Quality Control of 3D Geospatial Data

VOLKER WALTER, Stuttgart

ABSTRACT

New technologies for the acquisition of 3D geospatial data (in particular airborne and terrestrial laser scanner) as well as an increasing performance of CPUs and GPUs make it possible that 3D geospatial data can be collected for large areas and handled on standard PCs. 3D GIS technology is a young market with high growth potential and very new kind of applications are realizable. For example, the company Google introduced impressively that large volumes of 3D geospatial data can be provided in a user friendly way over the internet. However, the collection of 3D data is still a very time and cost intensive task. Quality control mechanisms are needed in order to guarantee that the capital investments are made on a sustained basis.

1. INTRODUCTION

Geospatial data are stored in Geoinformation Systems (GIS) that can be described with the IMAP model (**I**nput, **M**anagement, **A**nalysis and **P**resentation). In the last years, substantial technological progress could be observed at the Input and Presentation of 3D geospatial data. The Management and Analysis of 3D geospatial data is at the moment only particularly solved and will be an important research topic for the next years. However, with the increased abilities for the collection and visualization of 3D geospatial data very new applications are possible that open GIS technology to a broad number of users.

In the past, 3D data collection was done with stereo photogrammetry or geodetic measurements. With these techniques it was very time intensive to capture 3D data for large areas. Nowadays this situation has changed. The collection of 3D geospatial data with airborne laser scanner is meanwhile a fully developed technology. Digital Terrain Models are acquired all over the country with laser scanners for several years. With software tools it is possible to generate automatically 3D city models from laser scanner data (Brenner 2001). One of the limiting factors of airborne laser scanner is that only points on the terrain surface, building roofs or on other objects are captured and not on the side. This problem can be overcomed with terrestrial laser scanners that are commercial available for about ten years.

Terrestrial laser scanners can be used in different modes. In “static mode” the laser scanner is located at a fixed position and the laser beam is deflected vertically and rotated horizontally in order capture the surrounding area. The problem of this technique is that several acquisition positions are needed to collect data from 3D objects. The collection process can be accelerated with “stop-and-go” laser scanning, where the laser scanner is mounted on a mobile platform to speed up the transport time, but the technique is still the same. Another approach is “mobile” laser scanning, where the laser scanner is mounted on a mobile platform and the laser beam is deflected vertically with a high repetition rate while the platform is moving. This enables the very fast collection of large areas (Hunter, Cox and Kremer 2006).

The second factor for the increasing availability of 3D applications is the dramatic improvement of the performance of Graphical Processing Units (GPUs) in the recent years. The performance improvement of GPUs outpaces even Moore’s Law. For traditional microprocessors the yearly performance increase of CPUs is about 1.4 (Ekman 2004) whereas the computational capabilities of GPUs increase is about 2 (Owens et. al. 2005). Figure 1 shows the performance of GPUs compared to CPUs over the last four years.

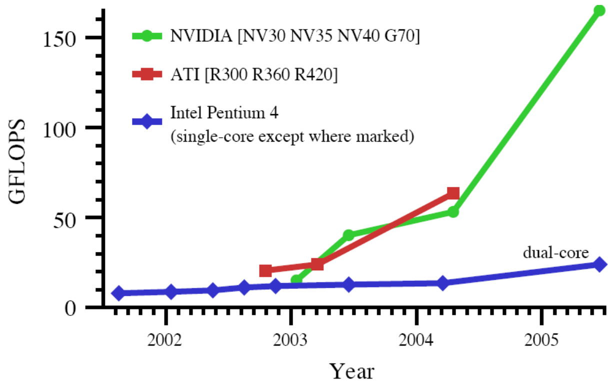

Figure 1: Comparison of the performance of GPUs and CPUs (J. Owens et. al. 2005)

The most important push for the impressive speed increase of GPUs comes from the development of computer games that demand for higher and higher GPU performance. In the recent years, computer games become an important part in recreational activities. Meanwhile, the sales figures of computer games outnumber partly even the sales figures of film industry. According to a market research of the NDP Group, sales for hard- and software for computer games in the first three months in the year 2007 were 1.1 billion Dollar (Gamesmarkt 2007).

In order to realize 3D applications, huge amount of data have to be collected. For example, a 3D city model of the city Stuttgart[1] contains more than 36,000 buildings which are represented with about 1.5 million triangles. In order to improve the visual appearance of this city model, façade textures were captured from 500 buildings that are located in the main pedestrian area. Approximately 5,000 ground based close-up photographs of the buildings were taken with a digital camera (Kada et. al. 2003).

