International Journal of Spatial Data Infrastructures Research, 2018, Vol.13, 131-171
Special Section: INSPIRE (Full Research Article)

DEMs: An Approach to Users and Uses from the Quality Perspective*

Francisco Javier Ariza-López[1], Elena Gabriela Chicaiza Mora[2], José Luis Mesa Mingorance[1], Jianhong Cai[3], Juan Francisco Reinoso Gordo[4]

[1] University of Jaén, Spain, {fjariza, jlmesa}@ujaen.es

[2] Pan American Institute of Geography and History (IPGH), Ecuador, gachisbar@gmail.com

[3] Beijing University of Civil Engineering and Architecture, China, cjh@bucea.edu.cn

[4] University of Granda, Spain, jreinoso@ugr.es

Abstract

The importance of Digital Elevation Models (DEMs) is great in geosciences, but a general view of users and uses which would bring the concept of quality closer to users is lacking. For this reason, the aim of this paper was three-fold: to obtain better knowledge of users, determine the main application domains of DEMs, and identify main use cases. For this purpose, we used data from two web questionnaires (MR1 and MR2), a search of the ScienceDirect database (MR3) and a Google search (MR4). The data coming from the MR1 resource have offered us a large number of cases in order to characterise the profile of users in Spain. The MR2 resource is an *ad hoc* designed survey which has allowed us, among other things, to identify those calculations that are more normal; determine that subjective evaluation of quality is of great importance for users; and conclude that there is a high percentage of users who do not use any quality index, and also that the majority of users do not know how to evaluate the influence of poor quality on their work. Through MR3, it was possible to analyse the relationships between relevant items and carry out a semantic analysis of a set of 950 abstracts. From MR4, it can be concluded that the formalisation of

DOI: 10.2902/1725-0463.2018.13.art12

applications as use cases is not normal. This paper identifies some research lines in order to offer users a better understanding of the issue of quality regarding DEMs.

Keywords: DEM, User, Uses, Fitness for use, Quality evaluation

1. INTRODUCTION

A Digital Elevation Model (DEM) is a generic term for digital topographic and/or bathymetric data and normally implies elevations of the terrain (bare earth) void of vegetation and man-made features (Maune et al., 2007). DEMs describe the morphological characteristics of the landforms of a terrain (topography) such as hills, ridges, cliffs, valleys, rivers, etc., and a large number of parameters can be derived (e.g., slope, aspect, curvature, etc.). DEM data can be generated by means of many diverse technologies (Maune, 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 as contour lines, triangular irregular networks, gridded surfaces, parametric surfaces, etc.

Information is of no value if it is not used or used properly. That highlights 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, the main uses (application domains). In a pioneering way, Weibel and Heller (1991) identified some application domains, and, after them many other 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), provide 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 elements 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 Geographical Information Systems and the growth of their use. Also, the existence of global digital elevation model (GDEMs) products has promoted a large number of global-scale applications. So currently, 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. Quality of spatial data is defined by the International Standard ISO 19157 (ISO 2013) with respect to 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 as topography (Oksanen and Sarjakoski, 2005; 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, 2013); glaciology (Racoviteanu et al., 2007), precision farming (Yao and Clark, 2000), solar irradiation (Tovar-Pescador, 2006), and so on.

The current way of dealing with the issue of 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 applications domains 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 (internal quality), and that more information needs to be included in order to achieve the purpose of 'fitness for use'. In this way, Boin and Hunter (2006, 2009) ask: "Do 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 for 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 studies which 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 the access and application of uncertainty analysis to DEM's users. Recently, Mesa-Mingorance et al. (2016) have analysed the 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) summarise four (4) years of free dissemination and use of a 10x10 metre DEM covering Italy entirely, and 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 studies (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 literature regarding practical cases is scarce. For this reason, Agumya and Hunter (1999) developed an example concerning the selection of a DEM for flood-extension estimation by means of a risk-based strategy where exposure to risks due to error in the data is a key element. So we can conclude that fitness for use is linked to use cases, and use cases are linked to application domains; thus if we want to advance the concept of "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 implemented until now. For this reason, the aim of this paper is three fold: obtain better knowledge of users, determine the main application domains of DEMs, and identify the main use cases.

This paper is organised as follows: First, the four different resources (MR1, MR2, MR3 and MR4) we have used for the recompilation of relevant data about users and uses are presented; second, all the results corresponding to each source, and specific methodological issues, are presented and discussed; and finally main conclusions are established.

2. MATERIAL, METHODS AND RESOURCES

In this study, we have three main aims (users, domains of use and use cases), and we have used several methods and available resources in order to obtain a better definition of all of them. We have used two surveys in order to detect user profiles: domains of application and specific requirements about the quality of DEMs. Also, we have used queries to a scientific database in order to obtain a better insight into domains, relations of domains and a semantic approach to domains. Finally, we have used internet queries, through Google, to obtain formal examples of use cases regarding DEMs. Table 1 summarises this approach.

In order to obtain information about users and uses of DEMs, our first choice was to search for existing web questionnaires linked to existing Spatial Data Infrastructures at a national level, and ask about registered data. During 2016, we located and analysed the websites of the Spatial Data Infrastructures of many countries (Argentina, Bolivia, Brazil, Canada, Chile, China, Colombia, Costa Rica, Cuba, Ecuador, France, Germany, Guatemala, Honduras, Italy, México, Nicaragua, Panamá, Paraguay, Perú the Dominican Republic, Spain, UK, Uruguay, USA, Venezuela), and we found only two countries with questionnaires for 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, where there is a questionnaire linked to each download.

