QUALITY ASSESSMENT IN GENERALIZATION: INTEGRATING QUANTITATIVE AND QUALITATIVE METHODS

Regula Ehrliholzer

Department of Geography
University of Zurich (GIUZ)
Winterthurerstrasse 190
CH-8057 Zurich (Switzerland)
Fax: +41-1-362 52 27
email: regula@gis.geogr.unizh.ch

Abstract

This paper presents a conceptual discussion of methods for quality assessment in the domain of cartographic generalization. The overall problem is addressed from different angles: the notions of quality existing in cartography (section 2), quality factors and criteria (section 3), components and levels of evaluation (section 4). Finally, a way to integrate qualitative and quantitative techniques into a coherent framework of quality assessment for generalization is proposed (section 5).

1 Introduction

With the advent of digital cartography and the attempts to automate generalization, the discussion of quality has become of increasing interest. However, until present quality assessment has been done rather informally and despite its great importance cartographic literature rarely addresses the topic other than implicitly. Few *methodological principles* are available for cartographic quality assessment and this is particularly true of generalization.

Using the assumption that cartography equals generalization, which places generalization as the most important cartographic process, it is clear that any comprehensive evaluation process cannot focus on generalization alone. For a complete evaluation, other aspects of the map production process have to be considered as well. On the other hand, the author argues that a generalization-specific evaluation methodology has to be developed to face the actual requirements within generalization research. In a sense, generalization offers a good opportunity to develop evaluation criteria and methods which can later be extended to other cartographic problems as well. Therefore, without neglecting the holistic nature of the generalization process, generalization assessment techniques have to be developed (by means of a certain analytical approach), that can be employed in practice as well as in the research domain. For this reason, quantitative and qualitative evaluation techniques have to be suitably combined.

In cartography, several notions of quality are commonly used what could be a reason for the cautions in dealing with this subject. Thus it is considered useful to address this topic in a broader context: the following section focuses attention on different and

sometimes contradictory quality concepts and their impact on quality assessment in cartography in general or in generalization in particular. The underlying five basic approaches are borrowed from Garvin [in 15].

2 Notions of Quality in Cartography

Cartographic generalization in a narrower sense is defined by the Swiss Society of Cartography as follows: 'By this we understand very simply everything which happens after deciding upon the seven influencing factors[1], namely the graphic generalization, the *continuous assessment* of this and any necessary *corrections* [13].

Hence it appears that quality control is part of the production process and that it is each cartographer's duty to grant quality at all times.
The so-called *process oriented approach* requires quality assessment on every level of production, resulting in an outcome that theoretically matches the previously defined requirements or specifications. In practice, quality control combined with a more organizational and institutionally oriented quality management (towards a quality assurance approach [14]), is crucial. Currently, this is a much discussed topic in different fields, especially in software engineering, where quality assessment of software products faces similar problems as cartographic products. A definition of 'quality' (ISO 8402) intended to use in a quality control environment is given in section 3.
In the field of digital generalization, such an approach implies that within an interactive environment supporting facilities have to be provided in order to allow continuous assessment. Ideally, these tools would be embedded in an increased quality *assuring* environment e.g., improved algorithms or expert systems. However, cartographic expert systems still 'fail to consider alternative solutions and have no capability for comparative evaluation' [3].

Often 'quality' is being linked to the *time* needed to complete a task or a product. Since production time manifests itself indirectly through the price of a product, one can talk of the *cost-benefit* approach (cf. [9, 1]). Just for economic reasons, map production would require the application of automated generalization techniques, which at present do not produce satisfying results without extensive refinements. The question remains whether quality loss has to be accepted because of this. A more precise outline of the term 'quality' is required nevertheless.

Quality can also be defined through the user's requirements. This approach is known as *fitness for use* and is in the GIS community, above all, discussed as a counterpart to data quality. Accordingly, the quality of a map would be defined through the degree of meeting user's requirements. In consideration of the fact that the production of maps as well as their interpretation requires training, this approach to carographic quality assessment is sometimes considered problematic.[2] Nowadays many people who are not trained in cartography are involvee in map production (e.g., GIS users). A standardized quality assessment approach may help increase the quality of products made by non-cartographers. The method for quality assessment proposed later on in this paper is meant to be carried out by experts, because a great amount remains dependent on professional expertise.

