

# PRIMARY DATA ANALYSIS AND PREPARATION FOR DTM GENERATION

G. Aumann, K. Eder, A. Pfannenstein, R. Würländer

Chair for Photogrammetry and Remote Sensing  
Technical University Munich  
Arcisstr. 21, D-8000 Munich 2, Germany  
Tel: +49-89-2105 2671; Fax: +49-89-280 95 73; Telex: 522854 tumue d  
E-mail: anton@photo.verm.tu-muenchen.de

## Commission IV

### ABSTRACT:

Experience in digital terrain modelling has shown, that analysis and preparation of the primary data is of great importance both for the productivity and the quality of DTM modelling.

The paper presents a semi-automatic procedure to analyze the primary data in a numerical and graphical way. After gross error detection and data structuring using a triangular irregular network (TIN) the density and distribution of the data is checked with the aim to get adequate parameters for the final DTM generation by the Finite Element Method.

In addition to this, tools for data completion and data refinement (e.g. automatic derivation of skeleton lines) can be supplied. The effect of the data preparation can be visualized quickly by updating the TIN and derived follow up products.

**Key Words:** Data Quality, DTM, Preprocessing.

## 1. INTRODUCTION

During the last decade digital terrain modelling technique has reached a rather high standard. Efficient program packages are available with sophisticated approaches for DTM interpolation and utilities for the derivation of various "follow-up" products (Ebner et.al., 1988, Köstli et.al., 1986).

An important component, however, the primary data analysis and preparation, is not yet solved to the benefit of practical use. Therefore a concept has been developed which supplies tools for gross error detection and quality control of the primary data, as well as for data

refinement and completion. The aim is to set up an optimal data set and to offer default values for the final DTM generation with the finite element method.

## 2. DTM PRIMARY DATA AND THEIR CHARACTERISTICS

According to the methods of data acquisition applied, a variety of input data sources and types has to be considered.

Three groups of data sources can be distinguished:

### photogrammetric data

Photogrammetric data acquisition is very common for medium and small scale DTM-projects. Data types are regular grids or equidistant profiles but also arbitrarily distributed reference points may be supplied. A peculiarity of photogrammetric data acquisition is the measurement of a variable grid (progressive sampling) where an initial grid is densified semi-automatically according to the terrain's undulation (Markarovic, 1973). Geomorphological information is given in form of break lines, skeleton lines (ridge and valley lines) and specific points (hilltops, hollows, saddle points). One of the main advantages of photogrammetric data acquisition is, that the geometric and geomorphological quality of the data can be checked by on-line verification with the stereo model (Reinhardt, 1991). Recently automatic procedures have been developed based on digital image correlation algorithms (Heipke, 1990). These approaches supply very dense point distributions but there may be areas without reference point coverage where no correlation was possible.

### tacheometric data

Direct measurement of the terrain's surface usually is carried out by self recording tacheometric equipment. The objective in the field is, to use a minimum of reference points for the surface description. Additionally planimetric objects (houses, roads, landscape elements) are measured. This points usually are added to the DTM primary data. The result are data sets with extremely inhomogeneous point distributions (figure 7), but a rather high point accuracy.

### digitized data

If DTM primary data are gained from existing maps, contours are digitized with optional distance or time increments or manual recording. In addition to that special points (height coded points) and elements (lake border lines) are recorded. The characteristic of these data sets is a very high point density along the contour lines but a deficit of information between the contours. An automatic procedure is given by scanning contour folios. After the raster/vector conversion a data set is available with about the same characteristics the digitized contours show.

### 3. DEMANDS FOR DTM PRIMARY DATA PREPROCESSING

An approach for DTM primary data preprocessing should be able to check the input data set without any a priori knowledge. There are several demands which can be formulated from a practical point of view:

- automatic determination of range in x,y and z
- plausibility checks
- information on point distribution within the project area
- interactive procedure for gross error detection and correction
- quality analysis of the input data set
- special treatment of digitized contours
- generation of reference points within defined areas (e.g. lakes)
- generation of artificial reference points within areas of sparse or none reference point coverage
- "on-line" check of the effect of the data manipulation
- proposal for parameters of DTM interpolation

For the user, it is of central importance that all tools are available within the DTM program environment.

### 4. CONCEPTION FOR PRIMARY DATA PREPROCESSING

An essential facilitation for DTM generation and also an improvement of the quality of the terrain description can be achieved by a preceding analysis and preparation of the primary data. Figure 1 shows a conception for primary data preparation. The components of this toolbox are embedded in an interactive working process which enables fast 2D and 3D data analysis.