The costs of data acquisition are in a rate of 5:1 compared with the other costs of the realization of a GIS (hardware, software, training and other services) (Stein 2006). For applications with very large datasets (like the example above) this ratio can get even worse. On the other hand, data have the longest life cycle in a GIS application.

Many data producers are collecting 3D geospatial data for local areas at the moment. This is no problem as long as the data are used only for the application for which they were acquired. If data from different producers and different areas should be integrated, measures are needed that guarantee that the quality requirements of the application are fulfilled. If those measures are not available, integration could be very difficult or even impossible.

[1] Stuttgart is a mid-sized town in the southern part of Germany and has about 600.000 inhabitants which is about 0.7% of all inhabitants of Germany (ca. 82 million)

In this paper we will discuss how 3D geospatial data can be described with quality measures. First an overview about existing and future 3D application is given. Then, general quality elements are introduced. These elements are subdivided into three classes and discussed on the focus on the handling of 3D geospatial data. A discussion with an outlook to future research concludes the paper.

2. 3D APPLICATIONS

A study of the department "Graphic Information Systems" of the Fraunhofer Institute identifies the 3D-GIS market as a sector with a high growth potential (Stein 2006). The following list shows typical applications that are based on 3D data:

- **Virtual Globes**: Google Earth and Microsoft Virtual Earth are already providing many 3D city models and it can be expected that most of all metropolises will be available in the near future. But also other 3D information can be visualized with virtual globes: for example (Tiede and Lang 2007) discuss an approach that incorporates 3D information layers from spatial analyses into virtual globes.
- **Architecture, city planning and landscape planning**: complex problems can be visualized in a much more reasonable way to decision maker and citizens with detailed 3D city and landscape models. An example for a virtual reality system for the decision support for 3D city planning can be found in (Steinicke, Hinrichs and Ropinski 2006).
- **Telecommunication**: mobile phone operating companies were one of the first driving forces for the acquisition of 3D city models in order to use them for the calculation of the propagation of electromagnetic waves (Siebe and Büning 1997).
- **Location Based Services (LBS)**: LBS are among the first applications that naturally should consider the third dimension (Verbree and Zlatanova 2007). Many users have problems to localize themselves in 2-dimensional maps. The understanding of a 3-dimensional representation is very often easier.
- **Mixed Reality Games**: new positioning technologies together with mobile computing and wireless communication enable the creation of new styles of games which combine real existing information with collected spatial data (e.g. city models) (Reimann and Paelke 2005). This could be a huge future market.
- **Civil protection**: this is an area that comprises of many different applications, like flood protection, earthquake simulations, pollutant dispersion, disaster management, etc. which have all in common that the underlying data has a 3D nature (Zlatanova, Oosterom and Verbree 2004).
- **Computer Aided Facility Management (CAFM)**: GIS and CAFM are currently undergoing the transition to storing and processing real 3D geospatial data (Fritsch and Kada 2004). Today, GIS and CAFM are based on different standards and software products. In the future these two systems will grow together and bring the indoor and outdoor world into one system.
- **Cultural Heritage**: Cultural monuments have an inestimable value. With virtual 3D models they can be renovated or even (if necessary) reconstructed. The models can also be used as basis for preservations and for interactive presentations (Wüst, Nebiker and Landolt 2004).
- **Second Live or other Web 2.0 applications**: Second Live is one of the most popular Web 2.0 applications. It represents a user interface to a virtual 3D world in that also real data can be incorporated. It is not sure at the moment whether Linden Lab (the developers of Second Live) will be a successful company in the future. However, the ideas of providing a 3D user interface to the World Wide Web will completely normal in foreseeable time (Fally

2007). The author of the present paper is convinced that a combination of an application like Second Live and a virtual globe like Google Earth will change the way how we interact with the internet completely and will be a huge market in the future.

The above list is not complete. There a many more application that are based on 3D geospatial data. At present, it can be observed that the focus of these applications is on buildings or city models. The reason for this is that the acquisition of this kind of objects is manageable with existing technology and there already exists software for an automatic acquisition of these objects. In the near future the spectrum will broaden and also other objects like trees or road furniture will be acquired.