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 ™ search		X	X	X	
MR4. Google ™ search					X

So the first resource we were able to use was data coming from the download center (http://centrodedescargas.cnig.es/CentroDescargas/) of the National Geographic Institute (IGN) of Spain (IGN, 2018). For each single free download, the users must fill in a questionnaire (Table 2). Of course, the data and the analysis derived thereof are strictly only valid for the Spanish case, but we think that this can be a starting point. This source will show results for a national level in a developed country, and we consider that these results can, in some way, be generalised to other countries of similar characteristics. A strong point of this source is the large number of survey responses available. 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 regarding application domains, and the scope. Some of these questions can be answered by users by checking only one or more options and the analysis is straight-forward. Other questions are answered with free text and the analysis is more complex. In this study, we only used the results of the first type of questions, specifically the questions numbered 1, 2, 4, 6, 7 and 9 in Table 2.

The IGN supplied us with 12,493 records coming from downloads of DEMs that took place during the entire year of 2015. Approximately 100 percent (99.99%) of all cases were downloads that took place in Spain, and there were 6,087 single users. These downloads were related to 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. Identification of the user profile
2. Opinion about the utility of the data
3. Observations regarding utility
4. Opinion about the need to update
5. Comments on the need to update
6. Location (province)
7. Country
8. Description of the intended use
9. Identification of the application domain
10. Comments and observations to application domains
11. Identification of the application spatial scope

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 designed a questionnaire and conducted a worldwide survey. Table 3 presents a summary of the topics of the questionnaire. This questionnaire (MR2) includes 18 questions that can be grouped as follows:

- Information about the user. In this section (Questions 1-3), we ask them about their occupation and experience because these are important issues in defining of users´ profiles.
- Information on the use of DEMs. In this section (Questions 4-9), we wish to gain insight into the use cases and kinds of applications, processes and data that users are developing and using.
- Information regarding relevant quality aspects. In this section (Questions 10-18), we ask users about quality aspects (exigencies, use of quality indexes, degree of satisfaction with those quality indexes, etc.).

In order to obtain an international view, we used a web form platform. We selected Cognito Forms™, a free online-form builder which allows the use of several kinds of answers and uploading of files. In order to obtain a widespread contribution to 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. The origin of the answers was 60.8% from Europe, 25.5% from America and 13.7% from Asia.

Table 3 – Web questionnaire #2

1. Identification of the occupation
2. Name of the institution
3. Years of experience using DEMs
4. Identification of the purposes of the use of DEMs
5. Identification of common calculations and the derivatives from DEMs
6. Upload a document explaining the use
7. A description of the process(es) in the use case(s)
8. Identification of the data model
9. Identification of the sources, scale/resolution and accuracy
10. Identification of the main purpose of the application
11. Identification of the expression of accuracy in the use case(s)
12. Identification of the use of one or several indices to express quality
13. Satisfaction with the current accuracy indices being used
14. Suggestions to improve when expressing quality
15. Suggestions if no index is used to express the quality
16. Explanation of the influence of bad quality on the results
17. Score the influence of the quality requirements of DEM on the results
18. Monetize the consequences of using DEMs of bad quality

Another important source has been ScienceDirect™. ScienceDirect is a data base which hosts over 2500 journals and more than 33,000 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) it offers a very large database of contents that we think can be considered sufficiently representative, b) it 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) it offers as a result a list of topics that can be considered related to use cases. We queried this database using journals and books from 2006 until 2016 ([2006, 2016]) by means of several expert searches of 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 physics, 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 implemented:

- 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"

As a result of such searches, we obtain the amount of documents that fulfill the query (theme query) and a list of topics with the frequency of appearance. The topics are the keywords extracted from the papers that ScienceDirect returns to the queries. The topics returned by ScienceDirect are those keywords provided by the authors of the papers. 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".

All the abstracts of the resulting papers were revised one-by-one in order to eliminate improper cases. Afterwards, all abstracts were clustered by means of a text clustering tool in order to establish groups sharing same semantic.

Finally, the last resource we used was to query the internet by means of Google in order to obtain use case examples presented in a formal way. Here "formal" means the existence of a certain structure in the contents. Here, we used impersonal searches (www.impersonal.me) so as not to be conditioned by our user profile. In this case, we used similar queries to those presented above – but including the term "use case". The result of these queries is a set of documents that we have analysed manually in order to localise the proper documents.

3. 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. Tables present aggrupation of the cases using grey scale levels. The groups presented in the tables by means of grey levels are derived from a cluster analysis of cases. The cluster is based on the Euclidean distance between classes using the SPSS hierarchical cluster analysis.

3.1. 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 groups can be identified. The most frequent profile is that of private users (individual users) (70.74%), where freelances are included. This is the first group and is very far from the others cases. The second, more frequent case corresponds to a university profile (17.71%), where researchers, teachers and students are included. Each one of the other profiles accounts for less than 5% of frequency and all of them are included in the third group. Not surprisingly, given their greater number, the smaller companies (micro-size enterprises) have a higher percentage of downloads than the larger ones. All the authorities together (National, Regional and Local) sum up 3.99% of the downloads, which clearly indicates that the main use (> 95%) is in the private sector.

The MR1 provides a list of application domains, but we must remember that this enables multiple selections, meaning that a user can mark one or more application domains. There were 72.8% of the users who 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. Nearly 95% of users marked *3*, *2* or only *1* application domain.

