

ERROR MANAGEMENT IN DIGITAL ELEVATION MODELLING¹

B. MÁRKUS

Department of Surveying
Technical University, H-1521 Budapest

Received: July 1, 1992

Abstract

The paper aims at the accuracy aspects of Digital Elevation Modelling, compares the most common data acquisition methods, data structures used to store elevation data and the biases that occur in extending the point data to derived information. General use techniques of interpolation will also be discussed, however, the greatest emphasis is given to the methods that are available as a pragmatic response to the wide spectrum of problems of error management.

Keywords: error management, digital elevation modelling, geographical information systems.

Introduction

The Digital Elevation Model (DEM) is a digital representation of the topographic surface. In a more general sense, a DEM may be used as a digital model of any single-valued surface. However, here attention is focused on digital models of the terrain surface. The Digital Elevation Model and/or the derived thematic surface layers (elevation zones, slope angle, slope aspect, etc.) are very frequently used components in the Geo-Information Systems (GIS) supporting decisions based on this technique. Digital Elevation Modelling is increasingly required and used in various GIS operations, such as visibility studies and routing, simulation models for landscapes and landscape processes, erosion studies, extraction of morphologic characteristics, etc. The paper aims at the accuracy aspects of Digital Elevation Modelling.

Any abstract of reality will contain discrepancies from its source. With traditional cartographic methods many of the problems are visible and the skilled topographer makes the necessary adjustments and knows how far the information can be relied upon. With a Digital Elevation Mod-

¹The work reported in this paper is part of a larger project which has been funded by the Hungarian National Science Foundation Grant No. 685

elling System the equivalent operations often are not transparent (black box effect), usually the operators are no longer so skilled and the problems are more or less completely invisible. The digital modelling has the potential to dramatically increase both the magnitude and importance of errors in the models. The results may be used for decision making and planning despite possessing levels of uncertainty that are completely unknown and usually cannot even be guessed. That is why the accuracy analysis is one of the most important problems in the development and applications of the system.

At several stages in the modelling process, errors are being introduced and propagated. Simplifying the problem, from the point of view of system management, the errors can be categorized into three groups (*Fig. 1*). In the first group are the errors of data acquisition (positional and attribute errors: errors by careless work, errors in measurement, inadequate sampling techniques, etc.) building up the model of reality. The second group occurs when the information is derived from the digital elevation model, from the database (modelling errors). At this level further abstractions from reality come into the system. At the end of the process: use errors usually related to the inappropriate use of the final products and therefore they could be considered the most difficult type of error grasp (errors of information interpretation). One may conclude that the accuracy of a product could be reduced to some function, dependent in the value on two variables, to wit source errors and processing errors, respectively. It is obvious that use errors cannot be taken into account; first of all, the user himself will be responsible for making sensible choices. If the value of these variables could be estimated before the actual integration takes place, the product accuracy could be produced. Unfortunately, the way in which the errors interact is not yet fully understood. Besides, information concerning the accuracy of source data is often lost during the data capture stage. As a consequence, it will be extremely difficult to estimate the final accuracy from the separate error sources (VAN DER WEL, 1991).

The structure of the paper is based on a hierarchy of needs for evaluating output product quality (VEREGIN, 1989): error source identification, error detection and measurement, error propagation modelling, strategies for error management, and strategies for error reduction.

This paper compares the most common data acquisition methods, data structures used to store elevation data and the biases that occur in extending the point data to derived information. Applicable techniques of interpolation will also be discussed, however, we focus on the methods that are available as a pragmatic response to the wide spectrum of problems of error management.

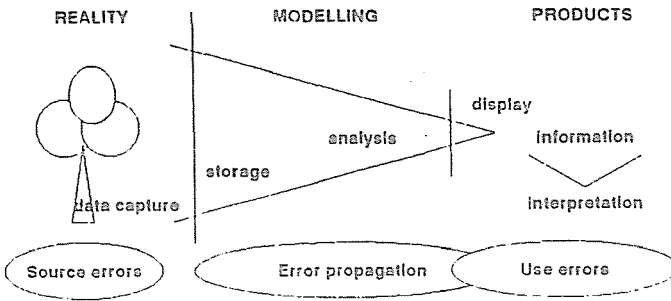


Fig. 1. Errors in digital modelling

Data Acquisition

Surfaces potentially have an infinite number of points which can be measured. Obviously it is impossible to record every point; consequently, a sampling method must be used to extract representative points to build a surface model that approximates the actual surface.