3. DATA QUALITY

Data quality requirements are dependent on the application (fitness for use). For example, a system for automatic vehicle guidance has other quality requirements as a tourist information system. In order to describe and evaluate the quality of a database, different quality elements can be defined. In the following, we first discuss different quality elements (that are also valid for 2D geospatial data) and afterwards we group them together and discuss them especially with emphasis on the modeling and managing of 3-dimensional data.

3.1. Quality elements

Quality is defined in (ISO 2000) as the totality of characteristics of a product that bear on its ability to satisfy stated and implied needs. In the GIS world, these characteristics are traditionally called "elements of spatial data quality" (Oort 2005). Different approaches, that describe spatial quality with different numbers of elements, can be found in the literature. The following list shows those elements which are the most discussed ones. This list is not a complete enumeration of all spatial data quality elements and there are elements in the list that describe similar characteristics. Therefore, the meaning of some elements is overlapping and some elements are dependent from others.

- **Lineage** contains information about the data producer, data sources, data capturing and data processing methods.
- **Accuracy** describes the (probable) difference between the values in the database and the "true" values. It can be divided into absolute and relative accuracy of position, quantitative and qualitative accuracy of attributes and temporal accuracy.
- **Availability** refers to the time and effort which is necessary in order to get access to the data.
- **Metadata** are used for the documentation and exchange of the quality characteristics of a database. Metadata can also be described with meta-quality characteristics. For example: if the positional accuracy is estimated only from a smaller sample size, then that estimate has smaller quality (Oort 2005).
- **Completeness** refers to the extent to which all objects are present in the database. Overcompleteness can happen for example if objects, that have an area below a minimum area or minimum width, are stored in the database.
- **Correctness** indicates whether the data is captured according the data model and how well the data match the real landscape.
- **Consistency** refers to the absence of apparent contradictions in a database. Consistency is a measure of the internal validity of a database and is assessed using information that is contained within the database (Veregin 1998).

- **Up-to-dateness** contains information about the date when the data was collected or was checked.

There exist further data quality elements. For example “security” (protection against unauthorized access) could be for some tasks an important characteristic, but it is only important for few applications.

In the following, we do not discuss all elements in detail but group the elements into three different aspects “Interoperability”, “Modeling aspects” and “Automatic update and quality control”. These aspects describe data quality from different points of view. However, they are not completely independent from each other and partly describe similar things.

3.2. Interoperability

Interoperability is the ability of software systems to exchange data and methods. Standardized interfaces for the lossless exchange of information must be available in order to guarantee the security of investments. Interoperability allows different applications from different vendors to work together seamlessly and has influence on the quality elements *lineage*, *availability*, *metadata* and *consistency*.

In the following we will concentrate on the exchange of data. The exchange of methods (for example with web services) is a relatively young technology and at the moment not widely-used in the GIS world. In order to achieve interoperability, it is at first necessary that systems use standardized exchange formats. The following standards are available for the exchange of 3D geospatial data:

- **VRML** (Virtual Reality Modeling Language) was developed as a standard for the exchange of 3D data in the internet. It does not only enable the exchange of geometrical data but has also functionality for the definition of animations, light sources, sensors for user interaction, multimedia extensions and more. The exchange of 3D geospatial data is problematic, because VRML does not support geodetic coordinate systems and has only single-precision floating-point numbers.
- **GeoVRML** is an extension of VRML and enables the georeferencing of objects and the representation of complex terrain models.
- **X3D** is an extension of VRML with more functionalities.
- **GML3** (Geography Markup Language) is a XML-based specification of the Open GIS Consortium (OGC) and is based on the ISO-Standard 19107 “Spatial Schema”. GML enables the exchange of geospatial objects with attributes, relations and geometries.
- **CityGML** is based on GML3 and defined especially for the exchange of 3D city models. The semantic data model of CityGML defines the most important object classes of city models, like buildings, terrain models, waterways, streets, vegetation or road furniture. CityGML supports different Levels-of-Detail (LOD).

A standardized exchange format does only guarantee syntactic interoperability and does not solve the problem of semantic differences between datasets. If data are captured by different institutions, the data will not fit together because they will have geometrical and topological differences. The reasons are the use of different data sources, different collecting processes and different interpretation of the data sources. This problem intensifies if the data are captured in different scales or even in different data models. Integration of data from different sources is an important research topic. A solution of this problem could be for example matching techniques (instance matching

(Walter and Fritsch 1999) or schema matching (Volz and Walter 2006)) or ontologies (Fonesca and Egenhofer 1999).

