Enhancing the usability of digital elevation data

Tomaž Podobnikar, Domen Smole

Dr Tomaž Podobnikar
Institute of Anthropological and Spatial Studies
Scientific Research Centre of the Slovenian Academy of Sciences and Arts
Novi trg 2, SI-1000 Ljubljana, Slovenia
Tel: +386 1 470 64 93, Fax: +386 1 425 77 95, tomaz@zrc-sazu.si

Domen Smole
DFG CONSULTING, d. o. o.
Pivovarniška 8, SI-1000 Ljubljana, Slovenia
Tel. +386 1 432 13 50, Fax: +386 1 432 22 81, domen.smole@dfgcon.si

Digital elevation model (DEM) is one of the most important datasets for the greater part of spatial-based studies and research. Generally, a high quality DEM could be used as all-purpose dataset, but unfortunately its production could be very expensive. If we know a nature of application that applies DEM, if our demands for the final result are clear, and if provider of DEM gives enough clear information for usability of DEM to potential user, then the producer can simplify its production or user can adjust the DEM selection. From the user's point of view, the selection of DEM can be greatly simplified if appropriate tools are available.

Keywords: DEM, modelling, interpolation, quality, usability, ontology

INTRODUCTION

Generation or selection of a digital elevation model (DEM) suitable for different spatial analyses or visualisation purposes is being discussed here. Firstly we should stress that DEM is only a model that is an approximation of the nature and its nominal ground. The models, in our case DEMs, might be different concerning their purpose of use, quality of data sources or interpolation algorithms, experiences of operator, etc. Our starting point is that the DEM should be carefully produced or chosen regarding purpose of our applications.

In general, more course analyses require lower accuracy DEM than detailed ones. We suppose that DEM for producing contour lines in any scale or to emphasize the main characteristics of the geomorphology, should be more course and smoother than the DEM for calculating slopes or aspects. Modelling of hydrographical networks requires geomorphologically correct DEM. To get overview of the geomorphology or visualisation of the whole Alps, the DEM could be carefully generalised from more detail data or appropriate modelled. For the analyses of natural landscape or ancient environment, the recent anthropogenic changes (stone quarries, dykes, etc.) should be eliminated. For the palaeo-landscape analyses, geological changes should be considered.

Anyway, we assume that the final user should be aware of the characteristics, cost and usability for particular application before to decide between using of the existing DEM or employing its own DEM production. Here exists also a third option, adopting existing DEM to own needs. For the correct decision is important also considering many other elements of the required DEM, like quality parameters including positional and temporal accuracy, completeness, lineage, etc.

If we are decided to produce our own DEM or to adopt existing DEM, then it is important to select suitable software for interpolation and to take into account quality of the data sources for DEM modelling. In the next chapters we will demonstrate simple tests of the different algorithms using the same basic data sources. Further on we will enhance the DEM production with fusion data sources of different quality and type. These procedures are educational for understanding the management of data sources and DEMs themselves to get the final suitable DEM for required needs. We will also apply ontology-based approach for resolving semantic heterogeneity of different vocabularies between DEM user's demands and producer's metadata information. The main problem for increase of the usability of DEM is how to get or produce a suitable DEM data for required application.

HINTS FOR THE DEM INTERPOLATION

DTM (digital terrain model) is mathematically defined as spatial distribution of heights that are described by continuous and regionally (within the segments) smooth surface. In the praxis good approximation of the DTM is digital elevation model (DEM) that is recorded as two-dimensional discrete matrix of data heights that is more common known as grid structure. In the next, we are going to centre to DEM.

DEM is usually produced form sampled data that are used as its source. Ideally the data sources would be used without of interpolation. For example, only contour lines themselves may represent a model of terrain. They can be acquired directly, for instance photogrammetrically from stereo model or indirectly, for example from analogue cartographic data, satellite images, by surveying, etc. Interpolation is also not necessary in the cases if data source is very precise and high density, and especially if the data is acquired directly into regular grid (DEM). But interpolation of data sources to produce DEM is necessary if the data sources themselves do not predict treated landscape phenomena.