A surface model should:

- i. accurately represent the surface,
- ii. be suitable for efficient data capture,
- iii. minimize data storage requirements,
- iv. maximize data handling efficiency,
- v. be suitable for surface analysis.

The choice of data acquisition strategy and technics is critical for the quality of the results. Input data should reflect adequate information in the modelling. The database should contain the significant surface points and structural features. Nowadays DEM data are derived usually from three alternative sources: field surveying, photogrammetry and map digitising.

Surveying data are often put in directly into the computer through data recorders which may be attached to field instruments. Since field data tend to be very accurate, and topographers tend to adapt the survey to the character of the terrain surface, the accuracy of the DEM is very high. However, as this particular data collection technique is relatively time consuming, its use is limited to small areas.

A very often used method is to collect the elevation data based on the stereoscopic interpretation of aerial photographs. It is possible to distinguish a number of sampling methods, describing grids, profiles or contours of the terrain surface. Progressive sampling has been proposed as a method to automate the sampling in response to a varied terrain, though there are

still problems of redundancy associated with raster encoding (MAKAROVIC, 1973). Each of these methods attempts to minimize the data collection effort, maximizing at the same time the accuracy of the model. Depending on the sampling method and imagery that are used, the resulting DEM accuracy will be medium or high. Photogrammetric methods are used in large engineering projects as both regional and as nationwide databases.

Perhaps the most frequent method is to digitize contour lines from existing topographic maps. These analogue data may be digitized through manual digitization, semi-automated line following, or by means of automatic raster scanning and vectorization. Due to the relatively high costs of surveying and photogrammetry and the large volumes of existing maps, this indirect method is predominant for large data collection projects. Despite their widespread use, contour data present some problems. The accuracy of each object depends on the cartographic processes which generated it, in the form of abstraction and generalization, and these are sensitive both to scale and to the nature of the object. Contours are mainly a form of surface visualization and are not particularly useful as a method for digital surface representation. An excessive number of points is sampled along contours (oversampling), and no data across contours (undersampling). Furthermore, errors may be introduced (in drawing, line generalization, reproduction, etc.) and a lot of information is lost in map making (WEIBEL, 1991).

During planning a nationwide DEM for Hungary the last two methods (photogrammetry and map digitizing) were compared on four different types of study area (MÁRKUS, 1986). They have the same size (one map sheet in the scale of 1:10 000, 24 sqkm); two of them are flat, the others are hilly and mountainous.

Table 1

	Flat1	Flat2	Hilly	Mountainous
height difference [m]	5	4	101	294
average slope [%]	1	1	8	25
average density Ph/D	1135/730	2215/400	1361/1277	3140/1140
RMSE Ph/D	0.33/0.33	0.28/0.30	0.79/0.79	1.81/+++
RMSE 'True'	0.10	0.05	0.21	0.13
contour interval [m]	1	1	2.5	2.5

Table 1 shows the main characteristics of the experiments. The average point density varies from 400 to 3140 [points/sqkm]. In Table 1 Ph means photogrammetry, D means map digitizing. The root mean square

errors (RMSE) were computed using the differences between a directly measured 'true' grid and the grid derived from the different DEMs. The accuracy of the derived grid elevations is nearly the same for both photogrammetry and digitizing, approximately 30% of the contour interval with the exception of mountainous area. The model built from map digitizing in this case contained incorrect contour values, in this manner RMSE was not evaluated (+++). Consequently, digitized contours of topographic maps provide a compromise method of obtaining a nationwide DEM, on a relatively low cost, with the specified accuracy.

