Eurographics Italian Chapter Conference (2007)
Raffaele De Amicis and Giuseppe Conti (Editors)

Methods for Spatial Data Quality of 3D City Models

Michel Krämer, Jörg Haist and Thorsten Reitz

Fraunhofer-Institut für Graphische Datenverarbeitung, Darmstadt, Germany

Abstract

3D city models are nowadays used in very different applications. Due to this, the commercial exchange and, thus, the demand for high quality data becomes more and more important. We describe a quality model that defines common parameters used for spatial quality measurement, especially when dealing with 3D city models. Therefore, we explore different representations of city models: the reality, the user's idea and the digital data set. A well-defined mathematical formalism which addresses the different quality parameters is presented. This formalism also helps to create algorithms for measurement and improvement of spatial data quality. The aim is to be able to define uniformed criteria which can be easily transformed into software. The implemented prototype serves as base for the evaluation in which we provide examples based on actual data sets using typical quality requirements.

Categories and Subject Descriptors (according to ACM CCS): I.3.6 [Computer Graphics]: Methodology and Techniques

1. Introduction

3D city models are used in a large number of practical applications, where high quality data has become indispensable. For example, when city models are – in conjunction with urban development – used to plan new buildings, it is important to have a data set that is as complete as possible (at least in the near area around the new buildings). So, the planner can figure out if the new buildings fit into their surrounding area or if they cause problems like unexpected shadowing. Another application is the preventive environmental protection, where city models are used to simulate everyday problems like noise pollution and also more heavy ones like flood catastrophees. In order to create procedures/action plans for flood and noise protection, one needs data sets where object geometries are as exact as possible (or at least as exact as it is needed for the specific application).

But completeness and exactness of object geometries are not the only quality characteristics. Think of the application of city models in tourism. Tourists, who use a geographical information system to inform theirselves about cities they may want to visit in the future, do not care about how exact the 3D coordinates are. Instead, they expect the visualization of the scene to be close to reality. So, object attributes have to be accurate, i.e. the building's color and texture if available. The name and address also have to be correct, as tourists may want to use a navigation system to find sights in reality later. Moreover, geographical information systems often offer 4D data. Therefore, it is important to have a data set with correct temporal information which is also very up to date.

Hence, incorrect data can lead to economic difficulties. Institutions and companies buying spatial data often realize that there are serious errors in their data sets or that their city models are out of date. Thus, they spend a lot of money and effort to correct these errors, although it would be easier to buy better data sets from 3^{rd} party companies. The problem with such data sets is that there is no common definition for spatial data quality. What do above-mentioned terms like *completeness of objects* or *attribute accuracy* exactly mean? Today we nearly do not find any exchange of harmonized data quality data. If there was a well-defined *quality model* describing common terms for spatial data quality, buyers and sellers would be able to determine the quality of their data sets in a common way.

The main contributions of this paper are as follows:

- We describe a quality model, which defines the terms mentioned above. It uses a well-defined mathematical formalism, so different ideas about quality parameters are eliminated.
- The formalism can be used to create algorithms for measurement and improvement of city model quality. Some

examples will be given in section 5. Herewith it is possible to accomplish automatic quality analysis.

2. Related work

In 1995 S. C. Guptill and J. L. Morrison published the book *Elements of Spatial Quality* [GM95] on behalf of the International Cartographic Association (ICA). It described seven categories of spatial quality: Lineage, Positional Accuracy, Attribute Accuracy, Completeness, Logical Consistency, Semantic Accuracy and Temporal Information, thus providing the elements for a unified quality model. Regarding applications of this model and other quality models, there have been projects and research in which at least portions of it were applied, often to single data types. In [MWLP03], ATKIS cadastral data was evaluated regarding their geometric accuracy, up-to-dateness and the semantic accuracy of the descriptions by comparing this data to grid data like orthophotos. Another technique to determine quality aspects of geodata, especially positional accuracy and completeness, was presented in [RW98]. Here, topological and geometrical differences are being calculated using region adjacency graphs and zone skeletons which are created for two different data sets and then correlated.

