se considera bajo otro término.
16.- Geología. Se engloban todas las actividades que se han identificado relacionadas con la

Recibido: 07/03/2016

Aceptada versión definitiva: 30/05/2016

© Los autores

www.geofocus.org

Mesa-Mingorance, J.L., Chicaiza, E.G., Buenaño, X., Jianhong, C., Rodríguez-Pascual, A.F., Ariza-López, F.J. (2016): "Análisis de los usuarios y usos de los MDE en España", *GeoFocus (Artículos)*, nº17, p. 133-160. ISSN: 1578-5157

geología, geodinámica, edafología, etc.
17.- Geomática. Se engloban todas las actividades que se han identificado relacionadas con la cartografía, generación de mapas, gestión de MDT, aplicaciones en geodesia, aplicaciones en fotogrametría y teledetección, aplicaciones topográficas, etc.
18.- Hidrología. Se engloban todas las actividades que se han identificado relacionadas con el agua superficial y costera.
19.- Historia. Se engloban todas las actividades que se han identificado relacionadas con la historia.
20.- ImpactoAmbiental. Se engloban todas las actividades que se han identificado relacionadas con estudios y proyectos de impacto ambiental. Se consideran fuera del término general de Medioambiente por su especificidad.
21.- Incendios. Se engloban todas las actividades que se han identificado relacionadas con incendios.
22.- IngCivil. Se engloban todas las actividades que se han identificado relacionadas con cualquier actividad propia de la ingeniería civil, excluidas la agronómica, forestal y telecomunicaciones.
23.- Investigación. Se engloban todos los registros en los que aparece este término si bien no especifica el campo de la investigación. Cuando aparece Docencia e Investigación se ha normalizado como investigación. Las tesis y doctorados se normalizan como investigación.
24.- JuridicoJudicialAdministrativoLicitacion. Se engloban todos los registros relacionados con actividades jurídicas, judiciales, administrativas o de preparación de licitaciones.
25.- Localización. Muchos usuarios utilizan este término para describir el uso que dan a los MDT.
26.- Medioambiente. Es una categoría amplia en la que se engloban todas las actividades relacionadas con el medioambiente y que no se han considerado bajo otro epígrafe.
27.- Minería. Se engloban todos los registros relacionados con actividades mineras.
28.- Modelizacion. Muchos usuarios utilizan este término para describir el uso que dan a los MDT. Dentro de esta categoría no se incluyen las modelizaciones hidrológicas (hidrología), ni las modelizaciones 3D (Visualización3D).
29.- Navegación. Se engloban todos los registros relacionados con la navegación.
30.- OcioDeporteTurismo. Se engloban todos los registros relacionados con actividades de ocio y deporte (senderismo, travesía, rutas con bici, etc.) y el turismo.
31.- Paisaje. Se engloban todos los registros relacionados con el paisaje.
32.- Planificación. Se engloban todos los registros relacionados con el paisaje.
33.- Problemas. Se engloban todos los registros relacionados con problemas con los datos, descargas, etc.
34.- PruebasDatosSistemas. Se engloban todos los registros relacionados con usos en los que se indica que son pruebas de los datos, de los sistemas de descarga, de herramientas software que se desarrollan sobre estos datos, etc.
35.- PublicacionesDivulgación. Se engloban todos los registros relacionados con usos en los que se indica la creación de folletos, uso en web, creación de imágenes para publicaciones, etc.

Mesa-Mingorance, J.L., Chicaiza, E.G., Buenaño, X., Jianhong, C., Rodríguez-Pascual, A.F., Ariza-López, F.J. (2016): "Análisis de los usuarios y usos de los MDE en España", *GeoFocus (Artículos)*, n°17, p. 133-160. ISSN: 1578-5157

36.- Redes. Se engloban todos los tipos de redes de suministro (p.e .agua, electricidad, etc.)
37.- RiesgosNaturales. Se engloban los casos que se indica este tipo de análisis sobre cualquier aspecto (p.e. sísmico, natural, etc.).
38.- SinEspecificar. Numerosos usuarios no aportan ningún comentario y otros muchos no aportan información adecuada para determinar el uso concreto (p.e. particular, el de siempre, etc.). Muchos casos aparecen como estudio(s), pero no se sabe si es un estudio/oficina/gabinete de ingeniería o que es una actividad docente, se ha considerado SinEspecificar. También se ha incluido aquí los registros en los que se indica "consulta", "informes", y otros con contenidos absurdos y ofensivos, etc.
39.- Telecomunicaciones. Se engloban todos los registros relacionados con actividades propias de las telecomunicaciones.
40.- TFCTFGTFM. Se engloban todos los registros relacionados con la actividad de elaboración de los proyectos fin de carrera, fin de grado o fin de máster. Las tesis se consideran investigación.
41.- UrbanismoOTerritorio. Se engloban todos los registros relacionados con la actividad de gestión y planificación urbana y territorial.
42.- Visualizacion. Se engloban todos los registros relacionados con la actividad de visualización (pero no mapas, que se consideran bajo Geomática).
43.- Visualizacion3D. Se engloban todos los registros relacionados con la actividad de obtener visualizaciones 3D, ya sean en pantalla, en forma de maquetas 3D, entornos virtuales, video juegos, simulación gráfica, etc.
44.- Valora el producto, realiza sugerencias, agradece su disponibilidad. Casos en los que se realizan comentarios en alguna de estas líneas.

Como se ha indicado, los valores registrados en el campo sectores pueden incluir un número variado de contestaciones. Para poder tratar adecuadamente este campo, se ha añadido a cada registro tantos campos como posibles opciones presenta el formulario de encuesta, y se ha creado una variable booleana que toma valor 1 cuando el usuario ha marcado esa opción.

Respecto al análisis de los datos, el objetivo es caracterizar los usuarios y usos de los MDT en España y para ello las herramientas a utilizar son las propias de la estadística descriptiva. Se analizarán campos de manera independiente mediante el uso de porcentajes, también se utilizarán técnicas de tabulación cruzada para obtener valores numéricos en el cruce de aquellas opciones que se consideren más relevantes.

3. Resultados y análisis de los datos.

A continuación se presentan y comentan los resultados del análisis del total de registros disponibles, organizados en las siguientes secciones:

- Lugar y descarga de producto
- Perfil del usuario
- Uso del producto
- Valoración del producto

Recibido: 07/03/2016

Aceptada versión definitiva: 30/05/2016

© Los autores

www.geofocus.org

Mesa-Mingorance, J.L., Chicaiza, E.G., Buenaño, X., Jianhong, C., Rodríguez-Pascual, A.F., Ariza-López, F.J. (2016): "Análisis de los usuarios y usos de los MDE en España", *GeoFocus (Artículos)*, n°17, p. 133-160. ISSN: 1578-5157

3.1. Lugar y descarga del producto

En este subapartado se presentarán los resultados relativos a la frecuencia de descarga de cada producto y al lugar donde el usuario ha indicado que trabaja.

En relación a las descargas por producto, los resultados de los 12 493 registros son los mostrados en la tabla 3.

Tabla 3. Datos referentes al tipo de MDT descargado

MDT	Nº de registros	% sobre el total
MDT200	1 377	11.02
MDT25	4 180	33.46
MDT05	6 936	55.52

Se puede apreciar que el mayor porcentaje de descarga es para el MDT05 y conforme aumentamos el paso de malla, y por tanto disminuye la resolución, la cantidad de descargas va disminuyendo. Consideramos que este hecho se puede deber a la tendencia que tenemos todos los usuarios a utilizar siempre el producto de mejores prestaciones, con independencia de que sea o no el más adecuado para un trabajo concreto.

En relación al lugar, la encuesta presenta varios campos para determinar el lugar dónde reside el usuario y dónde se produce la descarga. En cuanto a los países, la Tabla 4 presenta los valores. Como cabía esperar la mayoría de las descargas se producen en España. Fuera de España el número de descargas es residual. Los países que aparecen o son europeos o hispanoamericanos. Lo relevante es que se trata de descargas esporádicas y que todos ellos contabilizan sólo 1 o 2 descargas, a excepción de Francia, Portugal y Reino Unido, que contabilizan algún caso más.

Tabla 4. Descargas por países

País	Nº de registros
Alemania	1
Argentina	2
Bélgica	2
Bolivia	1
Brasil	1
Canadá	2
España	12 470
Francia	4
Gutemala	1
Honduras	1
Italia	1
México	1
Portugal	3

Recibido: 07/03/2016

Aceptada versión definitiva: 30/05/2016

© Los autores

www.geofocus.org

Mesa-Mingorance, J.L., Chicaiza, E.G., Buenaño, X., Jianhong, C., Rodríguez-Pascual, A.F., Ariza-López, F.J. (2016): "Análisis de los usuarios y usos de los MDE en España", *GeoFocus (Artículos)*, n°17, p. 133-160. ISSN: 1578-5157

Reino Unido	3

En la Figura 2 se presentan dos mapas con la distribución de las descargas, el primero de ellos como porcentaje de descarga en cada provincia respecto al total de las descargas registradas y el segundo, como tanto por mil de las descargas realizadas en esa provincia respecto a su población según los datos del Instituto Nacional de Estadística para 2014. Conviene aclarar la forma de confección de ambos mapas. Para el primero se ha escogido un sistema de clasificación de intervalos de roturas naturales (*Natural Break*), en el cual se agrupan los datos creando clases de tal forma que la varianza interna de las clases sea mínima en comparación con la varianza entre clases. Con este método conseguimos representar los grupos más significativos de porcentajes de descargas. Para el segundo mapa, el sistema de clasificación que se ha adoptado ha sido el de intervalos de igual amplitud, ya que el tanto por mil se comporta de una forma muy regular con una distribución normal y existe un buen reparto de los valores de tanto por mil de todas las provincias en todo el recorrido de la variable.

En la Figura 2.a se observa la existencia de un pico de descargas en Madrid ($\approx 18\%$). La elevada cantidad de descargas en Madrid puede parecer directamente relacionada con su población y peso económico, pero según este criterio también deberían destacar otras provincias como Barcelona, pero en ese caso consideramos que se ha de tener en cuenta que el Instituto Cartográfico de Cataluña también ofrece un servicio de descarga de MDE. El siguiente nivel de descargas está en el intervalo $[2.81, 4.39]$, donde se encuentran 7 provincias de carácter periférico (La Coruña, Barcelona, Valencia, Alicante, Murcia, Granada, Sevilla) más Zaragoza. En esta misma figura destaca el hecho de existir una especie de franja de territorio, de orientación predominante este-oeste, formada por provincias en las que se ha detectado el menor porcentaje de descargas.

Dado que la Figura 2.a presenta el total de descargas y que la población total de cada provincia es muy distinta, se ha elaborado la Figura 2.b en la que se puede ver el tanto por mil de descargas respecto a la población total de cada provincia. Con esta figura se obtiene una perspectiva complementaria al caso anterior. Así, en este caso destacan las provincias de Ávila, Cuenca y Albacete en el intervalo más alto. El caso de Ávila consideramos que se puede deber a la existencia de la Titulación Universitaria correspondiente al Grado de Ingeniería en Geomática y Topografía, con su correspondiente máster, pues como se verá con posterioridad, el mundo universitario es un gran usuario de MDE. En este caso también existe un efecto periferia, pero en el sentido contrario al caso anterior, lo cual es entendible pues aunque el número de descargas sea mayor, al matizarlas por mayores poblaciones los tantos por miles decrecen. Así, destaca el bajo nivel de descargas, en tanto por mil, en las provincias forales, las catalanas y valencianas. Pensamos que, además de lo indicado anteriormente, probablemente esto también puede deberse a la existencia de servicios autonómicos de descargas de productos similares. Finalmente, llama también la atención el caso de Málaga, que siendo una provincia muy poblada, y teniendo un alto nivel de desarrollo y actividad, sin embargo, tiene un nivel bajo de descargas en proporción a su población.

Mesa-Mingorance, J.L., Chicaiza, E.G., Buenaño, X., Jianhong, C., Rodríguez-Pascual, A.F., Ariza-López, F.J. (2016):
"Análisis de los usuarios y usos de los MDE en España", *GeoFocus (Artículos)*, n°17, p. 133-160. ISSN: 1578-5157

Figura 2. Distribución por provincias de las descargas en España: a) porcentaje respecto al total de descargas, b) descargas en tanto por mil respecto al total de la población de la provincia.

3.2. Perfil del usuario

Establecer el perfil del usuario es uno de los principales objetivos de este análisis. Para el sistema de descargas cada usuario es cada persona o institución que se da de alta en el sistema y,

Recibido: 07/03/2016

Aceptada versión definitiva: 30/05/2016

© Los autores

www.geofocus.org

Mesa-Mingorance, J.L., Chicaiza, E.G., Buenaño, X., Jianhong, C., Rodríguez-Pascual, A.F., Ariza-López, F.J. (2016): "Análisis de los usuarios y usos de los MDE en España", *GeoFocus (Artículos)*, n°17, p. 133-160. ISSN: 1578-5157

por tanto accede con una determinada identificación. De tal forma que, bajo una misma identidad pueden acceder personas distintas (p.e. en una empresa) y una misma persona puede acceder con identificaciones distintas. En el conjunto de datos se pueden identificar 6 087 usuarios distintos para el periodo de análisis considerado. Aclarada esta situación, el formulario de encuesta posee un campo que ofrece la posibilidad de elegir entre 9 opciones que determinan perfiles de usuario de distinta índole, cuyo análisis es directo, tal y como se muestra en la Tabla 5.

Respecto al total de los productos, como se puede observar, el perfil más importante, con gran diferencia sobre el resto, es el de los usuarios particulares, donde se incluyen los profesionales libres. La segunda tipología más importante de usuario se corresponde con las universidades, donde se incluyen las actividades docentes, de investigación y de realización de trabajos fin de carrera y similares. Como es lógico, dada su mayor cantidad, las empresas de menor tamaño presentan un mayor porcentaje de descargas que las de mayor tamaño. Las administraciones (general, autonómica y local) contabilizan un total del 3.99 % de las descargas, lo que indica claramente que el uso mayoritario (> 95 %) queda fuera de este ámbito. En cuanto a cada uno de los tres productos descargables, esta tabla también nos indica que existe una gran estabilidad en el comportamiento. Los porcentajes son muy similares entre cada uno de los casos y respecto al total. En esta comparación se advierte que las tres administraciones incrementan su porcentaje conforme disminuye el tamaño de malla del producto (mayor resolución) y que, sin embargo, en otros perfiles (p.e. universidad) ocurre lo contrario.

Tabla 5. Descargas por perfil de usuario

Perfil Usuario	% Todos productos	% MDT200	% MDT25	% MDT05
Administración general del Estado	1.79	1.53	1.10	2.26
Administración Local	0.97	0.44	0.81	1.17
Comunidad Autónoma	1.23	0.87	0.81	1.56
Gran Empresa (>250 empleados)	0.82	0.29	0.77	0.97
Mediana Empresa (entre 51 y 250 empleados)	0.74	0.73	0.50	0.89
PYME (<50 empleados)	1.16	0.94	0.89	1.37
Microempresa (<10 empleados)	3.39	1.45	2.61	4.25
Universidad	17.71	20.26	18.54	16.70
Usuario particular	70.74	71.90	72.42	69.51
Otros	1.43	1.60	1.56	1.33
Total	100.00	100.00	100.00	100.00

Un aspecto de especial interés es la actividad que realizan los usuarios, es decir, el número de descargas realizadas en el año de análisis. En esa línea, la Figura 3 presenta el número de usuarios (eje vertical) frente al número de descargas. Como puede verse, el rango de descargas va desde 1 hasta 39, pero la inmensa mayoría de los usuarios sólo realiza una descarga.

Mesa-Mingorance, J.L., Chicaiza, E.G., Buenaño, X., Jianhong, C., Rodríguez-Pascual, A.F., Ariza-López, F.J. (2016): "Análisis de los usuarios y usos de los MDE en España", *GeoFocus (Artículos)*, nº17, p. 133-160. ISSN: 1578-5157

Figura 3. Número de descargas realizadas por el mismo usuario.

El número más habitual de descargas está entre 1 y 3. Los usuarios que realizan 1 o 2 descargas contabilizan el 80 % de los casos y los usuarios que realizan hasta 5 descargas, el 95 % de los casos.

A la vista de estos resultados podríamos pensar que un gran número de usuarios son usuarios esporádicos, puesto que realizan 1 o 2 descargas. Pero esto debe ser matizado por los hábitos de descarga y por el funcionamiento del sistema de descargas y encuesta. Así, el sistema de encuestas sólo recoge un registro por descarga cuando, sin embargo, la descarga puede incluir varios ficheros. Por tanto, hábitos de descarga distintos, por ejemplo, fichero a fichero, o todos de manera conjunta, generan número de registros distintos si bien el número de ficheros finalmente descargados puede ser el mismo.

Tabla 6. Descargas según persona física o empresa

Sexo	N	%
EMPRESA	64	1.0
PERSONA	5 897	96.7
NoSeSabe	136	2.2
Total	6 097	100.0

3.3. Uso del producto

Con los datos disponibles, la caracterización del uso del producto se puede realizar a partir del campo "área temática" del formulario de la encuesta. Este campo ofrece opciones múltiples que

Recibido: 07/03/2016

Aceptada versión definitiva: 30/05/2016

© Los autores

www.geofocus.org

Mesa-Mingorance, J.L., Chicaiza, E.G., Buenaño, X., Jianhong, C., Rodríguez-Pascual, A.F., Ariza-López, F.J. (2016): "Análisis de los usuarios y usos de los MDE en España", *GeoFocus (Artículos)*, nº17, p. 133-160. ISSN: 1578-5157

permiten marcar los sectores donde se va a realizar la utilización de los datos descargados. Además, para complementar esta información se ofrece al usuario un campo de comentarios y el campo de texto libre "descripción de uso", que fueron normalizados con los términos de la Tabla 2. A continuación se presentan los resultados derivados de ellos.