---

[1]These factors are: scale, source material, special conditions for legibility, symbol specification, choice of colours, technical reproduction capabilities [13]. Note, that the term 'assessment' is rarely included in definitions of generalization.

[2]In addition to this fact, the user community can often not be determined exactly.

A purely *product related approach* looks at quality as a definable measure. From observable properties of products these can be valued and it is possible to make out quality differences. Quality can be defined as 'absence of defects'.

In contrast, quality can be looked at as something that is not definable, but recognizable through *experience*. Quality in this sense is equivalent to 'high standard' [15]. While emphasizing the intuitive, subjective, and creative work style of the cartographer his products are often declared as a sort of artwork, which withdraws them from objective judgement. The author argues that this view amplifies the impossibility to define cartographic quality and imposes an impediment for scientific treatement of the topic. A rather objective and structured assessment method is not intended to suppress the creative process, but rather tries to find a way to make statements about it in a consistent environment.

3 Quality Factors and Criteria

Starting out from a general definition of quality, this section focuses on the relationship between quality factors and criteria and furtheron a definition of generalization quality assessment is proposed.

The ISO Standard 8402 defines quality as:

> *The totality of features and characteristics of a product or service that bears on its ability to satisfy given needs* [7].

From this definition it follows that quality has to be defined according to specified needs. These quality needs can be formulated on different abstraction levels [14]. Quality factors represent the first level. Because they are not specific enough to be used operationally, quality criteria have to be derived by means of operationalization. Operationalization means the transmission of a term in an observable or measurable dimension. In the other direction, a quality accumulation of the sub-results will deliver a sort of 'overall' quality.[3]

In the literature of cartography no definition of quality like the above stated is found. Because it is formulated rather generally, however, this version could be transposed to the cartographic field.
All known generalization requirements can be regarded as quality factors. These can be subdivided into of quality factors for content, graphics, reproduction and traditional generalization needs. Still, the break-down of quality factors into quality criteria is a remaining problem in quality assessment of generalization solutions. This problem is especially valid for holistic factors like 'conservation of the overall structure'. Such factors are not easyly, or even not at all operizationable. However, a transmission in at least an observable dimension is important for a structured evaluation.

With regard to the theortical framework proposed later in this paper (cf. section 5), generalization quality assessment might be defined as:

> *The application of a set of selected criteria relevant to generalization on a cartographic product that intends to ascribe quantitative or qualitative attributes for which the rating scale has to be transformed in order to get a quality judgement that combines both quantitative and qualitative aspects.*

[3]Called the break-down/accumulation approach [14].

4 Assessment Contexts

Today, in the field of cartography there exist two counterparts: production/research and traditional/digital generalization. Because of the different goals within these contexts there exist different requirements on the evaluation methods.

4.1 Quality Assessment of Generalization in the Context of Production and Research

In the domain of production, evaluation is important for *quality assurance.* The goal is the *removal* of quality deficiencies during the ongoing production process and a final revision for product release.
In generalization research the focus is not on the production of a map but on *theoretical* aspects. Evaluation can therefore be an input for further evaluation of e.g., original data, algorithms, operators, generalization systems or techniques. A long term goal is the improvement of digital generalization systems. For the use of artificial intelligence techniques such as rule-based systems or genetic algorithms, methods for quality assessment are an essential prerequisite [16, 17]. In the case of rule-based systems, evaluation is needed to support the acquisition of explicit knowledge to be stored in the rule base, and to resolve conflicts between rules. In genetic algorithms [5], the evaluation of automatically generated alternatives against quality criteria forms an integral part of an underlying 'generate-and-test' strategy. In the domain of GIS, evaluation serves the validation of analytical results and supports users with little cartographic knowledge (cf. section 2).