Table 4 – MR1: User profiles in Spain

User profile	%
Private 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 = 6,087	

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 others very specific; and, because it allows multiple answers, it can be considered that users can express their application domain in an accurate manner. General terms like Cartography and Environment are clearly the dominant options, accounting for more than 38.35% together. A second group of more specific activities are those whose percentage is in the interval [2.86; 6.23], accounting for 46.43% altogether. Here Education & Training and Forest & Biodiversity are the two cases with the largest values. It is interesting to note that Leisure and Recreational applications comprise an important application domain, and that this domain can be considered outside the initial professional purposes of DEM use. This is a very good index of the spread of the 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 contains 18.17% of responses. It is a long tile of application 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 analyses (see Tarquini and Nannipieri, 2017), it is possible to observe a great level of coincidence in the domains and the 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 = 19,728 (multiple responses are possible)	

The MR1 asked users about usefulness (perceived utility) and the need for updating. 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 that this means a high level of satisfaction with the products. In relation to the need for updating, greater variation is observed in the results. Here there is no such unanimous agreement on the need of updating.

Table 6 – MR1: Assessment of perceived utility and updatedness needs in Spain

Utility assessment		Assessment of updating need	
Utility level	%	Updating 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=6,087			

3.2. 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 with information about users of DEMs, the use of DEMs, general features of the DEMs being used, and finally information about relevant quality aspects of DEMs. Following this order, here the more important findings are presented.

3.2.1. Users

The information about the users was queried in Questions 1 to 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, with 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 note that profiles linked to teaching, studying and research add up, altogether, to 63.8%, which means that the majority of DEM users belong to a very speculative sector. In relation to years of experience (Q3), the conclusion is that the respondents have, for the most part, a wide experience in the use of the DEM, since they have more than four years of experience. In relation to the organisation profile (Q2), we can distinguish three groups: The first group includes only the category University, which is the predominant organisation profile in very close to half of the cases. The second group also includes only one category, Mapping Agencies, which is very close to a quarter of the cases. Finally, a third group can be established including the remaining categories: Research institutions, Environmental Agencies and Companies. This result is, somewhat coincidental with the result of Q1.

Table 7 – MR2: Assessment of perceived utility and updating needs in Spain

User profile (Q1)	%	Organisation 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's, 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 organisation profile from Table 6 can be analysed in relation to the 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 purpose 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 two different views of DEM users. 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 DEMs in the world and their use by the more active and innovative users of DEMs. We think that both views are of great interest for understanding the users and uses of DEMS.

3.2.2. Use and general features of DEMs

The information about the use and general features of DEMs was queried in Questions 4 to 9 of MR2. Table 8 presents the summary in relation to the purpose of use. In the first category (accounting for 52.8%), two general purposes of use are included (Water and Environmental studies) and an item related to DEMs´ data capture techniques, which means a great relation with DEM generation. The second category (accounting for 27.6%) includes a very important field of application (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 DEM generation in the first two categories, accounting for 26.93% together, and this means that users are very close to DEM data generation and refinement. The last category includes a long list 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 for the largest percentage (75.5%) is the use of the elevation data (height). The calculation of the slope, another typical general analysis, is close to the previous one, but accounting for 71.6%. If we consider the third case (aspect), three direct topographic attributes are the three most common calculations. In the second category, we can see many calculations related to operations resulting in features (e.g., contours, view-shed, catchment and channels) and two other cases (aspect and flow accumulation) related to fields. The third category includes calculations related to surface (curvature, concavity and 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 most 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 explanatory document. In this case, 38.2% of respondents (N=102) upload a file. The majority (74.35%) of the documents are scientific papers (journal and congress papers), and the remaining a mix of different types (reports, presentations, and so on). Scientific papers were published in journals such as: Computers and 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 occur in their use case, but the answers received are not focused on the aim of the question itself. 20% of responses were void and the majority of the remaining were generic answers (e.g., DEM generation, DEM production, extraction of elevation and certain topographic features, etc.), with no proper information about process steps. For this reason, we considered that the results of this question are not valid.

The results in relation to the data model are presented in Table 10. The grid model (91.2%) is the predominant option; this result coincides with other analyses and is coherent because the majority of DEM datasets are offered as grid datasets. The TIN and contour models are in the same range order. The use of profiles is detected and accounts for 12.7%. This model is basically used for civil engineering applications and hydrological modelling in channels (e.g., HEC-

RAS). Others are in a class that includes other answers such as ASCII, Lidar Data, DEM, and similar answers; but these are 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 shows the relationship between data models being used by respondents. One conclusion of this table is that users use several data models but have a single preferred data model (e.g., Grid, TIN, Contours, Profiles), and this data model is used by this user more than the other options. This is very clear for the Grid case, but also for the other three cases.

Table 10 – MR2: Use of data models (Q8)

Data model	%
Grid	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)

%	Grid	TIN	Contours	Profiles
Grid	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 added 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 the 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 latter 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 broken down into one account for each element of the range. The percentage has been calculated for the total number of answers for each case (e.g., resolution, scale, etc.). It is interesting to point out that resolution is the preferred term for expressing this data property and accounts for 141 responses. Scale and accuracy are the least-used terms. In the case of resolutions, it is very clear that the 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 products available from mapping agencies and Earth observation systems. The responses using the term accuracy show only three levels (mm, cm, m), and the centimetres case accounts for the largest percentage (54.55%). This means a bias to greater 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 1.41m and 0.7m, which is related to the metric options of resolution and accuracy.