Spatial Data Infrastructures (SDIs) are very important in the context of data quality and interoperability, because they provide geospatial data with defined quality standards, harmonized interfaces and standardized exchange formats. Therefore they guarantee syntactic and semantic interoperability. Spatial data infrastructures are being built up at regional, national and even international level. Currently, most of them do offer only 2D or 2.5D (terrain models) data. An example for 3D spatial infrastructure can be found in (SDI-3D 2007). In this project a spatial data infrastructure for the city Heidelberg is implemented.

3.3. Modeling aspects

Geometric modeling in 2D is straightforward: spatial objects are represented in the vector world with points, lines and polygons and in the raster world with raster cells. Geometric modeling in 3D can be done with different modeling techniques which come historically from different applications. The kind of modeling has influence on the quality elements *accuracy*, *correctness* and *consistency*. The following list enumerates the most common geometric modeling techniques for 3D geospatial data:

- **Parametric Instancing**: objects are described with a fixed set of parameters and an external orientation. For example a house with a flat roof can be modeled with seven parameters: length, width, height, x, y, z, a. For each different object type a different set of parameters has to be defined. The resulting data structure is very compact. Parametric Instancing is very suitable for buildings, because most of all buildings can be represented with a small set of different object types. However, complex buildings are very difficult to describe with this technique. Parametric Instancing is only usable for 3D models that consist of a small set of different object types with a non-complex appearance.
- **Enumeration**: is the 3D correspondent to 2D raster data. Objects are modeled with volume cells. This results in a very simple data structure. The disadvantages of this modeling technique are that the objects can only be approximated (which leads to a non-aesthetical appearance) and the data volume is typically very large. Enumeration is not suitable for very large 3D geospatial models.
- **Boundary Representation (BREP)**: is the most used modeling technique for 3D data in the GIS world. 3D objects are modeled with 0-, 1-, 2- and 3-dimensional primitives (nodes, edges, meshes and bodies). This is a very flexible technique and corresponds to the typical modeling of 2D vector data. CityGML is based on BREP representation.
- **Constructive Solid Geometry (CSG)**: complex objects are modeled by combining primitives (cubes, cylinders, prisms, pyramids, spheres, etc.) with Boolean operations. The advantage is that also very complex 3D objects can be represented with a small set of primitives. The main disadvantages are that the topology is not explicitly available and spatial analyses are very complex. CSG modeling is often used in planning tools of architects and city planners.
- **Other modeling techniques**: in Computer Aided Facility Systems (CAFM) we can find often modeling techniques that describe 3D geometry only implicitly, like 2D floor plans or cross sections. Very often these plans are even not georeferenced which makes the integration of CAFM and GIS data difficult.

It is easy to see that the modeling technique has a direct influence to the accuracy and correctness of the data. The more flexible the modeling technique the more accurate and correct can be the objects represented. The modeling has also an influence to the consistency of the data. Here we have the inverse situation: the more flexible the modeling technique the more difficult is it to check the consistency of the data. For example, it is very easy to implement automatic consistency checks for Parametric Instancing models. It can be checked that the objects do not overlap and that the parameters "make sense" (for example: non-negative height). In Enumeration modeling similar checks can ensure consistency: objects are not allowed to overlap and volume cells of an object have to be connected. In BREP modeling more sources of error are possible: degenerated faces, holes, wrong orientation of faces, degenerated bodies, etc. Errors like that are not possible with Parametric Instancing or Enumeration. Automatic consistency checks of CSG models are mathematical challenging because of the manifold possibilities how an object can be constructed.

Inconsistencies can automatically be detected but in many cases not automatically corrected. For example: in a database are two buildings that are overlapping. This situation can easily be detected but normally not automatically corrected, because the software cannot decide which of the two buildings is collected wrongly. This involves the work of a human operator.

Another aspect of 3D data modeling is the representation of data in different Levels-of-Detail (LOD). (Gröger and Kolbe 2003) suggest five different LODs for city models:

- **LOD 0**: regional model, 2.5D DTM and 3D landmarks
- **LOD 1**: city model, block models without roof structures
- **LOD 2**: city model, houses with roof structures and textures, vegetation
- **LOD 3**: city model, detailed house structures, vegetation, road furniture
- **LOD 4**: indoor model, detailed architecture models

LOD representations are important for example for real-time visualizations, visualization on mobile devices with low CPU power or small screens or if data should be transferred through narrow-band networks (mobile phone network). Even though that the LODs are defined with further characteristics (for example minimal acquisition size or point accuracy) there are still lot variants possible how an object can be captured. Different operators will come to different results. The same problem exists with automatic generalization software. Because different generalization software are based on different algorithms, the results will differ. The open question is how these inconsistencies can be handled.

