GIS and GENERALIZATION

Methodology and Practice

GISDATA 1

Series Editors
Ian Masser and
François Salgé

Edited by
Jean-Claude Müller
Jean-Philippe Lagrange
and Robert Weibel

Taylor & Francis
Publishers since 1798

GIS AND GENERALIZATION

Methodology and Practice

Also in the GISDATA SERIES

II ***Spatial analytical perspectives on GIS in the environmental and socio-economic sciences***
edited by M. Fischer, H. Sholten and D. Unwin

III ***Geographic objects with undetermined boundaries***
edited by P. Burrough and A. Frank

Series Editors

I. Masser and F. Salgé

GIS AND GENERALIZATION

Methodology and Practice

Editors

J. C. Müller, J. P. Lagrange and R. Weibel

GISDATA I

Series Editors

I. Masser and F. Salgé

Taylor & Francis
Publishers since 1798

UK Taylor & Francis Ltd, 4 John St, London WC1N 2ET

USA Taylor & Francis Inc., 1900 Frost Road, Suite 101, Bristol, PA 19007

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN 0-7484-0318-3 (cased)
0-7484-0319-1 (paper)

Library of Congress Cataloging in Publication Data are available

Cover design by Hybert Design and Type

Typeset by Keyword Typesetting Services Ltd, Wallington, Surrey

Printed in Great Britain by Burgess Science Press, Basingstoke, on paper which has a specified pH value on final paper manufacture of not less than 7.5 and is therefore 'acid free'.

Contents

Series Editors' Preface

The GISDATA series

Over the last few years there have been many signs that a European GIS community is coming into existence. This is particularly evident in the launch of the first of the European GIS (EGIS) conferences in Amsterdam in April 1990, the publication of the first issue of a GIS journal devoted to European issues (*GIS Europe*) in February 1992, the creation of a multi-purpose European ground-related information network (MEGRIN) in June 1993, and the establishment of a European organization for geographic information (EUROGI) in October 1993. Set in the context of increasing pressures towards greater European integration, these developments can be seen as a clear indication of the need to exploit the potential of a technology that transcends national boundaries to deal with a wide range of social and environmental problems that are also increasingly seen as transcending national boundaries within Europe.

The GISDATA scientific programme is very much part of such developments. Its origins go back to January 1991, when the European Science Foundation funded a small workshop at Davos in Switzerland to explore the need for a European level GIS research programme. Given the tendencies noted above it is not surprising that participants of this workshop felt very strongly that a programme of this kind was urgently needed to overcome the fragmentation of existing research efforts within Europe. They also argued that such a programme should concentrate on fundamental research and it should have a strong technology transfer component to facilitate the exchange of ideas and experience at a crucial stage in the development of an important new research field. Following this meeting a small coordinating group was set up to prepare more detailed proposals for a GIS scientific programme during 1992. A central element of these proposals was a research agenda of priority issues grouped together under the headings of geographic databases, geographic data integration, and social and environmental applications.

The GISDATA scientific programme was launched in January 1993. It is a four-year scientific programme of the Standing Committee of Social Sciences of the European Science Foundation. By the end of the programme more than 300 scientists from 20 European countries will have directly participated in GISDATA activities and many others will have utilized the networks built up as a result of them. Its objectives are:

- to enhance existing national research efforts and promote collaborative ventures over coming European-wide limitations in geographic data integration, database design and social and environmental applications;
- to increase awareness of the political, cultural, organizational, technical and informational barriers to the increased utilization and inter-operability of GIS in Europe;
- to promote the ethical use of integrated information systems, including GIS, which handle socio-economic data by respecting the legal restrictions on data privacy at the national and European levels;

- to facilitate the development of appropriate methodologies for GIS research at the European level;
- to produce output of high scientific value;
- to build up a European network of researchers with particular emphasis on young researchers in the GIS field.

A key feature of the GISDATA programme is the series of specialist meetings that is being organized to discuss each of the issues outlined in the research agenda. The organization of each of these meetings is in the hands of a small task force of leading European experts in the field. The aim of these meetings is to stimulate research networking at the European level on the issues involved and also to produce high quality output in the form of books, special issues of major journals and other materials.

With these considerations in mind, and in collaboration with Taylor & Francis, the GISDATA series has been established to provide a showcase for this work. It will present the products of selected specialist meetings in the form of edited volumes of specially commissioned studies. The basic objective of the GISDATA series is to make the findings of these meetings accessible to as wide an audience as possible to facilitate the development of the GIS field as a whole.