Digital Elevation Model

The captured data must be structured to enable handling by subsequent modelling operations. A variety of data structures for DEMs has been in use over time. Today, however, the majority of DEMs conform to two data structures: the regular grid and the structural DEM (randomly distributed set of surface specific points) are the most common approaches of representing digital surface data.

The grid model is based on a regular sample of elevation points. It is probably the most common type of DEM in use today. Its advantages include a relatively compact storage and a form which lends itself to fast algorithms in data manipulation, modelling. The problems of regular grid representation primarily in its inability to represent the terrain variability. If the cell size is altered to become spatially more sensitive to areas of high terrain variability, data redundancy causes problems in storing the needlessly small grid size for the smooth portions of the area. Thus, regular grids cannot describe structural features, additional data have to be added for this purpose.

Another popular data structure is the set of surface specific points in an irregular, structural DEM. Perhaps the most important advantage of this method is that this approach reduces the data redundancy offering the capability of varying the spatial resolution where appropriate. The points stored are surface specific and therefore keep the geometry of the surface. Structural features of the surface can easily be incorporated into the model.

Besides the above mentioned methods, there exist a few other structures for surface representation. Structuring the progressive sampling data the quadtree approach fits well, where the model is recursively broken into quadrants in response to terrain variability. The models are very often hybrid, combining the possibilities, benefits of regular and irregular structures.

Height Interpolation

In order to define a continuous surface from the discrete elevation points the system must have a method to interpolate between these stored points. Abundant literature exists on methods for interpolation. LAM (1983) groups the methods into exact and approximate interpolation. The former preserves the values at the data points, while the latter smoothes out the data. Another popular way to classify interpolation methods is to do it by the range of influence of the data points involved. Global methods, in which all sample points are used for interpolation may be distinguished from local, piecewise methods, in which only data points nearby are considered (WEIBEL, 1991).

Often used methods of grid based interpolation are bilinear, inverse weighted distance on 9 neighbours and spline (shortest cord). These methods are easily integrated with raster GISs.

Triangulated Irregular Network (TIN) may be defined as a network of contiguous, planar triangles that vary in size, shape, elevation, slope and aspect in response to local surface geometry. These triangles are generally formed from a random distribution of points. The elements of the model can be varied in size to accommodate the need for greater spatial resolution over a more changeable portion of the terrain surface. A simple interpolation on the TIN structure uses planes fitted to the three nodes of triangles. The more sophisticated systems result in a continuously smooth faceted surface. TIN can be easily used by a vector GIS. Disadvantages of TIN modelling include the requirement of a more complex data structure and more involved algorithms (LEE, 1991).

There is no space here to introduce the other interpolations, rather it seems appropriate to mention briefly the characteristics and peculiarities of DEM interpolation:

- i. There is no best interpolation method that is clearly superior to all others and appropriate for all applications.
- ii. In literature the accuracy is defined as the closeness of results, computations or estimates to true values (or values accepted to be true). However, to compare the model with reality is a quite difficult, expensive and time consuming process. Accuracy of a DEM entails not only the average differences of points on the DEM from the real surface, but also involves the distribution of the non-random spatial component of errors.
- iii. The accuracy of DEM outputs are determined mainly by the density, distribution and the accuracy of data points: by the quality of database and by the roughness of the terrain surface itself; the effect

of the interpolation method and its parameters have much smaller importance. Using advanced mathematical algorithms the results are highly influenced on the applied parameters or conditions.

- iv. From the point of view of the cartographic quality in DEM products the effect of interpolation is much more significant.

Thematic Data

In GIS environment the thematic derivations of elevation (slope, aspect, curvature) are also very useful (THEOBALD, 1989).