4.2 Quality Assessement of Generalization in the Context of Traditional and Digital Generalization

Due to computer assisted generalization the situation for evaluation has changed because errors are produced that would never have been subject to evaluation in manual generalization where assessment and execution are closely related. Algorithms sometimes produce results that are obviously incorrect (notably topological errors). It therefore seems appropriate to correct Imhof's [6] statement that (with some reservations), the solution of a cartographic task cannot be true or false but only good or poor, respectively. This is valid - with some limitations - for manual generalization because a skilled cartographer certainly does not construct mistakes of that kind. The evaluation context of digital generalization is characterized by a variety of components and levels of evaluation, which all would require their own criteria sets (cf. 5.4).

Components of evaluation: in the domain of computer assisted generalization several elements can be subject to evaluation such as algorithms, operators, supporting facilities which depend on the digital generalization system. Given the fact that in current cartographic systems, generalization is commonly applied under interactive control, the following elements can be subject of evaluation as well: the *selection* of operators and algorithms, the *sequencing* of operators, and the editing which depend on the user's interactions. Evaluation of these elements is necessary because they influence the quality of a generalization solution. Research that concentrates on these interactions belongs to the field of knowledge acquisition.

However, it is considered important to have a critical look at quality assessment of overall generalization solutions alone because in the end *their* quality decides over the quality of the above mentioned elements. Is the evaluation of generalization solutions dependent on factors such as the quality of an operator one could get lost in circular argument. Because of this, efficiency or user-friendliness of operators are not taken into

account at the present stage of this work. But it has to be emphasized that these two factors dominate over others because only when user-friendliness and performance are guaranteed, the quality in terms of software usability can be fully exhausted.

Levels of evaluation: evaluation can be carried out on different map and feature levels.

On the *map level,* the overall map situation is subject to assessment of generalization quality. The temporal component of evaluation here is stressed, which means generalization solutions can be assessed at different finishing grades. In summary, these levels include source maps (source data), generalization solutions on screen (intermediate steps), and target maps (paper/screen).

The most important problem at this stage consists in an evaluation of *intermediate digital solutions.* Ad hoc evaluation [16] on the screen causes some difficulties for the following reasons: symbolization (e.g. single line/double line) can differ from the printed version, meaning that this fact has to be taken into account for visual or computational assessment; colouring may demand different minimal distances; visual interpretation does not make sense when the map is presented at target scale due to the screen resolution. In addition it causes problems to overview the entire map situation on screen.[4]

On a *feature level* the elements of evaluation can be simple geometry, objects, geometry in context, objects in context, and the entire situation. The more complex the feature level is, the more qualitative assessment processes seems to be suitable. Many different generalization solutions can be produced through operator sequencing or testing of the effects of different operators on the same features.

On the same map and feature level, evaluation should be possible in a global and local manner. This is particularly important while comparing inhomogeneous maps. Therefore a meaningful tessellation of the working space is a prerequisite.

For the long term, computer assisted cartography will perhaps have as a consequence the replacement of multi-task maps with more single task maps [1]. This means also that user specific preferences can be stated more precisely. For quality assessment this means that a certain amount of subjectivity is eliminated due to that criteria can be selected or specified according to the more concise formulated map purpose and not according to the interpretation of the evaluating expert (e.g., feature selection,).

5 Integration of Quantitative and Qualitative Methods

In the words of Fisher [3] 'we are not equipped to evaluate maps either quantitatively or qualitatively within a computer environment, and so distinghish one from another on the basis of its quality as a map'. In this section we present an *integrative* quality assessment approach which tries to combine qualitative and quantitative evaluation techniques with respect to their limitations. A combination of both qualitative and quantitative processes may overcome the defficiencies of each technique applied in isolation. Starting with a discussion of subjective and objective aspects of evaluation, the paper goes on to point out methodological requirements for quantitative and qualitative assessment techniques.

[4]The overall image as a must for analysing task is stressed by Spiess [12]. Several points relevant for presentation on screen are mentioned.

5.1 *Objective vs. Subjective Assessment*

It is known that the generalization process consists to a certain degree of subjective components. Given this, one may not conclude, however, that evaluation based on generalization steps taken on a subjective level has to be purely subjective itself. In a way it is important to distinguish subjectivity of generalization from subjectivity of evaluation. Evaluation is objectivizable thus far as a standardized or conducted procedure guaranteeing comprehension and comparability of the expert's statements.