For these reasons the work described in the series is likely to be of considerable importance in the context of the growing European GIS community. However, given that GIS is essentially a global technology most of the issues discussed in these volumes have their counterparts in research in other parts of the world. In fact there is already a strong US dimension to the GISDATA programme as a result of the collaborative links that have been established with the National Center for Geographic Information and Analysis through the National Science Foundation. As a result it is felt that the subject matter contained in these volumes will make a significant contribution to global debates on geographic information systems research.

Ian Masser and François Salgé

Editors' Preface

GIS and Generalization: methodology and practice

This book comes out as the result of a specialist meeting in Compiègne, France, in December 1993. The meeting was sponsored by the European Science Foundation (ESF) Social Science Program through the GISDATA Programme, and initiated by the GISDATA Steering Committee which recognized generalization of geographic information as one of the major unresolved issues in geographic information systems. The chapters of this book represent the contributions of the experts who gathered at the Compiègne meeting. As such, we feel that this compilation of articles represents a major effort in describing the state of the art in scientific generalization. It also provides a perspective on future research priorities, such as model generalization for the manipulation of geographic objects independently from their graphic representations.

This book comes at a time when the prospects for automation in the generalization of geographic information are still uncertain. There is no doubt that part of the generalization process can be automated, as shown by several authors reporting on various experiments and commercial software already implemented. But the extent to which automation will be achieved in the future is still unknown. So far, we have not gone much beyond the implementation of geometric operators which are used automatically to generalize individual geographic features such as street networks, coastlines and population settlements. Therefore, this book offers both a speculative view on ways to implement automated generalization, and practical examples of already implemented or directly implementable generalization software for the purpose of map production.

Generalization is motivated by the need to provide multiple views of geographical data at various scales and levels of resolution. It is a tool which has been in use for many years, particularly in cartography for the production of maps at smaller scales. What is new, however, is the context in which this tool operates. The introduction of geographical information systems (GIS) for planning purposes and automated map production in mapping agencies provides a new *raison d'être* to generalization, namely as a tool to model data to support either spatial analysis or the automated production of maps at multiple scales. Hence, the flavour of this book, which discusses generalization issues from a model building perspective. Generic problems, modelling and knowledge acquisition issues are presented first, before discussing data quality and production issues. A major difficulty for the implementation of automated procedures is the acquisition and the formalization of geographic and cartographic knowledge (essentially intuitive and fuzzy), which has been used for many years in manual generalization. Various views are expressed, some defending a holistic approach to generalization, others favouring a more practical stepwise approach. The opinion, however, that semantic object definition, geographical analysis and knowledge formalization is a necessary prerequisite to resolving graphical conflicts in map representation is strongly represented throughout the book. The concerns about loss of data quality, which potentially occurs through generalization, are also taken into consideration, since generalized information may lead to distorted views of reality.

The book is organized in six parts; Part I Chapter 1 by Müller *et al.* is an introduction to all following parts. It contains the position paper by the members of the organizing Task Force that was distributed to all participants prior to the meeting, requesting them to write their own discussion papers in response to the issues that were addressed in this article.

Part II on generic issues of generalization contains papers that provided the major elements for discussion at the meeting. Chapters 2 by Morehouse and 3 by Spiess are the two keynote presentations that were given at the meeting. The invited speakers were deliberately chosen so that their respective background and approach towards the automation of generalization would differ. Thus, even in these two contributions, some of the tensions, but also design options that are typical to generalization, become apparent. Based on their experience with previous projects, Grünreich (Chapter 4) and Weibel (Chapter 5) discuss the fundamental importance of including the relation between cartographic and model generalization, methods for knowledge acquisition, and quality assessment.

Part III deals with object-oriented methods and knowledge-based modelling. Chapter 6 by Ruas and Lagrange opens the discussion debating the implications that the choice of data representations (i.e. data models) may have on the knowledge that can be represented in the respective systems, and thus the tasks that can be fulfilled. Buttenfield (Chapter 7) stresses the need for appropriate data models to support map generalization, whereas Bundy *et al.* (Chapter 8) and Oosterom (Chapter 9) propose various alternatives to structure the data in order to facilitate automated generalization.