Table 13 – MR2: Resolution, Scale, Accuracy and Density (Q9)

Resolution (1)	**%**
mm	4.69
cm	9.38
m	36.72
x10m	42.97
x100m	3.13
x1000m	3.13

Scale (2)	**%**
E0,5K	3.57
E1k	7.14
E2K	10.71
E5k	14.29
E10k	10.71
E25K	17.86
E50K	17.86
E100k	7.14
E250K	7.14
E1000K	3.57

Accuracy (3)	**%**
mm	18.18
cm	54.55
m	27.27

Density (4)	**%**
2 p/m2	27.27
1 p/m2	27.27
0,5 p/m2	45.45

% for each option. Number of respondents = 102, number of responses = 141(1); 28(2); 22(3); 11(4)

3.2.3. Information about relevant quality aspects

Several aspects related to 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 to 18.

It is interesting to obtain 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 very common calculations were included (slope and watercourse and watershed delineation). The third class also includes 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 close relation to Q5, and for this reason results must be somewhat similar. For instance, if we join 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 expressing accuracy (Table 15). The option with the larger percentage (47.1%) corresponds to the use of a statistical approach. It is interesting to note here that the second option, with 26.5%, corresponds to subjective judgments (e.g., visual inspection). We do not know if this is due to ignorance or dissatisfaction with the most objective methods. The use of a 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. Depending on the case, the option includes cases that can be considered in any of the previous ones. It is interesting to note 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. Concerning the latter, 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 understanding the actual needs of users. Q12 is oriented towards obtaining some insight into this, and the results are shown in Table 16. First of all, it is relevant that 20.59% of respondents indicate that they do not use any indices; also 8.82% of cases give no answer, so 37.25% of responses indicate 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 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 indicates 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 toward achieving a better characterisation of quality, and closer to "fitness for use". So, we conclude that 42.16% of responses indicate some trend of using indices with a better description of quality than traditional measures based only on 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 with the accuracy indexes was addressed by Q13 (Table 17). Only a few cases (6.9%) indicate that respondents are really quite satisfied, with almost the majority of the respondents (49%) indicating that they are satisfied. On the other hand, 18.6% are dissatisfied and 6.9% very dissatisfied, so 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. It is also remarkable that almost 10% of the users give no answer to this simple question. This percentage is high and very close to that indicated by Q11 (related to the expression of accuracy), where the conclusion was that there were a lot of users of DEMs who have no concern about the quality of the data. So both questions, Q11 and Q13, would appear to demonstrate this.

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 with the actual information of quality is querying about possible improvements. Q14 moves in this direction (results in 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. Approximately 3.92% of answers indicated no improvement was needed. Also, 2.94% of respondents indicated that they have no criterion for answering this question. In relation to those answers directly related to this question, it is interesting to note that the majority of respondents demand more information, not better measures or indices. The demand for better information about altimetric (Z) and planimetric (XY) accuracies are the two answers with the largest percentages, both together reaching 30.39%. Approximately, 9.8% of respondents want to know what the spatial distribution of DEMs errors is at each point of the DEM grid, which is a very reasonable and useful demand for many purposes. 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 indicated by 11.76% (=6.86% +4.9%) of responses, but if we consider here other linage related issues (e.g., information about ground surveys, etc.), the lineage will be the topic with the largest 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 for 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 through the proposed measures, and metaquality elements, is proposed by the same percentage (1.96%). Finally, there is a large number of responses, each one accounting for only one response-case and which we have summarised in the case 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 that metadata are very complex because they have to be accessed and understood. This conclusion is derived because the majority of public spatial datasets of developed countries have metadata (e.g., following ISO 19115 (ISO 2014)), 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 obtain some suggestions from users who did not use any index. This question was not automatically linked with the response for Q11 and Q12, so responses could come from all respondents. In this way, we performed a cross tabulation of responses to this question with those of previous questions and found that respondents not included as initial targets gave a response. Finally, we found 11 different responses (Table 19) coming from all kinds of respondents. All answers are unique except for the seventh answer (drainage network), proposed by two respondents. Some of these proposals coincide with the results of Q14, which is logical. Nevertheless, we think that there are very interesting proposals in these suggestions for exploring 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 to the influence of bad quality on 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%) were not directly related to this question. The answers indicating estimation problems, no matter what parameter (e.g., slope, orientation, volume, etc.) are those that had the highest percentage (24.51%). 22.55% of answers indicated problems caused by data (e.g., rework, conflation, location, etc.). 15.68 % of answers indicated the influence of bad quality over decisions, costs and hazards. Finally, 5.88% of responses indicated that poor quality is not relevant to results. The situation described is very surprising to us. This implies that a high percentage of users do not understand the quality-result relationship in their use cases.

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 in 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 who think that the influence is low (7.8%) and even none (2%), also there are 2.9% of users who were not able to give an answer to this question. It is interesting to compare these results with those presented for Q16. For instance, for the Q17 case 2% indicated no influence, but 5.88% indicated no relevance in the case of Q16. Also, for Q16 there were 30.39% of no answer responses but here for the Q17 case there were only 2.9% of no answer responses. We conclude that people are not able to offer an explanation of the influence of bad quality (Q16), but they have the feeling that bad quality has relevance (Q17).