Another quality factor of 3D city models is the quality of the façades that should be captured for LOD 2 and higher. Façade images can be extracted from aerial images or from terrestrial images. There are a lot of factors that influence the quality of façade images: spatial and radiometric resolution, light conditions, accuracy of inner and exterior orientation, contrast, disturbing objects (like cars or pedestrian), etc. The quality of the façade images has a very strong influence to the visual appearance of a city model. At the moment we have no quality model to describe these characteristics.

3.4. Automatic update and quality control

Automatic quality control is needed in order to estimate automatically the *accuracy*, *completeness* and *correctness* of spatial data. This can be done by checking the data on sample areas or on the whole dataset. Automatic update is needed to improve automatically the *accuracy*, *completeness* and *correctness*. Both approaches work very similar: input data (laser data, aerial images, terrestrial images, etc. and already captured 3D data) are processed with image interpretation algorithms in order to control the existing objects and to find new objects that are not in the database.

One of the main problem of automatic update and quality control of 3D geospatial data is the availability of adequate input data: airborne laser scanner capture only points from the object surface; aerial photographs or satellite images can capture images from the side, but only from parts of 3D objects and with a limited image quality; terrestrial laser scanner data or terrestrial images are not available for large areas and with high repetition rate, because of the time intensive acquisition process.

An approach to overcome this problem could be the use of data from mobile sensor networks. Distributed imagery collection is one the hot research topics in the next years. The idea is to capture data with low quality but with high redundancy. Many redundant representations of the same object can be transformed in one high quality model. Data sources could be imagery and video streams from PDAs or mobile phones, from car-mounted sensors or even from smart dust sensor networks. First results of this research area are published in (Agouris, Duckham and Croitoru 2007).

4. DISCUSSION

The recent hardware developments will continue in the foreseeable future. GPUs and CPUs will profit more and more from parallel processing cores. Graphic accelerators will also be available for mobile devices and internet applications will profit from GRID technologies. Also we will see an increasing availability and use of 3D geospatial data. On the other hand there are an increasing number of users which are less aware of the quality of geospatial data because there is an increasing distance between those who use spatial data (the end users) and those who are best informed about the quality of spatial data (the producers) (Oort 2005). However, GIS applications are using any kind of data, independent from their quality. The problem is that the data quality is not integrated in data management and analysis and therefore no information about the quality of the result is available. The second problem is the lack of interoperability. Especially 3D city models are meanwhile available for many cities. But typically they are acquired and managed by different organizations, with different quality characteristics, in different systems, data models and exchange formats. This makes it difficult to integrate these datasets or to access them in a uniform way.

Data quality descriptions are a very important part of geospatial data. Producer should document the quality of their data and exchange them in a standardized format. Very often, geospatial analyses are based on different datasets with different data quality characteristics. In that case we need instruments in order derive quality information for the aggregated data. Also the results of other algorithms have to be evaluated. For example, automatic generalization tools transform data from one scale into a smaller scale. At the moment we have no approaches that measure the quality of the results of a generalization. A lot of research has still do be done in this area. Especially the quantification of data quality in spatial analyses is a very important topic. Other research topics are for example quality improvement with mobile sensor networks or visualization of data quality.

5. REFERENCES

Agouris, P., Duckham, M. and Croitoru, A. (2007): From Sensors to Systems: Advances in Distributed Geoinformatics (currently under preparation).

Brenner, C. (2001): City Models - Automation in Research and Practice. In: Fritsch, D. and Spiller, R. (eds.): Photogrammetric Week '01, Herbert Wichmann Verlag, Heidelberg, 149 – 158.

Ekman, M., Warg, F. and Nilsson, J. (2004): An In-Depth Look at Computer Performance Growth. Technical Report 2004-9, Chalmers University of Technology, Department of Computer Engineering, Göteburg.

Fally, G.M. (2007): Second Live. GIS Business, 6/2007, 42 – 45.

Fonseca, F. and Egenhofer, M.J. (1999): Ontology-Driven Geographic Information Systems. ACM-GIS, 14 – 19.