For a grid model these are usually found by a differencing operation, taking the partial derivatives with respect to x and y which are found by

$$\frac{\partial z}{\partial y} = \frac{z_{i,j+1} - z_{i,j-1}}{2\Delta h} \quad \text{and} \quad \frac{\partial z}{\partial x} = \frac{z_{i+1,j} - z_{i-1,j}}{2\Delta h},$$

where z is the elevation of the grid and Δh is the grid spacing. Thus slope is given as

$$s = \left[\left(\frac{\partial z}{\partial y} \right)^2 + \left(\frac{\partial z}{\partial x} \right)^2 \right]^{\frac{1}{2}},$$

and exposure (the compass direction of the maximum rate of change) is calculated as

$$\tan E = - \frac{\frac{\partial z}{\partial y}}{\frac{\partial z}{\partial x}}.$$

This technique uses the four immediate neighbours and thus an individual slope is based on a horizontal distance of twice the cell size, causing smoothing. Generally, the derived maps are noisier than the original surface and consequently, smoothing of the data (e. g. with a low pass filter) is a common processing step.

The derivation of slope and aspect from a TIN based DEM is a more straightforward technique. Here the trigonometric computation using three nodes will find the slope and aspect of the contained triangular plane. It is important to observe that any product which is derived from a TIN will be heavily dependent upon the quality of the TIN. Since a TIN is primarily a topological structure, quality also relates to the logical consistency of the TIN, that is whether the formation of the triangles complies with the geomorphological facts of the terrain surface being modelled or not.

Error Detection, Model Refinement and Filtering

It is vital to the data processing of any data input to DEM how the data were sampled. Clearly, if much of the data are insufficient, inadequate, not totally representative, or were collected at an inappropriate spatial scale, their use will reduce the chance to obtain good results. Along with DEM generation procedures, the DEM manipulation processes are of fundamental importance for the performance and flexibility of a modelling system. They are needed for the modification and refinement of existing models. DEM manipulation consists of processes for DEM editing, filtering and merging, and for the conversion between different data structures.

DEM editing involves updating and error correction. It is helpful if the system has intelligence to find the possible errors (e. g. flat triangles in TIN, or very steep slopes in general) and if an effective user interface is used (visualization and interactive editing). Interaction by direct manipulation supported by visual feedback will greatly simplify the editing task.

DEM filtering may serve two purposes: smoothing or enhancement, as well as data reduction. Smoothing and enhancement filters are best applied to gridded models. The effect of smoothing is to remove details, and make the DEM surface smoother. Enhancement has the opposite effect as surface discontinuities are emphasized. Smoothing filters have been used to eliminate blunders, while enhancing filters to find them out.

DEM filtering procedures are also used to reduce the database volume. A data reduction process of this kind may be desirable to eliminate redundant data points (e. g. within the digitizing tolerance), to save storage space and processing time, or to reduce a DEM's resolution. An efficient approach is called the VIP (very important point) method. The procedure is based on an algorithm that repeatedly adds the most significant point to the triangulation, until no additional points are needed to describe the surface within the specified tolerance. The importance of a point is given by the vertical distance between the point and the triangulated model without that point. At small VIP tolerance levels, only redundant data points are removed; at larger tolerance values data points are retained on the basis of their significance in reflecting the general morphology of the terrain surface.

In the following, we present a case study that illustrates many of the issues discussed throughout this chapter². Since the sampling process is highly data dependent, it is recommended to ensure that the representation of the surface is correct before finally accepting it for inclusion into the database. *Fig. 2* shows a contour plot for checking the digitizing process (3633 digitized points on 1 sqkm area). *Fig. 3* allows to make a pictorial

²The case study was elaborated with the help of ARC/INFO TIN.

