

MODELING DATA QUALITY WITH POSSIBILITY-DISTRIBUTIONS

G. Navratil

Institute for Geoinformation and Cartography, Vienna University of Technology, Gusshausstr. 27-29, A-1040 Vienna, Austria – navratil@geoinfo.tuwien.ac.at

KEY WORDS: GIS, Data Quality, Metadata, Modelling, Fuzzy Logic, Possibility Theory, Interoperability

ABSTRACT:

Description of data quality relies heavily on numbers. When dealing with data sets, which have been collected during longer periods, however, variation in the data quality will become evident. Older data may have different quality than newer data and other aspects including height or accessibility may influence the quality of parts of the data set. Precise numbers do not reflect this variation, thus I propose the use of fuzzy numbers to specify data quality.

Fuzzy numbers are based on the specification of a distribution function. The distribution function may be a probability or a possibility function. Probability is more difficult to determine than possibility. Still, the possibility function may provide all information relevant for the user. Thus, providing the possibility function may be sufficient to improve the data quality description.

The paper uses the Austrian cadastre as an example. The separation between legal and technical influences allows the specification of the possibility distributions. The example is restricted to temporal and positional accuracy and completeness. The assumption of quality requirements of two user groups finally allows assessing the fitness for use based on the possibility distributions.

KURZFASSUNG:

Die Beschreibung von Datenqualität basiert vorwiegend auf Zahlenangaben. Beim Arbeiten mit Datensätzen, die über einen längeren Zeitraum erfasst wurden, erkennt man schnell Unterschiede in der Qualität verschiedener Daten desselben Datensatzes. Alte Daten wurden mit anderer Qualität erfasst als neuere Daten und auch Rahmenbedingungen wie Höhenlage oder Zugänglichkeit können die Qualität von Teilen des Datensatzes beeinflussen. Scharfe Zahlenangaben können diese Schwankungen nicht vermitteln. Daher schlage ich die Verwendung unscharfer Zahlen zur Spezifikation der Datenqualität vor.

Unscharfe Zahlen basieren auf der Angabe von Verteilungsfunktionen. Das kann entweder eine Wahrscheinlichkeitsfunktion oder eine Möglichkeitsfunktion sein. Die Bestimmung der Möglichkeit eines Ergebnisses ist einfacher als die Bestimmung der Wahrscheinlichkeit. Trotzdem beschreibt die Möglichkeitsfunktion eventuell bereits alle Umstände, welche für den Nutzer eines Datensatzes relevant sind. Daher kann die Verwendung einer Möglichkeitsfunktion ausreichen, um die Beschreibung der Datenqualität zu verbessern.

Der Artikel verwendet den Österreichischen Kataster als Praxisbeispiel. Die Trennung der technischen und rechtlichen Einflüsse ermöglicht die Angabe von Möglichkeitsfunktionen. Die Behandlung im Artikel beschränkt sich auf zeitliche und räumliche Genauigkeit, sowie Vollständigkeit. Die Annahme von Qualitätsanforderungen für zwei unterschiedliche Nutzergruppen erlaubt es schließlich, die Nutzbarkeit der Daten anhand der Möglichkeitsverteilungen zu beurteilen.

1. INTRODUCTION

The amount of available data increased with the development of new technologies. Availability of data and capability of processing more data than before led to new applications like online route planning or visualizations in landscape planning and architecture. The outcome of the application depends on adequate selection of data sets. However, data quality varies with the source. Data quality descriptions like positional quality, temporal quality, or completeness have been defined to cope with that problem (Guptill and Morrison, 1995).

Automation of data processing requires automatic handling of data quality. Fitness for use is a method to describe, if a specific data set is suitable for a specific task (Chrisman, 1984). Automatic pre-selection of the data sets can simplify the selection process for the user. Automatic pre-selection requires automatic determination of the fitness for use. The necessity for the discussion of user needs has been pointed out by Byrom (2003). Approaches to define the user needs and to use them to specify the fitness for use have been presented by Jahn (2004), Grum and Vasseur (2004), and Pontikakis and Frank (2005). A first step in that process is the suitable description of data quality.

Data sets are often collected over long periods. Graphs of street networks, for example, are determined once and updated periodically to reflect changes in the network. Since the quality of the determination may change due to improved technology, the quality of the data varies within the data set. A solution to this problem is using the worst case for providing a number for the quality. However, this solution may give a wrong impression if a small part of the data is of significantly worse quality than the rest of the data set. The Austrian cadastre, for example, provides data on parcel boundaries. The data is either part of the coordinate-based cadastre and thus has a positional quality of 15 cm or stems from the traditional system based on evidence in the real world with a positional accuracy of up to a hundred meters (absolute). However, most

of the low quality boundaries are located in mountainous regions or forests where the determination of the boundary is less important than in urban areas. Thus specifying 100 m as positional quality of the Austrian cadastre would comply with the definition above but would give a wrong impression of the overall quality. Similar discussions are possible for other aspects of data quality and other data sets. Quality should thus not be described by a single value but by at least an interval.