En primer lugar conviene recordar que un usuario puede marcar una o más áreas temáticas. La Tabla 7 presenta la frecuencia con que se eligen una o más opciones en relación con estas áreas. Casi el 73 % de los usuarios marcan una única área temática, lo que indica que están muy focalizados en sus aplicaciones, con independencia de cuál sea su tipología, y que la clasificación propuesta en el formulario de la encuesta es funcional. Casi el 95 % de los usuarios está incluido en el grupo formado por las tres primeras categorías de esta tabla, es decir, que marcan 3, 2 o sólo 1 área temática. Hay casos de usuarios que llegan a marcar un número muy elevado de las opciones propuestas, incluso en algunos casos marcan todas las opciones. Sin embargo, el porcentaje de casos es realmente bajo y puesto que la combinatoria de los 33 casos ofertados es elevadísima no se puede realizar un estudio en más detalle sobre la interacción entre diversas áreas.

Tabla 7. Número de áreas marcadas por los usuarios

Número de áreas marcadas	%
1,00	72.8
2,00	15.2
3,00	6.3
4,00	2.2
5,00	1.2
6,00	.7
7,00	.3
8,00	.3
9,00	.2
10,00	.2
11,00	.1
12,00	.1
13,00	.1
14,00	.0
15,00	.1
16,00	.0
18,00	.0
19,00	.0
21,00	.0
22,00	.0
25,00	.0
32,00	.0

Recibido: 07/03/2016

Aceptada versión definitiva: 30/05/2016

© Los autores

www.geofocus.org

Mesa-Mingorance, J.L., Chicaiza, E.G., Buenaño, X., Jianhong, C., Rodríguez-Pascual, A.F., Ariza-López, F.J. (2016):
"Análisis de los usuarios y usos de los MDE en España", *GeoFocus (Artículos)*, nº17, p. 133-160. ISSN: 1578-5157

33,00	.0
Total	100.0

El conteo de casos que recibe cada área temática sí es algo factible y su procesado permite generar la gráfica que se observa en la figura de la Tabla 8 y que representa las frecuencias totales. La Tabla 8 presenta la misma información en forma de porcentajes, porcentajes calculados respecto a los 12 493 registros, ya que debemos tener en cuenta que hay mayor número de sectores registrados, puesto que el usuario en una descarga puede marcar más de un sector de uso, no obstante se muestra el porcentaje respecto a las descargas totales. Estos datos indican claramente las áreas principales. Las aplicaciones a la cartografía y al medio ambiente son claramente dominantes, con más de 3 000 conteos (60.57 %). un segundo grupo de actividades son aquellas que están en el entorno de los 1 000 (42.62 %), que son: hidrología, educación y formación, ocio y tiempo libre, ordenación del territorio, forestal y biodiversidad. Se podría establecer un tercer grupo, en torno al valor 500 (54.69 %) y uno último por debajo de este valor.

Mesa-Mingorance, J.L., Chicaiza, E.G., Buenaño, X., Jianhong, C., Rodríguez-Pascual, A.F., Ariza-López, F.J. (2016):
“Análisis de los usuarios y usos de los MDE en España”, *GeoFocus (Artículos)*, nº17, p. 133-160. ISSN: 1578-5157

Tabla 8. Porcentaje de descarga en cada uno de los sectores. (Ordenado por porcentaje)

	Sectores	%
Educación y Formación	Cartografía	35.25
Forestal/Biodiversidad	Medioambiental	25.32
Ocio y Tiempo Libre	Educación y Formación	9.83
Hidrología	Forestal/Biodiversidad	9.80
Ordenación del Territorio y Urbanismo	Ocio y Tiempo Libre	7.76
Infraestructuras e Ingeniería Civil	Hidrología	7.75
Geología	Ordenación del Territorio y Urbanismo	7.48
Agricultura	Infraestructuras e Ingeniería Civil	6.15
Arqueología y Estudios Temporales	Geología	5.76
Investigación, Ciencia e Innovación	Agricultura	4.95
Turismo	Arqueología y Estudios Temporales	4.58
Otros	Investigación, Ciencia e Innovación	4.57
Catastro	Turismo	3.44
Suelos (Edafología)	Otros	3.11
Energía y Recursos Minerales	Catastro	2.72
Entomología/Biodiversidad	Suelos (Edafología)	2.35
Zonas de Riesgo/Protección Civil	Energía y Recursos Minerales	1.63
Geofísica	Fauna, Entomología/Biodiversidad	1.51
Artística y Cultural	Zonas de Riesgo/Protección Civil	1.47
Cambio Climático	Geofísica	1.41
Navegación y Localización	Artística y Cultural	1.38
Oceanografía/Costas	Cambio Climático	1.38
Vivienda	Navegación y Localización	1.29
Transporte y Logística	Oceanografía/Costas	1.11
	Vivienda	1.03
	Transporte y Logística	1.02
	Telecomunicaciones	0.93
	Defensa y Seguridad	0.87
	Demografía, Expansión Urbana	0.77
	Sociología	0.55
	Jurídico	0.32
	Sanidad	0.22
	Marketing (Mercadotecnia)	0.17

Existe un campo de abierto de comentarios a las áreas que ha sido utilizado 580 veces por los usuarios para incluir sus aportaciones. Estos comentarios han sido normalizados según los criterios indicados anteriormente, y el resultado es el que se presenta en la Tabla 9.

Tabla 9. Número de casos por comentarios normalizados del apartado Sectores de uso

Comentario normalizado	Casos
No Contesta	11 913
Agricultura	11
Análisis SIG	6
Arqueología	14
Arquitectura	2
Curiosidad	1
Defensa, Seguridad y Protección Civil	3
Docencia y Estudios	80

Recibido: 07/03/2016

Aceptada versión definitiva: 30/05/2016

© Los autores
www.geofocus.org

Mesa-Mingorance, J.L., Chicaiza, E.G., Buenaño, X., Jianhong, C., Rodríguez-Pascual, A.F., Ariza-López, F.J. (2016): "Análisis de los usuarios y usos de los MDE en España", *GeoFocus (Artículos)*, nº17, p. 133-160. ISSN: 1578-5157

Forestal	3
Geografía y Territorio	1
Geología	5
Geomática	13
Hidrología	10
Historia	1
Incendios	2
Ingeniería Civil	3
Investigación	15
Jurídico, Judicial, Administrativo o Licitación	1
Medioambiente	15
Navegación	1
Ocio, Deporte y Turismo	35
Problemas	3
Pruebas	4
Redes	4
Sin Especificar	293
TFG y TFM	18
Urbanismo o Territorio	2
Valora, Sugiere, Agradece	34

Como observamos en la Tabla 9, no contestan a este apartado 11 913, lo cual supone el 95.36 % de las encuestas. Por su parte, de los 580 casos en los que se contesta, tan sólo especifican alguna cuestión 287 casos, los casos restantes (293) no incluyen ninguna información de valor (p.e. códigos extraños, textos absurdos, etc.). Esto supone que se dispone aproximadamente del 50 % de casos con los que poder trabajar. En la Figura 4, podemos observar el porcentaje que supone cada respuesta normalizada respecto al total de las 287 respuestas especificadas.

Mesa-Mingorance, J.L., Chicaiza, E.G., Buenaño, X., Jianhong, C., Rodríguez-Pascual, A.F., Ariza-López, F.J. (2016): "Análisis de los usuarios y usos de los MDE en España", *GeoFocus (Artículos)*, n°17, p. 133-160. ISSN: 1578-5157

Figura 4. Porcentaje de respuestas normalizadas respecto al total de respuestas consideradas (287).

Como podemos apreciar, y aunque en menor número, estos comentarios normalizados siguen una distribución bastante similar a los comentarios de la descripción de uso normalizado, donde el mayor porcentaje se dedica a Docencia y Estudios, Ocio y Deportes, TFG y TFM, Investigación, etc. Llama la atención que la categoría de comentarios “Valora, Sugiere y Agradece” toma un alto valor porcentual de los comentarios. Esta categoría no aparece en el caso de la descripción de uso normalizado. Es importante destacar que en muchos casos se trata de agradecimientos y valoraciones positivas de los productos.

Finalmente, se va a realizar un análisis del campo *Descripción de Uso Normalizado* con los 43 primeros términos resumen que se establecieron anteriormente. La tabla 10 presenta los términos normalizados de descripción generalizada y la frecuencia y número de veces correspondiente a su uso. Aquí se debe indicar, que a diferencia de lo que ocurría con las áreas temáticas, se ha asignado un único término a cada registro. Se debe aclarar que si bien la encuesta obliga a rellenar este campo, no siempre las contestaciones son útiles. Así, hay algo más de un 14 % de casos en los que los valores introducidos por los usuarios no indican nada (p.e. letras, números, tonterías, improprios, etc.).

Tabla 10. Descargas de MDT respecto a la descripción de uso

Descripción de uso normalizado	Nº de registros	% respecto al total
Acústico	28	0.2
Agricultura	190	1.5

Recibido: 07/03/2016

Aceptada versión definitiva: 30/05/2016

© Los autores

www.geofocus.org

Mesa-Mingorance, J.L., Chicaiza, E.G., Buenaño, X., Jianhong, C., Rodríguez-Pascual, A.F., Ariza-López, F.J. (2016): "Análisis de los usuarios y usos de los MDE en España", *GeoFocus (Artículos)*, nº17, p. 133-160. ISSN: 1578-5157

Análisis SIG	171	1.3
ArqueologíaPatrimonio	271	2.2
Arquitectura	29	0.2
Aviación	12	0.1
Biología	3	0.0
Catastro Mercado	103	0.8
Comprobación	19	0.2
Curiosidad	81	0.6
DefensaSeguridadProtecciónCivil	81	0.6
DocenciaEstudio	3 507	28.1
EnergíasRenovables	51	0.4
Forestal	180	1.4
GeografíaTerritorio	83	0.7
Geología	188	1.5
Geomática	926	7.4
Hidrología	515	4.1
Historia	62	0.5
ImpactoAmbiental	37	0.3
Incendios	35	0.2
IngCivil	184	1.5
Investigación	723	5.8
JurídicoJudicialAdministrativo Licitación	25	0.2
Localización	27	0.2
Medioambiente	347	2.8
Minería	20	0.2
Modelización	31	0.2
Navegación	39	0.3
OcioDeporteTurismo	1 189	9.5
Paisaje	40	0.3
Planificación	9	0.1
Problemas	2	0.0
PruebasDatosSistemas	116	0.9
PublicacionesDivulgación	63	0.5
Redes	46	0.4
RiesgosNaturales	22	0.2
SinEspecificar	1 762	14.1
Telecomunicaciones	31	0.2
TFCTFGTFM	839	6.7
UrbanismoOTerritorio	194	1.5
Visualización	44	0.4
Visualización3D	120	1.0

Recibido: 07/03/2016

Aceptada versión definitiva: 30/05/2016

© Los autores

www.geofocus.org

Mesa-Mingorance, J.L., Chicaiza, E.G., Buenaño, X., Jianhong, C., Rodríguez-Pascual, A.F., Ariza-López, F.J. (2016): "Análisis de los usuarios y usos de los MDE en España", *GeoFocus (Artículos)*, n°17, p. 133-160. ISSN: 1578-5157

En la Tabla 10 se han marcado en negrita los términos que tienen una frecuencia superior al 5 %, aquí no se ha incluido el caso "sin especificar", que se corresponde con aquellas entradas que no son válidas o que no especifican nada. Como se observa, "DocenciaEstudio" es, con diferencia, el caso de uso más descrito. Además, dada su especificidad el caso de los trabajos fin de carrera, de máster y grado se ha considerado aparte. Por ello, sumando ambos casos estas actividades se alcanza un 34.8 %, es decir, más de 1/3 de los usos tienen que ver con la formación, lo cual asegura, de cara al futuro, una gran potencialidad de usuarios con cierto nivel técnico. Esto es perfectamente coherente con el perfil mayoritario de usuarios particulares y de universidades que se ha descrito anteriormente.

El término de "Ocio, Deporte y Turismo" alcanza casi un 10 % del total y se conforma como el segundo término más utilizado por los usuarios. Es interesante confirmar que las actividades lúdicas tienen tanta importancia entre los usuarios de los MDT. Este uso sugiere la existencia de posibilidades de negocio mediante la creación de productos de valor añadido que deberán estar muy centrados en deportes concretos.

Bajo el término "Geomática" se han incluido todo el hacer cartográfico (geodesia, cartografía, topografía, fotogrametría, etc.), por lo que es lógico que se alcance un porcentaje elevado dada la alta funcionalidad de los MDT en este campo. Consideramos que se trata de un uso predominantemente profesional, desarrollado por profesionales libres y pequeñas empresas.

La investigación es el último término que contabiliza más de un 5 % de los casos, en concreto un 5.8 %. Esta actividad está muy relacionada con la docencia y estudio, y con los trabajos fin de carrera, máster y grado, que muchas veces incluyen una componente investigadora. Sin embargo, se ha considerado de manera independiente por estar así indicado, de manera explícita, por numerosos usuarios y por las notables implicaciones científicas y tecnológicas que tiene la investigación. En esta línea, de cara al futuro cabe esperar la existencia de personal muy formado y de mayor cantidad de publicaciones científicas y divulgativas basadas en datos MDT.

Una vez descritos los términos con mayores porcentajes, queda un conjunto diverso de aplicaciones profesionales de gran interés e importancia. Aquí queremos destacar las aplicaciones hidrológicas, pues contabilizan un 4.1 % de los casos, pero también aparecen las aplicaciones al medio ambiente, patrimonio, agricultura, etc. Todas estas aplicaciones se podrían agrupar bajo un epígrafe de aplicaciones profesionales, donde también deberíamos incluir la Geomática, y se obtendría que la importancia de los usos profesionales alcanza algo más del 28 % de los casos.

A la vista de lo expuesto en los párrafos anteriores, podemos agrupar los usos en tres tipologías principales, por un lado los usos docentes e investigación, luego las aplicaciones profesionales de diversa índole y, finalmente, los usos de ocio y deporte. Existen otras respuestas no tratadas hasta ahora y que también consideramos relevante mencionar aquí, por diversos motivos. En concreto son las que se han normalizado con los siguientes términos: problemas, curiosidad y visualización 3D. Son escasas, pero hay contestaciones que indican problemas con los datos, mayoritariamente con las descargas y formatos. Esto nos hace pensar que el usuario no sabe dónde debiera hacer estos comentarios o que, sabiéndolo, le resulta más cómodo usar la encuesta. Con el

Recibido: 07/03/2016

Aceptada versión definitiva: 30/05/2016

© Los autores

www.geofocus.org

Mesa-Mingorance, J.L., Chicaiza, E.G., Buenaño, X., Jianhong, C., Rodríguez-Pascual, A.F., Ariza-López, F.J. (2016): "Análisis de los usuarios y usos de los MDE en España", *GeoFocus (Artículos)*, nº17, p. 133-160. ISSN: 1578-5157

término curiosidad se han normalizado las entradas que indican una actitud de exploración del producto por los usuarios, es decir, en principio no se pretende más que un primer contacto con este tipo de datos para conocerlos mejor. Finalmente, bajo el término de visualizaciones 3D se incluyen entradas relacionadas con el desarrollo de entornos virtuales e inmersivos y que destacamos por su presumible importancia de cara al futuro.

3.4. Valoración del producto

La valoración del producto por los usuarios es el último aspecto que se va a analizar en este trabajo. Para ello la encuesta ofrece los campos de utilidad y necesidad de actualización, que se valoran numéricamente, y dos campos de texto libre para que el usuario realice sus comentarios. En primer lugar se atenderá a la utilidad y luego a la necesidad de actualización.

La Figura 5 presenta el diagrama de frecuencias relativas de la valoración del nivel de utilidad asignado por los usuarios a cada producto. Aquí debe recordarse que los valores más altos significan mayor utilidad.

Figura 5. Porcentajes de los niveles de utilidad en cada tipo de MDT.

Como podemos observar en la Figura 5, donde se muestran las frecuencias relativas de descargas en cada tipo de MDT, los usuarios marcan el máximo nivel de utilidad de forma mayoritaria, para todos los productos. Si nos fijamos en el MDT05, más del doble de las descargas considera un valor 5 de utilidad, mientras que otro tercio de usuarios considera valores entre 4 y 3. Algo similar sucede en los MDT200 y MDT25.

De forma conjunta a todos los productos MDT la Tabla 11 muestra la distribución porcentual de las valoraciones. El porcentaje de casos que considera el valor 5 de utilidad, la máxima, alcanza el 58.9 %. Además, los porcentajes que reciben las valoraciones más bajas son prácticamente residuales, no alcanzando ni un 4 % de manera conjunta. En muchas encuestas que

Mesa-Mingorance, J.L., Chicaiza, E.G., Buenaño, X., Jianhong, C., Rodríguez-Pascual, A.F., Ariza-López, F.J. (2016): "Análisis de los usuarios y usos de los MDE en España", *GeoFocus (Artículos)*, nº17, p. 133-160. ISSN: 1578-5157

aplican la escala de Likert los que responden tienden a evitar valores extremos, pero consideramos que este no es el caso. Realmente, para todo el que conoce las posibilidades que ofrece un MDT, es lógico pensar que no considere valores bajos en esta pregunta. Entonces ¿por qué aparecen valoraciones de nivel 1 y 2? Creemos que más bien se corresponden con respuestas atípicas, es decir, que se valora otro aspecto o que simplemente son resultado de equivocaciones. Como conclusión evidente se puede afirmar que existe una valoración generalizada de "muy útil" para los MDT.