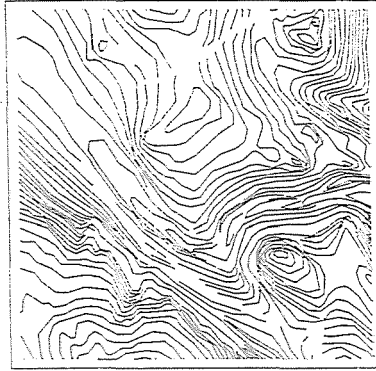


Fig. 2. Plot of the original contours from which the surface was generated

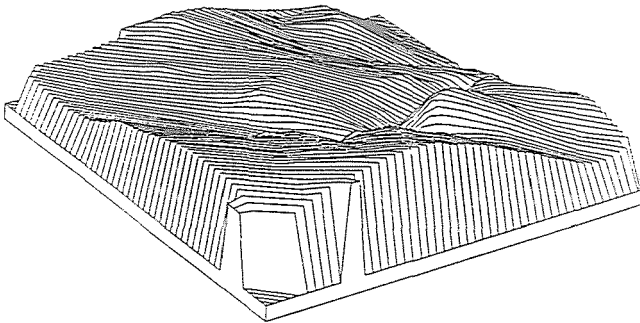


Fig. 3. Perspective surface representation

evaluation of the contour values (attributes) through a 3D surface representation.

Let us plot a contour map derived from the DEM's digital surface representation (*Fig. 4*), and compare with the original contours. At the North-East segment we can realize very significant differences because of the inconsistent surface description. The map of flat triangles (*Fig. 5*) helps to locate the problems in data capture.

Adding 45 surface specific points, placed where natural features cross a contour line and midway between contours, highly reduces the area of flat triangles. A VIP filtering is shown in *Fig. 6*: starting from a computed, gridded DEM (10,000 points), a TIN DEM with a tolerance of 1 m

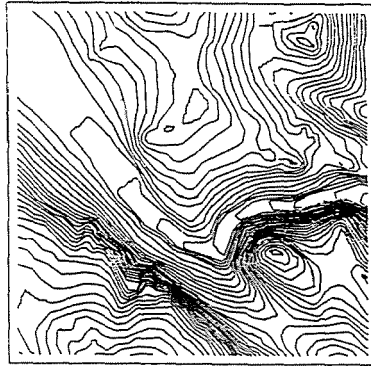


Fig. 4. Contours derived from DEM

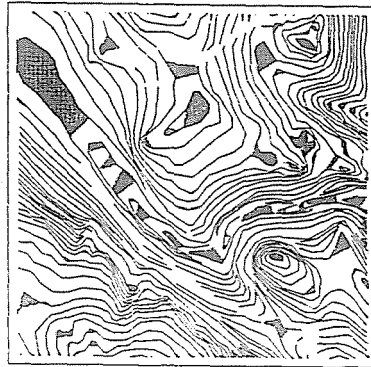


Fig. 5. Flat triangles are represented by grey-toned areas

(1800 points smaller than 50% of originally digitized points) were obtained. The database generated by VIP is an intelligent sample of the surface, since it is composed of points that have been selected to most accurately reflect the surface structure.

The task of converting a DTM of a certain representation into another structure can be handled by a combination of DEM generation and manipulation procedures.

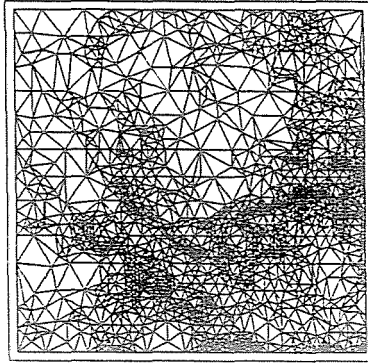


Fig. 6. VIP has selected significant points in regard to the surface

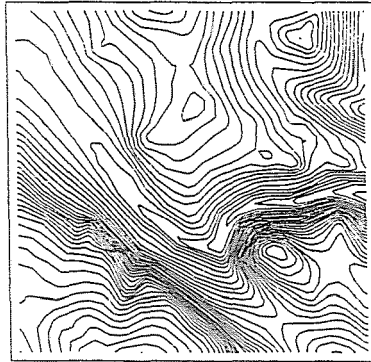


Fig. 7. Contours derived from the VIP filtered DEM

Error Propagation and Quality Model

The real problem is not the reduction of errors, instead, the more immediate and seemingly simpler task is purely to learn how to live with errors in elevation modelling. This task is seen as involving two major components; firstly, to develop an adequate means of representing and modelling the error characteristics of height data; and secondly, to develop methods and techniques that can explicitly take error into account during their operations with data. In many applications it is only necessary to be reasonable about the error assumptions rather than seek high degrees of accuracy. Furthermore, what many applications seem to require is not precise esti-

mates of error but some confidence that the error levels are not so high as to render in doubt the validity of the results (OPENSHAW, 1989).