The use of fuzzy numbers is a solution for the problem of varying data quality. Fuzzy numbers specify a range for the value. This would allow showing the range of quality within the data set if used for specifying data quality. The user will see the worst and best possible quality and can then decide if the data set fits his needs. Fuzzy numbers are defined using distributions. Probability distributions specify probabilities for the different possible values whereas possibility distributions only describe the possibility of the outcome. In this paper I discuss an approach using possibility distributions.

The paper uses the example of the Austrian cadastre to specify possibility distributions. Different aspects of quality are modeled with different distributions. Then I assume user requirements and present a method to compare the data quality with the user requirements. This allows assessing the fitness for use.

2. DATA QUALITY

How can we describe data quality? ISO 19113 “Quality Principles” (ISO 19113, 2002) defines the framework for a quality model. The data quality elements are

- completeness,
- positional accuracy,
- temporal accuracy,
- logical consistency, and
- thematic accuracy.

Each of these elements describes a specific aspect of geographic data. The description uses

numbers. We may assume normal distribution and use standard deviation as a measure for the positional accuracy of coordinates of points stored in the data set. I will use the first three elements to show that the use of precise numbers is not sufficient for complex data sets. The other elements could be treated in a similar way.

The determination of data quality must consider the creation process. The creation process is influenced by three different aspects, which also influence the quality of the resulting data (Navratil and Frank, 2005):

- technology,
- legality, and
- usability.

Technological possibilities limit the achievable quality. In general, there is a maximum quality as well as a minimum quality. The usual method to create a terrain model, for example, is either terrestrial 3D-measurements or evaluation of aerial images. In both cases the quality of the terrain model depends on the expenditure. Lower flight height will improve the quality of the model as well as condensing the network of terrestrial measured points. Different equipment used for measurement will result in different precision of the height points. The precision will not be arbitrarily high since there are always small changes in the terrain, e.g., footsteps. Such changes should not influence the terrain model and thus the precision will be such that these small changes do not affect the result. Reduction of quality is also limited. Having less than a single height point in a terrain model is useless.

Laws also have an impact on the quality of available data sets (Navratil, 2004). Laws directly influence quality and extent of data sets required for tasks specified in laws. Laws may prohibit the use of data with higher quality than specified due to data protection laws or security reasons. There is a proposal for a German law, for example, that shall regulate the spreading of high resolution satellite images. The argument for the necessity of the law is national security (German Bundesrat, 2007). In contrast to the

technological limitations legal influences are not ‘hard’. It is possible to specify the maximum technical quality of a length measurement. The quality is determined by the definition of the length unit and it is impossible to have measurements that exceed this quality. This is not true for laws. The Austrian law stipulates a minimum precision of 15 cm for boundary points in the coordinate-based cadastre. Since it is extremely difficult to prove the quality of coordinates, a point with a precision of only 20 cm may also be accepted. Thus legal rules on data quality can be often seen as guidelines to develop technical solutions. It is then assumed that the results of the process meet these rules.

Usability of data also affects data quality. Data used more often will usually have higher quality since it produces more revenue and thus more money is available for collecting the data. Maps provide an example: Nautical maps have higher quality in the areas where it is needed. Users do not need detailed nautical maps in the middle of the Atlantic Ocean but they do need them in coastal areas where the danger of hitting the ground is high. Thus more money is spent on mapping coastal areas than on mapping the ocean resulting in better data quality.

3. IMPRECISE NUMBERS

Many real world situations cannot be described precisely. A geographic example is the goal of a navigation process (Duckham, Kulik et al., 2003) Statistics on the number of cars waiting at a red traffic light seems to be a simple task but the definition of a ‘waiting car’ is difficult. A stopped car is definitely waiting but how about a car rolling slowly toward the traffic light? What is the maximum speed that a rolling car may have to be labeled ‘waiting’? Question like that cannot be answered with crisp definitions. This problem led to the development of mathematics with imprecise numbers.

Reasoning can be defined as testing the correspondence of a specified hypothesis with given statements. The statements can be data stored in

a database and the hypothesis is a query on these data. A typical example is a database containing the heights of persons and the question, if a specific person is 'tall'. Four different situations can be determined (Dubois and Prade, 1988):

- Both, the data in the database and the definition of 'tall' are crisp. The entry in the database for the person could be 1.7 m and 'tall' is defined as '>1.65 m'. This leads to traditional, two-valued logic.
- The data in the database is vague and the definition of 'tall' is crisp. Here the definition for 'tall' is the same as above but the entry in the database is expressed with a possibility distribution. The possibility distribution describes the degree of possibility that the height of the person corresponds with the value. This leads to possibility theory as published by Zadeh (1978; 1979) and expanded by Dubois and Prade (1988).
- The data in the database is crisp and the definition of 'tall' is vague. Here the entry in the database could be 1.7 m but the concept of 'tall' is uncertain. This leads to many-valued logic.
- Both, the data in the database and the definition of 'tall' are vague. This leads to fuzzy logic (Zadeh, 1975).