Tabla 11. Niveles de utilidad según los usuarios de los MDT

Nivel de utilidad	Porcentaje de descarga (%)
1	1.6
2	1.9
3	14.4
4	23.2
5	58.9

El otro aspecto que se incluye en este apartado es la valoración de la "Necesidad de actualización". Propiamente no se trata de una valoración directa del producto, pero sí se nos orienta sobre un rasgo que han de tener estos productos. La Figura 6 presenta el histograma de frecuencias relativas para cada producto y opción de valoración.

Figura 6. Porcentajes de los Niveles de necesidad de actualización en cada tipo de MDT.

Los datos que presenta la Figura 6 muestran un patrón distinto al caso anterior en cuanto al perfil de las valoraciones, pero semejante en cuanto a la valoración de los productos. En este caso, que este aspecto no es tan importante para todos los usuarios, pues en los valores bajos y muy bajos

Recibido: 07/03/2016

Aceptada versión definitiva: 30/05/2016

© Los autores

www.geofocus.org

Mesa-Mingorance, J.L., Chicaiza, E.G., Buenaño, X., Jianhong, C., Rodríguez-Pascual, A.F., Ariza-López, F.J. (2016): "Análisis de los usuarios y usos de los MDE en España", *GeoFocus (Artículos)*, n°17, p. 133-160. ISSN: 1578-5157

(1 y 2) hay una acumulación importante de frecuencias. El comportamiento en los tres tipos de MDT es muy parecido, a diferencia de los valores absolutos de descarga que son significativos entre tipos de MDT.

En relación con la valoración conjunta de los tres productos, la Tabla 12 presenta los datos agregados para esta valoración. En esta agregación se observa lo ya comentado para los productos individuales, los porcentajes están algo más repartidos que en la valoración de la utilidad. Así, el porcentaje de la valoración 5 es de 36,2 % y los valores más bajos (1 y 2), acumulan ya un 16,5%, cantidad nada despreciable. Además, aquí llama la atención que en la secuencia de valores relativa a los niveles 5, 4 y 3 no se sigue una tendencia de descenso progresivo, como en el caso de la evaluación de la utilidad, sino que la valoración del nivel 3 supera a la del nivel 4. Creemos que el pico de frecuencia que aparece en este valor intermedio se puede deber, en el caso de esta pregunta concreta, a cierto sesgo al valor central que pueden aparecer en las contestaciones a preguntas según la escala de Likerts (Friedman y Amoo, 1999).

Tabla 12. Niveles de necesidad de actualización según los usuarios de los MDT

Nivel de Necesidad de Actualización	%
1	8.1
2	8.4
3	26.3
4	21.0
5	36.2

4. Conclusiones

Se ha realizado una caracterización de los usuarios y usos de los MDT que se descargan del centro de descargas del CNIG. Para realizar este trabajo se ha utilizado un conjunto de 12 493 registros de encuestas voluntarias realizadas en línea en el momento de descarga y correspondientes al año 2014. Estas encuestas se refieren a los tres productos MDT disponibles (MDT05, MDT25 y MDT200). A la luz de los datos y de su análisis se puede concluir lo siguiente en relación con los siguientes aspectos:

- Lugar de descarga y producto. El lugar de descarga mayoritario es España, existen pocas descargas en otros países, pero con mayor frecuencia en nuestros vecinos (Francia y Portugal). Esta situación es lógica, si bien nos sorprende que el uso por nuestros vecinos sea tan limitado, cuando son numerosas las aplicaciones transfronterizas que pueden requerir este tipo de datos. En relación con el producto, el más descargado es el MDT05. Consideramos que esto es debido a la pretensión que tenemos todos los usuarios de acceder siempre a los datos de la mayor resolución posible, sea o no estrictamente necesario.
- Perfil de usuario. Los perfiles de usuario de mayor número de descargas son los usuarios particulares (71 %) y los relacionados con universidades (18 %). El 60 % de los usuarios sólo

Recibido: 07/03/2016

Aceptada versión definitiva: 30/05/2016

© Los autores

www.geofocus.org

Mesa-Mingorance, J.L., Chicaiza, E.G., Buenaño, X., Jianhong, C., Rodríguez-Pascual, A.F., Ariza-López, F.J. (2016): "Análisis de los usuarios y usos de los MDE en España", *GeoFocus (Artículos)*, nº17, p. 133-160. ISSN: 1578-5157

realizó una descarga a lo largo del periodo analizado. Uso de los productos. Los usos predominantes son, por orden de importancia: a) todos aquellos relacionados con el hacer docente e investigador, en todas sus modalidades (universidad, institutos, centros de investigación, academias, cursos diversos, etc.), que alcanzan el 41 % de los casos; b) los usos de carácter profesional (p.e. geomática, hidrología, ing. civil, etc.) que, en su conjunto computan un 28 %, y finalmente, c) los usos relacionados con el ocio, deporte y turismo que alcanza un 9.5 % de los casos.

- Valoración del producto. La valoración de la utilidad del producto es mayoritariamente (60 %) el valor más alto de la escala en la encuesta. La necesidad de actualización del producto no se valora tan alta como su utilidad. Esta situación parece del todo lógica, pues el primer interés de los usuarios de datos es que exista el dato. En la actualidad los MDT disponibles son relativamente recientes, por lo que consideramos que el aspecto temporal no es del todo crítico.

Entendemos que este trabajo no podría haber sido posible sin los datos procedentes de la encuesta que se contesta en línea, por ello consideramos que también se pueden extraer algunas conclusiones al respecto:

- Utilidad. Esta encuesta es breve y se focaliza en aspectos esenciales, y su utilidad es grande como se demuestra por el trabajo aquí expuesto. Sin embargo, si se revisan los sistemas de descarga de otras agencias cartográficas no está muy difundida la utilización de herramientas semejantes. Consideramos que, aunque los datos sean abiertos, el disponer de una buena encuesta opcional, y de un procesamiento adecuado posterior, son claves para entender mejor los usuarios y usos y proponer mejoras en los productos.
- Diseño. Consideramos que el diseño de la encuesta es básico y funcional. El sistema de registro de usuario podría incluir más información al objeto de permitir un análisis más profundo de los resultados. Los campos relativos a perfil y áreas temáticas se podrían dotar de una estructura jerárquica de opciones para que los usuarios pudieran elegir mejor su nivel de trabajo y permitir análisis más sencillos. Los campos que presentan mayor problemática son los que permiten la entrada de texto libre pero, sin embargo, aportan información muy relevante. Se debería disponer de un sistema de procesamiento de lenguaje natural para facilitar su explotación.

Respecto a este trabajo, consideramos que la mayor crítica que se le puede hacer proviene de presentar una perspectiva parcial, pues sólo se centra en los datos procedentes de la encuesta en línea de los usuarios de descarga que voluntariamente han elegido hacerla. Los usuarios que han declinado rellenarla y otros usuarios, como los de servicios web o los de pago, han quedado fuera de análisis. Sería interesante poder disponer de datos relativos a esos otros usuarios para disponer de un análisis más completo de los usuarios y usos de los MDT en España. Aún así, consideramos que este análisis es un primer paso en esta caracterización y que tiene valor tanto para el productor de datos, como para el sector, para realizar una primera caracterización de los usuarios y usos.

Finalmente, no queremos cerrar este trabajo sin hacer una reflexión sobre el interés de disponer de realimentación procedente de los usuarios. En el paradigma de las IDE, basado en la explotación de servicios estándar e interoperables, abiertos y libres, el usuario es esencialmente anónimo. Identificar al usuario, o exigirle rellenar una encuesta, rompe las condiciones de servicios abiertos (sólo se puede exigir reconocimiento y compartir igual), y supondría una barrera para la explotación automática de servicios desde aplicaciones informáticas. Todo ello dificulta

Recibido: 07/03/2016

Aceptada versión definitiva: 30/05/2016

© Los autores

www.geofocus.org

Mesa-Mingorance, J.L., Chicaiza, E.G., Buenaño, X., Jianhong, C., Rodríguez-Pascual, A.F., Ariza-López, F.J. (2016): "Análisis de los usuarios y usos de los MDE en España", *GeoFocus (Artículos)*, nº17, p. 133-160. ISSN: 1578-5157

enormemente, y posiblemente sesga, la obtención de resultados a partir de encuestas voluntarias, como las aquí utilizadas. En esta línea, son muy escasos los centros de descargas, vinculados a IDE nacionales o regionales, que han incluido encuestas voluntarias, y por ello consideramos que el CNIG ha tenido buen criterio al incluirla pues permiten conocer mejor el producto, los usuarios y sus apreciaciones, y así poder desarrollar acciones de mejora.

Agradecimientos

Al Centro Nacional de Información Geográfica por cedernos los datos para realizar este análisis. A los usuarios por rellenar las encuestas.

Referencias bibliográficas

- Decreto 141/2006, de 18 de julio, por el que se ordena la actividad cartográfica en la Comunidad Autónoma de Andalucía (BOJA nº 154, de 09/08/2006)
- Directiva 2007/2/CE por la que se establece una Infraestructura de Información Espacial en la Unión Europea (Inspire).
- EC (2008). The Socio-Economic Impact of the Spatial Data Infrastructure of Catalonia. European Commission Joint Research Centre
- ICECA (2011). NTCA 01-002. Modelo Aseguramiento Calidad de productos de IG.
- Ley 14/2010 de 5 de julio, sobre las infraestructuras y los servicios de información geográfica en España.
- Real Decreto 1545/2007 de 23 de noviembre, por el que se regula el Sistema Cartográfico Nacional.
- Friedman H., Amoo T. (1999). Rating the rating scales. In *Journal of Marketing Management*, Vol. 9:3, Winter 1999, 114-123.
- Piet Stroeven, Kai Li, Nghi L.B. Le, Huan He, Martijin Stroeven. (2015). Capabilities for property assessment on different levels of the micro-structure of DEM-simulated cementitious materials. *Construction and Building Materials*. Vol. 88, 105-117.
- P. Nila Rekha, R. Gangadharan, P. Ravichandran, P. Mahalakshmi, A. Panigrahi, S.M. Pillai (2015). Assessment of impact of shrimp farming on coastal groundwater using Geographical Information System based Analytical Hierarchy Process. *Aquaculture*. Vol 448, 491-506.
- Anders Juel, Geoffrey Brian Groom, Jens-Christian Svenning, Rasmus Ejrnaes. (2015). Spatial Application of Random Forest models for fine-scale coastal vegetation classification using object based analysis of aerial orthophoto and DEM data. *International Journal of Applied Earth Observation and Geoinformation*. Vol. 42, 106-114.
- Christopher Gomez, Yuichi Hayakawa, Hiroyuki Obanawa. (2015). A study of Japanese landscapes using structure from motion derived DSMs and DEMs based on historical aerial photographs: New opportunities for vegetation monitoring and diachronic geomorphology. *Geomorphology*. Vol 242, 11-20.
- Siddharth Saksena, Venkatesh Merwade. (2015). Incorporating the effect of DEM resolution and accuracy for improved flood inundation mapping. *Journal of Hydrology*. Vol. 530, 180-194.

A.4 Encuesta a usuarios especializados

Encuesta sobre calidad en Modelos Digitales de Elevaciones

Pertenecemos a un grupo de investigación de la Universidad de Jaén (España), trabajamos en temas relacionados con la Geomática: Cartografía, Sistemas de Información Geográfica. Modelos Digitales de Elevaciones, etc. Ya seas productor de MDE o usuario, nos gustaría que respondieras a una pequeña encuesta. Nuestro interés está relacionado con la calidad de MDE, la cual consideramos que, en algunos casos, es evaluada o expresada de manera poco adecuada y poco afín a los usuarios, y por ello buscamos una forma de evaluar y expresar la calidad de una manera más útil y asequible al usuario y más adecuada al uso. La encuesta se orienta a la recogida de información acerca de los casos de uso. Por favor completa el cuestionario para cada caso de uso de MDE en los cuales trabajes. La encuesta es muy simple; puedes seleccionar opciones y también escribir texto o subir ficheros. MUCHAS GRACIAS POR PARTICIPAR

1. Ocupación (Si ninguna de las opciones se adapta a tu caso, escribe en el cuadro final la correcta)

☐

Estudiante de doctorado

☐

Estudiante (otros)

☐

Investigador

☐

Profesional de la Geomática

☐

Profesional de otras áreas diferentes a la Geomática

☐

Profesor Universitario

☐

2. Institución u Organismo (Por favor escribe el nombre de la institución u organismo en el que trabajas)

3. Años de experiencia usando MDE

☐

Menos de un año

☐

Entre uno y cuatro años

☐

Más de cuatro años

4. ¿Para qué propósito usas los MDE? (Si ninguna de las opciones se adapta a tu uso, escríbela en el cuadro final)

☐

Arquitectura

☐

Catastro

☐

Ciencia de suelos y botánica

☐

Control geodésico de glaciares

☐

Estudios hídricos y redes de drenaje

☐

Estudios medioambientales

☐

Evaluación de desastres, protección civil, etc.

☐

Evaluación y estudio de cultivos agrícolas

☐

Fotogrametría, LiDAR, Láser escáner, UAV, Teledetección, etc.

- ☐ Geodesia
- ☐ Geología
- ☐ Ingeniería Civil
- ☐ Minería
- ☐ Monitorización de deformaciones, subsidencias, etc.
- ☐ Producción de Modelos Digitales de Elevación
- ☐ Sismicidad
- ☐ Telecomunicaciones, electrónica, redes energéticas y de comunicaciones
- ☐ Vulcanología

☐

5. ¿Qué parámetros derivados necesitarías a partir de MDE? (Selecciona una o más opciones de la lista, si ninguna es apropiada escribe la utilizada)

- ☐ Acumulación de flujo
- ☐ Análisis de costes
- ☐ Análisis de cuencas visuales
- ☐ Área de captación
- ☐ Concavidad y convexidad
- ☐ Curvatura
- ☐ Curvas de nivel
- ☐ Delineación de la red de drenaje
- ☐ Detección de canales
- ☐ Detección de divisorias
- ☐ Elevación
- ☐ Generación de parámetros del modelo
- ☐ Índice de humedad topográfica
- ☐ Índice Topográfico Compuesto
- ☐ Irregularidad
- ☐ Límites de cuencas
- ☐ Orientación
- ☐ Pendiente
- ☐

6. ¿Podrías suministrar un documento explicando el uso que haces de los MDE? (podría ser el artículo o trabajo más representativo del uso de MDE)

Puedes adjuntar a la presente encuesta los ficheros doc, docx, pdf, odt, zip

7. Por favor describe de una forma breve o esquemática el proceso o procesos que realizas en el uso de MDE. (Por favor, escribe tu respuesta)

8. ¿Qué tipo de modelo de datos usas?

☐

Curvas de nivel

☐

Perfiles

☐

Ráster

☐

TIN

☐

9. ¿Qué fuentes, escala/resolución y exactitud de datos has usado? (Por favor, escribe tu respuesta)

10. ¿Cuál es la principal exigencia de cálculo en tu caso de uso? (Selecciona una o más opciones de la lista, si ninguna se adecúa, escribe tu propuesta)

☐

Cálculo fiable de la pendiente

☐

Cálculo fiable de la orientación

☐

Cálculo fiable del volumen

☐

Cálculo fiable de la Z

☐

Cálculo fiable de los cursos de agua

☐

Cálculo fiable de insolación

☐

Cálculo fiable de cuenca visual

☐

Cálculo fiable de cuenca hidrográfica

☐

11. ¿Cómo evalúas la calidad de los resultados en tus casos de uso de MDE? (Si ninguna de las opciones de la lista se adecúa, escribe la adecuada)

☐

Ninguna

☐

Usas una evaluación de base estadística

☐

Sometes los resultados a pruebas de uso

☐

Realizas un juicio subjetivo

☐

12. ¿Usas uno o varios índices para expresar la calidad de los resultados? (Si ninguna de las opciones de la lista se adecúa, escribe la adecuada)

☐

Ninguna

☐

Sólo usas una medida de exactitud posicional

☐

Consideras medidas relacionadas con los resultados (Por ejemplo medida de desplazamiento de los cursos de agua derivados del MDE)

☐

Las dos anteriores, de manera conjunta pero independientemente

- ☐ Construyes un índice compuesto con ambas

☐

13. ¿Estás satisfecho con el índice(s) de calidad que utilizas actualmente?

- ☐ Nada Satisfecho
- ☐ Poco Satisfecho
- ☐ Satisfecho
- ☐ Bastante Satisfecho

14. ¿Mejorarías algo a la hora de expresar la calidad de los resultados en relación a tu caso de uso de MDE?. Por ejemplo, si fueras a adquirir un MDE para tu caso de uso, sobre qué aspecto de su calidad te gustaría que te informaran, de qué manera, con qué tipo de medida cuantitativa. (Por favor, escribe tu respuesta)

☐

15. ¿Si todavía no has usado ningún índice para expresar la calidad de tus datos MDE en tu caso de uso, pero esta encuesta te sugiere pensar en alguno, nos podrías sugerir un índice? (Por favor, escribe tu respuesta)

☐

16. Por favor explica la forma en la que afecta la falta de calidad del MDE en tu caso de uso. Por ejemplo, tu caso de uso o aplicación es muy sensible al valor de la pendiente, o deseas determinar los cauces de agua y te salen desplazados respecto a la realidad, o los datos generan que esos cursos no tengan continuidad espacial, etc.(Por favor, escribe tu respuesta)

☐

17. Podrías valorar la influencia de los requerimientos de calidad de los MDE en los resultados de tus procesos? (Por favor, selecciona una opción de la lista)

- ☐ Alto
- ☐ Medio
- ☐ Bajo
- ☐ Ninguno

18. Para tu caso de uso, ¿podrías cuantificar económicamente las consecuencias que un MDE de mala calidad tiene sobre los resultados? (Por favor, escribe tu respuesta, si tiene documentación y quieres compartirla, adjúntala también ésta puede ser cualquier material que desees compartir con nosotros sobre esta temática (explicación del caso de uso, especificaciones, procesos de trabajo, procesos de evaluación de la calidad, índices de calidad, valoración económica, etc.)