Unfortunately, results are often judged in terms of the quality of graphics used rather than their intrinsic value. Model builders and model users usually can calibrate and validate their models carefully and carry out sensitivity analysis that show how model outputs respond to variations in data inputs and control parameters (BURROUGH, 1991). Model validation and sensitivity analysis are time consuming, and much more work has to be done on the ways in which uncertainties in the values of control parameters and input data affect results.

The situations may be subdivided into error production and error propagation. The former refers to a situation in which errors in input products are attributable mainly to the operation itself. Thus, errors may be present in output products where none existed in the data used to construct them. Error propagation is the process in which errors present in height data are passed through an operation and accumulate in output products.

In principle, the problems can be tackled by using the existing theory of error propagation. In practice, the problems are usually too far complex for such a simple minded approach. Instead it is much simpler and also far more general to seek a universal solution based around a Monte Carlo simulation approach (OPENSHAW, 1989). This data uncertainty simulation is easily applied to DEM's point data. If the output coverages are rasterized, then frequency distributions can be displayed on the top of the deterministic and error free results.

Kriging was used as the spatial prediction technique, which is based on the theory of regionalized variables. Any variable, distributed in space is by definition regionalized. Kriging is carried out in two steps. The first step involves modelling the spatial structure of the regionalized variable. In the case of stationary conditions the spatial structure can be described by the semivariogram. In the case of non-stationary conditions the spatial structure can be described by the order of the drift and the generalized covariance function. In the second step the selected model for the spatial structure is applied to the data set to predict values at desired and unmeasured locations. An interesting by-product of Kriging is the estimated Kriging standard deviation.

The quality model can be an important part of a database, and its purpose is to provide system users with quality parameters which will help them assess the quality of the data they intend to use.

Design

There are many good reasons for wanting to know the quality of the results obtained from a DEM. Without adequate quality control, investments for development cannot be put on a sure footing and it is difficult to compare the results of different analyses. A lack of quality control affects data collection and sampling because no one has any idea of just how many, and what kind of observations are necessary to support a given analysis.

There is strong significance between the terrain type, data point density and the accuracy of DEM results. FREDERIKSEN (1980) suggests to use the Fourier transformation in finding out the optimal regular grid size. Several very high density profile measurements on the modelled area serve as input for the Fourier transformation. The results of the transformation may be extrapolated for a larger area, in this way they can be used for determining the shortest terrain surface wavelength, which will be taken into account using a given regular grid size.

Another approach is given by (SÁRKÖZY, 1986): In the civil engineering cut-and-fill design, the required accuracy of results is given in 5% relative error.

$$m_r V = \frac{\Delta V}{V}.$$

In the case of a gridded DEM for the minimal number of model points (N) the following formula can be deduced

$$N = 3600 \frac{m_{\Delta z}^2}{\Delta z_k^2},$$

where $m_{\Delta z}$ is the accuracy of height differences in model points, and Δz_k is the average of height differences. Consequently, in nearly flat areas the designer needs high accuracy models.

Formalization of the knowledge that we already possess and putting that in a knowledge base next to a system would help the user to choose the optimal data capture, the best set of tools to solve the problem within constraints of data quality and cost. A really intelligent system would be able to carry out error propagation studies before a major data crunching operation in order to estimate if the methods and data chosen were likely to yield the results intended. It would report to the user where the major sources of error come from and would present him with a set of options with which he could achieve better results (BURROUGH, 1991).

