UNITED
NATIONS

E

Economic and Social Council

Distr.
LIMITED

E/CONF.90/L.3
21 MARCH 1997

ORIGINAL: ENGLISH

SIXTH UNITED NATIONS REGIONAL
CARTOGRAPHIC CONFERENCE FOR
THE AMERICAS
New York, 2-6 June 1997
Item 5 (a) of the provisional agenda*

REVIEW OF THE LATEST TECHNOLOGY IN CARTOGRAPHIC DATA ACQUISITION, MANIPULATION, STORAGE AND PRESENTATION, WITH SPECIAL EMPHASIS ON POTENTIAL APPLICATIONS IN DEVELOPING COUNTRIES: AUTOMATED MAPPING PROJECTS: DEVELOPMENT AND APPLICATION OF DIGITAL CARTOGRAPHIC DATABASES, INCLUDING DIGITAL TERRAIN MODELLING

Automatic Generalization in Cartography

Paper submitted by Finland**

---

* E/CONF.90/1.

** Prepared by Olli Jaakkola, Finnish Geodetic Institute (FGI), Masala, Finland.

Map generalization is a process that aims to render only the most important characteristics and simplify the complexity of world phenomena. Maps try to represent units of phenomena having supposedly uniform content and clear boundaries. In generalization, we face two separate goals. First is the simplification of the representation of individual phenomena. Thus, we aim to produce either traditional maps or, nowadays more commonly, geographical information for the spatial analysis. The second goal is the compilation of separate phenomena into one comprehensive map representation. Facing these two goals, the map generalization has remained as one of the few tasks in mapping, which has resisted the attempts of automatization. It is a difficult task to automatize since, as a process, generalization is multidimensional, and thus involves a great deal of empirical knowledge of the map maker, which is very difficult to express precisely or mathematically for the computer. The rules of generalization involve with complex interactions between the size and patterns of ground features, the input and output scale information, and the user objectives.

In the research carried out at the Finnish Geodetic Institute during the years 1993-1996, overseeing the production and use of land cover maps, the methods of map generalization have been developed in order to solve the first goal of generalization of individual phenomena. The large amounts of satellite information combined with statistical registers and map data, provide a detailed and up-to-date source for land cover data. In order to avoid duplication efforts in producing multi-scale, -date, and -purpose land cover data the need for linking and joining existing datasets of different sources, content and scale has emphasized the need for automatic map generalization. The developed automatic generalization methodology is build on standard raster-based GIS modelling language called the Map algebra. The automatic generalization process is divided to several sequential operations using the Map algebra. By changing the pixel size, operations and parameters within operations we may handle different types and scale of areal features.

The research has provided interesting empirical applications. The generalization of European CORINE land cover data from Finland is automatically generalized with the Map algebra. The iterative generalization provides multiple representation of detailed land cover data in a series of multi-scale land cover maps with different minimum features sizes. The generalized land cover databases seem to fulfill the accuracy standards given for the manual production. In addition, the automatic generalization method is used for generalizing the areal features in small scale topographic maps of Finland. Also the methodology is equally suitable for soil and forest maps with categorical patches. There is a possibility to give different parameters for different map features, or to give topological constraints to linear features, like to the roads. In overall, all kinds of nominal scale areal features are equally suitable for the developed automatic generalization method.

/...

INTRODUCTION

Importance of map generalization has not been widely understood until lately. The users of Geographical Informations Systems (GIS) have largely ignored the concept of scale, and thus claimed that there is now a possibility to have scale free data at our GIS. The awareness and consequences of misusing map data have steadily increased and now there is a general tendency to study and developed methods for integrating and generalizing geographical data. Nevertheless, generalization is a difficult process to automate and translate to computer based operations, and until lately there are very few examples of completely automatically generalized maps. In fact, when it comes to land cover data, the results of this research are one of the first examples of completely automatically generalized land cover maps.