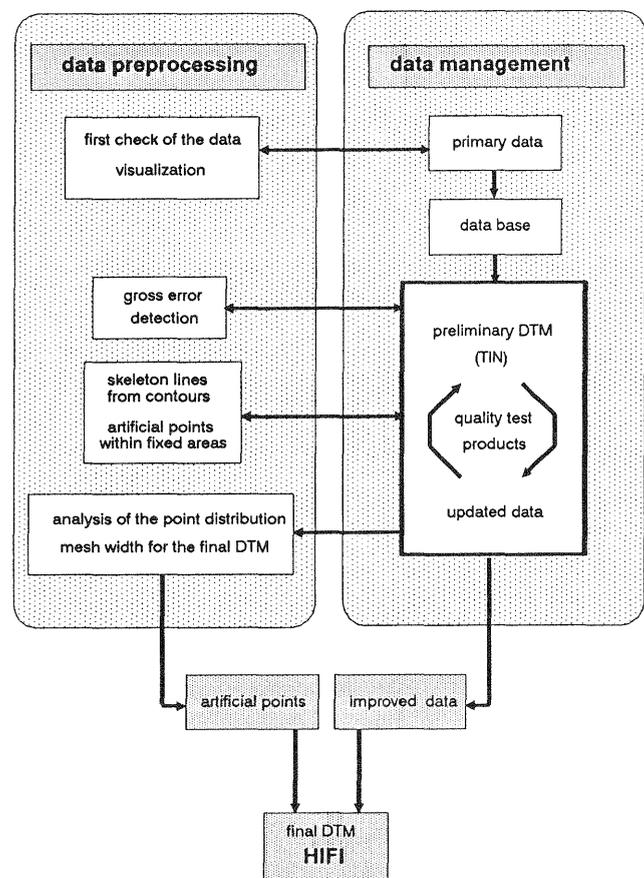


Figure 1: Architecture for data preprocessing

One part of such a toolbox is error detection and elimination. Also information about the terrain, embedded implicitly in the primary data, has to be made available for the final DTM generation. These information can be automatic derived skeleton lines when contours build the primary data set or points defining fixed areas (e.g. lakes) for the DTM surface description.

Further the primary data analysis should give an accuracy rating for the achievable terrain description out of a quality test and also an estimation of realistic parameters (mesh width) needed for the final DTM generation.

Within an other tool a homogenous point distribution can be generated to a certain extent by the interpolation of artificial points in areas without any information. By means of these artificial points an insufficient terrain description can be avoided for those critical areas.

## 5. REALIZATION OF COMPONENTS FOR PRIMARY DATA PREPROCESSING

### 5.1 Checking primary data

Primary data of large DTM's usually are submitted in several data files. The completeness of this data files for the whole DTM area has to be checked in a first step. This task can be done with tools for the presentation of data on a graphical screen or a plotter. In addition the range of the data in all three dimensions is tested to get the parameters for the data base area or to find points that are clearly out of range. Program tools for the detection of identical points or crossing lines have to be used as well.

### 5.2 Data management

Organizing primary data is important for program tools that need fast and uniform access to the data of large DTM's. These requirements are valid for programs deriving DTM follow-up products as well as for programs used to check and refine the primary data.

A DTM data base such as the HIFI data base (Ebner et.al.,1988) is able to manage large DTM areas. The primary data can be stored in a HIFI data base. For the data preparation steps as described in the following chapters it is necessary to build a DTM structure that can be updated easily. Using a preliminary DTM with the structure of a triangulated irregular network (TIN) powerful capabilities for updating data and DTM structure are available (Reinhardt, 1991). Both this TIN structure and the combined data structure of HIFI can be used within the GIS Interface of HIFI (Ebner et.al., 1990). Therefore this GIS Interface, consisting of interface subroutines for data handling, data editing and product derivation, is suited best to organize the primary data for data preparation purposes. Since the DTM of

the GIS Interface is handled within the main memory of the computer, data processing can take place interactively but is limited by the maximum number of points that can be stored within the main memory DTM (e.g. 8000 points for a main memory DTM with the size of one megabyte). The data of large DTM areas can be loaded patchwise from the HIFI data base into the main memory DTM and stored back into the HIFI data base after data correction and refinement.