[FGS02] shows that knowing about the lineage of positional data is very important for the assessment of data quality and argues that current systems are rather agnostic to this fact. The International Federation of Surveyors has also given the topic a higher level of importance by establishing a working group on the topic of Quality Management for Geodata, but so far mainly concentrates on positional accuracy (see [Sch06]).

To define guiding rules and standards regarding geodata and attached metadata has also been the aim of several standardization bodies. The Comité Européen de Normalisation (CEN), Technical Commitee 287 (CEN/TC287) started to develop the European draft standard DIN V ENV 12656 [Deu98] in 1998, but the work has been discontinued later and the results were sent to the International Organization for Standardization (ISO), since the Technical Committee 211 (ISO/TC211, "Geographic information/Geomatics"), addressed the same problems. In the meantime the ISO/TC211 has published a set of international standards related to spatial data quality: ISO 19113 (Quality principles, [Int02]) and ISO 19114 (Quality evaluation procedures, [Int03b]), which are parts of the standard ISO 19115 (Metadata, [Int03a]), contain definitions from DIN V ENV 12656 and some extensions. The draft standard ISO 19138 [Int05] will address quality measures in the future.

3. Conceptual reality

A 3D city model may not always contain all objects that can also be found in reality. Most of all, the user would like to have buildings and terrain, whereas cars, signs, traffic lights and the like are not always needed. Thus, there is a difference between reality and the user's idea about the city model. This idea is called the *conceptual reality* (see [Deu98]). Each object in this reality belongs to a certain class, i.e. a group like *Buildings*, *Trees*, etc.

There's another difference between the digital data set and the conceptual reality. If the data set has been captured by laser scanning (see [Maa05]), for example, there may be a lot of errors, most likely missing objects or misclassified ones.

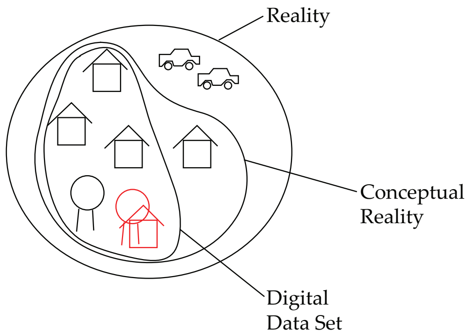

Figure 1: *The difference between reality, conceptual reality and the digital data set.*

Figure 1 shows the difference between the three ideas. Cars are not needed by the user and, thus, they don't appear in the conceptual reality. One building has not been captured by the laser scanner. So, the digital data set doesn't contain it. Moreover, one house has been mistakenly classified as a tree.

The conceptual reality may be used to measure quality. As mentioned above, it is the user's idea about the city model and it may be specified in different kinds: as another 3D model, as a simplified model (for example in 2D), as a simple list of attributes or as something else that can be compared to the digital data set. Previous research has shown, that there are six criteria for quality measurement (see [GM95], [Joo98] or [Int03a]):

1. **Positional Accuracy**: The 3D coordinates of all objects have to be as exact as possible (close the ones in the conceptual reality)
2. **Completeness**: Objects and attributes must be complete
3. **Semantic Accuracy**: Classification of objects must be correct and object attributes must have valid values
4. **Correctness**: Object attributes must have correct values
5. **Temporal Conformance**: Objects must be within defined time constraints (see below)
6. **Logical Consistency**: Logical rules (e.g. all object faces must be oriented clock-wise) have to be consistent for all objects.

These criteria are sometimes called the *Elements of Spatial Data Quality* [GM95] or *Quality Parameters* [Deu98].

4. Quality model

As mentioned above, a well-defined quality model is needed to describe common terms, i.e. the elements of spatial data quality. We use a mathematical formalism to define this model. The following set will be used:

$\mathbf{A} = \{\mathbf{O_1}, ..., \mathbf{O_n}\}$ This set contains all objects from the conceptual reality
$\mathbf{A'} = \{\mathbf{O'_1}, ..., \mathbf{O'_m}\}$ A set of all objects from the digital city model
$\mathbf{E} = \{\mathbf{P_1}, ..., \mathbf{P_r}\}$ This set contains all possible object attributes from the conceptual reality
$\mathbf{E'} = \{\mathbf{P'_1}, ..., \mathbf{P'_s}\}$ A set of all digital object attributes from the digital data set.