Generalization aims to render only the most important characteristics and simplify the complexity of world phenomena. Maps try to represent units of phenomena having supposedly uniform content and clear boundaries. In reality, the maps are always categorized representation of world phenomena. There are often no universally agreed categories, and in practice the categories and features mapped are defined operationally for certain procedures and applications.

In generalization, we face two separate goals. First is the simplification of the representation of individual phenomena. Thus, we aim to produce either traditional maps or, nowadays more commonly, geographical information for the spatial analysis. The second goal is the compilation of separate phenomena into one comprehensive map representation. The research described here concerns more of the first goal, namely, multiple representation of land cover data generalized to various levels of detail and scale. Until lately, the map generalization has remained as one of the few tasks in mapping, which has resisted the attempts of automatization. It is a difficult task to automatize since, as a process, generalization is multidimensional. Thus it involves a

/...

great deal of empirical knowledge of the map maker, which is very difficult to express precisely or mathematically for the computer. The rules of generalization involve with complex interactions between the size and patterns of ground features, the input and output scale information, and the user objectives. The second goal actually multiplies the number of difficulties involved with generalization, and is left out of this research.

At present, in many countries the manual mapping processes in overall, are being replaced with new automatic procedures. Nevertheless, the National Mapping Agencies are still forced to produce and store multiple scale versions of different maps. There are a number of reasons for this: there is no production tool for generalization that can derive the required data sets; there is no tool to propagate updates through a series of derived data sets; and the processes of regenerating data sets are expensive and require a long time (Müller et al. 1995). Thus, there is an urgent need for more powerful automatic generalization procedures.

The aims of the research have been related to the production of combined multiple databases. The costs of manually producing continent wide small scale land cover databases are enormous. There has been, and still is, a call for less costly automatic procedures (see e.g. Molenaar 1996). Also, the experiences in different land cover projects have confirmed that the manual procedure is very slow. The supervised classification combined with automatic generalization is considerably faster, and the limitations come mainly from the acquisition of suitable satellite images and auxiliary map data. As a practical task, there was a need for generalizing a detailed large scale National Finnish land cover data to coarser small scale CORINE (Coordination of Information on the Environment) land cover data (CORINE land cover 1992). Since there were not enough resources available for manual generalization, we were forced to develop an automatic generalization method. The developed method seemed to work well, and thus there was also a call for generalizing areal features for small scale topographic maps.

In both of these projects there was also a need for extending the present statistical testing methods to deal with the quality of generalized data (Jaakkola 1996). Some new testing methods based on discrete multivariate descriptions for revealing the accuracy of generalized land cover data have been developed (Jaakkola 1994b). They have been used for optimizing the generalization operations and parameters. The accuracy tests have shown that it is possible to meet the manual production quality specifications by using a supervised classification with auxiliary data and an automatic generalization. The methods have been tested formally and informally with several databases, some of which have been quite large.

GENERALIZATION IN CARTOGRAPHY

Traditionally, the making of maps is seen as combining the characteristics of both science and art. It is said that with generalization, art enters into the making of maps. With new techniques, such as automatic generalization, the art is diminishing, yet not totally disappearing. The choice of operations and parameters still includes some subjective or artistic components. Generalization may be viewed as an interpretation process which leads to a higher level view of some phenomena, looking at them 'in a smaller scale'. Also, generalization can be viewed as a series of transformations in some graphic representation of spatial information, intended to improve data legibility and understanding, and performed with respect to the interpretation which defines the end-product. Automatic generalization can be defined as the process of deriving a less detailed (small scale) data set from a detailed (large scale) geographic data source, through the application of spatial and attribute operations (Jaakkola 1995a).