### 5.3 Quality test and gross error detection

After structuring the data by a TIN the surface of this preliminary DTM can be used to estimate the data quality of the examined area as well as to detect gross errors. One algorithm for both purposes was developed and realized as a new subroutine of the GIS interface. The idea of the algorithm is to calculate the deviation between the surface represented by the TIN and the surface of the continuous terrain (see figure 2).

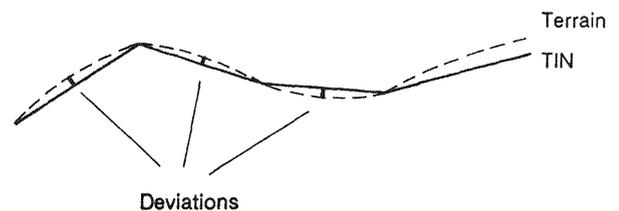


Figure 2: Deviations between TIN and continuous terrain

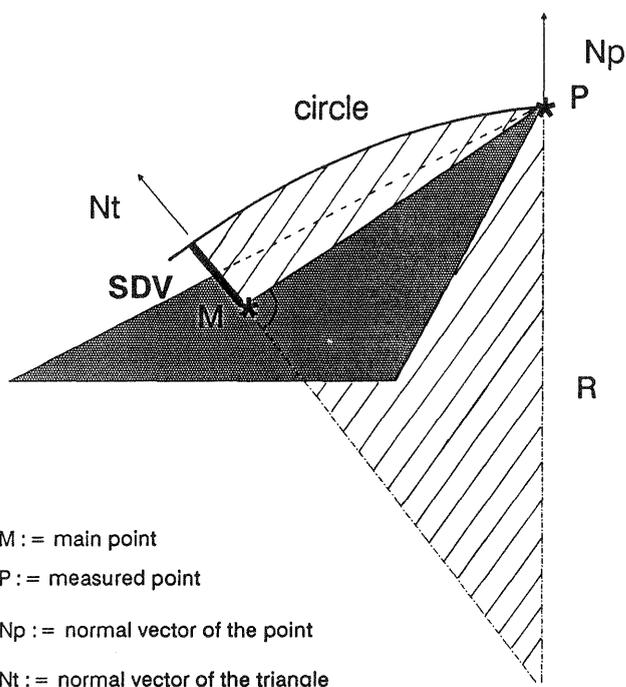
The amount of this deviation, called surface deviation value (SDV), can be used to assess the quality of terrain description by the primary data. As the surface of the continuous terrain is unknown, an assumption for the surrounding surface of a point is necessary. This surface can be described with the normal vector at the point and a constant curve assumed around the point. The normal vector of the specific point is calculated with a weighted mean value of the normal vectors of each surrounding triangle. Using a weight depending on the covered sector is important to consider the influence of the different triangles. A plane, set up by the normal vector of the triangle and the connection between the measured point and the main point of the triangle, is used for a simplified handling of the SDV computing (see figure 3). Now it is possible to calculate the constant curve defining a circle and the SDV for each triangle using the normal vector

#### 5.4 Adding information given by the data source implicitly

In case of existing contour maps as resource data an algorithm is necessary to extract the whole information from these contour maps. The digitalization of the contours yields the information on the contour lines directly. The geomorphological information, which is of essential importance, just is given by the contours implicitly. To extract this information automatically from the given contour set an approach was developed with the intention of generating DTMs from a given contour set (Aumann, et.al., 1991).

This approach will be presented shortly in the following. The automatic derivation of skeleton lines from a given contour set is founded on the calculation of special slope lines i.e. skeleton lines out of the aspect information. Arbitrary slope lines can be constructed by means of the aspect information. At the contour lines aspect vectors can be calculated as the unit vector of the bisector of the angle formed by two corresponding polygon sections. For calculating the aspect vectors in the area between two contour lines an interpolation is necessary. In this case a suited interpolation surface is given by triangle planes. This means that all contour line points are triangulated by a proper triangulation method e.g. a constrained Delaunay triangulation (TIN) (Reinhardt, 1991), already set up for other tools. As a result, the slope lines can be calculated by joining together the aspect vectors. For calculating the skeleton lines areas were chosen, where skeleton lines most probably occur. These areas, called critical areas, show up in the TIN in the form of horizontal triangles. In each critical area the skeleton line is calculated as a special slope line. The skeleton line is the one with minimum slope among slope lines in the surrounding and thus also the longest one. The starting point of the tracing should be the point which leads to the longest slope line in the area up to the next contour line. The result of the algorithm is illustrated in figure 6, where the contours and the automatically derived skeleton lines are shown. The whole information, the digitized contours and the automatically derived skeleton lines, lead to a sufficient surface description. Therefore the generation of a high fidelity DTM is possible (Aumann/Ebner, 1992).