Furthermore there are two functions used to map the conceptual reality to the digital data set:

$\mathbf{f} : \mathbf{A} \rightarrow \mathbf{A'}$ Maps objects from the conceptual reality to their digital equivalents
$\mathbf{g} : \mathbf{E} \rightarrow \mathbf{E'}$ Maps conceptual object attributes to their digital representations.

With these terms, it's now possible to define formulas for each element of spatial quality.

4.1. Positional Accuracy

Often it's not possible to achieve full exactness of 3D coordinates. Current data acquisition methods are too imprecise. Instead, the user is content with a certain probability that all coordinates are within a given confidence interval. For example, the user may define a constraint, that $\alpha = 90\%$ of all coordinates must have a maximum error of $\pm 50cm$ (that means, the confidence interval has an upper limit C_u of $50cm$ and a lower limit C_l of also $50cm$). Therefore the random error ε has to be calculated for all coordinates as shown in figure 2. The systematic error Δ_x is the same for all coordinates and can be ignored.

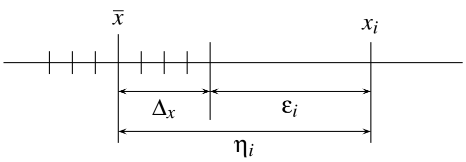

Figure 2: *There's an error η_i between a coordinate value $\bar{x}$ from the conceptual reality and the corresponding digital value x_i, whereas η_i is the sum of the systematic error Δ_x and the random error ε_i*

The probability $P_i(C_l \leq \varepsilon_i \leq C_u)$ that the error of a certain coordinate x_i is within the interval $[C_l, C_u]$ can then be calculated. Doing this for all coordinates finally leads to the mean probability $P = \frac{1}{n}\sum_1^n P_i$ which has to be checked against the value given by the user, that means P must be less than α. Otherwise the coordinates are not accurate.

The value of this quality parameter is a relative one. If you apply it to a large data set, you will most likely calculate P for each object. At the end you will determine how many objects match the user's constraint ($P < \alpha$) in relation to the number of all objects in the city model.

The positional accuracy is often separated into two parts: the horizontal and the vertical accuracy. The horizontal one is determined by calculating the error in x/y-direction, whereas the vertical accuracy uses the object's height.

4.2. Completeness

As mentioned above, there must be a way to check for completeness of objects and for completeness of attributes.

Definition 1 *(**Completeness of objects**): Let $f : A \rightarrow A'$ be the function that maps objects from the conceptual reality to the digital data set. All objects are complete if the following formula is true:*

$$\forall O \in A \;\; \exists O' \in A' \;\; \left(f(O) = O'\right) \;\; \wedge$$
$$\forall O' \in A' \;\; \exists O \in A \;\; \left(f^{-1}(O') = O\right)$$

That means, objects are complete if there is exactly one O' for each O and vice versa. Thus, f is a bijection.

Definition 2 *(**Completeness of attributes**): Let E_O be the set of all attributes of an object O and $E_{O'}$ the set of all attributes of a digital object O' respectively. Let $g : E_O \rightarrow E_{O'}$. All attributes are complete if*

$$\forall O \in A \;\; \exists O' \in A' \;\; \Big(\left(f(O) = O'\right) \;\wedge$$
$$\Big(\forall P \in E_O \;\; \exists P' \in E_{O'} \;\; (g(P) = P') \;\; \wedge$$
$$\forall P' \in E_{O'} \;\; \exists P \in E_O \;\; (g^{-1}(P') = P)\Big)\Big)$$

is true.

So, for each conceptual object O there's a digital O' which has exactly the same attributes. $g : E_O \rightarrow E_{O'}$ is a bijection.

There may be two results for this quality parameter. The number of missing objects/attributes and the number of objects/attributes that actually don't exists in reality. The last ones could have been added to a city model during data acquisition, for example, if the laser scanner was too imprecise and there were heaps classified as buildings by accident.