A.5 Artículo: DEMs: An approach to users and uses from the quality perspective

DEMs: AN APPROACH TO USERS AND USES FROM THE QUALITY PERSPECTIVE

Ariza-López, F.J.; Chicaiza-Mora, E.G.; Mesa-Mingorance, J.L., Cai, Jianhong

ABSTRACT

The importance of the MDEs is great in geosciences but it is lacking a general view of users and uses in order to approach quality to users. For this reason, the aim of this paper is three fold: obtain a better knowledge about the users, determine the main application domains of DEMs, and identificate main use cases. For this propose we use data from two web questionnaires (MR1 and MR2), a search of ScienceDirect data (MR3) base and a Google search (MR4). The data coming from the MR1 resource have offered us a great amount of cases in order to characterize the profile of the users in Spain. The MR2 resource is an *ad hoc* designed survey and has allow us, between others, to identify those calculations that are more usual, that subjective evaluation of quality is of great importance for users, that there is a high percentage of users who do not use any quality index, and that majority of users do not know how to evaluate the influence of poor quality on their work. By MR3 it has been possible to generate a graph that shows the relationships between relevant terms that are linked to publications on DEMs and a semantic analysis of a set of 950 abstracts. By MR4 it can be concluded that the formalization of applications as use cases is not usual. This paper identifies some research lines in order to offer users a better understanding about quality of DEMs.

Key words: DEM, USER, USES, QUALITY

INTRODUCTION

A Digital Elevation Model (DEM) is generic term for digital topographic and/or bathymetric data and normally implies elevations of the terrain (bare earth) void of vegetation and manmade features (Maune et al. 2007). DEMs describe the morphological characteristics of the landforms of a terrain (topography) such as: ups, downs, hills, ridges, cliffs, valleys, rivers, etc., and a lot of parameters can be derived (e.g. slope, aspect, curvature, etc.). DEM data can be generated by means or many diverse technologies (Manune, 2007) (classical surveying and photogrammetry, Lidar and Radar systems, etc.) and methods (stereoscopy, stereo correlation, range measurement, etc.), and the topographic surface modeling can be achieved by means of very different options such contour lines, triangular irregular networks, gridded surfaces, parametric surfaces, etc.

Information is of void value if it is not used and not properly used. That means the importance of a precise identification of users and the understanding of their needs. Data users are a key element in the case of spatial data applications because of the use of spatial data as an input to decision making processes. Nevertheless, to the best of our knowledge, only Wechsler (2003) identifies some profiles of users by means of a survey.

Because quality means fitness for use, for us it is very important to identify the uses of DEMs or, at least, main uses (application domains). In a pioneering way, Weibel and Heller (1991) identified some application domains, and after them other many authors (Sulebak, 2000; Maune, 2007; etc.). Also, Dewberry (2012) identifies 27 business uses in order to evaluate the benefit cost analysis of the National Enhanced Elevation Program of the USA. But only Wechsler (2003), and more recently Tarquini and Nannipieri (2017), provides quantitative data (percentages) in order to understand the actual share of each use.

In general, spatial data, and DEMs in particular, are very important data for earth sciences and engineering applications and, within the spatial data production domain, they are the most critical and cumbersome element of any large-scale digital mapping projects (Acharya and Fagerman, 2000). The spread of the use of DEMs is directly linked with the development of the Geographical Information Systems and the spread of their use. Also, the existence of global digital elevation model (GDEMs) products has promoted a large amount of global-scale applications. So, nowadays there are many DEMs dependent applications in many fields and at very different scales (micro, meso, macro) and many sources of DEMs data. In this context, the choice of a suitable elevation data set becomes important (e.g. Wise, 2007). For this reason, quality is a major concern (Ariza-López, 2002). Quality of spatial data is defined by the International Standard ISO 19157 (ISO 2013) by several quality components (positional accuracy, completeness, temporal quality, etc.). In this way, the causes, consequences and assessment of uncertainty in DEMs have received a lot of attention in a general context (Li 1994; Carlisle, 2005; Fisher and Tate, 2006) and in specific application contexts such topography (Oksanen and Sarjakoski 2004; Raaflaub and Collins, 2006), view shed (Ruiz, 1997; Riggs and Dean, 2007), hydrology (Wechsler, 2007; Li and Wong 2010, Charrier and Li, 2012), soil mapping (Cavazzi, xxx); glaciology (Watson et al. XXX), precision farming (Yao and Clark 2000), solar irradiation (Tovar-Pescador, 2006), and so on.

The current way of treating quality is not adequate; it is far from users understanding and from fitness for use (Boin and Hunter, 2009), which means that a better knowledge of domains of applications and use cases is needed. For instance, Devillers et al. (2005) argue that the majority of parameters being used to express quality are focused on characteristics relating to data production (the internal quality), and that more information needs to be included to reach the purpose of 'fitness for use'. In this way, Boin and Hunter (2006, 2009) ask themselves: if consumers of spatial data really understand data quality information?, and what communicates quality to the spatial data consumer?. In relation to the first question, they propose: the use of terminology of the data consumer instead of that of producers which is overly technical and industry-specific, and the need of focusing the quality information on describing product suitability and reliability instead of the production method. In relation to the second question they concluded that it is strongly affected by the context.

Specifically, in the field of DEMs there are some works that address the issue of a better understanding of quality by users. For instance, Wechsler (2003) explored the perceptions of DEM uncertainty by users, Podobnikar (2009) proposed some tools for visual quality assessment, and Darnell et al. (2008) presented a tool for facilitating to DEM's users the access and application of uncertainty analysis. Recently, Mesa-Mingorance et al. (2016) has analyzed of users and uses of some official DEMs products in Spain, this paper has identified the predominant user profiles and uses, and presents an assessment of the utility of the products by users. Tarquini and Nannipieri (2017) summarize 4 years of free dissemination and use of a 10x10 m DEM covering entirely Italy, they present a complete vision of the areas of use and some remarks about the appropriateness of this product to specific applications. On the other hand, producers desire a better understanding of their products by users and have created user's guides for DEMs products (e.g. OS 2013, ICSM 2008), and electronic metadata. But this is not fitness for use. Fitness for use is mentioned in several works (e.g. Lemmens 1999, Fisher and Tate 2006), and it is clear that it can only be assessed relative to an intended use (Devillers et al. 2002), but exist scarce literature in relation to practical cases. In this way, Agumya and Hunter (1999) developed an example concerning the selection of a DEM for flood-extend estimation by means of a risk-based strategy where the exposition to risks due to error in the data is a key element. So that, fitness for use is linked to use cases, and use cases are linked to application domains; thus if we want to advance the way of the "fitness for use", we must first determine application domains and use cases and, in a previous

step, try to gain a better understanding of users. But this approach has not been developed until now. For this reason the aim of this paper is three fold: obtain a better knowledge about the users, determine the main application domains of DEMs, and identificate main use cases.

This paper is organized as follows, first they are presented the four different resources (MR1, MR2, MR3 and MR4) we have used for the recompilation of relevant data about users and uses; second all the results corresponding to each source, and specific methodological issues, are presented and discussed; and finally main conclusions are established.

MATERIAL, METHODS AND RESOURCES

In this research we have three major aims (users, domains of use and use cases), and we have used several methods, and available resources, in order to get a better definition of all of them. Basically we have used two surveys to people in order to detect user profiles, domains of application and specific requirements about quality of DEMs. Also, we have used queries to a scientific data base in order to get a better insight about domains, relations of domains and a sematic approach to domains. Finally, we have used internet queries, through Google, to get formal examples of use cases about DEMs. Table 1 summarizes this approach.

Table 1 – General view of the methods/resources and scopes of analysis

Method / Resource	Targets				
	Users	Domains			Use cases
		Identification	Relations	Semantic	
MR1. Existing questionnaires	X	X			
MR2. Ad hoc survey	X	X			
MR3. ScienceDirect ^{TM} search		X	X	X	
MR4. Google ^{TM} search					X

In order to obtain information about users and uses of DEMs, our first choice was to look for existing web questionnaires linked to existing Spatial Data Infrastructures at national level, and ask about registered data. We located and analyzed the SDI's websites of many countries (Argentina, Bolivia, Brazil, Canada, Chile, China, Colombia, Costa Rica, Cuba, Ecuador, France, Germany, Guatemala, Honduras, Italy, México, Nicaragua, Panamá, Paraguay, Perú, República Dominicana, Spain, UK, Uruguay, USA, Venezuela), and we find only two countries with questionnaires to users. In the USA (USGS) registration of users is required and a "demographic" profile is requested: This profile is for all downloads, there is no specific questionnaire for each DEM download, so this is useless for us. The other country is Spain, here there is a questionnaire linked to each download.

So, the first resource we were able to use are data coming from the download center (<http://centrodedescargas.cnig.es/CentroDescargas/>) of the National Geographic Institute (IGN) of Spain (<http://www.ign.es>). For each one free download the users must fulfil a questionnaire (Table 2). Of course, the data and the derived analysis are, strictly, only valid for the Spanish case but we think that this can be a starting point in our way. This source will show results for a national level in a developed country, and we consider that these results can be, in some way, generalized to other countries of similar characteristics. A strong point of this source is the large number of available survey responses. This questionnaire was designed by the IGN several years ago and we had no option to modify it. In the context of this research, relevant questions for us are those related to the user profile, description of the intended use, application domains, comments and observations to application domains, and the scope. Some of these questions can be answered by users by checking only one or more options, in both cases the analysis is straight forward, but other questions are answered by free text. In this study we only use the results of the first kind of questions, specifically

questions numbered as 1, 2, 4, 6, 7 and 9 in Table 2.

The IGN has supplied us 12493 records coming from downloads of DEMs that took place during the whole year 2015. 99.99% of cases are downloads that took place in Spain and there are 6087 unique users. These downloads are related with three products: MDE05, MDE25 and MDE200 (the grid resolution is 5×5 m, 25×25 m and 200×200 m, respectively).

Table 2 – Web questionnaire of the IGN (Spain)

1. User profile Select one: [particular use; micro enterprises (<10 employees); small enterprises (<50 employees); medium enterprises (between 51 and 250 employees); large company (> 250 employees); Central Government; Regional Government; University; Local Government; Others]   2. Utility Select one: [1 = little use, 2, 3, 4, 5 = very useful]   3. Observations to utility Free text   4. Need to update Select one: [1 = little need, 2, 3, 4, 5 = much need]   5. Comments on the need to update Free text   6. Province Select one: [Álava, Albacete, Alicante...Zamora, Zaragoza]   7. Country Select one: [Afghanistan, Albania, Algeria...Zambia, Zimbabwe]   8. Description of the intended use Free text   9. Application domain Free selection: [Forest/Biodiversity, Climate change, Oceanography/Coasts, Cadastre, Infrastructures and Civil Engineering, Geophysics, Defense and National Security, Legal, Telecommunications, Living place, Mapping, Planning and Urban Planning, Research/Science and Innovation, Geology, Leisure and free time, Energy and mineral resources, Risks/Civil Protection, Archaeology and temporary studios, Sociology, Tourism, Education and formation, Agriculture and farming, Demography and urban sprawl, Hydrology, Soils (Soil Science), Transport and logistics, Navigation and location, Fauna/Entomology and biodiversity, Marketing, Health, Arts and culture, Others]   10. Comments and observations to application domains Free text   11. Scope Select one: [International, European, National, Regional, Provincial, Other (local, regional, watershed, etc.)]

The identification of users' profiles and fields of use (domains) can be approached using existing sources (e.g. the referenced papers of Mesa-Mingorance et al. (2016), Tarquini and Nannipieri (2017), Wechsler (2003), etc.); but in order to obtain some information, in relation to specific quality aspects of DEMs, we design the questionnaire presented in Table 3 and a worldwide survey. This questionnaire (MR2) includes 18 questions that can be group as follows:

- Information about the user. In this section (questions 1-3) we ask them about their occupation and experience because they are important issues in order to define profiles of users.
- Information about the use of DEMs. In this section (questions 4-9) we ask them in order to get insight into the use case and kind of applications, processes and data that they are developing and using.
- Information about relevant quality aspects. In this section (questions 10-18) we ask them in relation to quality aspects (exigencies, use of quality indexes, degree of satisfaction with those quality indexes, etc.).

To get an international view we used a web form platform. We selected Cognito Forms™ which is a free online-form builder, which allows the use of several kinds of answers and uploading files. In order to get a widespread contribution of our survey, we collected more than 2500 email addresses from diverse sources: authors of scientific papers, proceedings, documents and web pages; and fellows of different professional associations (ASPRS, FIG, ICA, etc.). The field work extended from November 2015 to December 2016, and we finally obtained 102 answers from 27 countries. Obtained answers' origin was 60.8% from Europa, 25.5% from America and 13,7% from Asia.

Table 3 – Web questionnaire #2

1. Occupation (Please select an option from the list, if any is appropriate, write your proposal) Select one: [Emeritus Professor/Readers; PhD Students; Geomatics Professional; Professional in other different fields Geomatics; Professional production of cartographic products; Research staff; University professor; Visiting lecturers]   2. Institution (Please write the name of your institution) Free text.   3. Years of experience using DEM Select one: [Less than 1 year; Between 1 and 4 years; More than 4 years]   4. For what purposes do you use the DEMs (Please select an option from the list, if any is appropriate, write your proposal) Select one: [Architecture; Cadastre; Civil Engineering; Control of glaciers Geodesy; Deformation monitoring, subsidence, etc.; Disaster assessment, civil protection, etc.; Enviromental studies; Evaluation and study of agricultural crops; Geodesy; Geology; Mining; Photogrammetry, LiDAR, Laser Scanner, UAV, Remote Sensing, etc.; Production of Digital Elevation Models; Seismicity; Soil Science and Botany; Telecommunications, electronics, power grids and communications; Volcanology; Water studies and river network]   5. Please, select one or more options in relation to the more common calculations and the derivatives you compute from DEMs Select one: [Aspect; Catchment area; Channel detection; Contours; Concavity and convexity; Cost analysis; Curvature; Drainage delineation; Elevation; Flow accumulation; Model parameter generation; Ridge detection; Roughness; Slope; Topographic wetness index (TWI)/ Compound Topographic Index (CTI); View-shed analysis; Watershed boundaries]   6. Could you supply us a document explaining the use you are doing? (It could be the most representative paper or work using of DEM, but this not required to complete the survey)   7. Please write the description, in a short or diagrammatic form, of the process or processes that take place in your use case of DEM   8. What model did you used? Select one: [Raster; TIN; Contours; Profiles]   9. What sources, scale/resolution and accuracy of data did you used respectively? (Please write your answer)   10. What is the main exigency of your application to calculate? (Please check one or more options from the list, if any is appropriate, write your proposal) Reliable calculation of slope; Reliable calculation of orientation; Reliable estimate of volume Reliable calculation of Z; Reliable estimates of watercourses; Reliable calculation isolation, watershed, view shed, etc.   11. How do you express accuracy in your use cases? (Please select an option from the list, if any is appropriate write your proposal) You use a statistical assessment base; You use the results in tests of use; You make a subjective judgment   12. Do you use one or several indices to express quality? (Please select an option from the list, if any is appropriate write your proposal) You only use measures of positional accuracy (for example RMSEz); You consider measures related to the results (for instance, you consider a measure of the displacement of water courses derivates of the DEM); You use the two previous measures (position + other), but in an independent way; You build a composite index using both kind of measures   13. Are you satisfied with the actual accuracy indices being used? (Please select an option from the list) Not at all satisfied; Dissatisfied; Satisfied; Quite satisfied   14. Is there something to improve when expressing quality in relation to the use case of your

interest? For instance, if you were to purchase a DEM for your use case, on what aspect of the quality you would like to inform you, how and what kind of quantitative measure. (Please write your proposal)

15. If you have not used any index to express the quality of your DEM data on your use case, but this survey suggests you think of any, what would you suggest to us? (Please, write your proposal)

16. Could you explain the influence of bad quality in the results of your works with DEMs. (For instance, if your use case is very sensitive to the value of the incline, or you want to determine watercourses and get displacements from the reality or generated data of DEM these courses do not have spatial continuity). (Please, write your answer)

17. Could you score the influence the quality requirements of DEM in results of your work with DEMs? (Please select an option from the list)

High; Medium; Low; None

18. Could you monetize the consequences of using DEMs of bad quality in the results of your works? (Please write your answer)

Another important source has been ScienceDirect™. ScienceDirect is a data base which hosts over 2500 journals and more than 33000 books (over 13 million peer-reviewed publications) from Elsevier imprints and partners. We selected this source to help us to determine application domains and use cases because: a) offers a very large database of contents that we think can be considered enough representative, b) offers a system enabling expert searches (e.g. you can use connectors and proximity operators, wildcard characters, search specific fields of the database or search for phrases), and c) offers as result a list of topics that can be considered related to use cases. We have queried this database using journals and books from 2006 until 2016 by means of several expert searches on the contents of the following fields: title, abstract, and author or publisher's keywords. In order to use these fields we have to use the “*tak()*” query of the system. All searches have two main parts, the first one in order to disambiguate the term “digital elevation model”, because the same words can be used in other scientific fields (e.g. in particle physic, heat transfer and mass transfer), and a second one in order to establish a specific field of application (e.g. hydrology, remote sensing, agriculture, etc.). The terms used in the second part of the query come from the list of application fields obtained from MR1 and MR2. These terms are considered here as “themes”. Thus, the structure of a query is: *tak*((disambiguation part of the query) AND (specific field for the search)). Following this structure, here are some examples of the searches we have developed:

tak ((DEM or {elevation model} or {digital elevation}) AND “hydrology”)

tak ((DEM or {elevation model} or {digital elevation}) AND “drainage”)

tak ((DEM or {elevation model} or {digital elevation}) AND “slope”)

tak ((DEM or {elevation model} or {digital elevation}) AND “landslide”)

tak ((DEM or {digital elevation model}) AND “remote sensing”)

Figure 1.a presents an example of a query and Figure 1.b the results of the same. The left margin of the page of the results includes a list of topics and a value indicating the frequency of each one. As a result of such searches we get the amount of documents that fulfill the query (theme query) and a list of topics with the frequency of appearing. With this information we will be able to derive the frequency of each topic and a force-directed graph to represent relations between themes and topics. It is important to note that these themes and topics are more or less the same thing; the only difference is that “themes” are used for the queries and the results are obtained as “topics”.