Generalization is one of the most challenging aspects of automating maps and geographical information production. The most significant difference between manual and digital

/...

generalization is that the manual process is holistic in its perception and execution. In comparison, the automatic generalization process operates much like the finite logic of a serial computer, manipulations are treated independently, and applied in a predetermined, sequential fashion. (McMaster & Shea 1992.) Most of the recent research in this field is concerned with line simplification or smoothing in a vector environment. Also, there is a group of theories on modelling the knowledge for generalization, which aims to transfer the knowledge 'skill' of a holistic manual process to the digital environment. From certain cases we are supposed to form rules, structure them for knowledge based systems, and get the generalization procedures and parameters for automatic processing (see Richardson & Muller 1991, Lee 1993). Nevertheless, the knowledge of manual generalization itself is often either too informal or too specific to be used as a control for automatic generalization.

The automatic generalization systems for land cover features have been proposed e.g. by Goffredo (1995), which includes low-level and high-level generalization using both raster and vector domain procedures. Other vector domain proposals for automatic generalization have been given e.g. by Fuller & Brown (1994). Schylberg (1993) describes a new approach to generalization, namely an object-based raster generalization. Simpler raster approaches are given by McMaster & Monmonier (1989), and Su & Li (1995). Eventually, the automation of generalization will lead towards a new way to compile map databases at different scales (Jakobson 1995, Kilpeläinen 1995).

Generalization modelling

Modelling of the generalization of the map data combines two aspects: the modelling of the map databases, its concepts and definitions at different scales, and the modelling of the generalization process. Both of them have to be combined. The modelling of the databases has been based on database descriptions for the input data and for the output data, the definitions of features, the desired minimum accuracy level, and the semantic relations of the features. Thus, in the generalization modelling we have analysed and extracted the data models of different scales, used the database descriptions as parameters of the change between different scales.

The modelling of the generalization processes has been a more difficult task since the original theory concerns more of a vector domain. However, the concepts and procedures have been defined purely using a raster-domain modelling language. The model language used for generalization operations is called the Map algebra (Tomlin 1990). The concepts of operations are adopted from the generalization literature (Aasgaard 1992, McMaster & Shea 1992, Weibel 1995, Ruas & Lagrange 1995) by using techniques originally invented by Schylberg (1993). We have partly simplified, partly extended these methods, and combined these operations in a more faster and easily modifiable manner.

A comprehensive conceptual model for a digital generalization is proposed by McMaster & Shea (1992). The model describes the philosophical objectives - 'why to generalize', cartometric evaluation - 'when to generalize', and spatial and attribute transformations - 'how to generalize'. Theoretically, this generalization model really is quite comprehensive, although there are some problems incorporating the new raster based generalization methods to the model. This the reason for redefining the generalization operations used in here. The approach used is more a posteriori than a priori and the comprehensive theory of generalization is still developing.

Different views of map generalization

There are two views to automatic generalization (e.g. Müller et al. 1995, Grünreich 1995), model view and cartographic view. The cartographic generalization is based on visual thinking and spatial reasoning, and it comprises of scale-dependent and design-

/...

dependent model generalization and (carto)graphic modelling which is specific for each map (Grünreich 1995). On the other hand, spatial objects may need to have multiple digital representation in which internal representations 'models' should be distinguished from visualizations 'cartographic representations' (Müller et al. 1995). For this we need the model generalization, which can be defined as the derivation of primary models of lower semantic and geometric resolution (Jaakkola 1995b) from a basic digital object model. In practice, these two views of generalization are not always differentiated, and the same database could be used for both. Both the model and cartographic generalization require similar generalization operators and parameters.

The functions of land cover data generalization may be analysed through the purpose, scale, map details, and quality requirements (João 1991). General requirements of model generalization are listed by Weibel (1995). If we consider the generalization procedures developed for land cover data in this research, most of these requirements are fulfilled. The generalized products are predictable, and we can repeat exactly the same procedures, since they are fully automatic. To a certain extent the deviations of the generalized data from the original are minimized, e.g. changes of shares of classes (Jaakkola 1994b). The overall data volume is reduced greatly, when the data is vectorised. The integrity of different scale products is kept, because of the same input data. The used parameters are quite self-explaining, and there are not too many of them. Finally, the procedure is fast, and completely automatic, and therefore quite competitive.