Fixed areas (e.g. lakes) should be treated in a special way. Artificial points, which describe these areas, have



- M := main point
- P := measured point
- Np := normal vector of the point
- Nt := normal vector of the triangle

Figure 3: Definition of the surface deviation value (SDV)

of the point and the normal vector of the triangle set up at the main point. Points defining break lines are not investigated, because the terrain explicitly is not continuous at these points. The mean value of this SDV's all over the examined region, considering the area of the triangles as weight, is called the quality estimation value (QEV) and gives an idea about the accuracy of terrain description by the measured points. In case of a regular point distribution the SDV's of different regions can be used to analyze roughness of terrain. Further on this explained procedure to examine the surface of a TIN will be called quality test.

Defining a threshold value for the ratio of the maximum SDV of all triangles surrounding a point and the QEV is a new method for gross error detection. As shown in chapter 6 it performs this task very well. Graphical representation of the detected points together with the structure of the TIN or follow-up products like contours, all possible with subroutines of the GIS Interface, is a real way to check this points. Adding tools for interactive graphical editing will make this a powerful method for gross error detection and primary data update.

to be generated to avoid an insufficient terrain description caused by the interpolation algorithm of the final DTM. For that purpose the boundaries of these areas must be located. Then the planar triangles from the preliminary TIN forming these areas afford a sufficient surface description to interpolate the artificial points.

### 5.5 Data update and follow-up products

Within the GIS Interface of HIFI subroutines for data editing are available. In addition to the update of data the TIN is updated in real time. Thus after each data modification step as explained above an updated TIN is present and the modification actions can be checked at any time deriving DTM follow-up products, e.g. contours or shaded relief representations. Also the quality test is always available to check the improvements of the terrain description.

### 5.6 Point distribution

The final DTM generation with the finite element method realized in the program HIFI needs two main parameters a priori:

- the mesh width of the resulting grid for the interpolated points and
- the interpolation weight for the primary data.

Up to now not enough attention is paid to the determination of these parameters. Usually they are chosen empirically or due to the conditions of a project without knowing whether the parameters meet the requirements or not. Therefore the quality of a created DTM could be overestimated by the user or the potential of the primary data is not used fully for the DTM generation.

In the following a method for a realistic estimation of the mesh width is described. The lowest hierarchic level of the data organization in the HIFI data base is called a subarea, which includes 8\*8 meshes. Within every subarea there should exist a minimum of primary data. This minimum is the only a priori value for the following calculations.

In a first step a mesh width is automatically proposed, which is derived from the ratio of the whole DTM area and the primary data within this area. This mesh width can be changed optionally by the user. Using this mesh width a test of the homogeneity of the point distribution can be started. For that all data must be organized by subareas and the subareas by point classes (e.g. class  $i$  :

$n$  to  $m$  points). The result is represented and can be analyzed by an image with colour coded subareas or in a summary way by a histogram (figure 4).

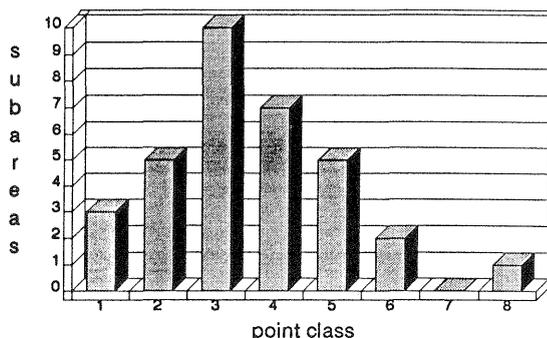


Figure 4: Histogram of a point distribution

If a great inhomogeneity within the point distribution is detected artificial points optionally are interpolated by means of the TIN within areas of sparse or none reference point coverage to support the interpolation algorithm for the final DTM.