From an economic perspective, Q18 addressed the same issue as Q17. The majority of the answers (45.63%) gave the “no answer” option, and this is the “no answer” case with the largest 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 with 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 indicated that the use of DEMs of bad quality has no economic consequence. For us this answer is very contradictory with respect to Q17. We do not see a direct relation between this result and those of Q17; the only justification we find is that "influence" in Q17 was understood as influence on results, and that results of bad quality have no economic consequence for these users. A similar percentage, 23.30% of respondents, gave no quantification but wrote an explanation of consequences. This means the acknowledgement of the influence of bad quality but also the incapacity to translate bad quality problems into money. Only 7% of respondents were able to monetise 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). Finally, other important issue to take into account is that many respondents are related to scientific applications, where it is very often impossible to evaluate economic impacts.

Table 21 – MR2: Valuation of the influence of bad quality (Q17 & Q18)

Score of influence (Q17) (1)	**%**
High	56.9
Medium	30.4
Low	7.8
None	2.0
No answer	2.9

Monetization of consequences (Q18) (2)	**%**
No answer	45.63%
None	23.30%
No quantification is possible (Explanation of consequences).	23.30%
x1.000.000 USD	2.91%
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

3.3. Application domains: identification, relations and semantics 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.

Identification is a redundant task 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 out by users; the searches in ScienceDirect have a bias to scientific and technical applications, but the spectrum of possibilities is also 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 best short description of a task or application. Abstracts are a key element in the peer review system of scientific journals, so they must be clearly explanatory of the uses/applications and achievements.

3.3.1. 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 7,856 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 reorganised 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 note here that quality terms account for 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

Domains	%
Slope	11.85
Soil	7.73
River	6.61
Drainage	5.47
Forest	4.65
Flow	4.56
Water	4.06
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

Tools and data	%
Remote sensing	6.58
Lidar	3.72
Landsat	3.13
GPS	2.06
Aster	1.9
SAR	1.69
GDEM	0.48

Quality	%
Accuracy	5.82
Quality	3.03

% of total cases. N= 7,856

3.3.2. Relational analysis

As mentioned 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 a weight less than 40 were eliminated. In the same way, the number of cases of topics citations was used as the weight for the topics, and topics with weights lower than 15 were deleted in order to show only the most relevant. As result of the explained filtering we obtained 34 themes, 39 topics and 391 relations between them.

For the relational analysis between themes and topics we prepared a bipartite graph with two disjoint sets: the themes and the topics. Table 23 presents the statistical analysis of the three elements (theme nodes, topic nodes and Theme-Topic relations) of the graph. Due to the manner in which the consultation was carried out, the sum of the weights of themes is larger than the sum of topic weights and the weight of topics is equal to the weight of relations because the latter is derived from the former. The mean number of relations from themes and topics is very similar (≈10-11.5). This value is the mean degree of the graph vertex (nodes). But there is a 4:1 relation when analysing the mode value for the same. Also the deviation of the topic's relations is 1.5 times greater than the deviation of the theme's relations. We consider these differences between themes and topics as an indication that the themes have been well chosen. The mean of the relations' weights is 7.7 which indicates that a topic appears, in mean terms, 7.7 times related to a theme.

Table 23 – Statistical analysis of relations between Themes and Topics

		N	Sum	Min	Max	Mean	Median	Mode	Deviation
Themes	Weights	34	7,856	38	931	231.1	156	50	197.7
	Relations from Themes	391	391	1	17	11.5	12	16	4.3
Topics	Weights	39	2,998	15	463	76.9	44	18	86.2
	Relations from Topics	391	391	2	29	10.0	8	4	6.6
Relations	Weights	391	2,998	1	70	7.7	5	3	7.3

3.3.3. Semantic analysis: words and clustering 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 richer way. In this case nine hundred and fifty abstracts were used, manually revised and coming from all the queries. This analysis was performed in two directions: by a) a word-frequency analysis and b) a text clustering.

In relation to the word-frequency analysis of the wording of the set of abstracts, there were more than 245,000 words and more than 13,000 different words in this set. After the elimination of stop words, we finally obtained a set of 74,009 words (2,010 different words).

Table 24 shows a list of 64 word items with the terms in which absolute frequency is greater than 300. This list of words gives us a very clear idea of the main concepts that papers are dealing with. As expected, all the words are concerned with the previously indicated application domains, themes and topics.