Which of these types of logic shall we use for modeling data quality? The data describes the world. Since the world changes, the data must change, too. Thus the data acquisition is a continuous process. Data quality parameters shall describe the quality of this data. It will not be possible to use a crisp description. Crisp descriptions are possible for single elements within the data set. However, the quality will vary throughout the data set and this variation should be reflected by the data quality description. Thus description of the whole data set will be vague. There may be general statements like:

- The positional accuracy of a point is 15 cm.
- The last update of the topographic map was done in 2001.

These expressions only set limits for the quality. The first proposition does not say that each point in the data set has a positional accuracy of 15 cm. The proposition may be meant as an average value or (which is more likely) as a bounding value. The proposition could then be rephrased as "The value of the positional accuracy for an arbitrary point in the data set is unknown but it is less than 15 cm." The second statement is similar. It only defines that changes in the real world that happened after 2001 will not be included in the map. Thus we deal with uncertain data.

The questions we raise are crisp or can be made crisp. From a user's perspective we have two different questions:

- I need a data set with a specific quality. Is it available?
- There is a data set with a specific quality. Can I use it for the purpose at hand?

Both questions are crisp. In the first case there may be several parameters for the data quality. All of these parameters must be fulfilled. Thus a data set either fulfills the quality specification or it does not. This gives a crisp answer to the question. The second question is more complex. Like in the first case data quality issues must be considered but in addition a cost-benefit-analysis is necessary. According to Frank (1995) and Krek (2002) the value of a data set emerges from better decisions. The value can be compared to the costs of acquisition and processing of the data. The data set is applicable if the costs are lower than the benefits and there is no other possible outcome than using or not using the data set. Thus both questions are crisp and we must use possibility theory to describe the problem.

4. POSSIBILITY FUNCTIONS

A discussion of processes requires a method to describe the outcome of processes. Possibility functions (Zadeh, 1978) are such a method. In general, the use of fuzzy methods is suitable for

the results of precise observation processes and they can be used to make statistical analysis (Viertl, 2006). Viertl uses probability functions, which assign probabilities to each possible outcome. This requires detailed knowledge since the probabilities must be determined. Possibility distributions avoid that problem by not providing probabilities. Possibility distributions only specify the possibility of the result: The value 0 shows impossibility and 1 shows possibility. Values between 0 and 1 provide information on the plausibility of the outcome. Thus, a result with value 0.4 is possible but less plausible than a result with 0.8. It does not state, however, that a result with 0.8 is twice as probable as a result with 0.4.

An introduction to the mathematical definition of possibility distributions based on propositions has been shown by Wilson (2002): The most common way to express propositions is using a set Θ of mutually exclusive and exhaustive possibilities. A possibility distribution π assigns a value of possibility to each element of the set. If there is an element with value 1 then the function is said to be normalized.

$$\pi: \Theta \rightarrow [0,1] \quad (1)$$

Values between 0 and 1 express orderings of propositions.

5. QUALITY OF CADASTRAL DATA

The Austrian cadastral data is used in this paper as an example for a large data set collected over an extended period. The data set includes parcel identifiers, parcel boundaries, and current land use. Details on the Austrian cadastral system can be found in different publications (Angst, 1969; BEV, 1991; Twaroch and Muggenhuber, 1997). An important aspect is the definition of boundary. Whereas evidence in reality (like boundary marks, fences, or walls) defines the boundary in the traditional Austrian cadastre, the new system uses coordinates to specify the

position of the boundary. This change allows the creation of data sets reflecting reality since the data provides the legal basis for the boundaries.

The elements of data quality as listed in section 2 must be defined in order to specify the quality of the Austrian cadastre. Completeness of the boundary data in the cadastre can be defined as the percentage of boundary lines, which are either missing or classified in a wrong way. Positional accuracy connects to the elements defining the boundary lines. The Austrian cadastre uses boundary points to define the boundary. Thus the positional accuracy of the boundary points stipulates the positional accuracy of the data set. Temporal accuracy of the cadastre is the time between a change in reality and its reflection in the cadastral data set.