Interpolation techniques base on the principles of spatial autocorrelation, which assumes that objects close together are more similar than objects far apart. On the edges of the interpolated area extrapolation is also reasonable. Unfortunately no one of the interpolation techniques is universal for all data sources, geomorphologic phenomenon or purposes. We should be aware that in the praxis, different interpolation methods and interpolation parameters on the same data sources lead to different results. The best chosen algorithms on fair data sources should not differentiate much from nominal ground that is idealisation of our desired model and which is commonly similar to actual Earth's surface. Divergences between results of interpolation and from nominal ground are especially consequences of the following circumstances:

- available data sources do not approximate terrain – distribution, density, accuracy, etc. of the sources is not appropriate)
- selected interpolation algorithm is labile (is not enough robust) on the employed data sources
- chosen interpolation algorithms or data structure are not suitable for selected terrain geomorphology or application
- perception or interpretation of Earth's surface (better: nominal ground) is not the same when more DEM operators work on the same problem; operator's own imagination is common and reasonable problem in DEM production

Application requirements play an important role to expected characteristics of the used DEM. For example, we do not need high geomorphologic quality of DEM for regional, small scale analyses and for calculating average altitudes. But geomorphologic accuracy is more sensitive for visibility analyses and even more for analyses that uses algorithms bases on derivates like slope, aspect, cost surface, drainage, path simulation, etc.

In the most cases, a very high quality DEM should cover all application demands. So it is preferable to find a good and robust interpolation algorithm, what unfortunately difficult task is. Even if more generalised surface is required, DEM with high detail can be simplified to the required quality. It should be noticed that appropriate generalisation methods are very important for producing required DEM. Commonly these methods are complex.

DEM MODELLING WITH COMMON INTERPOLATION ALGORITHMS

We were tested some most common interpolation algorithms based on inverse distance weighted (IDW), kriging and spline using the same data sources. The IDW methods apply the idea that influence decreases with increasing the distance from particular points. The method could be good for interpolation of geomorphologically smooth areas. Kriging methods take into consideration autocorrelation structures of elevations in order to define optimal weights for different distances from a point and then automatically evaluate the results. The method requires a skilled user with geostatistical knowledge. Spline-based methods fit a minimum-curvature surface through the input points. The interpolation ensures continuous and differentiable (smooth) surface. Rapid changes in gradient or slope may occur in vicinity of the data points.

We employed all of three described algorithms using contour lines data sources on the study area, which is geomorphologically variable (see Figure 1). All of the algorithms were used on standardised way and with default parameters. First of all we decided to asses the results with visual approach, which is suitable for general overview of consequences of the interpolation methods.