Table 24 – MR3: Semantic content

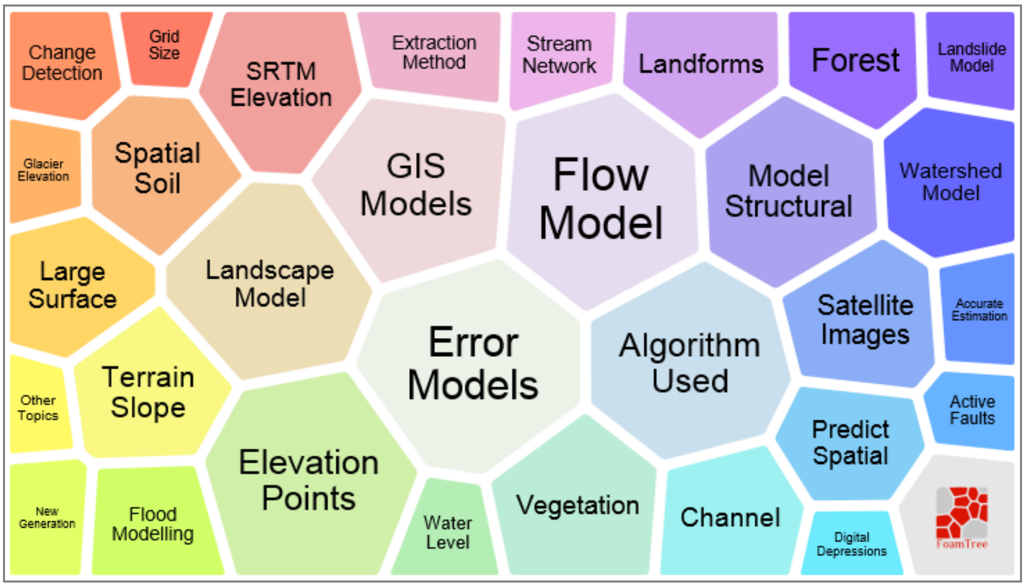
Term	N	Term	N	Term	N
DEM	2745	APPLICATION	645	WATER	397
MODELS	2730	FLOW	584	MEASURE	395
DATASET	1851	SCALE	573	PARAMETER	391
ELEVATION	1608	HYDROLOGICAL	554	EFFECTS	379
AREA	1379	PROCESSING	544	POINT	375
DIGITAL	1352	TERRAIN	516	GENERATION	372
MAP	1081	ERROR	472	RIVER	371
METHOD	1030	CLASSIFICATION	464	GLACIER	366
TOPOGRAPHY	961	CHANGE	463	NETWORK	364
SOIL	935	DISTRIBUTION	457	LANDSCAPE	363
HEIGHT	916	ESTIMATION	443	ASSESSMENT	355
ANALYSIS	882	ALGORITHM	433	LIDAR	352

STUDY	868	TIME	430	LAND	349
RESULTS	865	PREDICTION	427	EROSION	348
SURFACE	858	FIELD	426	VALUES	333
RESOLUTION	797	INFORMATION	425	CHARACTERISATION	331
DIFFERENT	734	SYSTEM	420	GEOMORPHOLOGY	322
SLOPE	720	SENSING	414	GIS	317
ACCURACY	716	FLOOD	403	VEGETATION	317
IMAGERY	704	TECHNIQUES	403	DRAINAGE	315
SPATIAL	661	PRODUCT	403	STRUCTURE	311

In relation to text clustering, there are some algorithms available for carrying out this kind of process. We 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: first text segmentation is performed, and if terms-language is recognised the stemming application and stop-words development are 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 Factorisation ED Factory method for matrix model, with a maximum matrix size of 100,000 and a maximum word document frequency of 0.90. A word document frequency threshold of 10% was also considered. Following this procedure, a total of 31 clusters with their labels were identified (Figure 1). The size of each Voronoi polygon depends on the number of occurrences of phrases. The core of clusters is around cases including the word “model” (e.g., flow model, GIS model, algorithms, etc.). The majority of clusters have a meaningful interpretation, according to the theory objective of the Lingo algorithm.

Figure 1 – 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 bear 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, the context of analysis and also the tools for the analysis are different. We believe that this approach provides a more focused view of research presented in scientific papers and use cases.

3.4. Approximation to use cases by means of MR4

Use cases are a method used in software and system engineering to identify, clarify, and organise requirements. A use case is a formal definition, a structured document explaining for a specific requirement the actors, actions, inputs outputs and decisions needed in order to achieve a goal. Unified Modelling Language (UML) diagrams, or other graphical language, can be used in order to depict use cases in a standard way. Weibel and Heller (1991) presented a table of DTM applications domains and their functional requirements desegregated into a list of issues that can be understood as use cases in a software engineering project. For us use cases are on a higher level, and not related to software capabilities development but rather to 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 DEMs which have been somewhat formalised; and second, if use cases are sufficiently detailed they will inform us about quality-related issues.

The identification of use cases concerning DEM applications or uses on the Internet using Google searches has been difficult. A lot of non-relevant documents were found and only very few with the structure of a use case. There are many documents using the term “uses case” dealing with applications of DEMs, but very few of them are formal presentations of an actual use case. Finally, only four (4) documents were considered as proper use case explanations. All documents came from European initiatives funded by diverse public projects (e.g., Inspire, 6th and 7th framework programs, etc.). Not all the documents have the same in-depth and detailed description of the explained use case, and the point of view is also different, for instance the use cases presented in the standard SS 637008 (SS, 2006) are more application schemes (data models) than application use cases. Smith (2005) presents five use cases (glacial geomorphological mapping, search for blandings, winter snow accumulation, soil erosion estimation and river characterisation) directly related to the use of DEMs. In the RISE project (Brönner, 2007) there are several use cases related to hydrology but only one elevation use case scenario. In this case, it is focused on the comparison of a real drainage network with a theoretical one. There is a technical guideline with data specifications on elevation (Inspire, 2013) where specific use cases of DEMs appear. These are high level use cases that cover only four specific scopes: two of the use cases are related to hydrological applications (Flood mapping and Maintenance of fairways), while the third use case is the application of elevation data as input for other data production (Orthoimagery production); the last use case (elevation mapping) is focused on the generation of DEMs, so that is not a proper application of DEMs.

In summary, we have identified 13 use cases related to the following scopes: water management (3 cases), glaciology, channel network and spatial data production (2 cases each) and decision-making, soil, flood risk and navigation (1 case each). These use cases cover only a few of all the many possible use cases, but are centered on important current applications. All of them are consistent with previous findings of this study. These use cases are not a harmonised set of use cases, but all of them present a formal structure and can be used as a starting point for further developments.