Frank, Grum et al. (2004) show that the quality elements completeness, positional accuracy, and temporal accuracy are not independent. Positional accuracy and completeness dilute over time and thus depend on temporal accuracy. The assumption used by Frank, Grum et al. is that data is obtained once and there is no update process. This is not true for the Austrian cadastre. Changes in the real world do not affect the position of the boundaries since the boundaries are based on the position of boundary points. Coordinates define the position of the boundary points and these coordinates are not influenced by changes in the real world. They can only be changed by cadastral processes (Navratil and Frank, 2003). Thus the positional accuracy of the boundaries is independent of the time. However, this is only true for the coordinate-based system. The traditional cadastre does have a dilution of precision over time because the boundaries are defined by evidence (boundary stones, fences, walls, etc.) in the real world, which can change. The same is valid for completeness. In the traditional cadastre land can be acquired by adverse possession. Adverse possession creates a new boundary, which is not part of the cadastral data. This, too, is not possible in the coordinate-based cadastre.

6. MODELING DATA QUALITY WITH POSSIBILITY DISTRIBUTIONS

6.1 Technological Influence

From a technological point of view the Austrian cadastral data can be up to 100% complete. The cadastral data set can be kept complete if the original cadastral data set is complete and each change is included in the data set. Completeness of the original data set cannot be proven although there are tests, which can be used to check it. Cadastral maps of the country, for example, must not have areas without defined ownership. Ownership on land is defined by the land register and the land register uses the cadastre as spatial reference. Thus missing areas or boundaries in the cadastre would result in inconsistent or incomplete data sets in the land register because a piece of land would have two different ownership situations or no defined owner.

Reduction in completeness would be technically possible by removing boundary lines. This is a simple process and thus each level of completeness from 0% (all lines removed) to 100% is possible. The possibility distribution for the completeness is defined for the different levels of percentage thus ranging from 0% to 100% (see Figure 1).



Figure 1. Possibility distribution for technological influence on completeness

Positional accuracy for cadastral boundaries depends on the accuracy of boundary points, which depends on the precision of point determination and the point definition itself. Modern

technical solutions for point determination use GPS and high-precision measurement equipment. This results in a standard deviation of 1 – 5 cm. This accuracy can be reached if the whole data set is re-measured to eliminate influences of outdated measurement methods like triangulation networks as performed in Austria in the 19th century. Reduction of quality is possible, e.g., by using cheaper equipment. The lower limits are reached if the standard deviation affects the topology described by the data set. In case of the Austrian cadastre, effects will start with an accuracy of approximately 1m and the data set may become unusable in large parts of Austria with an accuracy of approximately 10 m.

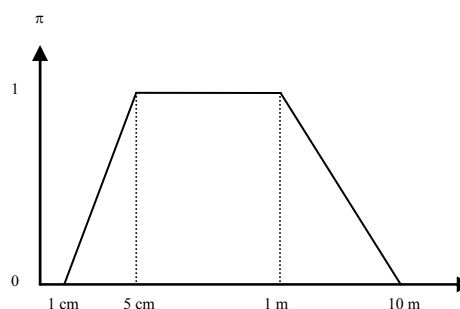


Figure 2. Possibility distribution for technological influence on positional accuracy

Temporal accuracy depends on the processing time for changes. Changes in the real world require documentation and must then be inserted in the cadastral data. This takes time since it must be guaranteed that there are no errors in the data. In addition, publication must be secure. Thus, the database visible outside the cadastral organization is a copy of the original database. This copy is updated daily to keep the update process simple. This restricts the temporal accuracy of the accessible data. We can assume that the process from the definition of the boundary to the published change requires at least a week. It is possible, however, that the process takes longer. The maximum age of the data is the age of the cadastre itself, which is less than 200 years. This could be the case if an existing boundary has not yet been included in the data set although it is shown on legally valid map

documents. Figure 3 shows the resulting possibility function.

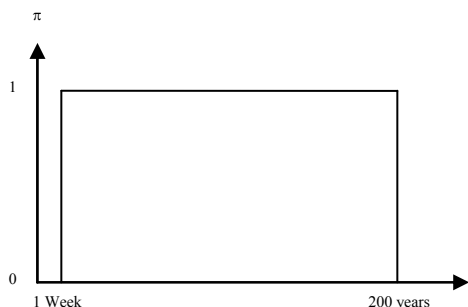


Figure 3. Possibility distribution for technological influence on temporal accuracy

6.2 Legal Influence

The cadastral system shall guarantee private rights and allow fair taxation. Thus land owned by the public was originally not included. This should have been corrected with the introduction of the digital cadastral map. Still, there is the possibility that boundaries between areas owned by the public are missing. This is represented in Figure 4.