ScienceDirect

Search all fields
Author name
Journal or book title
Volume
Is

All
Journals
Books
Reference Works
Advanced search | Expert search

? Search tips

Search for (Enter terms using Boolean connectors e.g. "heart attack" AND stress)

tak ((DEM or {elevation model} or {digital elevation}) AND "drainage")

Refine your search

☒ Journals
☐ Open Access articles only
☒ Books

Economics, Econometrics and Finance
Energy
Engineering
Environmental Science

Hold down the Ctrl key (or Apple Key) to select multiple entries.

☐ All Years
☒ 2006 to: 2016

Search

a

Search all fields
Author name
Journal or book title
Volume
Issue
Page
Advanced search

Search results: 298 results found for pub-date > 2005 and pub-date < 2017 and tak ((DEM or (elevation model) or (digital elevation)) AND "drainage") [All Sources(Agricultural and Biological Sciences,Computer Science,Decision Sciences,Earth and Planetary Sciences,Economics, Econometrics and Finance,Energy,Engineering,Environmental Science,Social Sciences)].

Save search alert
RSS

Refine filters

Year
☐ 2016 (41)
☐ 2015 (36)
☐ 2014 (38)
☐ 2013 (24)
☐ 2012 (30)
View more >>

Publication title
☐ Geomorphology (81)
☐ Computers & Geosciences (41)
☐ Journal of Hydrology (17)
☐ CATENA (12)
☐ Remote Sensing of Environment (11)
View more >>

Topic
☐ drainage network (13)
☐ drainage (12)
☐ soil (12)
☐ basin (9)
☐ cell (9)
☐ channel (8)
☐ lake (8)
☐ drainage area (7)
☐ fault (7)
☐ river (7)
☐ stream (7)
☐ study area (7)
☐ drainage basin (6)
☐ fault zone (6)
☐ flow (6)
☐ river basin (6)
☐ water (6)
☐ catch (5)
☐ drainage density (5)
☐ flow direction (5)
View less >>

Purchase
Export

Relevance
All access types

☐ Parallelizing drainage network algorithm using free software: Octave as a solution Original Research Article
Mathematics and Computers in Simulation, In Press, Corrected Proof, Available online 29 September 2016
J.F. Reinoso-Gordo, M.J. Ibáñez, R. Romero-Zalitz
Abstract Purchase PDF - \$35.95

☐ Derivation and validation of supraglacial lake volumes on the Greenland Ice Sheet from high-resolution satellite imagery Original Research Article
Remote Sensing of Environment, Volume 183, 15 September 2016, Pages 294-303
Mahsa S. Moussavi, Waleed Abdalati, Allen Pope, Ted Scambos, Marco Tedesco, Michael MacFerrin, Shane Grigsby
Abstract Research highlights Purchase PDF - \$35.95

☐ Structural pattern of the Saiss basin and Tabular Middle Atlas in northern Morocco: Hydrological implications Original Research Article
Journal of African Earth Sciences, Volume 119, July 2016, Pages 150-159
O. Dauteuil, F. Moreau, K. Qarqori
Abstract Research highlights Purchase PDF - \$35.95

☐ Improving the identification of hydrologically sensitive areas using LiDAR DEMs for the delineation and mitigation of critical source areas of diffuse pollution Original Research Article
Science of The Total Environment, Volume 556, 15 June 2016, Pages 276-290
I.A. Thomas, P. Jordan, P.-E. Mellander, O. Fenton, O. Shine, D. Ó hUallacháin, R. Creamer, N.T. McDonald, P. Dunlop, P.N.C. Murphy
Abstract Graphical abstract Research highlights PDF (3357 K) Supplementary content

☐ Landscape response to recent tectonic deformation in the SW Pannonian Basin: Evidence from DEM-based morphometric analysis of the Bilogora Mt. area, NE Croatia Original Research Article
Geomorphology, Volume 263, 15 June 2016, Pages 132-155
Bojan Matoš, José Vicente Pérez-Peña, Bruno Tomljenović
Abstract Research highlights Purchase PDF - \$35.95

☐ Assessment of drainage network extractions in a low-relief area of the Cuvelai Basin (Namibia) from multiple sources: LiDAR, topographic maps, and digital aerial orthophotographs Original Research Article
Geomorphology, Volume 260, 1 May 2016, Pages 32-50
F.C. Persend, C. Gomez
Abstract Research highlights Purchase PDF - \$35.95 Supplementary content

☐ An efficient variant of the Priority-Flood algorithm for filling depressions in raster digital elevation models Original Research Article
Computers & Geosciences, Volume 90, Part A, May 2016, Pages 87-96
Gulyun Zhou, Zhongxuan Sun, Suhua Fu
Abstract Graphical abstract Research highlights Purchase PDF - \$19.95

b)

Figure 1 – Example of a query (a) and a query result (b) in ScienceDirect

All the abstracts of the resulting papers, one by one, were revised in order to eliminate improper cases and after this, all abstracts have been clustered by means of a text clustering tool in order to establish groups sharing same semantic.

Finally, the last resource we have used is to query internet by means of Google in order to get use case examples presented in a formal way, here “formal” way, here "formal" means the existence of a certain structure in the contents. Here we used impersonal searches (www.impersonal.me) to be not conditioned by our user profile. In this case we used similar queries as presented above but including the term "use case". The result of these queries is a set of documents that we have analyzed manually in order to localize the proper documents.

RESULTS AND DISCUSSION

Here we present, in the same order as in the previous section, the results obtained from each of the four methods/resources (MR1 up to MR4) described in the previous section. When needed, remarks about tools and methods will be incorporated.

Approximation to users and identification of domains by means of MR1

The questionnaire of the MR1 (see Table 1) has 11 questions, but we centered our analysis on the results of those questions (Q1, Q2, Q4 and Q9) that are of interest from the perspective of this paper.

In relation to the user's profiles (Table 4) three groups¹ can be identified. The most frequent profile is that of particular users (70.74%), where freelances are included. This is the first group and is very far from the others cases. The second case more frequent case corresponds to a university profile (17.71%), where researchers, teachers and students are included. Each one of the other profiles accounts less than 5% of frequency and all of them are included in the third group. Not surprisingly, given their greater number, that the smaller companies (micro-size enterprises) have a higher percentage of downloads that larger ones. All the authorities together (National, Regional and Local) sum 3.99% of the discharges, which clearly indicates that majority use (> 95%) is in the private sector.

Table 4 – MR1: User profiles in Spain

User profile	%
Particular users	70.74
Universities	17.71
Micro-size enterprises (<10 employees)	3.39
Administration of the National Government	1.79
OTHERS	1.43
Administration of Regional Governments	1.23
Small-size enterprises (<50 employees)	1.16
Administration of Local Governments	0.97
Large-size enterprises (>250 employees)	0.82
Medium-size enterprises (between 51 and 250 employees)	0.74
% of total cases, N = 6087	

The MR1 provides a list of application domains, but we must remember that enables multiple selections, it means that a user can mark one or more application domains. 72.8% of users marked only one application domain, meaning that they are very focused on their application domains, irrespective of their type, and that the survey's proposed classification is functional.

¹ The groups presented in tables by means of colors are derived from a cluster analysis of cases. The cluster is based on the Euclidean distance between classes using the SPSS hierarchical cluster analysis.

Nearly 95% of users marked 3, 2 or only 1 application domain.

The Table 5 shows the percentage of cases recorded for each application domain (relative to the total of responses). Because this question offers more than 30 different options for the answer, some of them very general but other very specifics, and because allows multiple answers, can consider that users can express their application domain in a good manner. General terms as Cartography and Environment are clearly the dominant options, accounting more than 38.35 % together. A second group of more specific activities are those which percentage is ranging in the interval [2.86; 6.23], and accounting 46.43% all together. Here Education & training and Forest & biodiversity are the two cases with greater values. It is interesting to notice that Leisure and recreational applications is an important application domain and that this domain can be considered out of the initial professional proposes of use of a DEM. This is a very good index of the spread of use of DEMs in society. Of course, Hydrology, which can be considered as a classical and dominant application domain of DEMs, is included in this group. The last group includes 21 different fields and sums 18.17% of responses. It is a large tail of applications domains identifying some interesting application domains (e.g. Legal, Health, marketing, etc.). The case “Others” (1.97%) takes into account other minority options (e.g. 3D printing, virtual reality and games, etc.). If the results shown by Table 5 are compared with other similar analysis (e.g. see Tarquini and Nannipieri 2017) it is possible to observe a great level of coincidence in the domains and order of importance of such domains.

Table 5 – MR1: Application domains in Spain

Application domains	%
Cartography	22.32
Environment	16.03
Education and training	6.23
Forest/Biodiversity	6.20
Leisure and recreation	4.91
Hydrology	4.91
Land use and planning	4.74
Infrastructures and Civil Engineering	3.89
Geology	3.65
Agriculture	3.14
Archaeology	2.90
Research/Science and Innovation	2.89
Tourism	2.18
OTHERS	1.97
Cadaster	1.72
Soils	1.49
Energy and mineral resources	1.03
Fauna/Entomology and biodiversity	0.96
Risks/Civil Protection	0.93
Geophysics	0.89
Arts and culture	0.88
Climate change	0.87
Navigation and location	0.82
Oceanography/Coasts	0.70
Living place	0.65
Transport and logistics	0.64
Telecommunications	0.59
Defense and National Security	0.55
Demography and urban development	0.49
Sociology	0.35
Legal	0.20
Health	0.14
Marketing	0.11
% of total cases, N = 19728 (multiple responses are possible)	

The MR1 asked users about the usefulness (perceived utility) and the updateness needs. Table 6 presents the results of these two questions. More than 82% of users graded the utility value as high or very high, and the percentage of low grading is residual. We considered it means a

good level of satisfaction with the products. In relation to the need of updateness, greater variation is observed in the results, here there is no such unanimous agreement on the need of updateness, perhaps this is due to the fact that the products evaluated are relatively recent.

Table 6 – MR1: Assessment of perceived utility and updateness needs in Spain

Utility assessment		Updateness need assessment	
Utility level	%	Updateness need	%
1 (low)	1.6	1 (low)	8.1
2	1.9	2	8.4
3	14.4	3	26.3
4	23.2	4	21.0
5 (very high)	58.9	5 (very high)	36.2
% of total cases, N=6087			

Approximation to users and identification of domains by means of MR2

As stated before, MR2 is an *ad hoc* designed worldwide survey. The aim of this survey is to supply us information about users of DEMs, the use of DEMs, general features of DEMs being used and, finally, information about relevant quality aspects of DEMs. Following this order, here are presented the more important findings. Tables include the number

Users

The information about the users was queried in questions 1-3 of MR2, and a summary of the answers is presented in Table 7. In relation to the user profile (Q1), we can distinguish two main groups, the first one including three categories: teachers, researchers and professionals of the geomatics sector; and a second one including: postgrad students and professionals of other sectors than geomatics. It is interesting to notice that profiles linked to teaching, studying and research sum, altogether, 63.8% which means that the majority of DEMs users belongs to a very speculative sector. In relation to years of experience (Q3) conclusion is that the respondents have, for the most part, a wide experience in the use of the DEM, since they have more than 4 years of experience. In relation to the organization profile (Q2), we can distinguish three groups. The first group includes only the category University, which is the predominant organization profile, very close to the half of the cases. The second group includes also only one category, Mapping Agencies, very close to the quarter of the cases. Finally, a third group can be established by the rest of categories: Research institutions, Environmental agencies and Companies. This result is, somewhat, coincident with the result of Q1.

Table 7 – MR2: Assessment of perceived utility and updateness needs in Spain

User profile (Q1)	%	Organization profile (Q2)	%	Years of experience (Q3)	%
University teacher (professor, emeritus, etc.)	27.5	University	47	More than 4	85.3
Researcher	27.5	Mapping Agencies	23.5	Between 1 and 4	13.7
Professional (Geomatics)	24.4	Research institution	9.8	Less than 1	1
Professionals (Other fields)	10.8	Environmental agencies	9.8		
Students (PhD, master, etc.)	8.8	Companies	5.9		
OTHERS	1.0	Professionals	2		
		OTHERS	2		
% of total cases, N=102					

In the comparison of these results to those results of MR1, we have to consider that only the organization profile from Table 6 can be analyzed in relation to results of Table 4. Here the answers and figures are different, of course, and we have to take into account that the

questionnaires do not have the same questions and classification scheme, that the propose and channels for contact were very different and that: i) in the case of MR1, there is a specific product that is biasing the population of respondents, ii) in the case of the MR2, there is a specific source of the e-mails (papers in international journals and professional associations), biasing the population of respondents. So we can conclude that we have to different views of users of DEM. The first one, provided by MR1, is a view related to specific DEM products and their use by society (in a specific country), and the second one related to a general scientific use of DEM in the world and their use by the more active and innovative users of DEM. We think that both views are of great interest for understanding the users and uses of DEMS.

Use and general features of DEMs

The information about the use and general features of DEMs were queried in questions 4-9 of MR2. Table 8 presents the summary in relation to the purpose of use. In the first category (accounting 52.8%), there are included two general proposes of use (Water and Environmental studies) and an item related DEM's data capture techniques, which means a great relation with DEM generation. The second category (accounting 27.6%) includes a very important field of application (the Geology) and the Others case, which is the grouping of many specific identified purposes, but each one with a low percentage (less than 1%). The production of DEMs is another item in this category and is related to data processing techniques. So, we have two specific purposes related to DEMs generation in the first two categories, they account 26.93% together, and it means that users are very close to DEM's data generation and refinement. The last category includes a large tail of purposes with percentages between 1.63% and 4.88%.

Table 8 – MR2: Purpose (Q4)

Purpose	%
Water studies and river network	19.51
Photogrammetry, LiDAR, Laser Scanner, UAV, Remote Sensing, etc.	18.70
Environmental studies	14.63
Production of Digital Elevation Models	8.13
Geology	6.50
OTHERS	7.31
Evaluation and study of agricultural crops	4.88
Disaster assessment, civil protection, etc.	4.07
Soil Science and Botany	4.07
Civil Engineering	2.44
Glaciology	2.44
Volcanology	2.44
3D printing	1.63
Architecture	1.63
Geodesy	1.63
% for each option. Number of respondents = 102, number of responses = 123	

In relation to the calculations performed with the DEM data (Q5, with multiple responses), Table 9 presents the summary. Here we have classified results in four categories. Because DEMs are elevation data, the option accounting the largest percentage (75.5%) is the use of the elevation data (height). The calculation of slope, other typical general analysis, is close to the previous one, but accounting for 71.6%. If we consider the third case (aspect), three of direct topographic attributes are the three more common calculations. In the second category

we can see many calculations related with operations resulting in features (e.g. contours, view-shed, catchment and channels) and two other cases (aspect and flow accumulation) related with fields. The third category includes calculations related with the surface (curvature, concavity & convexity, roughness), secondary topographic attributes (TWI – Topographic Wetness index–, and CTI –Compound Topographic Index–), and other calculations. The fourth category includes cost analysis and hill shading. So, if we do not consider the three first cases (elevation, slope and aspect), that can be considered as general calculations for a DEM, the majority of calculations are centered on hydrological features and properties (drainage, flow, channel, catchment, etc.), which confirms that water related issues are the more important use of DEMs.

Table 9 – MR2: Calculations (Q5)

Calculation	%
Height	75.5
Slope	71.6
Aspect	52.9
Drainage delineation	51.0
Contours	48.0
Flow Accumulation	46.1
View-shed analysis	44.1
Catchment Area	36.3
Channel detection	33.3
Curvature	29.4
Indexes (TWI, CTI)	28.4
Concavity & Convexity	26.5
Ridge detection	26.5
Roughness	27.5
Model parameter generation	23.5
Cost analysis	10.8
Hillshade	3.0
Others	11.0
% for each option. Number of respondents = 102, number of responses = 675	

Q6 was not a question but a request for an explaining document. In this case, 38.2% of respondents (N=102) upload a file. The majority (74.35%) of the document are scientific papers (journal & congress papers), and the remaining a mix of different types (reports, presentations, and so on). Scientific papers were published in journals such as: Computers & Geosciences, Earth surface processes and landforms, Ecological Modelling, Environmental Modelling & Software, Geomorphology, International journal of Geographical Information Science, ISPRS Journal of Photogrammetry and Remote Sensing, Journal of Archaeological Science, Journal of Hydrology, Journal of South American Earth Sciences, Journal of Structural Geology, Landscape and Urban Planning, Plant soil Environ, The Cryosphere, and several national journals. In some way, these results corroborate those related to the user profile (Q1) and those related to purpose (Q4).

Q7 is asking respondents a short description of the process(es) that takes place in their use case, but received answers are not focused on the aim of the question itself. 20% of responses were void and the majority of the remaining was generic answers (e.g. DEM generation, DEM production, extraction of elevation and certain topographic features, etc.), with no

proper information about process steps. By this reason we considered that the results of this question are no valid.