4. CONCLUSIONS

The importance of DEMs is great and is recognised in many documents, and for this reason knowledge about the users, uses and facts related to the quality of DEMs is important. In this paper several resources (MR1, MR2, etc.) and activities (surveys, database queries, etc.) have been deployed in order to characterise 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 large amount of cases which help to characterise in detail the profile of the users in Spain. In this case, 70.74% of users are private users and we consider that this profile can be extrapolated 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* survey designed to obtain specific information related to quality. This survey is biased by the source of contact information, which comes from scientific papers. At first glance some information (e.g., user's profiles and application domains) coming from this survey is redundant compared to that from the MR1, but both are complementary perspectives which allow us to know the profiles of general users (MR1) and advanced users that have a more scientific character and demands (MR2). Because of this, the results are not the same (e.g., compare Table 4 versus Table 7 (case 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 normal (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 grid and that the most used resolutions 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 indicating that they are satisfied or quite satisfied. Users do not know how to evaluate the influence of poor quality on their work, and this is a critical aspect that should be taken into account in future research. In addition, MR2 (Q11 and Q15) has provided a set of ideas that can serve to guide future work on quality in DEMs.

MR3 is related to queries to the ScienceDirect database. This resource allows us to search millions of scientific papers and discover papers dealing with research focused on DEMs. From queries to the ScienceDirect database, it has been possible to analyse the graph of the relationships between relevant terms that are linked to publications on DEMs. Also, we have performed a semantic analysis of a set of 950 abstracts, and this analysis shows the parts of sentences present more frequently. This analysis demonstrates a novel perspective as a way to summarise thousands of documents, and ScienceDirect is a resource that can allow interesting analysis.

From MR4 it can be concluded that the formalisation of applications as use cases is not normal. Very few cases have been found and all are focused on very concrete projects (e.g., Inspire, GRADE, etc.). This indicates that it is not an obvious necessity. The majority of use cases are related to water/hydrology, which indicates the maturity of hydrological applications, and also their social and legal importance.

Regarding the uses, and from a combined perspective of the four methods / resources, there is a great direct match between MR1 and MR2, the main uses are of an environmental nature and are related to spatial data. The MR3 case also reinforces this idea. Table 22 shows environmental applications (first column) and aspects related to spatial data (second column). The MR4 provides very few cases, but again environmental applications and applications focused on spatial data appear. Within the environmental applications, those related to water are the most important – an aspect that is confirmed with MR4.

Finally, we can conclude that the results of this paper have evidenced that the quality-related issues analysed (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 into quality issues focused on specific use cases, and the outreach and dissemination of findings and developments in specific user communities.

REFERENCES

Acharya B., J. Fagerman and C. Wright (2000). Accuracy assessment of DTM data: A Cost Effective Approach for a Large Scale Digital Mapping Project. In Dowman I. and L. Janssen (ed) XIXth ISPRS Congress July (ISPRS Archives – Volume XXXIII Part B2), Amsterdam, The Netherlands, pp: 16-23.

Agumya, A. and G.J. Hunter (1999). Assessing ‘fitness for use’ of geographic information: What risk are we prepared to accept in our decisions?. *In Spatial Accuracy Assessment, Land Information Uncertainty in Natural Resources*, Edited by: Lowell, K and Jaton, A. 35–43. Chelsea, MI: Ann Arbor Press.

Amii R.D., N.J. Tate, and C. Brunsdon (2008). Improving user assessment of error implications in digital elevation models. *Computers, Environment and Urban Systems*, 32:268–277.

Boin A.T. and G.J. Hunter (2006). “Do spatial data consumers really understand data quality information?”, *Proceedings 7th International Symposium on*

Spatial Accuracy Assessment in Natural Resources and Environmental Sciences. Edited by M. Caetano and M. Painho.

Boin, A.T. and G.J. Hunter (2009). What communicates quality to the spatial data consumer?. In Stein, A., Bijker,W., & Shi,W., (Eds). *Quality aspects in spatial data mining*, Boca Raton, pp. 285–296.

Brönner, C. (2007). Comparison of real with theoretical drainage network. Project: RISE (Reference Information Specifications for Europe).

Carlisle, B.H. (2005). Modelling the Spatial Distribution of DEM Error. Transactions in GIS. 9(4): 521–540

Cavazzi S, R., Corstanje, T. Mayr, J. Hannam and R. Fealy R. (2013). Are fine resolution digital elevation models always the best choice in digital soil mapping? *Geoderma*. 195–196: 111–121.

Charrier R. and Y. Li (2012). Assessing resolution and source effects of digital elevation models on automated floodplain delineation: a case study from the Camp Creek Watershed (Missouri). *Applied Geography*, 34:38-46.

Darnell, A.R. N.J. Tate and C. Brunsdon (2008). Improving user assessment of error implications in digital elevation models. *Computers, Environment and Urban Systems*, 32 (4):268-277.

Devillers, R., Bédard, Y., and R. Jeansoullin (2005). Multidimensional management of geospatial data quality information for its dynamic use within GIS. *Photogrammetric Engineering & Remote Sensing*, 71(2), pp. 205–215.

Devillers, R., Gervais, M., Bédard, Y. and R. Jeansoulin (2002), "Spatial Data Quality: From Metadata to Quality Indicators and Contextual End-User Manual". Proceedings of OEEPE/ISPRS Joint Workshop on Spatial Data Quality Management, March 21-22, 2002, Istanbul. pp. 45–55.