If criteria are specified precisely enough then comparison among them is objective per se. Therefore the consistent application of the criteria 'preserving the black/white ratio' e.g., is clearly objectively assessable. The problem is that it is not always possible to stick to this rule due to other generalization requirements which have priority. Hence one has to deal with special cases or exceptions from the rule. If a violation of the criteria could be found there has to be examined afterwards whether it was justified or not. At this point a certain subjectivity comes into play; an assignment to good/poor (*how* good/poor, respectively) or correct/incorrect has to be made. For the example given above this can only be handled by the application of qualitative procedures.
This implies that the application of quantitative procedures is not equivalent to an objective assessment. In the same way qualitative procedures do not exclusively mean subjective assessment. There are situations where no computational solutions are proposed, but that can be judged objectively nevertheless. The combination of qualitative and quantitative evaluation procedures will be discussed further in the following subsection.

5.2 *Quantitative vs. Qualitative Methods*

Quantitative methods are comprised of e.g., counting, measures of length, areas and relations. A remaining problem for quality assessment is the development, adaptation or selection, of adequate quantitative measures for generalization criteria. For some sub-problems, such as line generalization, measures have been proposed [8, 2, 11].
Qualitative methods are those which are not based on the application of measurements but on statements of evaluating experts. These statements can be in text-form and may include screen snapshots or sketches.

The *issues* of both quantitative and qualitative techniques can consist in a description or a qualifying judgement (see figure 1). Description and judgement is not only a semantic differentiation of quality which exist in the everyday linguistic context [4], but also – starting from these two meanings – but also issues of different meaningfulness, since they are found on different *rating scales* (i.e. ratio, interval, symbolic/text).
Description covers all results that do not allow an unambiguous, rated quality judgement yet. Supporting facilities' output of a generalization system are mentioned here as an example: they provide the descriptive foundation to use in the generalization quality assessment and they assist in detecting problematic situations. Detection of obvious quality lacks can also be carried out visually, but support provided by a system can make the detection process easier and more effective. Spatial distribution of errors makes also part of quality description as well as other statistical tools that support qualitative inspection (i.e. issues of error detection does not encompass a qualifying statement yet).
Is a quality attribute such as good/poor (or in grades) is assigned to a generalization solution, it is a *judgement* we are dealing with.

As mentioned above, the results of quantitative and qualitative methods are on different rating scales. In order to integrate both methods the rating scales have to match (by means of scale transformation).
Therefore results of quantitative methods have to be reduced to an ordinal level. This reduction theoretically has to be carried out accordingly to *quality limits*. But these quality limits do not exist in the domain of generalization [1]. As long as this is the case, reduction remains a rather subjective process.[5]
Results of qualitative methods have to undergo scaling in the inverse direction, what can be called a classification resp. assignment. This involves, that e.g., an expert's judgement in textual form has to be assigned to an ordinal scale (e.g., five grades).
The criteria-driven assessment method is characterized by the fact that quantitative and qualitative techniques are based on a *set of criteria* that are extracted from a criteria catalogue. Every quantitative process has to be accompanied by a qualitative one but the inverse must or can not always be the case.