4.3. Semantic Accuracy

Semantic Accuracy is given when

a) all objects are correctly classified and
b) all attributes have valid values.

The first point is quite obvious, since classification is one of the most common problems in semantics. The second one can be explained with an example: The attribute “type” could have the values “Ferrari F430” or “Porsche 911 Turbo”, but also “flat” or “hipped”, depending on if the object is a car or a roof.

This results in the following two definitions:

Definition 3 ***(Classification)***: *Let C_O be the class of an object O, and $C_{O'}$ the class of a digital object O' respectively. Classification of objects is correct if the following formula is true:*

$$\forall O \in A \;\; \exists O' \in A' \;\; \Big((f(O) = O') \;\; \wedge \;\; (C_O \; = C_{O'}) \Big)$$

So, for each object O its digital representation O' has to belong to a class, equal to the one of O.

Definition 4 ***(Semantic accuracy of attributes)***: *Let $V_{P'}$ be the set of all valid values for the property $P' \in E_{O'}$. Let $val(P')$ be the value of P'. All attributes are valid, if*

$$\forall O' \in A' \;\; \forall P' \in E_{O'} \;\; \Big(val(P') \in V_{P'} \Big)$$

is true.

For large data sets a relative result can be calculated for this quality parameter by counting the number of objects/attributes that are accurate in relation to the number of all objects/attributes.

4.4. Correctness

Object attributes have to be correct. That means, that their digital values must be equal to the ones in the conceptual reality. Again, a relative value can be calculated.

Definition 5 ***(Correctness)***: *Digital values are correct if the following formula is true:*

$$\forall O \in A \; \exists O' \in A' \left(\Big(f(O) = O' \Big) \; \wedge \right.$$
$$\left. \Big(\forall P \in E_O \; \exists P' \in E_{O'} \; (g(P) = P' \wedge val(P) = val(P')) \Big) \right)$$

That means for each object O and its digital representation O', corresponding attributes must have the same value.

4.5. Temporal Conformance

Temporal Conformance consists of four points, referring to the accuracy of time measurement and also to the data’s actuality (see [Deu98]):

a) Accuracy of time measurement
b) Date of the last data update
c) Update frequency
d) Temporal validity

The user can define constraints according to these points. For example, the accuracy of time measurement (a) should be better than a certain value. The date of the last update (b) tells something about the quality of spatial data. If a data set is rather old, it may not reflect reality any longer. Thus, a temporal validity (d) is often given, meaning, that if a data set gets older than a certain number of days, months or years it becomes invalid. The update frequency (c) can also be very interesting when a data set is used in geographical information systems which can handle temporal (4D) data.

As mentioned above, the user defines constraints. Temporal Conformance will be given, if the data set matches these constraints.

4.6. Logical Consistency

Logical Consistency describes certain rules from the following categories:

1. Geometrical Consistency (e.g. “*All points must be 3D*”)
2. Topological Consistency (e.g. “*All line strings must be closed*”)
3. Semantic Consistency (e.g. “*All churches must have the same map symbol*”)
4. Format Consistency (“*The data set’s format (file format, etc.) must match the given specifications*”)

So, the user defines rules from these categories and all objects must follow them:

Definition 6 ***(Logical Consistency)***: *Let $R_{O'_i}$ be the set of all rules for a digital object O'_i. Logical Consistency will be given, if all objects follow the same rules:*

$$\forall (O'_i, O'_j) \in A' \;\; (R_{O'_i} \; = \; R_{O'_j})$$

5. Evaluation

Algorithms for the quality parameters described in this paper were implemented using the *CityServer3D* technology [Rei05]. This Java client/server application supports several file formats like VRML, GML or CityGML [KGP05] which are often used to save large city models.

In this paper we present two examples: Calculation of the completeness of objects and positional accuracy.