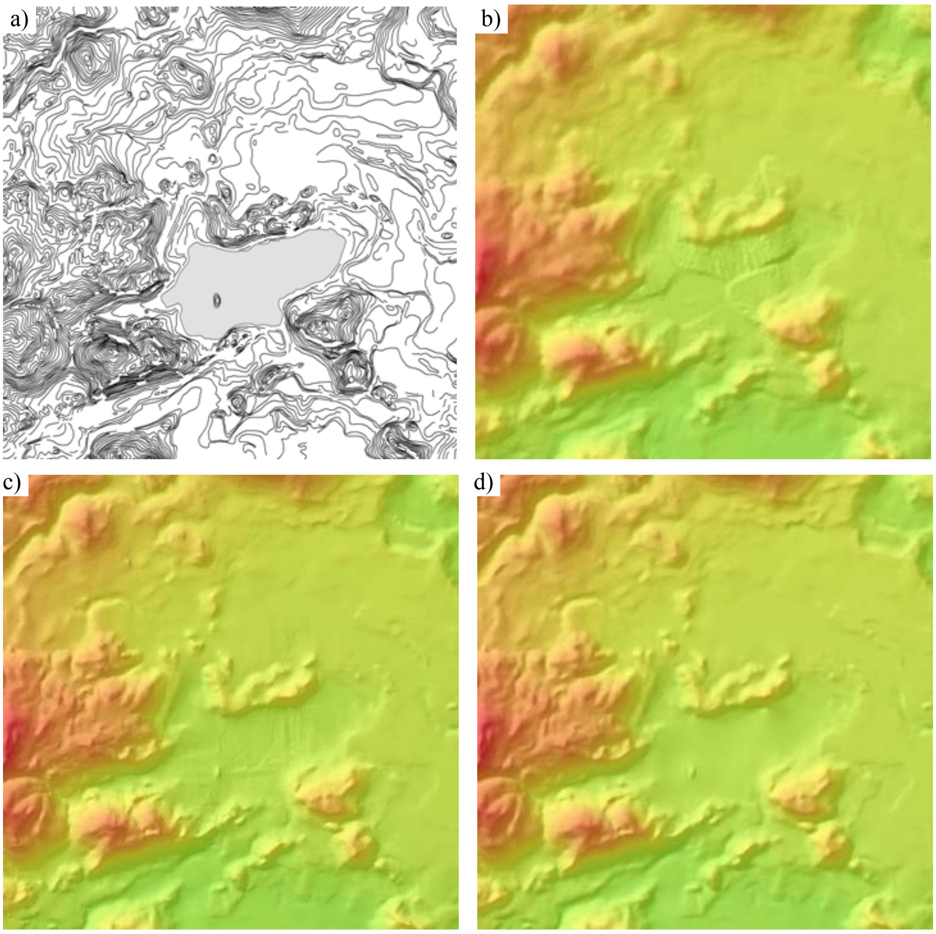

Figure 1: Contours with interval of 10 m and lake of Bled in the western Slovenia (upper left). DEM is produced with IDW – smooth (upper right), kriging – more details (lower left) and spline based method – smooth but with recognisable characteristic features (lower right) (area of 5000 by 5000 m).

For general purpose it is difficult to decide which algorithm produces the best DEM form contour lines. IDW algorithm is optimal if we need results produced in a short time and if the real terrain is smooth. Kriging method is useful in this case but some problems occur mainly on the areas with low density of data sources. Spline-based algorithm produces smooth surface and fortunately without many of badly interpolated areas. If we would like decide to use only one of three basic methods of the contour lines interpolation, then we can think on following way:

- to get optimal result without much effort: use spline algorithm
- to get the best general result for more advanced analyses and visualisation: use kriging algorithm
- to get the fastest result: use IDW

We can stress that there are no bad DEM interpolation algorithms (Beex 2003). Some of them have simply more advantages in certain circumstances. The algorithms are actually the most flexible part of the whole modelling process. It is because usually nobody has opportunity to use the ideal data and one can therefore only select the algorithm that is the most suitable for the used data sources and application.

If operator or user knows a purpose of the DEM's application, then he can decide about importance of particular quality parameters. Generally, the optimal is the DEM that requires good results after evaluation of many geomorphologic and statistical quality parameters. Let's propose to allow combination of the three proposed basic algorithms. Then the best DEM from contour lines would be produced as combination of kriging and spline. The kriging would be applied for the areas around the characteristical features like peaks, sinks, valleys, ridges, edges, etc., but the spline algorithm would be preferred on the other areas.

PROPOSED METHOD FOR ALL-PURPOSE HIGH QUALITY DEM MODELLING

We are going to put more energy to enhanced DEM modelling and then visually compare the results with previous interpolations. The main idea here is to produce as good as possible DEM that would be useful for most of applications. The main data source is still the same, contour lines, but the DEM is being enhanced with fusion of some other data sources, like local ones, lower quality DEMs, geodetic network points and others. One could even use datasets without a height attribute such as lines of the hydrological network, roads, railways, standing water polygons, etc.

But the most important in the data fusion is still quality of particular data sources for fusion. If the main source is dataset of contour lines, it should be interpolated very carefully. The algorithms, mentioned in the previous chapter are not good enough. We did many more experiments with contour data interpolation. The result is that the most suitable is combination of more methods of interpolation and producing additional characteristical areas, lines and points.

For data fusion, a method of weighted sum of data with geomorphologic enhancement was developed (Podobnikar 2005). With this method, DEM is modelled through averaging and fusion of individual datasets considering their quality. Grid based datasets are thus overlaid as regards the weights of particular grid cells. After overlaying, geomorphologic enhancement is applied. At the beginning a unique grid size for all data sources is determined – the same as for the final DEM.