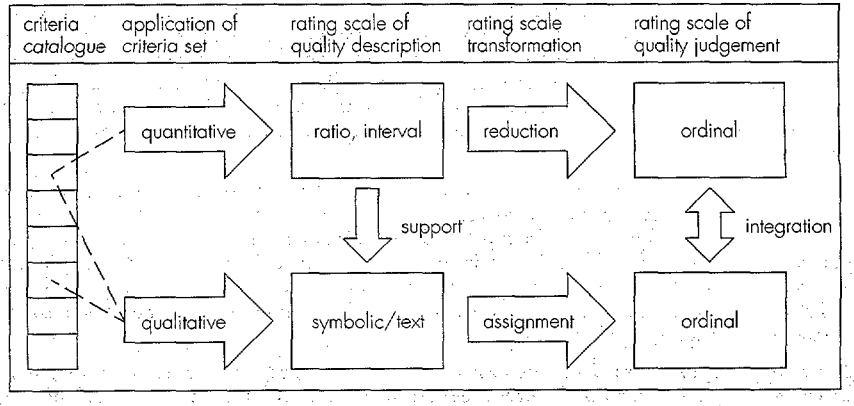

Figure 1: Framework for an integrated criteria-based assessment method

The final goal of evaluation is always a judgement (a qualifying statement), which implies that qualitative methods have priority over quantitative ones. The reason for quantitative methods often playing the role of supporting facilities only is given by the above mentioned facts (problem of scaling). Furthermore, results of quantitative and qualitative methods do not always correspond. This might be caused by several reasons as illustrate the following examples: (1) problems of validation: measures are inappropriate for the problem (e.g., measures that are preferably applicable to certain object classes), (2) problems of verification: computational inherent errors (e.g., pseudo errors due data structure[6]) (3) if there is no handling for exceptions provided (4) quantitative methods do not consider semantics (there are no measures for holistic factors e.g., gestalt), (5) many distances falling *nearly* below the minimal distance could

[5]This relates to the example given in section 5.1.
[6]This refers mainly to topological errors.

be worse than only a few tolerance violations (good legibility has priority, which is a more subjective quality factor).
Qualitative methods have another important property: positive aspects can be emphasized, whereas quantitative ones tend to be more suitable for detecting quality deficiencies, however, as Muller [10] states, 'bad maps are prehaps more easily recognized.'

A *prioritization* of a generalization solution from a range of solutions would accordingly be based on a comparement of ordinal scaled values, and the hole process as it is illustrated in figure 1 has to be carried out for each generalization solution.

Given the complexity of the proposed procedure, an evaluation of the method has to be carried out on behalf of a prototypical implementation which is outlined in the following subsection.

5.4 Platform for the Integration of Quantitative and Qualitative Methods

We believe that the format of a standardized questionnaire and checklist that asks the evaluating expert to give grades for a variety of assessment items is the most effective way to rigorously and consistently handle subjective aspects of generalization, also providing the most direct form of integration with quantitative measurements. For instance, as a counterpart of global quantitative measures, qualitative aspects on the global level such as 'maintenance of the overall character of the original map' can be assessed.

Of course, in some cases, an expert will find it hard to assign a grade that can only reflect an average for a particular assessment question. It must be possible for him/her to also document specific problems, typical situations, or details that are handled particularly well. A possible solution is the use of hypermedia techniques to implement an evaluation report (a flexible structure could be provided with an implementation in HTML[7]), allowing one to integrate screen snapshots, annotations, and sketches as a means of illustrating specific points.

A *prototype report form* would be based on a list of criteria, from which a set of criteria is extracted according to different generalization problems that depend on map purpose, scales, feature and map level, as well as on the assessment context. Modules of measures for quantitative procedures have to be integrated. This means that in a particular evaluation context, a specific report form could be compiled from relevant modules.

But there remain some points to discuss which probably are connected to more institutional questions: Firstly, who fixes the specifications (for criteria and quality limits)? Both actions are due to a certain relativity or subjectivity, respectively, since specifications and quality limits have to be proposed by persons who are embedded in a certain cartographic tradition. The example of minimal dimension illustrates that different specifications exist for the same criteria. In the case of an internal evaluation project, the specifications already used in the production process should be taken as reference for the assessment. More problems arise with international assessment projects.[8] If an evaluation procedure gives specifications this would mean a

[7]HTML: HyperText Markup Language, that is used to define components of a WWW document.
[8]A currently ongoing project on the international level is carried out by the OEEPE WG on Map Generalization. The activities focus on the evaluation of commercial generalization software.

standardization. However, in such cases it cannot be avoided that certain values are fixed.[9]
Secondly, what is the best way to design a criteria catalogue? This catalogue has to be as flexible as possible in order to ensure that the appropriate set of criteria can be chosen for a specific generalization context.