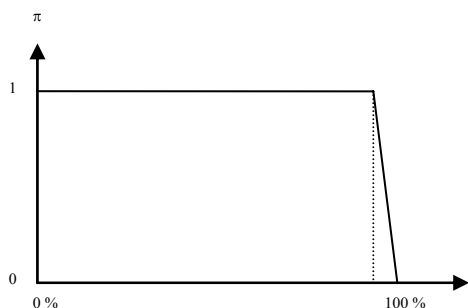


Figure 4. Possibility distribution for legal influence on completeness

Positional accuracy of boundaries depends on the cadastral system used, the coordinate-based cadastre or the traditional cadastre. The traditional cadastre uses the concept of adverse possession. A person acquires ownership of land by using the land for 30 years in the belief that the person is the lawful owner. Adverse possession is only detected in the process of boundary reconstruction or in case of dispute. Thus it is certain that parts of the data set are not describing the correct boundaries. These boundaries will

have a low positional accuracy since they are wrong. The percentage of these boundaries cannot be estimated but the assumption seems plausible that the number is not high since many boundaries are fixed by walls or fences. The possibility distribution will be similar to that shown in Figure 5b but the values may be 1 m and 100 m.

The decree for surveying (Austrian Ministry for Economics, 1994) stipulates a minimum positional accuracy for boundary points. The limit for the uncertainty of the position is 15 cm and thus theoretically the possibility distribution for the positional accuracy looks like in Figure 5a. This rule is strict as the law disregards statistical measures like standard deviation for decision making (Twaroch, 2005). However, as stated in section 3 it is difficult to control the actual accuracy of a boundary point. The existence of points with lower accuracy is possible. This is modeled in Figure 5b by gradual reduction of the value of the possibility distribution. I assume that accuracies of less than 20 cm will not be possible since they should have been detected.

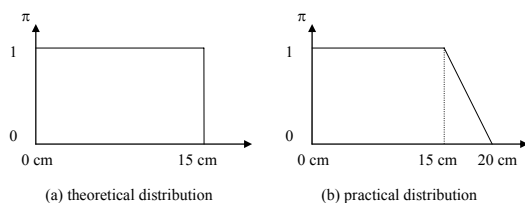


Figure 5. Possibility distribution for legal influence on positional accuracy in the coordinate-based system

Assessment of the temporal accuracy of cadastral data must consider the cadastral processes. Changes of boundaries emerge mainly from subdivisions and require surveys of the parcel boundaries. The resulting map must be registered in the cadastral office to be inserted in the cadastral map. The subdivision map is checked on correctness and the proof of correctness is documented. The owner then must register the subdivision at the land register and only after this step the new boundary is shown consistently in all cadastral data sets. The legally important action for the boundary definition, how-

ever, is the agreement of the neighbors on the boundary definition, which is expressed during the survey by a signature. The period between the date of signature and the visibility of the new line in the cadastral data set has been assessed as a week in section 4.1. However, the owner of the land must also submit the subdivision map to the land register and this must be done within a year. Thus the time span between survey and registration may be up to 1 year.

Temporal accuracy deteriorates with land consolidation. The purpose of land consolidation is to improve the efficiency of agriculture by increasing the parcel areas. Land consolidation is based on the productivity of the land and not the area of the land. Soil quality may vary even within small areas. Therefore, the amount of land owned by a person may change during land consolidation. This may lead to dissatisfied land owners who go to court to get their wishes fulfilled. Such trials can take a long time and if there are several of those trials, the time between survey (and definition of the boundaries in the field) and update of the data set may be decades. During this period the situation in the world (the new boundaries) and the cadastral data (the old boundaries) will differ. The number of land consolidations per year is small and most of them do not involve courts. Still, there is the possibility for such a case, and this shows in the possibility distribution.

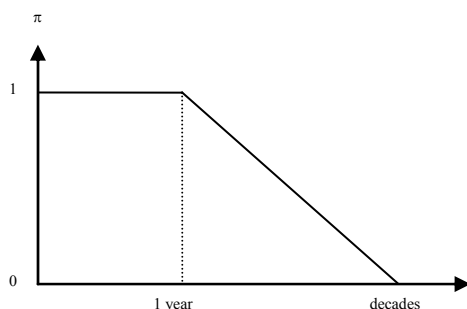


Figure 6. Possibility distribution for legal influence on temporal accuracy

7. MODELING THE USER NEEDS

I assume that cadastral data are used by two different groups of users (Navratil, Twaroch et al., 2005):

- Users of the boundary itself: Owners of land need data on their parcel and the neighboring parcels with high accuracy. Data on these parcels must be complete. Data on the remaining land is of no greater relevance to them.
- Users of the positional reference in general: The cadastre is the only large-scale map available for the whole area of Austria and thus it is often used to provide spatial reference. These users are not interested in a high quality of the boundary description. However, they need complete data sets in the area of interest.

Since these two groups have different requirements we have to treat them differently. The differences will show in the possibility functions. In contrast to the technological and legal influence the possibility functions are not based on the specifications of the data set but on the intended application. The possibility distribution shows if it is possible to use the data set for the specific application.

7.1 User of the boundary

Owners of land need complete data on their parcel. This knowledge, for example, is necessary to plan the construction of a building. Complete knowledge on the boundaries of neighboring parcels may also be necessary, e.g., for subdivisions. Thus the data set is useful as long as all necessary data is included.