Furthermore, each particular data source should be precisely evaluated by a reference dataset (points) regarding the standard test areas delineated by standard regionalised layers. The result is a predicted quality for each grid cell denoted with a random error (σ). For the sake of simplicity we will continue with our discussion with only two data sources. Height of DEM (H_{i+j}) regarding weights w_i and w_j and variances σ_i^2 and σ_j^2 are then

$$H_{i+j} = \frac{w_i H_i + w_j H_j}{w_i + w_j} = \frac{\sigma_j^2}{\sigma_i^2 + \sigma_j^2} H_i + \frac{\sigma_i^2}{\sigma_i^2 + \sigma_j^2} H_j$$

Weighted sums of pairs of surfaces $(i+j)_k$ are calculated iteratively by adding independent datasets to previous ones (Figure 2). The random error of the computed DEM (σ_{i+j}) incrementally decreases with every iteration. For two datasets it is calculated with the differentiation of heights as (Heuvelink 1998, Burrough and McDonnell 1998)

$$\sigma_{i+j} = \sqrt{\left(\frac{\sigma_j^2}{\sigma_i^2 + \sigma_j^2}\right)^2 \sigma_i^2 + \left(\frac{\sigma_i^2}{\sigma_i^2 + \sigma_j^2}\right)^2 \sigma_j^2} = \frac{\sigma_i \sigma_j}{\sqrt{\sigma_i^2 + \sigma_j^2}} = \frac{1}{\sqrt{w_i + w_j}} \le \min\left(\sigma_i,\ \sigma_j\right)$$

The best practical solution is to start DEM modelling with data sources of the lowest quality (lowest weights) and to finish with the best data.

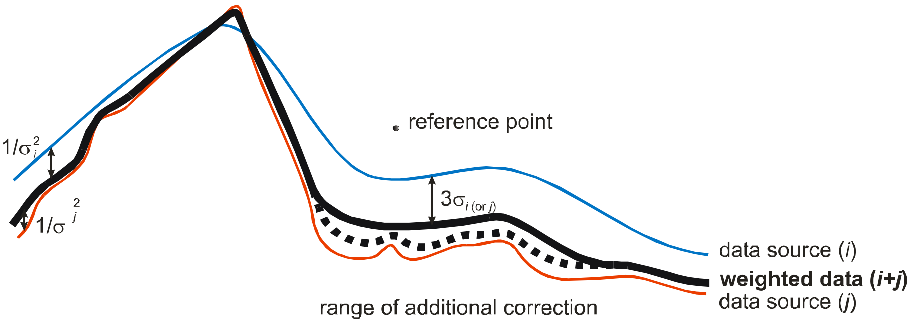

Figure 2: DEM, interpolated as weighted sum of data sources

DEM, derived iteratively with weighted sums of data is smoother than the geomorphologically highest quality data source (Figure 3). This is usually a consequence of the nature of the weighted sum. Geomorphologic enhancements of such a derived DEM are therefore required. The best solution seems to be to apply the enhancements only when the DEM is already derived from all weighted data sources.