5.1. Completeness of Objects

We used a city model of Darmstadt, Germany which was used in practise to evaluate an algorithm for calculating the completeness of objects. The data set has been created using orthophotos in conjunction with data from the land registry office, whereas most of the 3D objects have been modeled by hand. A medium completeness was expected, since many small, less important objects have not been modeled. The data set contained buildings and vegetation (see figure 3). A very up-to-date data set of ground plans from the land registry office, which could be considered quite complete, was used as the "conceptual reality".

Figure 3: *A city model of Darmstadt, Germany used in practise*

The algorithm created ground plans from all 3D objects by intersecting them with a virtual terrain. After that each ground plan was spatially intersected by the land registry office data. If the result set was not empty, the current object was considered existing in reality.

The algorithm found 19,849 objects, whereas the land registry office data set contained 50,541 ground plans. This led to a relative completeness of $\frac{19849}{50541} = 39.27\%$. For some 3D buildings, more than one ground plan could be found in the land registry office data set (especially for terrace houses). So, the algorithm was changed and the final results were much more realistic: A completeness of 66.63% was calculated, whereas the algorithm also determined that 6.4% of the digital objects did not exist in reality. This is because the data set from the land registry office contained building ground plans only. There was no information about vegetation.

5.2. Positional Accuracy

The same data sets were used to calculate the horizontal accuracy. Assuming that all digital objects could be found in the land registry office data set, the algorithm calculated a so-called error band (ε-band) around each ground plan (see figure 4). As mentioned above, coordinates which were within this band were considered accurate. The other ones were considered errors.

A rather high accuracy of at least 90% could be expected, because a large amount of 3D objects were modeled by hand. The algorithm was run three times with different values for ε:

- $\varepsilon = \pm 1m \longrightarrow accuracy = 72.94\%$
- $\varepsilon = \pm 2m \longrightarrow accuracy = 86.47\%$

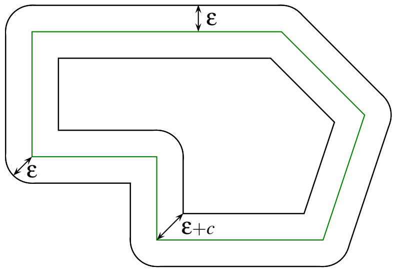

Figure 4: *Error band (ε-band) around a ground plan*

- $\varepsilon = \pm 4m \longrightarrow accuracy = 91.62\%$

So, the expected accuracy of 90% could be achieved with a maximum error of $\pm 4m$.

6. Conclusions

We present a quality model defining common parameters used for spatial quality measurement. Several quality parameters for 3D city models are described by a well-defined formalism. The developed algorithms provide measurement and also automatic improvement of data quality.

The documents mentioned in section 2 define several quality parameters. This paper includes all of them and adds the parameter "Correctness". Moreover, it defines a mathematical formalism, which can be used to create algorithms, like we did in section 5. The well-defined rules prevent users from having differing concepts about quality parameters.

The quality parameter "Completeness of Objects" may be interpreted in several ways: a city model may include different objects, but a bench in a park may be less important than a high-rise building for some applications. So, the question arises, if the completeness should be calculated for all objects or just for a single class. However, this paper uses ideas from well-known industry standards like [Deu98] or [Int02], which also do not differentiate between object classes. There are two reasons for this:

1. Practise has shown that current city models are often optimized for a special application. For example, data sets which are used for urban planning often contain buildings and vegetation only. Irrelevant objects like cars or traffic lights are not included.
2. The data set can be filtered very easily. If, for a certain application, only buildings are needed, all other objects can be removed. The only requirement is that the objects' classes are known. Modern geographic information systems don't allow objects without classes anymore, so this requirement does not represent a real problem.

7. Outlook

There are two topics not addressed in the actual work. On the one hand, texture quality is of main interest in visualizationcentered applications, and on the other hand, a user interface was not realized.

We do not address errors that can occur when textures are used with city models: if textures are created by photographing an object from a large distance, other objects (like persons, trees, etc.) appear on it. Removing these artifacts is a common problem in city model acquisition. Other common errors are distortion and forged colors. These errors cannot be fixed without complex algorithms. They should be addressed in a future work. More about textured city models can be found in [FZ03] and [FSZ04].