The results in relation to the data model are presented in Table 10. The raster model (91.2%) is the predominant option; this result coincides with other analysis and is coherent because of majority of DEM datasets are offered as raster datasets. The TIN and contour models are in the same range order. The use of profiles is detected and accounts 12.7%, this model is basically used for civil engineering applications and hydrological modelling in channels (e.g. HEC-RAS). Others is a class including others answers such as ASCII, Lidar Data, DEM, and similar answer but not adequate. There are 59.8% of users that indicate only one data model, 20.58% indicating the use of two data models, and 10.78% and 6.86% indicating the use of three and four data models, respectively. Table 11 is showing the relation between data models being used by respondents. One conclusion of this table is that users use several data models but have a preferred data model (e.g. Raster, TIN, Contours, Profiles), and this data model is used by this user more than the other options. This is very clear for the Raster case, but also for the other three cases.

Table 10 – MR2: Use of data models (Q8)

Data model	%
Raster	91.2
TIN	30.4
Contour	26.5
Profile	12.7
OTHERS	4.0
% for each option. Number of respondents = 102, number of responses = 164	

Table 11 – MR2: Cross tabulation of the used data models (Q8)

%	Raster	TIN	Contours	Profiles
Raster	91.18	26.47	22.55	11.76
TIN	26.47	30.39	15.69	8.82
Contours	22.55	15.69	26.47	8.82
Profiles	11.76	8.82	8.82	12.75
% for each option. Number of respondents = 102, number of responses = 164				

Table 12 presents the results in relation to sources (Q9). In this case, all respondents have indicated, at least, one source, and many of them several sources or a range of sources (e.g. Satellite – Total Station; Satellite – UAV). In these last cases we have add one count to each possible option of the range. Satellite data and data products of official mapping agencies (national or regional/state) are the options with largest percentages. In the next group, with similar percentages, are included a technology (Lidar) and a very ambiguous answer: Produced. This answer is confusing because this was an open question and respondents were able to indicate any specific source. Existing global products and photogrammetry are the two cases in the next group. The last category includes two cases related to existing data sets (Data Warehouse and Cartography) and a technique for data capture. If we group cases in relation to the use of existing products (e.g. GDEM, official products, etc.) and sources for producing their own product (e.g. satellite, Lidar, produced, photogrammetry, bathymetry), we obtain 37.1% and 62.9%, respectively which indicates that majority of respondents are also producing their own DEM data sets.

Table 12 – MR2: Source of DEM data (Q9)

Source	%
Satellite	22.47
Official products from Mapping agencies	19.66
Lidar	14.61
Produced	14.61
Global products	8.43
Photogrammetry	7.30
Data Warehouse	5.06
Bathymetry	3.93
Cartography	3.93
% for each option. Number of respondents = 102, number of responses = 178	

Q9 asked also about scale, resolution, accuracy and density, the last for the case of Lidar data. The results for all these cases are shown in Table 13. In this question the respondents can give one or more responses. Also, some respondents have given a range. Ranges have been decomposed in one account for each element of the range. The percentage has been calculated for the total amount of answers for each case (e.g. resolution, scale, etc.). It is interesting to point out that resolution is the preferred term in order to express this data property and accounts 141 responses. Scale and accuracy are less used terms. In the case of resolutions, it is very clear that majority of users are working with metric (m) or decametric (x10m) resolutions, which is coincident with the distribution of responses for the case of scale and with the majority of available products from mapping agencies and Earth observation systems. The responses using the term accuracy show only three levels (mm, cm, m), and the centimeters case is accounting for the larger percentage (54.55%), this means a bias to grater scales/resolutions in relation to the answers obtained with the terms resolution and scale. In the case of density, answers are only valid for Lidar data. In this case, the point spacing equivalent to density is between 1.41 m and 0.7 m, which is related the metric options of resolution and accuracy.

Table 13 – MR2: Resolution, Scale, Accuracy and Density (Q9)

Resolution (1)	%	Scale (2)	%	Accuracy (3)	%	Density (4)	%
mm	4.69	E0,5K	3.57	mm	18.18	2 p/m2	27.27
cm	9.38	E1k	7.14	cm	54.55	1 p/m2	27.27
m	36.72	E2K	10.71	m	27.27	0,5 p/m2	45.45
x10m	42.97	E5k	14.29				
x100m	3.13	E10k	10.71				
x1000m	3.13	E25K	17.86				
		E50K	17.86				
		E100k	7.14				
		E250K	7.14				
		E1000K	3.57				
% for each option. Number of respondents = 102, number of responses = 141(1); 28(2); 22(3); 11(4)							

Information about relevant quality aspects

Several aspects related with relevant quality issues (e.g. exigencies of the application, use of quality indexes, degree of satisfaction with those quality indexes, etc.), were aims of questions 10 up to 18.

It is interesting to get knowledge about the main exigency of calculus of the applications. This

issue was addressed by Q10, and results are shown in Table 14. Of course, the main exigency for a reliable calculus is the height. In a second category they have been included very common calculations (slope and watercourse and watershed delineation). The third class includes also common calculations (view shed and orientation), but two new options appear in this survey: insolation and volume. The Others category includes a large set of options (e.g. anisotropic artifacts, fault mapping, break lines, etc.), each one accounting only for one count. This question has a great relation with Q5, and for this reason results must be somewhat similar, for instance, if we joint the cases Orientation and Insolation, the four first elements of Table 14 are coincident with those of Table 9. All of these options are commonly included in software tools, but for many users the underlying algorithms and their reliability are unknown, as stated in recent studies (Mei-Po, 2016; Kitchin, 2017).

Table 14 – MR2: Main exigency of calculus (Q10)

calculation issue	%
Height	69.61%
Slope	42.16%
Watercourses	34.31%
Watershed	32.35%
View Shed	24.51%
Orientation	22.55%
Insolation	21.57%
Volume	20.59%
OTHERS	8.00%
% for each option. Number of respondents = 102, number of responses = 282	

Q11 is centered on how users are accustomed to express accuracy (Table 15). The option with the larger percentage (47.1%) corresponds to the use of a statistical approach. It is interesting to notice here that the second option, with 26.5%, correspond to subjective judgments (e.g. visual inspection). We do not know if it is due to ignorance or dissatisfaction with the most objective methods. The use of DEM in test of use is a pragmatic option between a statistical assessment (e.g. estimation or control of processes), and a subjective judgement. Also the subjective judgement can be based on a test of use. The option Depending on the case includes cases that can be considered in any of the previous one. It is interesting to notice here that: i) there are 5.9% of cases where no expression of accuracy is used, and ii) 4.9% of respondents have given no answer. For the last, we can conclude that, in the order of 11% of the cases, users of DEMs have no concern about the quality of the data.

Table 15 – MR2: Expression of accuracy (Q11)

Expression	%
Statistical assessment base	47.1
Subjective judgment	26.5
Test of USE	14.7
None	5.9
No answer	4.9
Depending on the case	1.0
% of total cases. Number of respondents = 102.	

The use of indices to express quality is a key element in order to understand actual needs of users. Q12 is oriented to get some insight and results are shown at Table 16. First of all, it is

relevant that 20.59% of respondents indicate that they do not use any indices; also there is 8.82% of cases giving no answer, so 37.25% of responses are indicating no use of indices. The majority, almost a third, use positional accuracy related indices, which is the most traditional way to express accuracy in DEMs. The third group of responses includes two cases, the first one supposes the use of positional accuracy indices plus other indices but in an independent way, so that several indices have to be considered altogether by the user; and the second one is indicating the use of measures related to the results, which is very interesting because it supposes a kind of “fitness for use” of the indices. Almost 4% of cases indicate some adaptation to the case which means “fitness for use”. A similar percentage of responses indicate the use of composite indices which is considered by us a step forward to achieve a better characterization of quality, and closer to “fitness for use”. So, we conclude that 42.16% of responses indicate some trend in order to use indices with a better description of quality than traditional measures based only on the positional accuracy.

Table 16 – MR2: Type of indices to express quality (Q12)

Option	%
Positional accuracy related indices	28.43
None	20.59
Positional accuracy + other, in an independent way	17.65
Measures related to the results	16.67
No answer	8.82
Depending on the case	3.92
Composite indices	3.92
% of total cases. Number of respondents = 102	

The degree of satisfaction about the accuracy indexes was dealt by Q13 (Table 17). Only few cases (6.9%) indicate that respondents really are quite satisfied, almost the majority of the respondents (49%) indicate that they are satisfied. On the other hand, 18.6% is dissatisfied and 6.9% very dissatisfied, so that, the percentage of satisfied users is in the order of 2.5 times the percentage of dissatisfied users, and only a small percentage is very dissatisfied. Also is remarkable that almost 10% of users give no answer to this simple question. This percentage is high and very close to that pointed out by Q11 (related to the expression of accuracy), where the conclusion was that there were a lot of users of DEMs which have no concern about the quality of the data. So that, both questions, Q11 and Q13, are according.

Table 17 – MR2: Degree of satisfaction (Q13)

Option	%
Satisfied	49.0
Dissatisfied	18.6
Quite satisfied	15.7
No answer	9.8
Not at all satisfied	6.9
% of total cases. Number of respondents = 102.	

Another way to approach user's satisfaction about actual quality information is querying about something to improve. In this direction was addressed Q14 which (results at Table 18). A large number (31.37%) of respondents did not respond and 14.71% of respondents gave answers not directly related to this question. 3.92% of answers are indicating no improve is needed. Also, 2.94% of respondents indicated that they do not have criterion to answer this question. In relation to those answers directly related to this question, it is interesting to notice

that majority of respondents are demanding more information, not better measures or indices. The demand of better information about altimetric (Z) and planimetric (XY) accuracies are the two answers with the larger percentages, both together rising 30.39%. 9.8% of respondents want to know how is the spatial distribution of DEMs errors in each point of the DEM's grid, which is a very reasonable and useful demand for many proposes. There is also a concern (7.84%) about the ground check survey (number of control elements, sample size, distribution, etc.). In a specific way, lineage information (sources and process steps or history describing transformations), is a demand pointed out by 11.76% (=6.86% +4.9%) of responses, but if we would consider here other lineage related issues (e.g. information about ground survey, etc.), the lineage will be the topic with the larger percentage. Resolution is a very basic characteristic of DEMs, but there are answers (3.92%) demanding more information about it. Some users suggest as an improvement the inclusion of information about the quality of break lines (e.g. water channels), which can be considered as a very interesting positional accuracy demand that can be satisfied using line-based methods. Other interesting proposals, each one accounting 1.96%, are to provide information about relative positional accuracy and to provide accuracies segmented by zones or other key factors (e.g. land cover types). The use of ISO 19157 by means of the proposed measures, and metaquality elements, is proposed by the same percentage (1.96%). Finally, there is a large tail of responses each one accounting only one response-case, that we have summarized in the case so called "others" which are related to many issues (e.g. date of capture, data format, coordinate system, etc.). A general feeling generated by these results is that either there is no metadata of the datasets respondents are using, or existing metadata are very bad, or that respondents do not use the metadata that exists for the datasets they use, or metadata are very complex for they to be accessed and understood. This conclusion is derived because of majority of public spatial datasets of development countries have metadata (e.g. following ISO 19115), and all this information can be introduced as metadata elements.

Table 18 – MR2: Something to improve (Q14)

Improvements	%
No answer	31.37
Better information on Z accuracy	17.65
Improper answers	14.71
Better information on XY accuracy	12.75
Information on spatial distribution of error (per point of DEM)	9.8
Information on the ground check survey	7.84
Other (data format, date, coordinate systems, etc.)	7.84
Information on all steps of the DEM generation process	6.86
Information on data source (including accuracy)	4.9
Information on resolution	3.92
No need	3.92
Information about the accuracy of break lines	2.94
No criterion	2.94
Information about relative accuracy	1.96
Accuracies segmented by zones/types	1.96
Metaquality	1.96
ISO 19157	1.96
% for each option. Number of respondents = 102, number of responses = 138	

Q15 was initially addressed to get some suggestions from users that did not use any index.

This question was not automatically linked with the response of Q11 and Q12, so that responses can come from all respondents. In this way we perform a cross tabulation of responses of this question with those of previous questions and find that respondents not included as initial target gave a response. Finally, have find 11 different responses (Table 19) coming from all kind of respondents. All answers are unique except for the 7th answer (drainage network), proposed by two respondents. Some of these proposals are coincident with results of Q14, which is logical. Nevertheless, we think there are very interesting proposals in these suggestions in order to explore new “fitness for use” approaches.

Table 19 – MR2: Proposals from respondents (Q15)

“Average difference (for a single cell) between true higher and lower elevation values of targets that fall inside the grid-cell and the value of the cell itself.”
“Completeness of filtered data”
“Indices from image analysis (e.g. horizontal displacement with Optical Flow and PIV -Particle Image Velocimetry)”
“Network properties such as path length between adjacent pixels”
“Positional measures from a regular grid of points”
“Quantity of pixels with disproportionate effects on the outcome”
“Related to the positional accuracy of the drainage network and break lines”
“Related to outliers labeling”
“Relative accuracy between different, neighboring points or cells, i.e. the precision of differences between points rather than on the absolute value itself.”

A descriptive approach about the influence of bad quality in the results was addressed by Q16 (Table 20). As happened for Q14, a large number (30.39%) of respondents did not respond, which is surprising. Also, here a lot of responses (13.73%) are not directly related to this question. The answers indicating estimation problems, no matter what parameter (e.g. slope, orientation, volume, etc.) are those that have the highest percentage (24.51%). 22.55% of answers indicate problems caused by data (e.g. rework, conflation, location, etc.). 15.68% of answers indicate the influence of bad quality over decisions, costs and hazards. Finally, 5.88% of responses are indicating that poor quality is not relevant to results.

Table 20 – MR2: Explanation of the influence of bad quality (Q16)

Options	%
No answer	30.39
Estimation	24.51
Data problems	22.55
Not related answer	13.73
Bad decision	7.84
No relevant	5.88
Cost	4.90
Hazards	2.94
% for each option. Number of respondents = 102, number of responses = 115	

The valuation of the influence of bad quality of DEMs is surveyed in Q17 and Q18 from a qualitative and quantitative (economic) perspective, respectively (results at Table 21). In relation to the score of the influence, the majority of the answers (56.9%) indicate that it is high, and 30.4% that influence is medium. There are users thinking that the influence is low (7.8%) and even none (2%), also there is 2.9% of users that are no able to give an answer to this question. It is interesting to compare this results with those presented for Q16. For

instance, for the Q17 case there is 2% indicating none influence but there was 5.88% indicating no relevance in the case of Q16. Also, for the Q16 there were 30.39% of no answer responses but here for the Q17 case are only 2.9% of no answer responses. We conclude that people are not able to offer an explanation of the influence of bad quality, but they have the feeling that bad quality has relevance.

From an economic perspective, Q18 proposed the same issue as Q17. The majority of the answers (45.63%) goes to “no answer” option, and this is the “no answer” case with the larger value in all the MR2 responses. This means that users are really not able to give an answer; they have no quantitative idea about the economic impact of bad quality. They routinely use DEM datasets but they do not know the economic consequences of poor data. This is in accordance to the results obtained for Q16 (no answer option) but now the percentage is higher because we are not demanding an explanation but a figure. Also, 23.30% of respondents indicate that the use of DEMs of bad quality has no economic consequence. For us this answer is very contradictory with respect Q17. We do not see a direct relation between this result and those of Q17; the only justification we find is that influence Q17 was understood as influence in results and that results of bad quality have no economic consequence for these users. A similar percentage 23.30% of respondents give no quantification but write an explanation of consequences. This means the acknowledgement of the influence of bad quality but also the incapacity of translate bad quality problems to money. Only 7% of respondents are able to monetize the consequences of bad quality given an order of magnitude, or range, of the economic impact. Here it is interesting to remember that respondents have some experience (85.4% more than 4 years and 13.7% between 1 and 4 years).

Table 21 – MR2: Valuation of the influence of bad quality (Q17 & Q18)

Score of influence (Q17) (1)	%	Monetization of consequences (Q18) (2)	%
High	56.9	No answer	45.63%
Medium	30.4	None	23.30%
Low	7.8	No quantification is possible (Explanation of consequences).	23.30%
None	2.0	x1.000.000 USD	2.91%
No answer	2.9	x100.000 USD	0.97%
		x1.000 USD	0.97%
		x100 USD	0.97%
		x10 USD	0.97%
		≈0 USD (Insignificant)	0.97%
(1) N=102; (2) N = 103, Multiple responses allowed			

Application domains: identification, relations and semantic by means of MR3

Table 1 has presented a general view of the methods and targets of this study; and in relation to MR3 there were three different targets, all of them related to domains: a) identification, b) relations between domains and iii) a semantic approach.

The identification is a redundant work in relation to previous MR1 and MR2 approaches, but we consider that redundancy is not bad in this case because this identification shows a somewhat different perspective from that of previous approaches. Here the results are not coming from questionnaires filled by users; the searches in ScienceDirect have a bias to scientific and technical applications but also the spectrum of possibilities is wider. The analysis of relations can be afforded because we have pairs of “themes” and “topics” which can be seen as relations between themes, and the themes as an approximation to domains. Finally, a semantic analysis is possible because abstracts can be considered as the better short

description of a work or application. Abstracts are key element in the peer review system of scientific journals, so that they must be clearly explanatory of the uses/applications and achievements.

Identification

Topics (or themes) are not exactly coincident with application domains but are close to them. Topics are proposed by authors of the papers, some journals offer a close list for the selection of topics and in other journals the proposal is completely free. As previously explained, a first result of this approach is the frequency of appearance of each topic (theme) in the results. A total of 7856 topics were retrieved and Table 22 shows the results as percentages. Topics present a mix of very different things: data capture systems (Lidar, GPS, SAR, LandSat, ASTER), products (GDEM), features (water, river, glacier), sciences (Geology, Hydrology), etc. For this reason, they have been divided into three categories, the first one related to domains of application in a wide sense, the second one related to data capture tools and products, and the third one related to quality issues. The terms of the first category could be reorganized in a way more closed to the domains of application, but as it would be something subjective and complex we prefer to continue working with them. It is interesting to notice here that quality terms sum 8.85% of the total which means a great general concern with these issues. All topics presented in Table 22 are, in some way, related to several previous questions (Q4, Q5, Q9 and Q10), and this confirms the importance of all of them for the scientific community.