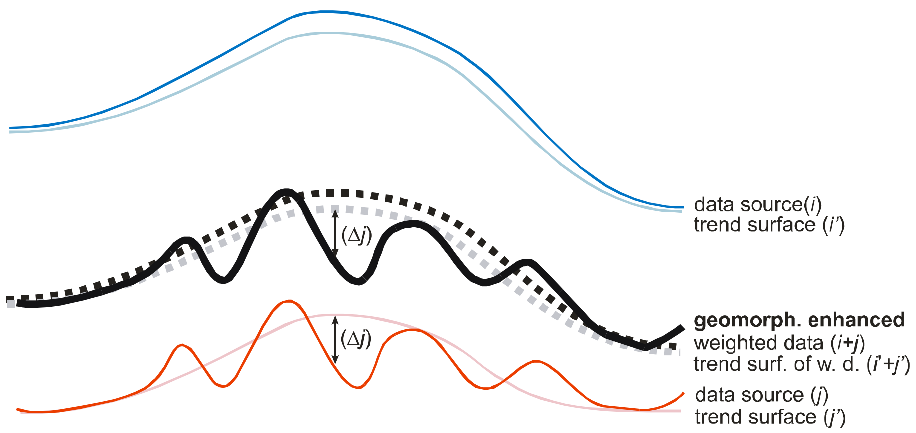

Figure 3: DEM, geomorphologically enhanced from weighted sum of data sources

The main step of geomorphologic enhancement is the generation of trend surfaces as low frequency functions – with generalising DEMs. Trends are produced with the same conditions for datasets of the statistically best DEM derived by weighting ($i+j$), and the DEM with appropriate geomorphology (j) (Figure 3). Relative elevations (Δj) as a high frequency part are computed from j and then added to the trend surface $i'+j'$ of the dataset $i+j$. In this way the final geomorphologically enhanced DEM is produced. Statistically it is slightly worse than the weighted one ($i+j$), but geomorphologically it is much better. The main problem of the described enhancement lies in finding a suitable filter to calculate the appropriate trend surfaces. The optimal solution is a compromise between geomorphologic improvements and retaining statistic quality.

Described method of DEM modelling was tested on the same study area as in the previous chapter. The modelled DEM looks visually geomorphologically high quality with clear and reasonable details (Figure 4). As it had been tested before, the method serves also DEM with high precision and accuracy. If we are going to produce geomorphologically and statistically highest quality DEM, we do not need to think much about required application.

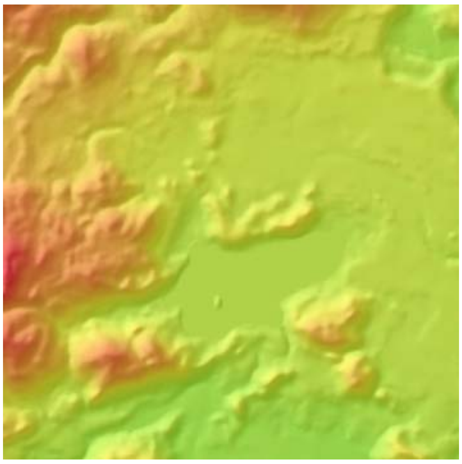

Figure 4: DEM, generated with more advanced modelling using weighted sum of data with geomorphologic enhancement (area of 5000 by 5000 m around lake Bled).

Effective and suitable DEM modelling from variable datasets is complex, rather iterative process that cannot be achieved intuitively or via a single step. In this sense experience is connected to the execution of stacks of tests and analyses, as well as a better understanding of the nature (parameters) of data. The quality of the DEM is evaluated for every data element and the portion of every data source element used for DEM modelling is known. With purposed approach, we can have full control of the production and effectively inform the final user about the characteristics of the DEM.

ONTOLOGICAL APPROACH FOR ENHANCING COMMUNICATION BETWEEN PRODUCER AND USER

Regarding experiences of producing a suitable DEM for user of any application, we are going to discuss about communication between producer of DEM and user. Usually there are two types of actors involved in the process:

- *a DEM producer*, who is usually an expert in DEM capturing, processing and visualization techniques etc. but she usually lacks in-depth know-how specific for other areas and
- *a DEM user*, who is usually an expert on a given (scientific) area, like for example land management, natural disaster monitoring etc., but has limited knowledge on available DEM sources, and/or corresponding DEM generation and/or selection techniques.

Not surprisingly, there is a gap in a communication between the DEM producer and the DEM user. If they communicate at all, they speak their own language, consisting of different vocabularies. The question is whether there exists an efficient approach that could be applied to bridge this gap? Currently a lot of research effort in the GIS community is oriented towards finding methods that would enable automatic interpretation of the meaning of different vocabularies, vocabularies of producers and vocabularies of users in particular. Ontology-based approach for resolving semantic heterogeneity problems are amongst most commonly referred ones (Karalopoulos et al., 2004).

Following is given an example of a communication scenario in which system based on the ontology approach plays a vitally important role. For example, there is an investor interested in building a golf course in the suburbs of the capital city that happens to be near the city's dumping area. She is interested in the availability and characteristics of a variety of spatial data for given area. Beside DEM, there may be available many other sources describing for example anthropogenic objects like buildings and traffic infrastructure, measurements of noise and smell, meteorological/climatological data like precipitation or solar radiation, data predicting the possibility and severity of flood accidents, earthquakes and landslides, not to mention statistical data that are very useful for marketing and other purposes. These are some of the parameters that can make planned business a success or a disaster if not analyzed in advance. Unfortunately, the investor is usually not aware of all the available datasets and its possible uses, processing algorithms etc. Besides, the vocabulary of the investor is different of the datasets producer what makes it very hard to find the needed information.