We have realized the presented ideas in a server environment which enables us to provide these mechanisms within web services. However, there are still a lot of use cases in which user interaction is needed. So, an integration into a client software is planned. With a graphical interface, users will be able to choose quality measurement mechanisms, to receive data quality reports and to see data quality parameters in the 3D visualization. Therefore, we plan to use standards like symbology encoding and filters.

With this, it will be up to the user if all parameters are used to estimate the quality of a city model. Of course, if different data sets shall be compared to each other, the same parameters must be used. So, they will be storable as configuration.

8. Acknowledgments

This research was supported by Prof. Dr.-Ing. Aristovoulos Christidis from the University of Applied Sciences Gießen-Friedberg, Germany.

References

[Deu98] DEUTSCHES INSTITUT FÜR NORMUNG E. V.: *Geoinformation – Datenbeschreibung – Qualität*, DIN V ENV 12656, 1998.

[FGS02] FISHER P., GOODCHILD M., SHI W. (Eds.): *Spatial Data Quality*. Taylor & Francis, 2002.

[FSZ04] FRÜH C., SAMMON R., ZAKHOR A.: Automated texture mapping of 3D city models with oblique aerial imagery. *3dpvt* (2004), 396–403.

[FZ03] FRÜH C., ZAKHOR A.: Constructing 3D city models by merging aerial and ground views. *IEEE Computer Graphics and Applications 23*, 6 (Nov/Dec 2003), 52–61.

[GM95] GUPTILL S. C., MORRISON J. L.: Elements of spatial quality, *Elsevier Science, Kidlington, Tarrytown, Tokyo*, 1995.

[Int02] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION: *Geographic information – Quality principles*, ISO 19113, 2002.

[Int03a] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION: *Geographic information – Metadata*, ISO 19115, 2003.

[Int03b] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION: *Geographic information – Quality evaluation procedures*, ISO 19114, 2003.

[Int05] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION: *Geographic information – Data quality maeasures*, ISO 19138, 2005.

[Joo98] JOOS G.: *Zur Qualität von objektstrukturierten Geodaten*. PhD thesis, Universität der Bundeswehr München, 1998.

[KGP05] KOLBE T. H., GRÖGER G., PLÜMER L.: CityGML – Interoperable Access to 3D City Models. In *Proceedings of the Int. Symposium on Geo-information for Disaster Management* (Delft, March 2005), Oosterom, Zlatanova,, Fendel, (Eds.), Springer Verlag.

[Krä06] KRÄMER M.: *Methoden zur Bestimmung und Steigerung der Datenqualität von 3D-Stadtmodellen*. Diplomarbeit, Fachhochschule Gießen-Friedberg, Fachbereich MNI, 2006.

[Maa05] MAAS H.-G.: Akquisition von 3D-GIS-Daten durch Flugzeuglaserscanning. *Kartographische Nachrichten 55*, 1 (2005), 3–11.

[MWLP03] MÜLLER S., WEIS M., LIEDTKE C.-E., PAHL M.: Automatic Quality Surveillance of GIS Data with GEOAIDA. In *ISPRS Archives*, vol. XXXIV, part 3/W8. Munich, 17.-19. Sept. 2003.

[Ope01] OPEN GIS CONSORTIUM: Topic 11: OpenGIS Metadata (ISO/TC 211 DIS 19115); Version 5, 2001.

[Rei05] REITZ T.: *Architektur eines interoperablen 3D-GIS unter besonderer Berücksichtigung von Visualisierungsanwendungen*. Master's thesis, Fachhochschule Furtwangen, Fakultät Digitale Medien, 2005.

[RW98] RAGIA N., WINTER S.: Contributions to a quality description of areal objects in spatial data bases. In *ISPRS Comm. IV Symposium* (Stuttgart, 1998).

[Sch06] SCHEU M.: Quality management for geodata. In *FIG Congress 2006 - ICM Munich* (Oct. 9^{th} 2006).

[vO05] VAN OORT P.: *Spatial data quality: from description to application*. Optima Grafische Communicatie, Rotterdam, The Netherlands, Dezember 2005.