Conclusions

The accuracy of DEM outputs is determined mainly by the density, distribution and the accuracy of data points: by the quality of database and by the roughness of the terrain surface itself; the effect of the interpolation method and its parameters have much smaller importance. Using advanced mathematical algorithms the results are highly influenced on the applied parameters or conditions. However, from the point of view of the cartographic quality in DEM products, the effect of interpolation is much more significant. Consequently, it is central to the data processing of any data input to DEM how the data were sampled. It is helpful if the system has intelligence to find the possible errors and if an effective user interface is used for model refinement. To save storage space and processing time, or to reduce a DEM's resolution, a data reduction method may be desirable to eliminate redundant data points (e. g. VIP method). The database generated by VIP is an intelligent sample of the surface, since it is composed of points that have been selected to most accurately reflect the surface structure.

In principle the accuracy problems can be tackled by using the existing theory of error propagation. In practice, the problems are usually too complex for such a simple minded approach. From the point of view of the system management, the real problem is not the reduction of errors, instead the more immediate and seemingly simpler task is purely to learn how to live with errors in elevation modelling. In many applications it is only necessary to be reasonable about the error assumptions rather than seek high degrees of accuracy (Taylor series, Monte Carlo simulation, quality model).

A lack of quality control affects data collection and sampling because no one has any idea of just how many, and what kind of observations are necessary to support a given analysis. Formalization of the knowledge that we already possess and putting that in a knowledge base next to a system would help the user to choose the optimal data capture, the best set of tools to solve the problem within constraints of data quality and cost.

Acknowledgement

This research is part of a larger project, which is sponsored by the Hungarian National Science Foundation under Grant No. 685. I am grateful to Prof. Ferenc Sárközy supporting this study.

References

- BURROUGH, P. A.: The Development of Intelligent Geographical Information Systems. EGIS Conference, Brussels, 1991, pp. 165-174.
- JACOBI, O.: Digital Terrain Model, Point Density, Accuracy of Measurements, Type of Terrain and Surveying Expenses. ISP Congress, Commission IV. Hamburg, 1980.
- FREDERIKSEN, P.: Terrain Analysis and Accuracy Prediction by Means of the Fourier Transformation. ISP Congress, Commission IV. Hamburg, 1980.
- LAM, S. N.: Spatial Interpolation Methods: A Review. The American Cartographer, 1983, pp. 129-149.
- LEE, J.: Comparison of Existing Methods for Building Triangular Irregular Network Models of Terrain from Grid Digital Elevation Models. International Journal of GIS, 1991, pp. 267-285.
- MAKAROVIC, B.: Progressive Sampling for Digital Terrain Models. ITC Journal, 1973, pp. 397-416.
- MÁRKUS, B.: Development of the National DEM of Hungary (Országos magassági adatbázis kialakítása). Research Report 1986.
- OPENSHAW, S.: Learning to Live with Errors in Spatial Databases. Accuracy of Spatial Databases. (Edited by Goodchild, M. - Gopal, S.), Taylor & Francis, London, New York, Philadelphia, 1989, pp. 263-276.
- SÁRKÖZY, F. - MÁRKUS, B.: Computer Aided Design in Surveying (Geodéziai AMT). Lecture Notes, Budapest, Tankönyvkiadó, 1986.
- THEOBALD, D. M.: Accuracy and Bias Issues in Surface Representation. Accuracy of Spatial Databases. (Edited by Goodchild, M. - Gopal, S.), Taylor & Francis, London, New York, Philadelphia, 1989, pp. 99-106.
- VEREGIN, H.: Error Modelling for the Map Overlay Operation. Accuracy of Spatial Databases. (Edited by Goodchild, M. - Gopal, S.), Taylor & Francis, London, New York, Philadelphia, 1989, pp. 3-18.
- WEIBEL, R. - HELLER, M.: Digital Terrain Modelling. Geographical Information Systems. Principles and Applications, (Edited by Maguire, D.J. - Goodchild, M. F. - Rhind, D. W.), Longman Scientific & Technical, 1991.
- VAN DER WEL, F. J. M.: The Integration of Remotely Sensed Data into a GIS: Estimation of the Accuracy, EGIS Conference, Brussels, 1991, pp. 1219-1227.

Address:

Dr. Béla MÁRKUS, Associate Professor
Technical University of Budapest
Department of Surveying
H-1521 Budapest, Hungary