Dewberry (2012). National Enhanced Elevation Assessment. http://www.dewberry.com/docs/default-source/documents/neea_final-report_revised-3-29-12.pdf?sfvrsn=0 [accessed 26 October 2017].

Fisher, P.F. and N.J. Tate (2006). Causes and Consequences of Error in Digital Elevation Models. *Progress in Physical Geography*, 30, 467-489.

H. Yao, and R.L. Clark (2000). Evaluation of Sub-Meter and 2 to 5 Meter Accuracy GPS Receivers to Develop Digital Elevation Models. *Precision Agriculture* 2(2), pp. 189-200.

ICSM (2008). ICSM Guidelines for Digital Elevation Data v.1. Intergovernmental Committee on Surveying and Mapping (ICSM). http://www.icsm.gov.au/elevation/ICSM-GuidelinesDigitalElevationDataV1.pdf [accessed 26 October 2017].

IGN (2018). Centro de descargas. Centro Nacional de Información Geográfica. Instituto Geográfico Nacional. http://www.ign.es/web/ign/portal [accessed 6 May 2018]

INSPIRE (2013). D2.8.II.1 INSPIRE Data Specification on Elevation – Technical Guidelines. http://inspire.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_EL_v3.0.pdf [accessed 26 October 2017].

ISO (2013). *Geographic information -- Data quality. ISO 19157:2013*. ISO, Switzerland.

ISO (2014). *Geographic information -- Metadata -- Part 1: Fundamentals*. ISO 19115-1:2014. ISO, Switzerland.

Kitchin, R. (2017). Thinking critically about and researching algorithms. *Information, Communication & Society*, 20(1):14-29.

Lemmens, M.J.P.M. (1999). Uncertainty in automatically sampled digital elevation models. In Lowell, K., Jaton A. (Ed). *Spatial accuracy assessment: Land information uncertainty in tatural resources.* Sleeping bear Press, Inc. pp. 339-407.

Li, J., and D.W.S. Wong (2010). Effects of DEM sources on hydrologic applications. *Computers, Environment Urban Systems*, 34:251-261.

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.

Maune, D.F. (2007). DEM user Applications. In Maune, D.F. (Ed). *Digital Elevation Models Tecnologies and Applications: the DEM users manual*, 2nd Edition. ASPRS, Bethesda. pp. 391-424.

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., E.G., Chicaiza, X., Buenaño, C., Jianhong, A.F., Rodríguez-Pascual, and F.J. Ariza-López (2016). Análisis de los usuarios y usos de los MDE en España, *GeoFocus* 17: 133-160.

Oksanen, J. and Sarjakoski, T., (2005). Error propagation of DEM-based surface derivatives. *Computers & Geosciences*, 31:1015-1027.

OS (2013). *OS TERRAIN 5 User guide and technical specification*. Ordnance Suevey, UK.

Osiński, S., J., Stefanowski, and D. Weiss (2004). Lingo: Search results clustering algorithm based on singular value decomposition. In

Proceedings of the International IIS: IIPWM´04 Conference Advances in Soft Computing, Intelligent Information Processing and Web Mining, Zakopane, Poland, 2004, pp. 359—368.

Podobnikar T. (2009). Methods for visual quality assessment of a digital terrain model. *SAPIENS*. 2(2):1-10. https://sapiens.revues.org/738 [accessed 26 October 2017].

Raaflaub, L.D. and M.J., Collins (2006). The effect of error in gridded digital elevation models on the estimation of topographic parameters. *Environmental Modelling & Software*, 21:710-732.

Racoviteanu A.E., W.F. Manley, Y. Arnaud and M.W. Williams (2007). Evaluating digital elevation models for glaciologic applications: An example from Nevado Coropuna, Peruvian Andes. *Global and Planetary Change*, 59:110-125.

Riggs, P.D. and D.J. Dean (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*. *Scoping a Geospatial Repository for Academic Deposit and Extraction*. JISC project.
http://wwwold.edina.ac.uk/projects/grade/usecasecompendium.pdf

SS (2006). *SS 637008:2006 Geographic information - Surface water systems - concept and application schema*. Swedish Standards Institute. Stockholm.

Sulebak J.R. (2000). Applications of Digital Elevation Models - DYNAMAP White paper, SINTEF. http://www.gisknowledge.net/topic/terrain_modelling_and_analysis/sulebak_dem_applications_00.pdf [accessed 26 October 2017].

Tarquini, S., and L. Nannipieri (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* 281:108-115.

Tovar-Pescador, J., Pozo-Vazquez D., Ruiz-Arias, J. A., Batlles J., Lopez G. and 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.

Wechsler, S.P. (2003). Perceptions of Digital Elevation Model Uncertainty by DEM Users. *Urban and Regional Information Systems Association Journal* 15(2):57-64.

Wechsler, S.P. (2007). Uncertainties associated with digital elevation models for hydrologic applications: a review. *Hydrol Earth Syst Sci* 11:1481-1500.

Weibel R, and M. Heller, (1991). Digital Terrain Modelling. In Maguire, D.J., M.F. Goodchild and D.W. Rhind (Eds.). *Geographical Information Systems: Principles and Applications*. London: Longman, pp. 269-297.

Wise, S.M. (2007). Effect of different DEM creation methods on the results from a hydrological model. *Comput. Geosciences*, 33:1351-1365.