We believe that natural language processing techniques based on machine learning algorithms have promising future in this field of research. It is clear that understanding different vocabularies goes beyond comparing words (e.g. string matching). Currently we are engaged in a research work which aim is to outline the characteristics of an incrementally learning system that would provide (semi)automatic mapping between different vocabularies of spatially related concepts offering capabilities similar to »question/answering system«.

An analysis done so far has shown that this communication should be managed as a sequence of questions and answers interactively exchanged between the user and the system. Generally speaking, this kind of communication process could be grouped into three parts. In the first, introductory part, the user must provide the system with answers regarding the area of given application, either as a list of land parcels or as a list of coordinates. Under these circumstances, the system should able to find all available datasets for the given area. This part of the communication is usually independent of user's type of application. What follows in the communication process is a segment of questions and answers regarding the purpose and the intention of the user's application. In our case, the investor is interested in changing the landscape into the golf course. This kind of environmental interference usually requires DEM data of high resolution and of high positional accuracy, whereas it is not that important to be geomorphologically very accurate. To be able to answer questions like this, the system must be aware of specific knowledge characterising each available spatial dataset. As a result, the user is given a list of most appropriate and available spatial datasets for given application. At last, the user usually requires information like the format of the datasets, the name and the phone number of the contact person etc. These kinds of questions are again independent of the type of the spatial data and/or user's application.

As part of this research project we will continue with building and analyzing a list of most frequent questions users are interested in regarding spatial datasets. As we have seen, this is not an infinite list of possible questions. What we would like to enhance further on is the process of finding questions to those answers. Currently this is done manually but should be replaced with a (semi)automatic process since huge amounts of spatial datasets and its descriptions in the form of metadata emerge on a daily basis. A system, when it is given a description of the spatial dataset in a natural language, should through the process of information extraction (IE) be able to find a set of answers for given answers from the raw text (Aitken, 2002). Building system like this, with an approach of inductive logic programming, is a complex task (Cussens et al., 1997), but we believe it could, among other things, help in bridging the gap in communication between the DEM producer and the DEM user. Preliminary results have shown that rules for answer extraction from raw text in Slovene language (pre-processed by POS tagger) could be learned. However, there is a plenty of research work yet to be carried out. We intend to focus on improving the accuracy of learned information extraction rules. When this is satisfied, the process of finding appropriate DEMs will be much easier what will in turn improve its reusability and justify its enormous production costs.

CONCLUSIONS

Prior to starting modelling of our DEM, we should ask ourselves: Do we need to produce an all-purpose high quality DEM? Or: Do we need to produce DEM just for our application? It is known that high quality DEM could consume even more than 100 times more time then production with using basic algorithms like IDW, or spline are. Furthermore we need advanced software, hardware of higher quality, experienced team and advanced know-how. For our decision it is important to have a review over existent DEMs which can be used unchanged for our applications or they are just sources within more advanced DEM production processes.

Specialisation in the information society is currently so high that we can not master all the processes on enough high level, so it is necessary to trust to the particular specialists and organise them to a reasonable team. It is nice to hear that quantity of digital spatial data is increasing, but unfortunately quality of data doesn't follow the quantity. Many of the producers do not test produced data sources or models enough carefully. The result of such work is data of unpredictable quality. Similarly, the software always offers more than hardware can manage.