Table 22 – MR3: Topics

Topics	%	Tools and data	%	Quality	%
Slope	11.85	Remote sensing	6.58	Accuracy	5.82
Soil	7.73	Lidar	3.72	Quality	3.03
River	6.61	Landsat	3.13		
Drainage	5.47	GPS	2.06		
Forest	4.65	Aster	1.9		
Flow	4.56	SAR	1.69		
Water	4.06	GDEM	0.48		
Planning	3.59				
Environment	2.32				
Urban	2.16				
Landslide	2.09				
Glacier	1.91				
Hydrology	1.79				
Geology	1.76				
Coast	1.53				
Climate change	1.5				
Agriculture	1.36				
Risk	1.3				
Biodiversity	1.16				
Solar radiation	0.88				
Mineral	0.75				
Infrastructure	0.65				
Navigation	0.65				
Engineering	0.64				
Subsidence	0.64				
% of total cases. N= 7856					

Relational analysis

As said before, the analysis of relations can be addressed because we have pairs of “themes” and “topics” which can be seen as relations between themes, and themes as an approximation to domains. Fifty-three themes were used for the queries to ScienceDirect and a list of two

hundred and forty-three different topics was obtained. With this amount of cases an analysis is very complex, so a selection process was afforded. The number of cases of topics citations obtained for each theme was used as the weight for the themes. Themes with weight less than 100 were eliminated. In the same way, the number of cases of topics citations was used as the weight for the topics, and topics which weights lower than 10 were deleted in order to show only the most relevant. As result of the explained filtering we obtained 21 themes, 39 topics and 222 relations between them.

For the relational analysis between themes and topics we selected a force-directed graph in order to represent the existing relations. Specifically, we selected a bipartite graph as appropriate tools for this propose. Newman (2010) affirms that in a network bipartite graph there are two kinds of vertices, one representing the original vertices and the other representing the groups to which they are related. For this purpose, we used R (R Core Team, 2013) with the packages igraph (Csardi and Nepusz, 2016) and statnet (Handcock et al. 2016). By this way, a network bipartite was built with two disjoint sets: the themes and the topics. The network visualization algorithm employed is named Large Graph Layout and it is recommended for big size networks Ognyanova (2016). Some network parameters (measures of centrality), were computed in order to analyze the structure of the relations. In this case, closeness centrality, betweenness centrality and cut-points have been analyzed.

Figure 2 shows the bipartite network graph generated, where the size of nodes represents the number of edges incident in each one, and width of edges indicate the topics frequency into themes. Density is 0.125, that means that the network has the enough ties to let a correct visualization (Luke, 2015). Closeness centrality declares that a vertex is more important if it is “close” to many other vertices. In the approach proposed by Sabidussi (1966) the centrality vary inversely with a measure of the total distance of a vertex from all others. For comparison purposes, the measure is normalized in the interval [0, 1]. For instance, closeness centrality for “Soil” is 0.562 which means that it is most important topic in this network. Betweenness centrality attempts to summarize the extent to which a vertex is located “between” other pairs of vertices. In this case, the “importance” is related to where a vertex is located with respect to the paths in the network graph (Kolaczyk et. al. 2014). Betweenness centrality of “Soil” is 220, and 204 for “Environment”, which shows the strong relation that exist between them and the queries about DEMs. The network cutpoints, also known as articulation points, are nodes that, if dropped, would increase the number of components in a network. Therefore, they usually occupy an important position connecting different parts in the network. In our case, the cutpoints are: Biodiversity, Environment and Solar Radiation (R_S). If these nodes were separated of the network, they will generate new components. To identify cutpoints, the cutpoint() function in statnet library in R was employed.

Figure 2 – Bipartite network graph (Themes and topics)

Semantic analysis Clustering analysis of papers' abstracts

Another approach developed was the semantic analysis of all papers' abstract obtained from the results of all queries in order to define domains in a much rich way. In this case were used nine hundred and fifty abstracts manually revised and coming from the all queries. This analysis is done by text clustering. There are some algorithms available to carry out this kind of process. We have employed the Lingo algorithm, one of the most reliable (Osiński et.al 2004). The main goal of this algorithm is the capability to capture thematic lines, that is discover groups of related documents and describe the subject of these groups in a meaningful human way (Osiński et.al 2004).

We used the Lingo algorithm implemented in Carrot2 Workbench software (<http://project.carrot2.org/>). In summary, the process is as follows, preliminary a text segmentation is developed and if terms-language is recognized, the stemming application and stop-words making is carried out. After this, frequent phrase extraction, cluster label induction, cluster content discovery and final cluster formation are executed. Finally, a Voronoi map is generated with the clusters identified by the Lingo algorithm. In relation to the

parameters and specific settings, we used the Nonnegative Matrix Factorization ED Factory method for matrix model, with a maximum matrix size of 100.000 and a maximum word document frequency of 0.90. Also a word document frequency threshold of 10% was considered. Following this procedure, a total of 31 clusters with their labels were identified (Figure 3). The size of each Voronoi polygon depends of the number of occurrences of phrases. The core of clusters is around of cases including the word “model” (e.g. flow model, GIS model, etc.). The majority of clusters have a meaningful interpretation, according the theory objective of Lingo algorithm.

Figure 3 – Cluster analysis of abstracts

This result is very different from that of the relational analysis and can be considered surprising for some of the voronoi polygons (e.g. error models is in center), but here we have to get in mind that the relational analysis is based on relations between terms and this analysis is based on the analysis of sentences that are present in abstracts. Sources are different, context of analysis is different and also the tools for the analysis.

Approximation to use cases by means of MR4

Use cases are a method used in software and system engineering in order to identify, clarify, and organize requirements. A use case is a formal definition, a structured document, explaining for a specific requirement, actors, actions, inputs outputs and decisions in order to achieve a goal. Unified Modelling Language (UML) diagrams, or other graphical language, can be used in order to depict uses cases in a standard way. Weibel and Heller (1991) presented a table of DTM applications domains and their functional requirements desegregated in a list of issues that can be understood as use cases in a software engineering project. For us use cases are in a higher level, and not related with software capabilities development but with applications development. The interest of use cases in our survey is twofold; first existing use cases will inform us about important and specific applications of DEM, that have been somewhat formalized and second; if use cases are sufficiently detailed they will inform us about quality-related issues.

The identification of use cases about DEM applications or uses in Internet using Google searches has been difficult. A lot of not relevant documents were found and only very few with the structure of a use case. Table 23 presents a list with the identified use cases. First, it is noticeable that the amount of found documents in Table 23 is much reduced. There are many documents using the term “uses case” dealing with applications of DEMs, but very few of them are formal presentation of a use case. Second, all documents of Table 23 are coming from European initiatives funded by diverse public projects (e.g. Inspire, 6th and 7th

framework programs, etc.). Third, not all the documents have the same in-depth and detail description of the explained use case, in this way the documents which reference is SS (SS, 2006) are more application schemes than use cases.

Table 23 – MR4: Use cases

Title of the use case	Theme	Reference
Glacial geomorphological mapping	Glaciology	Smith, M.J. (2005)
Search for blandings	Decision taking	Smith, M.J. (2005)
Winter snow accumulation	Glaciology	Smith, M.J. (2005)
Soil erosion estimation	Soil erosion	Smith, M.J. (2005)
River characterization	Channel network	Smith, M.J. (2005)
Comparison of real with theoretical drainage network	Channel network	Brönnner, C. (2007).
Flood mapping	Flood risk	INSPIRE (2013)
Orthoimagery production	Geomatics	INSPIRE (2013)
Elevation mapping	Geomatics	INSPIRE (2013)
Maintenance of fairways	Navigation	INSPIRE (2013)
EC water directive		SS 2006 *
Flow calculation		SS 2006 *
Calculation of the dispersion of pollutants and dissolved substances in water		SS 2006 *
		SS 2006 *

CONCLUSIONS

The importance of the MDEs is great and is recognized in many documents, and for this reason the knowledge about the users, uses and facts related to the quality of the DEMs is important. In this paper several resources (MR1, MR2, etc.) and activities (surveys, database queries, etc.) have been deployed to characterize these aspects.

The MR1 resource is the questionnaire of the DEM data download center of the Instituto Geográfico Nacional (IGN, Spain). The data coming from the MR1 resource have offered us a great amount of cases in order to characterize in detail the profile of the users in Spain. In this case, 70.74% of users are particular uses and we consider that this profile can be extrapolable to countries with a similar context to Spain. More than 82% of Spanish users graded the utility of DEMs as high or very high. Also, MR1 has provided us with a large list of application domains and their corresponding importance, in this case the main application domains are Cartography and Environment.

The MR2 resource is an *ad hoc* designed survey in order to obtain specific information related to quality. This survey is biased by the source of contact information, which is coming from scientific papers. At a first view, some information (e.g. user's profiles and application domains) coming from this survey is redundant with that from the MR1, but both are complementary perspectives and allow us to know the profiles of general users (MR1) and profiles of advanced users that have a more scientific character and demands. Because of this reason, the results are not the same (e.g. compare Table 4 vs Table 7(Q1) or Table 5 vs Table 8), but they have a general consistency. Many questions of the MR2 supply new and very interesting information. Now we have identified those calculations that are more usual (e.g. slope and height) and those on which the users have a greater exigency (e.g. height, slope, watercourses). It is confirmed that the most used data model is the raster and that the resolutions most used are in the order of 1m -10m. Users of DEMs express quality mostly by means of statistical evaluations, but subjective evaluation is of great importance. There is a

high percentage of users who do not use any quality index and the degree of satisfaction with the quality indexes used is not full, only 64.7% of users indicate that they are satisfied or quite satisfied. Users do not know how to evaluate the influence of poor quality on their work, this is a critical aspect that should be taken into account in future researches. In addition, MR2 (Q11 and Q15) has provided a set of ideas that can serve to guide future work on quality in DEM.

MR3 is related to queries to ScienceDirect database. This resource allows us to search millions of scientific papers and discover papers dealing with researches focused on DEMs. From queries to the ScienceDirect database it has been possible to generate a graph that shows the relationships between relevant terms that are linked to publications on DEM. Also, we have made a semantic analysis of a set of 950 abstracts, and this analysis outputs the parts of sentences present more frequently. This analysis shows a novel perspective as a way to summarize thousands of documents, and ScienceDirect is a resource that can allow interesting analysis.

From MR4 it can be concluded that the formalization of applications as use cases is not usual. Very few cases have been found and all focused on very concrete projects (e.g. Inspire). This indicates that it is not an obvious necessity. Majority of use cases are related to water/hydrology, which means the maturity of hydrological applications and also their social and legal importance.

Finally, we can conclude that the results of this paper have evidenced that the analyzed quality related issues (expression of accuracy, type of indices, satisfaction, possible improvements, the influence of bad quality and the monetization of consequences.) need some attention. There is a need for empowerment on quality issues by users and this can only be achieved through research on quality issues focused on specific use cases, and the outreach and dissemination of findings and developments in specific user's communities.

REFERENCES

- Agumya and Hunter (1999). Assessing fitness for use of geographic information: what risk are we prepared to accept in our decisions?
- Amii R. Darnell, Nicholas J. Tate, Chris Brunsdon (2008). Improving user assessment of error implications in digital elevation models. *Computers, Environment and Urban Systems* 32 (2008) 268–277.
- Anna T. Boin and Gary J. Hunter (2006). Do spatial data consumers really understand data quality information? In 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences. Edited by M. Caetano and M. Painho.
- Bishwa Acharya, Ph.D.* Jeffrey Fagerman, RLS (2000). ACCURACY ASSESSMENT OF DTM DATA: A Cost Effective Approach for a Large Scale Digital Mapping Project. IAPRS, Vol. XXXIII, Amsterdam, 2000
- Boin, A.T., & Hunter, G.J. (2009) What communicates quality to the spatial data consumer?. In: Stein, A., Bijker, W., & Shi, W., eds. *Quality aspects in spatial data mining*, CRC Press, Boca Raton, pp. 285–296.
- Brönnner, C. (2007). Comparison of real with theoretical drainage network. Project: RISE (Reference Information Specifications for Europe)
- Carlisle, Bruce H (2005). "Modelling the spatial distribution of DEM error". En: *Transactions in GIS* 9.4, p. 521-540
- Csardi G, Nepusz T: The igraph software package for complex network research, *InterJournal, Complex Systems* 1695. 2006. <http://igraph.org>
- Devillers et al. 2002. Spatial data quality: from metadata to quality indicators and contextual end user manuals
- Devillers, R., Bédard, Y., and Jeansoulin, R., 2005, Multidimensional management of geospatial data quality information for its dynamic use within GIS. *Photogrammetric Engineering & Remote Sensing*, 71(2), pp. 205–215.
- Fisher, Peter F y Tate Nicholas J (2006). "Causes and consequences of error in digital elevation models". En: *Progress in Physical Geography* 30.4, p. 467-489.
- H. Yao, R. L. Clark (2000). Evaluation of Sub-Meter and 2 to 5 Meter Accuracy GPS Receivers to Develop Digital Elevation Models. *Precision Agriculture*. Volume 2, Issue 2, pp 189-200

A.5. Artículo: DEMs: An approach to users and uses from the quality perspective

- Handcock M, Hunter D, Butts C, Goodreau S, Krivitsky P, Bender-deMoll S and Morris M (2016). statnet: Software Tools for the Statistical Analysis of Network Data. The Statnet Project (<http://www.statnet.org>). R package version 2016.9, CRAN.R-project.org/package=statnet.
- ICSM (2008). ICSM Guidelines for Digital Elevation Data v.1. Intergovernmental Committee on Surveying and Mapping (ICSM).
- INSPIRE (2013). D2.8.II.1 INSPIRE Data Specification on Elevation – Technical Guidelines
- J. R. Sulebak (2000). - Applications of Digital Elevation Models - DYNAMAP "White paper", 2000 1
- J. Tovar-Pescador, D. Pozo-Vazquez, J. A. Ruiz-Arias, J. Batlles, G. Lopez & J. L. Bosch (2006). On the use of the digital elevation model to estimate the solar radiation in areas of complex topography. *Meteorol. Appl.* 13, 279–287.
- Kitchin, R. (2017). Thinking critically about and researching algorithms. *Information, Communication & Society* Vol. 20 , Iss. 1,2017
- Kolaczyk, E. D., & Csárdi, G. (2014). Statistical analysis of network data with R Springer.
- Lemmens M.J.P.M. 1999. Uncertainty in automatically sampled digital elevation models. In *Spatial Accuracy assessment...*
- Li, Z. 1994. A comparative study of the accuracy of digital terrain models based on various data models. *ISPRS Journal of Photogrammetry and Remote Sensing*. 49(1):2-11.
- Luke, D. A. (2015). A user's guide to network analysis in R.
- Maune, D.F. (2007). DEM user Applications. In *Digital Elevation Models Technologies and Applications: the DEM users manual*, 2nd Edition. Edited by Maune, D.F. ASPRS, Bethesda, USA.
- Mei-Po Kwan (2016) Algorithmic Geographies: Big Data, Algorithmic Uncertainty, and the Production of Geographic Knowledge, *Annals of the American Association of Geographers*, 106:2, 274-282
- Mesa-Mingorance, J.L., Chicaiza, E.G., Buenaño, X., Jianhong, C., Rodríguez-Pascual, A.F., Ariza-López, F.J. (2016). "Análisis de los usuarios y usos de los MDE en España", *GeoFocus (Artículos)*, nº17, p. 133-160. ISSN: 1578-5157
- Newman, Mark (2010). \Networks: an introduction". En: United States: Oxford University Press Inc., New York, p.ags. 1-2.
- OS(2013). OS TERRAIN 5 User guide and technical specification. Ordnance Survey, UK.
- OKSANEN, J. and SARJAKOSKI, T., 2004. Characterization of the digital elevation model (DEM) error by using airborne laser scanning elevation data. *GIScience 2004, The Third International Conference on Geographic Information Science: extended abstracts and poster summaries*, Oct 20-23, 2004, Adelphi, MD, pp. 165-167.
- Osiński, S., Stefanowski, J., & Weiss, D. (2004). Lingo: Search results clustering algorithm based on singular value decomposition. In *Intelligent information processing and web mining* (pp. 359-368). Springer Berlin Heidelberg.
- Raaflaub, L.D. and Collins, M.J., 2006, The effect of error in gridded digital elevation models on the estimation of topographic parameters. *Environmental Modelling & Software*, 21, pp. 710-732.
- R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Riggs, P.D. and Dean, D.J. 2007. An investigation into the causes of errors and inconsistencies in predicted viewsheds. *Transactions in GIS*. 11(2):175-196.
- Ruiz, M.O. 1997. A causal analysis of error in viewsheds from USGS digital elevation models. *Transactions in GIS*. 2(1):85-94.
- Smith, M.J. (2005). Use Case Compendium of Derived Geospatial Data. Project: Scoping a Geospatial Repository for Academic Deposit and Extraction. JISC project.
- SS (2006). SS 637008:2006 Geographic information - Surface water systems - concept and application schema.
- Tarquini, S., Nannipieri, L. (2017). The 10 m-resolution TINITALY DEM as a trans-disciplinary basis for the analysis of the Italian territory: Current trends and new perspectives. *Geomorphology* 281 (2017) 108-115.
- Tomaz Podobnikar (2009). Methods for visual quality assessment of a digital terrain model SAPIENS. Vol.2 / nº2 Special issue
- Wechsler S.P. (2003). Perceptions of Digital Elevation Model Uncertainty by DEM Users. *URISA* vol 15(2):57-64.
- Wechsler, SP (2007). \Uncertainties associated with digital elevation models for hydrologic applications: a review". En: *Hydrology and Earth System Sciences* 11.4, p.ags. 1481-1500.
- Weibel R, Heller M (1991). Digital Terrain Modelling. In Maguire, D.J., Goodchild, M.F. and Rhind, D.W. (eds.). *Geographical Information Systems: Principles and Applications*. London: Longman, 269-297.
- Wise, S.M. (2007). Effect of different DEM creation methods on the results from a hydrological model. *Comput. Geosciences*. 33. 1351-1365.