6 Conclusion

Both topics discussed in this paper - quality assessment and generalization - are not easy to deal with on a theoretical level. Nevertheless an approach has been presented here that tries to cope with and provide a framework for a consistent and comprehensive scientific treatment of quality assessment in generalization. The author advocates a clear distinction between objective and subjective aspects of quality assessment. When integrating quantitative and qualitative techniques, their interdependencies have to be closely examined. The proposed theoretical framework may serve as a structuring scheme for subjective and objective components of the evaluation, as well as for quantitative and qualitative assessment techniques. The level of abstractness of the framework allows to use it as a basis for the development of a prototype system for an integrated assessment approach. Initially, however, the scope of this prototype should be limited to a restricted, tractable problem.

References

[1] Brassel, K., 1990. Computergestützte Generalisierung. in: *Kartographisches Generalisieren.* Schweizerische Gesellschaft für Kartographie. No. 10. pp. 37-48.

[2] Buttenfield, B. P., 1991. A Rule for Describing Line Feature Geometry, in: Buttenfield, B. P., and R. B. McMaster (eds.): *Map Generalization, Making Rules for Knowledge Representation.* Longman Group. pp. 150-171. Essex.

[3] Fisher, P., 1993. Formalizing the Evaluation of Cartographic Knowledge. A Position Paper for the Specialist Meeting of the NCGIA Initiative 8 on Formalizing Cartographic Knowledge.

[4] Frühauf, J., Ludewig, J., and H. Sandmayr, 1991. *Software-Projektmanagement und -Qualitätssicherung,* vdf, Zurich.

[5] Goldberg, -d. E., 1989. *Genetic Algorithms in Search, Optimization, and Machine Learning.* Addison-Wesley, Reading, MA.

[6] Imhof, E., 1965. *Kartographische Geländedarstellung,* Walter de Gruyter, Berlin.

[7] ISO 8402 1986, *Quality - Vocabulary,* International Standards Organisation.

[8] McMaster, R. B., 1986. *A Statistical Analysis of Mathematical Measures for Linear Simplification.* The American Cartographer 13(2), pp. 103-116.

[9] McMaster, R. B., and K. S. Shea, 1992. Generalization in Digital Cartography, Association of American Geographers, Washington.

[9]However it should be noted that there is no consensus on this topic. Assessment can either be tailored to the local environment or the cartographic community may define more universal criteria (i.e. top-down or bottom-up approach).

[10] Muller, J.-C., 1991. Prospects for and Impediments against a New Cartography in the 1990s, in: J. C. Muller (ed.): *Advances in Cartography.* Elsevier Science Publ.

[11] Plazanet, C., 1995. Measurements, Characterization, and Classification for Automated Line Featrue Generalization. *Auto-Carto 12,* 27.2.-1.3. 1995.

[12] Spiess, E., 1995. The Need for Generalization in a GIS Environment, in: Müller, J. C., J. P. Lagrange and R. Weibel (eds.): *GIS and Generalization, Methodology and Practice,* GISDATA 1. Taylor & Francis, pp. 31-46, London, Bristol.

[13] Swiss Society of Cartography, 1987. *Cartographic Generalization, Topographic Maps,* Cartographic Publication, No 2.

[14] Tervonen, I., 1994. *Quality-Driven Assessment: A Pre-Review Method for Object-Oriented Software Development,* Research Papers, Series A19, Departement of Information Processing Science, University of Oulu, Finland.

[15] Wallmüller, E., 1990. *Software-Qualitätssicherung in der Praxis,* Hanser, München, Wien.

[16] Weibel, R., 1995. Three Essential Building Blocks for Automated Generalization. in: Müller, J. C., J. P. Lagrange and R. Weibel (eds.): *GIS and Generalization, Methodology and Practice,* GISDATA 1. Taylor & Francis, pp. 56-69, London, Bristol.

[17] Weibel, R., St. Keller, and T. Reichenbacher, 1995. Overcoming the Knowledge Acquisition Bottleneck in Map Generalization: The Role of Interactive Systems and Computational Intelligence. *Proceedings of the Conference on Spatial Information Theory 95,* to appear as volume of Lecture Notes in Computer Science, Springer-Verlag.