Knowledge of generalization: parameters

The knowledge of generalization may concern many kinds of space concepts. The basic categories of space found in GIS literature, namely metric, topological, and structural, can be used to describe various levels of abstraction for spatial objects. Metric space describes distance relations and constitutes the lowest level of abstraction. Topological space, deals with the existence of spatial relations between points in space. The highest level of abstraction is reached through the structural space which only deals with entities and relations (Sowa 1984). The automation of generalization requires implementation of all these three types of spaces, and in addition, procedural knowledge concerning generalization operations and sequencing of operations. The generalization methods in here use mainly measures of metric space for individual features, and structural space for semantic relations between classes. However, the topological relations are implicitly also taken into account via the use of semantic relations.

The (geo)metric measures may be associated with single features, e.g. size, shape, or direction, or with multiple features, e.g. distance, density, or distribution on a one layer or on multiple layers including elementary and derived measures. Other landscape or overall characteristics could be calculated using e.g. average feature size, standard deviation of features sizes, feature diversity, or feature interspersion. The topological measures may concern of connectedness, inclusion, intersection or adjacency (e.g. Ruas & Lagrange 1995). Topological constraints may be expressed in terms of maximizing, minimizing or maintaining certain relationships. We may also be concerned with the semantic relations of the data, i.e. structural measures. The simple description of semantic relations between land cover classes is described in the hierarchy of land cover, e.g. in CORINE land cover nomenclature based on three levels of classes (Jaakkola 1995b). In the classification nomenclature the semantically close classes have a common superclass at a lower level in the hierarchy, whereas semantically distant classes have a common superclass at a higher level in the hierarchy. These relations can be used as control parameters in the generalization.

From all this information available in generalizing land cover data, We have used the information that have been described in the database descriptions. It includes geometrical minimum dimensions of feature size, shape, and distance between features, and also the priorities 'semantic relations' within hierarchical classification.

Process of generalization: operations

The great majority of generalization research is mainly dealing with vector-based methods (McMaster & Shea 1992, Weibel 1995, Ruas & Lagrange 1995 etc.), thus the operators used in the raster generalization have to be defined again. From the basic local, focal, and zonal raster analysis types we may form the generalization operations needed for a particular task, e.g. for the generalization of a land cover data.

In the generalization, there are two types of operations, namely attribute and spatial operations. One attribute operation has been used in the generalization of land cover data, namely reclassification. Seven spatial operations have been used in the generalization of land cover data, namely aggregation, merging, amalgamation, smoothing, and simplification (Jaakkola 1995b). In addition to those, we have used two other spatial operations in generalizing areal features in topographical maps, namely resampling and exaggeration.

It is possible to include topological limits for certain generalization operations. This can be done by selecting certain classes before executing the operation, and combining these classes together with generalization result. Also, the amalgamation procedure can be re-executed after all other operations, and thus we could check the database definitions of different classes, and correct features that are not fulfilling these definitions. Nevertheless, it is probable that the operations defined in this research do not fully satisfy the needs of generalizing land cover data. Also, it is likely that there is a need for using more global or extended focal properties of land cover features as parameters.

These operations have different statistical properties (Jaakkola 1994a). The effects on the positional and attribute accuracy, and on the accuracy of various derived statistics

/...

vary according to these properties. When we combine these operations, the effects vary again according to the order of the combined generalization operations. Overall, the most important statistical property of generalization operators is the change of scale. The generalization simplifies the data structure and necessarily adds error to the database, i.e. quality is always deteriorated in favor of simplicity and legibility. Thus, the error rate in generalized database includes 'the degree of the generalization level', more generalization, more error, and 'the actual error', the bias in statistical summary measures and unintended positional and attribute errors produced by generalization (Jaakkola 1994b. We may evaluate the quality with qualitative or quantitative methods (Aspinall 1995, Ehrliholzer 1995).