A.6 Script de la simulación de errores

```
---
title: "Río Garrapata"
output:
  pdf_document: default
---
```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE, results = FALSE, fig.path='Figures/')
```

Pasos previos
```{r librerias}
library(knitr)
library(rmarkdown)
library(rgdal)
library(raster)
library(sp)
library(ggplot2)
# library(plotly)
library(gstat)
library(dplyr)
```

Consideraciones

Los objetos han sido creados en base a la siguiente convención:
* Modelos digitales de elevación:
  + Dem Evaluado: DEM_eval
  + Dem de Referencia: DEM_ref
* Rasters:
  + Errores: errores
* Dataframes:
  + Dataframe propiamente dicho: _df
  + Spatial Pixels Dataframe: _spdf

Lectura de datos
Lectura de DEM del SIG-Tierras, que constituiría el DEM evaluado.
```{r import_DEM_eval}
DEM_eval <- raster("demc.tif")
```

```{r obj_proyecc}
UTMWGS84 <- DEM_eval@crs@projargs
```

```{r DEM_eval_spdf}
DEM_eval_spdf <- as(DEM_eval, "SpatialPixelsDataFrame")
```

```{r DEM_eval_df}
DEM_eval_df <- as.data.frame(DEM_eval_spdf) %>%
  rename(value = demc)
```

```{r plot_DEM_eval_df}
ggplot() + geom_raster(data = DEM_eval_df, aes(x = x, y = y, alpha = value))
```

```{r import_DEM_ref}
DEM_ref <- raster("demmr1.tif")
```

```{r sistem_referenc}
crs(DEM_ref) <- UTMWGS84
```

```{r import_clip}
```

```

clip_area <- readOGR(".", "area_simcorte")
...

```{r DEM_ref_spdf}
DEM_ref_spdf <- as(DEM_ref, "SpatialPixelsDataFrame")
DEM_ref_clip <- crop(DEM_ref, clip_area)
DEM_ref_clip_spdf <- as(DEM_ref_clip, "SpatialPixelsDataFrame")
...

```{r DEM_ref_df}
DEM_ref_df <- as.data.frame(DEM_ref_spdf) %>%
  rename(value = demmr1)
...

```{r plot_DEM_ref_df}
ggplot() + geom_raster(data = DEM_ref_df, aes(x = x, y = y, alpha = value))
...

```{r resampl_DEM_eval}
DEM_eval_r <- resample(DEM_eval, DEM_ref, method="bilinear")
...

```{r extens_espacial}
extent(DEM_ref)
extent(DEM_eval_r)
...

```{r errores}
errores <- (DEM_eval_r) - (DEM_ref)
plot(errores)
...

```{r mascara}
mascara <- readOGR(".", "masc_chone") #antes leia un archivo sin enmascarar (original)
crs(mascara) <- UTMWGS84
errores_mask_orig <- mask(errores, mascara)
errores_mask <- mask(errores, mascara)
errores_mask[errores_mask > 5 | errores_mask < (-5)] <- NA
...

```{r plot_errores_mask}
plot(errores_mask)
...

```{r corte_errores}
# Calculo de errores de partida
errores_partida <- crop(errores_mask, clip_area)
...

```{r muestreo_errores}
set.seed(12345)
errores_sample <- errores_mask
errores_sample[runif(729805) >= 0.005] <- NA # Randomly *unselect* 99.9% of the data
#plot(errores_sample)
...

```{r errores_df}
errores_sample_sp <- rasterToPoints(errores_sample, spatial = TRUE)
plot(errores_sample_sp, cex=0.01)
errores_df <- as.data.frame(errores_sample_sp) # errores_df se basa en errores_sample
...

```{r corte}
errores_sample_sp_clip <- errores_sample_sp[clip_area, ]
plot(errores_sample_sp_clip)
...

```

```
# Análisis estadístico
```

```
## Análisis gráfico
```

```
``{r distrib_datos}
cairo_pdf(filename = "graficas_explo_Chone.pdf", width = 9, height = 6)
par(mfrow=c(1,3))
histerrores_df$layer, main="", xlab="Error vertical (m)", ylab="Frecuencia")
qqnorm(errores_df$layer, main="", xlab="Cuantiles teóricos", ylab="Cuantiles muestrados")
boxplot(errores_df$layer, ylab="Error vertical (m)")
dev.off()
...

``{r test_normalidad}
library(nortest)
lillie.test(errores_df$layer)
#ks.test(errores$layer, "pnorm", mean(errores$layer), sd(errores$layer))
...

## Test de correlación espacial
``{r test_correlacion}
library(ape)
errores.dists <- as.matrix(dist(cbind(errores_df$x, errores_df$y)))
errores.dists.inv <- 1/errores.dists
diag(errores.dists.inv) <- 0
Moran_calculo<-Moran.I(errores_df$layer, errores.dists.inv)
Moran_calculo
...

## Revisión de tendencia espacial
``{r rev_tendencia_espacial}
library(geoR)
cairo_pdf(filename = "analisis_tendencia_Chone.pdf")
errores_geodata<-as.geodata(errores_df,coords.col=2:3, data.col=1)
plot(errores_geodata)
dev.off()
...

``{r vario_exper}
variogr<-variogram(layer ~ 1, errores_sample_sp)
cairo_pdf(filename = "variogr_exper_errores_Chone.pdf")
plot(variogr)
dev.off()
...

``{r vario_exper_nube}
variogr_exp_nube<-variogram(layer ~ 1, errores_sample_sp, cloud=TRUE)
cairo_pdf(filename = "nube_variogr_exper_errores_Chone.pdf")
plot(variogr_exp_nube)
dev.off()
...

``{r vario_exper_mapa}
variogr_exp_mapa<-variogram(layer ~ 1, errores_sample_sp, map=TRUE, cutoff=1000, width=50)
cairo_pdf(filename = "mapa_variogr_exper_errores_Chone.pdf")
plot(variogr_exp_mapa)
dev.off()
...

``{r vario_direcc}
variogr_exp_dir<-variog4(errores_geodata)
vmf <- fit.variogram(variogr, vgm(0.332, "Exp", 147.638, add.to=vgm(0.196, "Sph", 83.975, add.to=vgm(0.040, "Gau", 366.785))))
```

```

kable(vmf)
...

```{r variograma_direcc}
cairo_pdf(filename = "variogr_exper_direccionales_Chone.pdf")
plot(variogr_exp_dir)
dev.off()
...

```{r variograma_anidado}
print(plot(variogr, pl = F, model = vmf))
...

```{r RGeostats}
library(RGeostats)
erroresdb<-db.create(x1=errores_df$x, x2=errores_df$y, z1=errores_df$layer)
variog_exp<-vario.calc(erroresdb)
...

```{r erroresdb}
plot(erroresdb)
...

```{r varioexp}
plot(variog_exp)
...

```{r vario_ajuste}
Z.svar <- variog(errores_geodata, max.dist=2000, messages=FALSE)
# WLS fitting:
Z.vgm <- variofit(Z.svar, ini=c(0.565, 358.703), cov.model="spherical", fix.nugget=F)
env.model <- variog.model.env(errores_geodata, obj.var=Z.svar, model=Z.vgm)
Z.vgml <- likfit(errores_geodata, ini=c(0.565, 358.703), fix.nug=TRUE, lik.met = "REML")
...

```{r variogr_plot}
plot(Z.svar, envelope=env.model); lines(Z.vgm, lwd=2)
...

```{r modelaje}
model1 <- model.auto(variog_exp, struct=melem.name(c(2,3,4)))
print(model1)
...

# Simulación geoestadística

```{r grilla}
# Se usará las coordenadas del DEM evaluado
...

```{r simulacion}
N.sim <- 100
DEM.sim <- krige(layer ~ 1, errores_sample_sp_clip, DEM_ref_clip_spdf, vmf, nmax=30, nsim=N.sim)
...

```{r simulacion_plot}
cairo_pdf(filename = "simulaciones_Chone.pdf")
spplot (DEM.sim[c("sim1", "sim3", "sim8", "sim15")],
        col.regions = bpy.colors(), main="")
dev.off()
...

```{r seccion_transv_def}
# Cross-section at y=9928514:
gridcell <- DEM_eval_spdf@grid@cellsize[1]
cross.s <- data.frame(X=seq(DEM_eval_spdf@bbox[1,1]+gridcell/2,DEM_eval_spdf@bbox[1,2]-gridcell/2,
gridcell), Y=rep(9928514, DEM_eval_spdf@grid@cells.dim[1]))
coordinates(cross.s) <- ~ X+Y
proj4string(cross.s) <- DEM_eval_spdf@proj4string

```

```
cross.ov <- aggregate(DEM.sim, cross.s, mean)
...

```{r seccion_transv_plot}
cairo_pdf(filename = "seccion_longitudinal_Chone.pdf")
plot(cross.ov@coords[,1], cross.ov@data[[1]], type="l", xlab="X", ylab="Z", col="grey")
for(i in 2:N.sim-1){
  lines(cross.ov@coords[,1], cross.ov@data[[i]], col="grey")}
lines(cross.ov@coords[,1], cross.ov@data[[N.sim]], lwd=2)
dev.off()
...

```{r probabilidad}
MC <- N.sim
errores_new <- numeric()
for (i in 1:MC)
  errores_new <- cbind(errores_new, DEM.sim[[i]])

# Media de los errores calculadas a partir de las simulaciones
DEM.sim$media <- apply(errores_new, 1, mean)

# Cálculo de error medio de simulacion vs errores de partida
errores_partida_df<-as.data.frame(errores_partida)
comparac_errores<-DEM.sim$media - errores_partida_df$layer
hist(comparac_errores)

# Cálculo de probabilidades tomando como referencia un valor critico y positivos y negativos
DEM.sim$prob_1_5 <- apply(ifelse((errores_new > 1.5 | errores_new < -1.5), 1, 0), 1, mean)

DEM.sim$prob_mayor_1_5 <- apply(ifelse((errores_new > 1.5), 1, 0), 1, mean)
DEM.sim$prob_menor_1_5 <- apply(ifelse((errores_new < -1.5), 1, 0), 1, mean)
DEM.sim$prob_2 <- apply(ifelse((errores_new >= 2 | errores_new <= -2), 1, 0), 1, mean)
DEM.sim$prob_mayor_2 <- apply(ifelse((errores_new >= 2), 1, 0), 1, mean)
DEM.sim$prob_menor_2 <- apply(ifelse((errores_new <= -2), 1, 0), 1, mean)
DEM.sim$prob_3 <- apply(ifelse((errores_new >= 3 | errores_new <= -3), 1, 0), 1, mean)
DEM.sim$prob_mayor_3 <- apply(ifelse((errores_new >= 3), 1, 0), 1, mean)
DEM.sim$prob_menor_3 <- apply(ifelse((errores_new <= -3), 1, 0), 1, mean)
spplot (DEM.sim["prob_1_5"], col.regions = bpy.colors(),
        main="Probabilidad de errores mayores a +- 1.5 metros")

# Conversión de mapa de probabilidades a formato raster

DEM.sim_media<-raster(DEM.sim, layer=101, values=TRUE)
DEM.sim_prob_1_5<-raster(DEM.sim, layer=102, values=TRUE)
DEM.sim_prob_mayor_1_5<-raster(DEM.sim, layer=103, values=TRUE)
DEM.sim_prob_menor_1_5<-raster(DEM.sim, layer=104, values=TRUE)
DEM.sim_prob_2<-raster(DEM.sim, layer=105, values=TRUE)
DEM.sim_prob_mayor_2<-raster(DEM.sim, layer=106, values=TRUE)
DEM.sim_prob_menor_2<-raster(DEM.sim, layer=107, values=TRUE)
DEM.sim_prob_3<-raster(DEM.sim, layer=108, values=TRUE)
DEM.sim_prob_mayor_3<-raster(DEM.sim, layer=109, values=TRUE)
DEM.sim_prob_menor_3<-raster(DEM.sim, layer=110, values=TRUE)

# Exportación de raster a formato TIF para lectura en ARCGIS
writeRaster(DEM.sim_media, filename = "DEMsim_media.tif")
writeRaster(DEM.sim_prob_1_5, filename = "DEMsim_prob1_5.tif")
writeRaster(DEM.sim_prob_mayor_1_5, filename = "DEMsim_prob_mayor_1_5.tif")
writeRaster(DEM.sim_prob_menor_1_5, filename = "DEMsim_prob_menor_1_5.tif")
writeRaster(DEM.sim_prob_2, filename = "DEMsim_prob2.tif")
```

```
writeRaster(DEM.sim_prob_menor_2, filename = "DEMsimsim_prob_menor_2.tif")
writeRaster(DEM.sim_prob_mayor_2, filename = "DEMsimsim_prob_mayor_2.tif")
writeRaster(DEM.sim_prob_3, filename = "DEMsimsim_prob3.tif")
writeRaster(DEM.sim_prob_mayor_3, filename = "DEMsimsim_prob_mayor_3.tif")
writeRaster(DEM.sim_prob_menor_3, filename = "DEMsimsim_prob_menor_3.tif")
spplot (DEM.sim["prob_2"], col.regions = bpy.colors(),
        main="Probabilidad de errores mayores a 2 metros")
spplot (DEM.sim["prob_3"], col.regions = bpy.colors(),
        main="Probabilidad de errores mayores a 3 metros")
...
```

A.7 Informe de procesamiento del MDE Condado - Quito

Agisoft PhotoScan

Informe de procesamiento

Datos del levantamiento

Fig. 1. Posiciones de cámaras y solapamiento de imágenes.

Número de imágenes: 135

Posiciones de cámara: 135

Altitud media de vuelo: 87.1 m

Puntos de enlace: 65,818

Resolución en terreno: 3.27 cm/pix

Proyecciones: 394,951

Superficie cubierta: 1.91e+05 m²

Error de reproyección: 0.674 pix

Modelo de cámara	Resolución	Distancia focal	Tamaño de píxel	Precalibrada
FC330 (3.61 mm)	4000 x 3000	3.61 mm	1.56 x 1.56 micras	No

Tabla 1. Cámaras utilizadas.

Fig. 2. Gráfico de residuales para FC330 (3.61 mm).

FC330 (3.61 mm)

135 imágenes

Resolución 4000 x 3000	Distancia focal 3.61 mm	Tamaño de píxel 1.56 x 1.56 micras	Precalibrada No
Tipo:	Cuadro	Skew:	-0.373405
Fx:	2338.08	Cx:	2002.32
Fy:	2344.94	Cy:	1475.11
K1:	0.000823758	P1:	-0.000209182
K2:	-0.00882666	P2:	0.000105916
K3:	0.00359771	P3:	0
K4:	0	P4:	0

Puntos de control terrestres

Fig. 3. Posiciones de puntos de apoyo.

Nombre	XY error (m)	Error en Z(m)	Error (m)	Proyecciones	Error (pix)
2	0.0324915	-0.00317421	0.0326461	14	0.411
4	0.0232668	0.00154779	0.0233182	9	0.548
5	0.0187187	-0.00223141	0.0188512	16	0.724
7	0.0352328	0.00449594	0.0355185	13	0.556
RMS	0.0282346	0.00306855	0.0284009		0.580

Tabla 2. Puntos de apoyo.

Modelo digital de elevaciones

Fig. 4. Modelo digital de elevaciones.

Resolución: 13.1 cm/pix

Densidad de puntos: 58.578 puntos por metro cuadrado

Parámetros de procesamiento

Generales

Cámaras	135
Cámaras orientadas	135
Marcadores	4
Sistema de coordenadas	WGS 84 / UTMzone 17S (EPSG::32717)

Nube de puntos

Puntos	65,818 de 93,614
RMS error de reproyección	0.243516 (0.674126 pix)
Error de reproyección máximo	2.98344 (26.1415 pix)
Tamaño promedio de puntos característicos	2.58241 pix
Superposición efectiva	6.26646

Parámetros de orientación

Precisión	Alta
Pre-procesar emparejamiento de imágenes	Referencia
Puntos claves por foto	40,000
Puntos de enlace por foto	4,000
Restricción de máscara activa	No
Tiempo búsqueda de puntos homólogos	12 minutos 16 segundos
Tiempo de orientación	2 minutos 34 segundos

Parámetros de optimización

Parámetros	fx, fy, cx, cy, oblicuidad, k1-k3, p1, p2
Tiempo de optimización	2 segundos

Nube de puntos densa

Puntos	11,872,560

Parámetros de reconstrucción

Calidad	Media
Filtrado de profundidad	Agresivo
Duración del procesamiento	2 minutos 18 segundos

Modelo

Caras	770,088
Vértices	386,773

Parámetros de reconstrucción

Tipo de superficie	Bajorrelieve / terreno
Datos fuente	Densa
Interpolación	Habilitada
Calidad	Media
Filtrado de profundidad	Agresivo
Número de caras	791,503
Duración del procesamiento	34 segundos

DEM

Tamaño	5,740 x 5,931
Sistema de coordenadas	WGS 84 / UTMzone 17S (EPSG::32717)

Parámetros de reconstrucción

Datos fuente	Nube de puntos densa
Interpolación	Habilitada

Ortomosaico

Tamaño	17,169 x 17,696
Sistema de coordenadas	WGS 84 / UTMzone 17S (EPSG::32717)
Canales	3, uint8
Modo de mezcla	Mosaico

Parámetros de reconstrucción

Superficie	Malla
Permitir la corrección de color	No