The chapters of Part IV then focus on knowledge acquisition and representation, issues that are surely gaining in importance. Keller in Chapter 10 reviews techniques of artificial intelligence as they are available in computer science, and discusses how they could be made useful in the context of generalization. Chapter 11 by Heisser *et al.* presents a particular study in conjunction with the German ATKIS project. It shows how in a sufficiently defined and constrained situation, knowledge can be deduced from the definitions and constraints and integrated with generic cartographic knowledge. McMaster (Chapter 12) concentrates on procedural knowledge (i.e. generalization operators and their choice and parametrization) and on possibilities of using interactive systems for acquisition of this particular type of knowledge.

Part V deals specifically with data quality issues related to generalization activities. João (Chapter 13) discusses the importance of quantifying and controlling the effects of model and cartographic generalization on GIS map manipulations, and Painho (Chapter 14) presents a case study where the effects of generalization on the attribute accuracy of a small-scale vegetation map are analysed.

Part VI concludes the book with operational and implementation issues. Kilpeläinen and Sarjakoski (Chapter 15) propose an incremental approach, where updates after generalization can be propagated through the entire geographic database. Lee (Chapter 16) shows that commercial software can be successfully developed for semi-automated generalization. This software comes in the form of a tool box where the tools are generalization operators, which can be used to manipulate individual map features. Robinson (Chapter 17) and Vickus (Chapter 18) give their view on generalization strategies, requirements and development needs in two mapping agencies (Ordnance Survey, and Survey and Mapping in North-Rhine Westphalia).

As editors of this book, we take the opportunity to thank all authors for their contribution, as well as the various anonymous reviewers who helped in improving the quality of the articles. Books dealing with GIS and software issues have necessarily a

short life. We hope, however, that it provides for the next few years both encouragement and strong impulse to further theoretical and practical research in this important area, and that the concerns expressed in this book will influence the design of future commercial geographic information systems. We hope in particular that it leads the way to generalization solutions which will make it possible to concentrate on the application of generalization in GIS, once the theory has been mastered.

Jean-Claude Müller, Jean-Philippe Lagrange and Robert Weibel

Contributors

Geraint Bundy
Department of Computer Studies, Pryfysgol Morgannwg/University of Glamorgan, Pontypridd, Mid Glamorgan CF37 1DL, UK

Barbara Buttenfield
NCGIA State University of New York, Buffalo, NY 14261, USA

Edmund Furse
Department of Computer Studies, Pryfysgol Morgannwg/University of Glamorgan, Pontypridd, Mid Glamorgan CF37 1DL, UK

Dietmar Grünreich
Institute of Cartography, University of Hannover, Appelstrasse 9a, 30167 Hannover 1, Germany

Michael Heisser
Institut für Kartographie und Topographie, Universität Bonn, Meckenheimer Allee 172, 53115 Bonn, Germany

Elsa Maria João
Department of Geography, London School of Economics, Houghton Street, London WC2A 2AE, UK

Chris B. Jones
Department of Computer Studies, Pryfysgol Morgannwg/University of Glamorgan, Pontypridd, Mid Glamorgan CF37 1DL, UK

Stefan Keller
Unisys (Schweiz) AG, Zuercherstrasse 59–61, CH-8800 Thalwil, Switzerland

Tina Kilpeläinen
Department of Cartography and Geoinformatics, Finnish Geodetic Institute, Imalankatu 1A, SF-00240 Helsinki, Finland

Jean-Philippe Lagrange
IGN, 2 Avenue Pasteur, 94160 Saint Mandé, France

Dan Lee
Mapping Science Division, Intergraph Corp., 2051 Mercator Drive, Reston, VA 22091, USA

Robert B. McMaster
Department of Geography, University of Minnesota, 414 Social Sciences Building, Minneapolis, MN 55455, USA

Scott Morehouse
Environmental Systems Research Institute, 380 New York Street, Redlands, CA 92373, USA

Jean-Claude Müller
Geographisches Institut, Ruhr Universitaet Bochum, Universitaetsstrasse 150, Gebaeude NA, D-44780 Bochum, Germany

P.J.M. van Oosterom
TNO, Physics and Electronics Laboratory, PO 96864, 2509 JG The Hague, The Netherlands

Marco Painho
The Higher Institute for Statistics and Information Management (ISEGI), New University of Lisbon (UNL), Travessa Estêvão Pinto, Campolide, 1000 Lisboa, Portugal