## 6 . EXAMPLES AND EXPERIENCES

The approach has been tested by two practical examples. As a first data set digitized contours were preprocessed. The extension of the project area is about 3\*3 km and

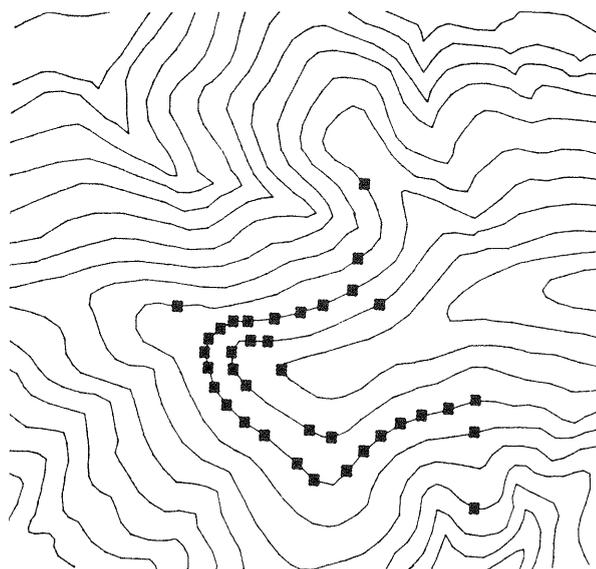


Figure 5: Detected gross errors along a piece of a contour line

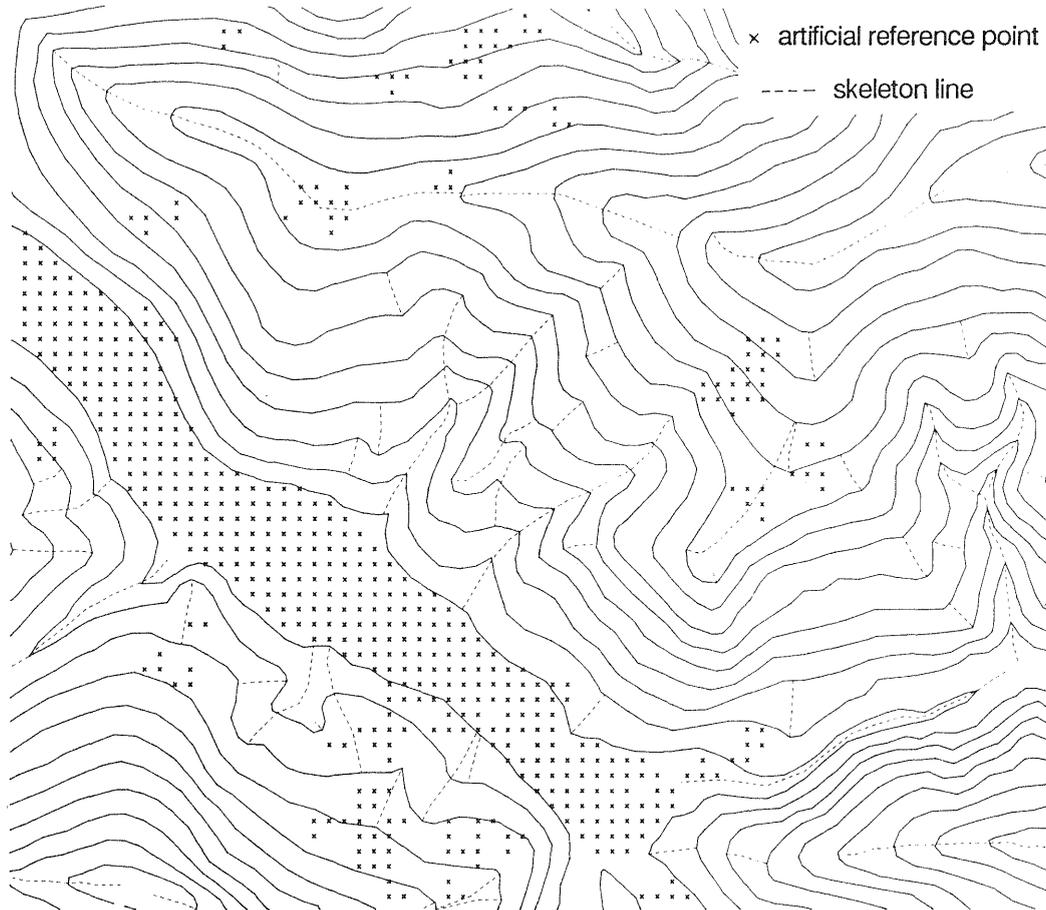


Figure 6 : Digitized contours, prepared for final DTM generation with HIFI

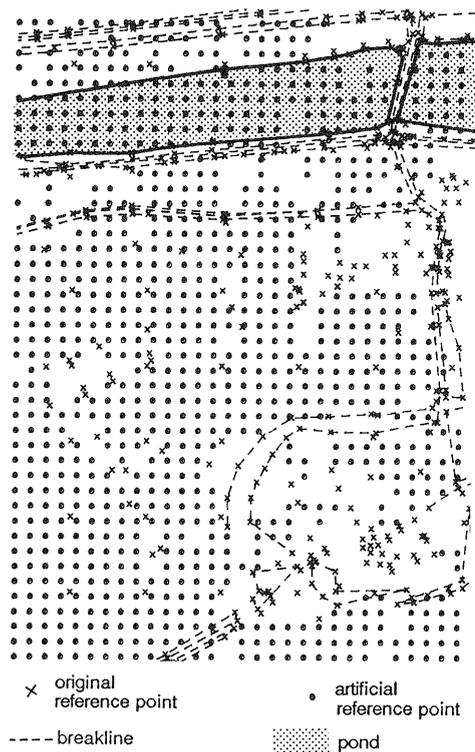


Figure 7 : tacheometric data set after primary data preprocessing



Figure 8 : final contours derived with HIFI from data represented in figure 7

the contour interval 20m. For gross error detection different errors had been introduced (single point error, pieces of contour lines with wrong height attribute).

With the utility for quality control and error detection errors greater than the contour interval were located. Figure 5 shows the SDVs greater than  $3 \cdot \text{QEV}$  along a piece of a contour line with a simulated error of 40m. Thereafter the data set was processed for automatic derivation of skeleton lines. The algorithm supplied the skeleton lines represented in figure 6. Finally the tool for generating artificial reference points was applied. During the different preparation steps, the quality estimation value (QEV) improved from 2.1m to 1.6m. The proposed grid width for DTM interpolation was 25m. Figure 6 shows the example after preprocessing and it can be considered as optimal prepared for further processing using the program package HIFI.

As a second example a tacheometric data set was tested. The point distribution was very inhomogeneous, because many planimetric elements have been measured and included into the data set. On the other side the real reference points describing the surface in the open field are rather sparse. The data set also contains a lot of breaklines and border lines of ponds.

The following results have been obtained from the analysis and preparation of the primary data :

Quality estimation value (QEV)	: 0.5 m
threshold value for gross error detection	: 2.0 m
number of artificial reference points	: 2843
proposed DTM grid width	: 12 m