MAJOR RESULTS AND THEIR CONTRIBUTION

The automatic generalization of land cover data is resolved with raster methods using the formal language of the Map algebra. The developed generalization operations are not directly linked to the more general generalization theory, and thus it is difficult to assess the theoretical validity of the generalization modelling approach used. The methods have shown, that the present generalization theory is not comprehensive, there is open space for further developments. Hopefully this research have clarified some of the concepts related to the automatic generalization of land cover data. Nevertheless, the practical validity is easier to evaluate with the quality testing methods developed. The formal quality evaluation for the automatically generalized CORINE land cover data confirmed that it is possible to fulfill the quality standards specified for the manual procedure. The generalization results have been tested with quantitative methods of multivariate statistical description, of error matrix and of derived measures e.g. for class areas and shares (Jaakkola 1994b). Also, the difference maps, overlay maps of vector and raster, describe the spatial distribution of uncertainty and accuracy of generalization procedures.

/...

In addition, the used operations and parameters are easily modifiable according to the needs of generalization, e.g. it is possible to change the size (Jaakkola 1995a) or the width of the minimum feature etc. Similar operations can be used for correcting databases, and for checking the homogeneity of different input data. Automatic generalization provides a tool to propagate, i.e. generalize automatically, updates for maps at different scales. The integrity of different scale products is kept. The automatic generalization provides an economic and fast method to produce consistent and accurate land cover data. Besides the spatial parameters, developed generalization methods can handle semantic information, and therefore, also topological relations (Jaakkola 1995b). In addition to land cover data, a similar automatic generalization has been used in generalizing areal features for small scale topographic maps. Also, the method has been used in generalizing land cover data for monitoring agricultural areas. Some of the maps have covered large areas, and still the generalization is done with one overnight batch process using a modern computer workstation. It is therefore considerably faster than manual generalization.

One of the major results is that the generalization procedures have proven that the Map algebra can be extended to the production of data. It provides a unified approach for handling and analysing spatial data. It is especially suitable for areal features, since we may change the pixel size from infinitely small to very large, and influence the positional accuracy of the data, and the speed of processing. There are many possibilities for extending the present methods for different kinds of land cover data.

REFERENCES

Aasgaard, R. (1992). Automated cartographic generalization, with emphasis on real-time applications. Ph.D. thesis at Institutt for kart og oppmaling, Universitetet i Trondheim. Norway. 119 p.

/...

Aspinall, R. (1995). Redefining 'scale' for digital geographic data. EUROCARTO XIII proceedings. Joint Research Centre, Ispra(VA), Italy. pp. 69-79.

CORINE land cover (1992). Brochure by European Environment Agency Task-Force. 22 p.

Ehrliholzer, R. (1995). Quality assessment in generalization: integrating quantitative and qualitative methods. Proceedings of the 17th International Cartographic Conference, vol 2, pp. 1442-1451.

Fuller, R.M. & N.J. Brown (1994). A CORINE map of Great Britain by automated means: a feasibility study. Institute of Terrestrial Ecology (Natural Environment Research Council). ITE project T02072J1. 37 p.

Goffredo, S. (1995). Knowledge-based system for automatic generalization of satellite-derived thematic maps. Proceedings of the 17th International Cartographic Conference, vol 1, pp. 108-117.

Grünreich, D. (1995). Development of computer-assisted generalization on the basis of cartographic model theory. In GIS and Generalization - methodology and practice, eds. Müller, Lagrange & Weibel. GISDATA 1, European Science Foundation, Taylor & Francis, London. pp. 47-55.

Jaakkola, O. (1994a). The quality of spatial data and generalization of a categorical coverage - a study for CORINE land cover Finland. University of Helsinki, licentiate thesis. 97 p.