Based on our experiences we propose that some kind of instructions should come together with interpolation algorithms. They would suggest the most important steps of DEM production for required application. The instructions should be prepared on basic and on higher level and may include the following steps: preparation for DEM production, pre-processing of data sources, processing DEM from sources and managing the DEM data. Furthermore we suggest that user's manuals of the DEM interpolation algorithms should propose more tips and tricks for the common users. Most often they include only general information as description of the algorithm with parameters, common purpose of use, and some simple examples. We suggest at least hints regarding appropriate algorithm and parameters implementation if data sources are differently distributed or different type. For example contour lines interpolation algorithms are different to the algorithms for scattered points. The tip that suggests which parameters might be used for production of high quality DEM from particular datasets might also be important. These kinds of user manuals should be treated as an extension to the ISO metadata standard for describing spatial datasets. They serve as a crucial source on which we apply information extraction rules for the purpose of finding answers describing different aspects of DEMs that users are interested in.

Significant aim of the DEM's nature is to find a balance between users' demands and capability of the developed realisation process. High quality DEM production using advanced methods could be very expensive. However, users always demand higher quality then it is offered, but this is not always reasonable. We should stress that even if we produce more sophisticated DEM and if more experienced producers are to be employed in the job, we would get different models. The solution proposed in this paper was confirmed through applied experimentation that enables cost-effective, high quality production and assumes higher collaboration between producers and users. Better DEM we can produce or choose regarding higher knowledge of the terrain characteristics and if we are aware of the application for that DEM is being used. Nevertheless the model should look reasonable!

REFERENCES

Aitken, J. S., 2002. Learning Information Extraction Rules: An inductive Logic Programming Approach. The Proceedings of the 15th European Conference on Artificial Intelligence ed. van Harmelen, F., IOS Press, Amsterdam, http://www.aiai.ed.ac.uk/~stuart/Papers/ilp-for-ie.pdf.

Beex, W., 2004. Use and abuse of digital terrain / elevation models. In Ausserer K.F., Börner W., Goriany M., Karlhuber-Vöckl L. (ed.). Enter the past, The e-way into the four dimensions of cultural heritage, CAA 2003 (BAR International Series 1227), 240–242.

Burrough, P. A., and McDdonnell, R. A., 1998. Principles of Geographical Systems Information Systems. Spatial Information Systems and Geostatistics, Oxford (Oxford University Press).

Cussens, J., Page, D., Muggleton, S., Srinivasin, A., 1997.Using Inductive Logic Programming for Natural Language Processing. (W. Daelemans, T. Weijters, and A. van der Bosch, editors, ECML'97 -- Workshop Notes on Empirical Learning of Natural Language Tasks, pages 25--34, Praga), http://citeseer.ist.psu.edu/cache/papers/cs/933/ftp:zSzzSzftp.cs.unimaas.nlzSzpubzSzecml97zSzcussens-ftp.pdf/cussens97using.pdf.

Heuvelink, G. B. M., 1998. Error Propagation in Environmental Modelling with GIS. London, Bristol (Taylor & Frances).

Karalopoulos, A., Kokla, M., Kavouras, M., 2004. Geographic knowledge representation using conceptual graphs, AGILE 2004 Conference Proceedings (Eds. F. Toppen, P. Prastacos).

Podobnikar, T., 2005. Production of integrated digital terrain model from multiple datasets of different quality. International journal of geographical information science (Taylor & Frances), 19/1, 69–89.

BIOGRAPHY

Tomaž Podobnikar is currently employed as a researcher in the Institute of Anthropological and Spatial Studies at the Centre for Scientific Research of the Slovenian Academy of Sciences and Arts. He holds a B.Sc. (diploma) in geographical information systems and its applications in environment and archaeology, M.Sc. in Monte Carlo methods and its applications in geographical information systems, Ph.D. in digital terrain modelling from various data sources of different quality, all the University of Ljubljana. He has authored several papers on GIS, DTM/DEM, and its applications. He was research fellow at the Technical University Delft (The Netherlands) and Technical University Vienna (Austria).

Domen Smole is currently employed as a researcher at DFG CONSULTING, d.o.o., a private company. He graduated in 2002 at the Faculty of Civil and Geodetic Engineering (University of Ljubljana) with the thesis on web geographic information systems. Shortly thereafter he started his postgraduate study at the Faculty of Computer and Information Science (University of Ljubljana) focusing on the research of ontological approaches in geodatasets based on machine learning principles. He is a co-author of the award winning poster at the AGILE 2004 conference held in Heraclion. Apart from the academic research he has been involved in various projects including digital photogrammetry, laser scanning, video data processing etc.