Completeness of less than 100% does not eliminate the possibility that all necessary information is contained in the data set. However, reduced completeness increases the danger that some data are missing. Thus we have to assume that there is a level of completeness, below which the data set is unusable for the owner because the danger is too high that at least one piece of data will be missing. Starting with this level the possibility increases to 1, which is

reached with the complete data set. Figure 7 shows the resulting distribution.

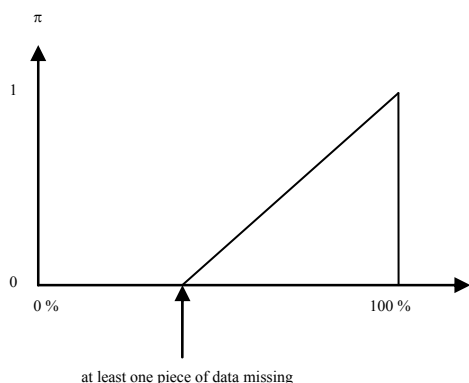


Figure 7. Completeness for users of boundary

Positional accuracy is important for land owners. Land owners want to use their land and one of the common types of use is the creation of a building. In Austria buildings must comply with legal rules specifying for example the maximum building height, the construction style, or the distance from the parcel boundary. The last point requires high positional accuracy to fit the strategies of courts. Thus, although an accuracy of 20 cm may be sufficient for some tasks of land owners, most tasks require a positional accuracy of at maximum 10 cm.

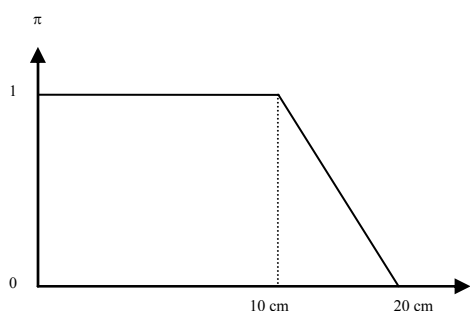


Figure 8. Positional accuracy for users of boundary

The demands of users for temporal accuracy are low because as owners of the land they are involved in all processes concerning their land or neighboring parcels. Thus they can validate, if the cadastral data set is up-to-date. The possibility function will thus be independent of the temporal quality and only depend on the correspondence with the knowledge of the owner.

The only exception emerges if the ownership changes. The buyer must be able to confirm the statements of the current owner and thus the cadastral data must contain all changes. In this case the temporal accuracy should be as high as possible, e.g., 1 day. Checking the plausibility of the propositions becomes more difficult with deteriorating temporal quality.

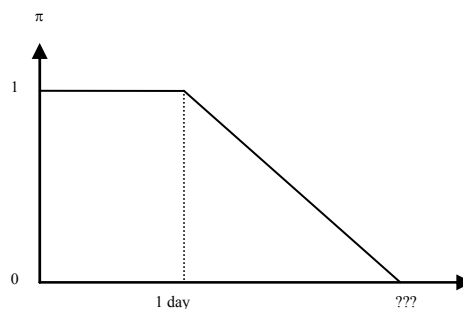


Figure 9. Temporal accuracy for users of boundary

7.2 Users of the spatial reference

If cadastral data is used for spatial reference, completeness is more important. Spatial reference is usually provided as background graphics. The maps produced will have a scale of 1:10.000 or smaller. Thus the level of detail provided by the map is low. Too many lines will confuse the viewer and thus the map maker will have to select lines for display. However, the removal will not be done arbitrarily. Lines will be removed mainly in areas with a high density of lines. Percentage of completeness must be high to allow the user to differentiate between essential and unnecessary lines. Decreasing completeness will also decrease usability. Figure 10 shows the resulting possibility distribution.

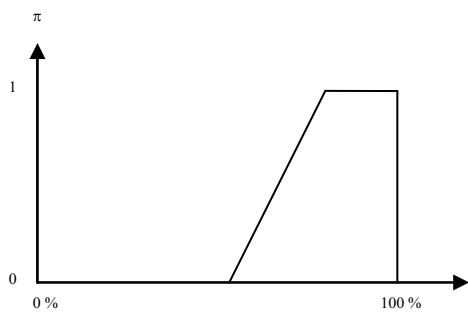


Figure 10. Completeness for spatial reference

Spatial reference has limited demands for positional accuracy. Assuming the above scale of 1:10.000 and accuracy on the map of 1/10 mm then the accuracy of the points should be 1 m. Higher mapping accuracy leads to higher accuracy demands but accuracy better than 0.5 m is not needed for positional reference. The lower limit of accuracy depends on the type of visualization. Accuracy of less than 10 m in built-up areas may result in less plausible data sets because it will not be possible to determine on which side of a street a point is and the number of display errors will be high due to small parcels.