Jaakkola, O. (1994b). Finnish CORINE land cover - a feasibility study of automatic generalization and data quality assessment. Reports of the Finnish Geodetic Institute, 94:4, 42p.

Jaakkola, O. (1995a). Automatic generalization of land cover data. Proceedings of the 17th International Cartographic Conference, vol 2. Barcelona, Spain. pp. 1984-1989.

Jaakkola, O. (1995b). Theme-based automatic hierarchic generalization for the European Land Cover data. EUROCARTO XIII proceedings, Workshop on Scale and Extent. JRC, Ispra, Italy. pp. 80-93.

Jaakkola, O. (1996). Quality and automatic generalization of land cover data. Publications of the Finnish Geodetic Institute, n:o 122, 39 p.

Jakobson, A. (1995). The Topographic Data System - A new way to compile and update topographic information. Proceedings of the 17th International Cartographic Conference, vol 1, pp. 993-1000.

João, E.M. (1991). The role of the user in generalization within geographic informations systems. In Cognitive and linguistic aspects of geographic space, eds. Mark & Frank. NATO ASI Series. Kluwer Academic Publishers, Dordrecht, The Netherlands. pp. 493-506.

Kilpeläinen, T. (1995). Updating multiple representation geodata bases by incremental generalization. Geo-Informations-Systeme, 8:4. pp. 13-18.

Lee, D. (1993). From master database to multiple cartographic representations. Proceedings of the 16th International Cartographic Conference, Cologne, Germany, vol 2, pp. 1075-1085.

McMaster, R. & M. Monmonier (1989). A conceptual framework for quantitative and qualitative raster-mode generalization. Proceedings of GIS/LIS '89. ACSM ASPRS, Falls Church, Virginia, U.S.A. pp. 390-403.

McMaster, R. & K. Shea (1992). Generalization in digital cartography. Association of American Geographers, Washington. U.S.A. 134 p.

/...

Molenaar, M. (1996). Multi-scale approaches for geodata. International Archives of Photogrammetry and Remote Sensing, vol. XXXI, part B3. Vienna. pp. 542-554.

Müller, J.-C., R. Weibel, J. Lagrange & F. Salgé (1995). Generalization: state of the art and issues. In GIS and Generalization - methodology and practice, eds. Müller, Lagrange & Weibel. GISDATA 1, European Science Foundation, Taylor & Francis, London. pp. 3-17.

Richardson, D. & Muller, J.-C. (1991). Rule selection for small-scale map generalization. In Map Generalization - Making rules for knowledge representation, eds. Buttenfield & McMaster. Longman Scientific & Technical. pp. 136-149.

Ruas, A. & J. Lagrange (1995). Data and knowledge modelling for generalization. In GIS and Generalization - methodology and practice, eds. Müller, Lagrange & Weibel. GISDATA 1, European Science Foundation, Taylor & Francis, London. pp. 73-90.

Schylberg, L. (1993). Computational methods for generalization of cartographic data in a raster environment. Royal Institute of Technology, Department of Geodesy and Photogrammetry. Doctoral thesis. Stockholm, Sweden. 137 p.

Sowa, J. F. (1984). Conceptual structures - information processing in mind and machine. The systems programming series, Addison-Wesley publishing company, Reading, Massachusetts. 481 p.

Su, B. & Z. Li (1995). An algebraic basis for digital generalization of area-patches based on morphological techniques. The Cartographic Journal, vol 32. pp. 148-153.

Tomlin, C.D. (1990). Geographic information systems and cartographic modeling. Prentice Hall, Englewood Cliffs, N.J., U.S.A. 249 p.

Weibel, R. (1995). Three essential building blocks for automated generalization. In GIS and Generalization - methodology and practice, eds. Müller, Lagrange & Weibel. GISDATA 1, European Science Foundation, Taylor & Francis, London. pp. 56-69.