Figure 7 shows a section of the primary data after preprocessing and figure 8 represents the final contours derived from the DTM generated with HIFI.

## 7. CONCLUSION

Preparation of primary data for final DTM generation by the Finite Element Method needs a number of tools as shown above. The use of a preliminary DTM with a data structure that can be updated easily enables the integration of all these tools into one program environment. Such a toolbox, based on the capabilities of the GIS Interface of HIFI for data structuring, data update and follow-up product derivation, was realized and tested at the chair of photogrammetry and remote sensing. Advantages of this integrated approach are

efficient controlling mechanism after all preprocessing steps, interactive manipulation by the user and real time update of primary data and DTM structure.

## REFERENCES

- Aumann G., Ebner, H., Tang, L., 1991: Automatic derivation of skeleton lines from digitized contours. *ISPRS Journal of Photogrammetry and Remote Sensing*, 46, 259-268.
- Aumann G., Ebner, H., 1992: Generation of high fidelity digital terrain models from contours. In preparation for XVII ISPRS Congress, Washington.
- Ebner, H., Reinhardt, W., Hössler, R., 1988: Generation, management and utilization of high fidelity digital terrain models. In: *Int. Arch. Photogramm. Remote Sensing*, Vol. 27, Part B11, pp. 556-566.
- Ebner, H., Hössler, R., Würländer, R., 1990: Integration of an Efficient DTM Program Package into Geographical Information Systems. In: *Int. Arch. Photogramm. Remote Sensing*, Vol. 28, Part 4, pp. 116-121.
- Heipke, Ch., 1990: Integration von Bildzuordnung, Punktbestimmung, Oberflächenrekonstruktion und Orthoprojection innerhalb der digitalen Photogrammetrie. *Deutsche Geodätische Kommission, Reihe C, Heft 366*.
- Köstli, A., Sigle, M., 1986: The random access data structure of the DTM program SCOP. In: *Int. Arch. Photogramm. Remote Sensing*, Vol. 26, Part 4, pp. 45-52.
- Markarovic, B., 1973: Progressive Sampling for Digital Terrain Models, *ITC Journal*, pp. 397-416.
- Reinhardt, W., 1991: Interaktiver Aufbau hochqualitativer digitaler Geländemodelle an photogrammetrischen Stereosystemen. *Deutsche Geodätische Kommission, Reihe C, Heft 381*.