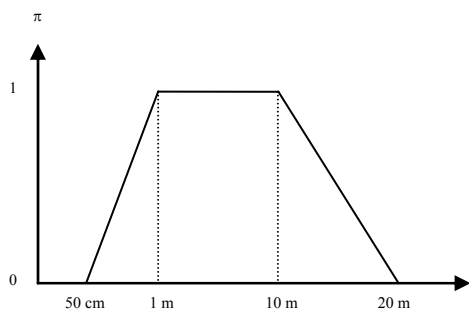


Figure 11. Positional accuracy for users of boundary

Temporal accuracy is not important for spatial reference. A data set will be useful for spatial reference as long as the parcels can be matched with reality. The number of significant changes in an area is usually rather small. The most significant changes emerge from land consolidation, which is rare. In addition, land consolidation is only performed for agricultural areas. Still the plausibility of a generally unchanged data set will reduce with the age of the data set.

However, it can be assumed that a cadastral data set can be used for providing spatial reference for 5 years without significant problems. Then the number of changes may start changing the boundaries so that the possibility of using the data set reduces. After 10 years we may have to obtain a new data set.

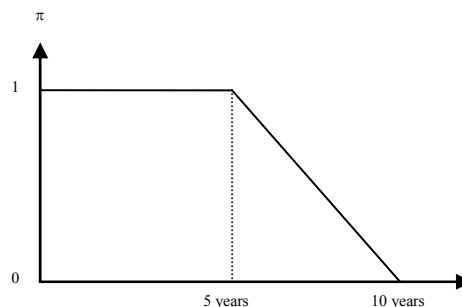


Figure 12. Temporal accuracy for users of boundary

8. COMBINATION OF POSSIBILITY DISTRIBUTIONS

A reasonable condition for data quality is that several conditions must be met. These conditions can be combined with a logical 'and'-relation. The minimum-function combines possibility distributions according to the logical 'and'. The 'or'-relation would lead to the maximum-function for the combination of possibility distributions (Viertl and Hareter, 2006). Figure 13 shows the combination of the possibility distributions for positional accuracy. The full line represents the influence of technology, the line with short dashes the legal influence and the line with long dashes the needs of land owners. The grey area marks the overlap of the possibilities. Figure 14 shows the same combination for users of the spatial reference. This combination has no solution.

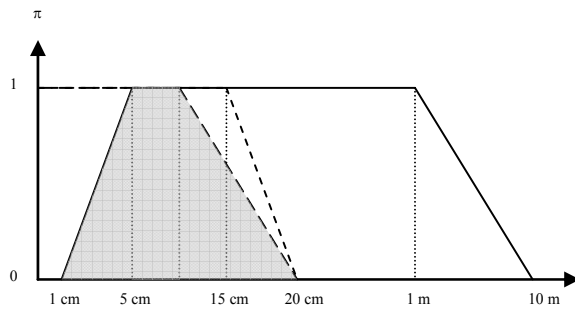


Figure 13. Combination of possibility distributions for positional accuracy for users of boundary

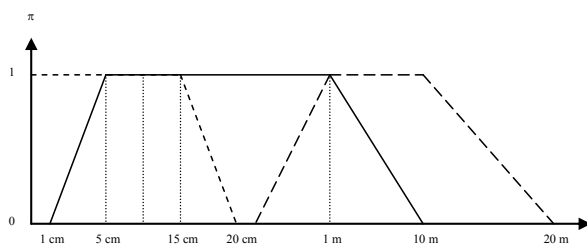


Figure 14. Combination of possibility distributions for positional accuracy for users of spatial reference

The example shows that the technical solutions and legal rules for cadastral systems do meet the demands of owners of land. Other types of use have different demands and thus the possibility function is different. Users who need spatial reference only require a different technical solution and different legal rules. This matches the empirical results reported by Navratil, Twaroch et al. (2005).

9. CONCLUSIONS

As we have seen it is possible to model the influences on data quality with possibility distributions. It was possible to specify all necessary possibility distributions. The combination of influences produced a result that can be verified by practical experience. The method thus can be used to assess the correspondence of the influences on data quality.

Left for future investigation is the application for data set selection. The paper showed how to model possibility distributions for influences on data quality. It also showed a simple method to combine these influences. A general method will be needed to create the possibility distribution for more general examples. These distributions might require more sophisticated methods of combination.