Fritsch, D. and Kada, M. (2004): Visualisation using game engines. In: Proceedings of the XXth ISPRS Congress, Commission V, WG 6.

Gamesmarkt (2007): US-Markt wächst weiter. Nr. 8, 7. Jahrgang, p 1.

Gröger, G. and Kolbe, T.H. (2003): Interoperabilität einer 3D-Geodateninfrastrukur. In: Bernhard, L., Sliwinski, A. and Senkler, K. (Hrsg.): Geodaten- und Geodienste-Infrastrukuren – von der Forschung zur praktischen Anwendung. Beiträge zu den Münsteraner GI-Tagen, Münster.

Hunter, G., Cox, C. and Kremer, J. (2006): Development of a commercial laser scanning mobile mapping system – Streetmapper. Second International Workshop: The Future of Remote Sensing, Antwerp, 17-18 October.

ISO (2000): Quality Management Systems - Fundamentals and Vocabulary (ISO 9000:2000).

Kada, M., Roettger, S., Weiss, K., Ertl, T. and Fritsch, D. (2003): Real-Time Visualization of Urban Landscapes using Open-Source Software. In: Proceedings of ACRS 2003 ISRS, Busan, Korea.

Oort, P. (2005): Spatial data quality: from description to application. Publications of Geodesy 60. NCG, Nederlands Commissie voor Geodesie, Netherlands Geodetic Commission, Delft.

Owens, J.D., Luebke, D., Govindaraju, N., Harris, M., Krüger, J., Lefohn, A.E. and Purcell, T.J. (2005): "A Survey of General-Purpose Computation on Graphics Hardware." In: Eurographics 2005, State of the Art Reports, August 2005, 21 – 51.

Reimann, C. and Paelke, V. (2005): Adaptive mixed reality games. In: ACM SIGCHI International Conference on advances in computer entertainment technology, Valencia, Spain, 302 – 305.

SDI-3D (2007): SDI-3D, Spatial Data Infrastructure for 3D-Geodata. i3mainz - Institut für Raumbezogene Informations- und Messtechnik, Forschung, Entwicklung, Dienstleistungen und mehr (http://www.i3mainz.fh-mainz.de/Article291.html).

Siebe, E. and Büning, U. (1997): Application of Digital Photogrammetric Products for Cellular Radio Network Planning. In: Fritsch, D. and Spiller, R. (eds.): Photogrammetric Week '97, Herbert Wichmann Verlag, Heidelberg, 159 – 164.

Stein, C. (2006): 3D GIS Markt – Studie 2005. Fraunhofer Institut für Graphische Datenverarbeitung IGD, Abteilung Graphische Informationssysteme, Darmstadt.

Steinicke, F., Hinrichs, K. and Ropinski, T. (2006): A hybrid decision support system for 3D city planning. ISPRS Technical Commission II Symposium, Vienna, 103 – 108.

Tiede, D. and Lang, S. (2007): Analytical 3D views and virtual globes - putting analytical results into spatial context. Joint Workshop "Visualization and Exploration of Geospatial Data", June 27-29, Stuttgart (Germany).

Verbree, E. and Zlatanova, S. (2007): Positioning LBS to the third dimension. In: Gartner, Cartwright and Peterson (eds.): Location Based Services and TeleCartography. Springer-Verlag, Heidelberg, 107 – 116.

Veregin, H. (1998): Data Quality Measurement and Assessment. NCGIA Core Curriculum in GIScience (http://www.ncgia.ucsb.edu/giscc/units/u100/u100.html).

Volz, S. and Walter, V. (2006): Linking Different Geospatial Databases by Explicit Relations. Geoinformation Science Journal, Vol. 6, No. 1, 41 – 49.

Walter, V. and Fritsch, D. (1999): Matching Spatial Data Sets: a Statistical Approach. International Journal of Geographical Information Science (IJGIS), 13(5), 445 – 473.

Wüst, T., Nebiker, S. and Landolt, R. (2004): Applying the 3D GIS DILAS to archeology and cultural heritage projects – requirements and first results. In: Proceedings of the XXth ISPRS Congress, Commission IV.

Zlatanova, S., van Oosterom, P. and Verbree, E. (2004): 3D technology for improving disaster management: GeoDBMS and positioning. In: Proceedings of the XXth ISPRS Congress, Commission VII, WG VII/5.