10. REFERENCES

- Angst, J., 1969. Das neue Vermessungsgesetz. *Österreichische Juristenzeitung* **337**.
- Austrian Ministry for Economics, 1994. Verordnung des Bundesministers für wirtschaftliche Angelegenheiten über Vermessung und Pläne (Vermessungsverordnung 1994 - VermV). **BGBLNr. 562/1994**.
- BEV, 1991. Database of Real Estates. Vienna, Austria: 27.
- Byrom, G. M., 2003. Data Quality and Spatial Cognition: the perspective of a National Mapping Agency. In: *International Symposium on Spatial Data Quality (ISSDQ)*, Hong Kong, The Hong Kong Polytechnic University, pp. 465-473.
- Chrisman, N. R., 1984. The Role of Quality Information in the Long-Term Functioning of a Geographical Information System. *Cartographica* **21**: 79-87.
- Dubois, D. and H. Prade, 1988. An Introduction to Possibilistic and Fuzzy Logics. *Non-Standard Logics for Automated Reasoning*. P. Smets, E. H. Hamdani, D. Dubois and H. Prade (Eds.). London, Academic Press Limited: 287-326.
- Dubois, D. and H. Prade, 1988. *Possibility Theory: An Approach to Computerized Processing of Uncertainty*. New York, NY, Plenum Press.
- Duckham, M., L. Kulik, and M. Worboys. Imprecise Navigation. *Geoinformatica* **7** (2): 79-94.
- Frank, A. U., 1995. Strategies for the Introduction of GIS. *Basic Concepts of GIS (ISPRS)*, Budapest.

- Frank, A. U., E. Grum, and B. Vasseur, 2004. Procedure to Select the Best Dataset for a Task. In: *Third International Conference on Geographic Information Science (GIScience)*, Maryland.
- German Bundesrat, 2007. Entwurf eines Gesetzes zum Schutz vor Gefährdung der Sicherheit der Bundesrepublik Deutschland durch das Verbreiten von hochwertigen Erdfernerkundungsdaten (Satellitendatensicherheitsgesetz). *SatDSiG*: 67.
- Grum, E. and B. Vasseur, 2004. How to Select the Best Dataset for a Task? In: *International Symposium on Spatial Data Quality (ISSDQ)*, Bruck a. d. Leitha, Austria, Department of Geoinformation and Cartography.
- Guptill, S. C. and J. L. Morrison, Eds., 1995. *Elements of Spatial Data Quality*, Elsevier Science, on behalf of the International Cartographic Association.
- ISO 19113, 2002. *Geographic Information - Quality Principles*.
- Jahn, M., 2004. User needs in a Maslow Schemata. In: *International Symposium on Spatial Data Quality (ISSDQ)*, Bruck a. d. Leitha, Austria, Department of Geoinformation and Cartography.
- Krek, A., 2002. *An Agent-Based Model for Quantifying the Economic Value of Geographic Information*. Vienna, Austria, Department of Geoinformation/TU Vienna.
- Navratil, G., 2004. How Laws affect Data Quality. In: *International Symposium on Spatial Data Quality (ISSDQ)*, Bruck a.d. Leitha, Austria, Department of Geoinformation and Cartography.
- Navratil, G. and A. U. Frank, 2003. Modeling Processes Defined by Laws. In: *6th AGILE Conference on Geographic Information Science*, Lyon, France, Presses Polytechniques et Universitaires Romandes.
- Navratil, G. and A. U. Frank, 2005. Influences Affecting Data Quality. In: *International Symposium on Spatial Data Quality (ISSDQ)*, Peking, China.
- Navratil, G., F. Twaroch, and A. U. Frank, 2005. Complexity vs. Security in the Austrian Land Register. In: *CORP 2005 & Geomultimedia05*, Vienna, Austria, Selbstverlag des Institutes für EDV-gestützte Methoden in Architektur und Raumplanung.
- Pontikakis, E. and A. U. Frank, 2004. Basic Spatial Data according to User's Needs-Aspects of Data Quality. In: *International Symposium on Spatial Data Quality (ISSDQ)*, Bruck a.d. Leitha, Austria, Department of Geoinformation and Cartography.
- Twaroch, C., 2005. Richter kennen keine Toleranz. In: *Intern. Geodätische Woche*, Obergurgl, Wichmann.
- Twaroch, C. and G. Muggenhuber, 1997. Evolution of Land Registration and Cadastre. In: *Joint European Conference on Geographic Information*.
- Viertl, R., 2006. Fuzzy Models for Precision Measurements. In: *Proceedings 5th MATHMOD*, Vienna, ARGESIM / ASIM.
- Viertl, R. and D. Hareter, 2006. *Beschreibung und Analyse unscharfer Information*. Vienna, Springer.
- Wilson, N., 2002. A Survey of Numerical Uncertainty Formalisms, with Reference to GIS Applications. Annex 21.1 to REV!GIS Year 2 Task 1.1 deliverable.
- Zadeh, L. A., 1975. Fuzzy Logic and approximate Reasoning (In Memory of Grigore Moisil). *Synthese* **30**: 407-428.
- Zadeh, L. A., 1978. Fuzzy Sets as a Basis for a Theory of Possibility. *Fuzzy Sets and Systems* **1**: 3-28.
- Zadeh, L. A., 1979. A Theory of Approximate Reasoning. *Machine Intelligence*, Vol. 9. J. E. Hayes, D. Michie and L. I. Mikulich. New York, Elsevier: 149-194.