Gary J. Robinson
NERC Unit for Thematic Information Systems, Department of Geography, University of Reading, Reading RG6 2AB, UK

Anne Ruas
Service de la Recherche-IGN, 2 Avenue Pasteur, 94160 Saint Mandé, France

François Salgé
IGN, 2 Avenue Pasteur, 94160 Saint Mandé, France

Tapani Sarjkoski
Department of Cartography and Geoinformatics, Finnish Geodetic Institute, Imalankata 1A, SF-00240 Helsinki, Finland

Johannes Schoppmeyer
Institut für Kartographie und Topographie, Universität Bonn, Meckenheimer Allee 172, 53115 Bonn, Germany

Ernst Spiess
Institut für Kartographie, ETH Hönggerberg, CH-8093 Zurich, Switzerland

Georg Vickus
Landesvermessungsamt Nordrhein-Westfalen, Muffendorfer Strasse 19–21, 53177 Bonn, North-Rhine-Westphalia, Germany

Robert Weibel
Department of Geography, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

The **European Science Foundation** is an association of its 55 member research councils, academies, and institutions devoted to basic scientific research in 20 countries. The ESF assists its Member Organizations in two main ways: by bringing scientists together in its Scientific Programmes, Networks and European Research Conferences, to work on topics of common concern; and through the joint study of issues of strategic importance in European science policy.

The scientific work sponsored by ESF includes basic research in the natural and technical sciences, the medical and biosciences, and the humanities and social sciences.

The ESF maintains close relations with other scientific institutions within and outside Europe. By its activities, ESF adds value by cooperation and coordination across national frontiers and endeavours, offers experts scientific advice on strategic issues, and provides the European forum for fundamental science.

This volume is the first of a new series arising from the work of the ESF Scientific Programme on Geographic Information Systems: Data Integration and Database Design (GISDATA). This four-year programme was launched in January 1993 and through its activities has stimulated a number of successful collaborations among GIS researchers across Europe.

Further information on the ESF activities in general can be obtained from:
European Science Foundation
1 quai Lezay Marnesia
67080 Strasbourg Cedex
tel: +33 88 76 71 00
fax: +33 88 37 05 32

EUROPEAN SCIENCE FOUNDATION

This series arises from the work of the ESF Scientific Programme on Geographic Information Systems: Data Integration and Database Design (GISDATA). The Scientific Steering Committee of GISDATA includes:

Dr Antonio Morais Arnaud
Faculdade de Ciencas e Tecnologia
Universidade Nova de Lisboa
Quinta da Torre, P-2825 Monte de Caparica
Portugal

Professor Hans Peter Bähr
Universität Karlsruhe (TH)
Institut für Photogrammetrie und Fernerkundung
Englerstrasse 7, Postfach 69 80
(W) 7500 Karlsruhe 1
Germany

Professor Kurt Brassel
Department of Geography
University of Zurich
Winterthurerstrasse 190
8057 Zurich
Switzerland

Dr Massimo Craglia (Research Coordinator)
Department of Town & Regional Planning
University of Sheffield
Western Bank, Sheffield S10 2TN
United Kingdom

Professor Jean-Paul Donnay
Université de Liège, Labo. Surfaces
7 place du XX août (B.A1-12)
4000 Liège
Belgium

Professor Manfred Fischer
Department of Economic and Social Geography
Vienna University of Economic and Business
Administration
Augasse 2-6, A-1090 Vienna, Austria

Professor Michael F Goodchild
National Center for Geographic Information
and Analysis (NCGIA)
University of California
Santa Barbara, California 93106
USA

Professor Einar Holm
Geographical Institution
University of Umea
S-901 87 Umea
Sweden

Professor Ian Masser (Co-Director and Chairman)
Department of Town & Regional Planning
University of Sheffield
Western Bank, Sheffield S10 2TN
United Kingdom

Dr Paolo Mogorovich
CNUCE/CNR
Via S. Maria 36
50126 Pisa
Italy

Professor Nicos Polydorides
National Documentation Centre, NHRF
48 Vassileos Constantinou Ave.
Athens 116 35
Greece

M. François Salgé (Co-Director)
IGN
2 ave. Pasteur, BP 68
94160 Saint Mandé
France

Professor Henk J. Scholten
Department of Regional Economics
Free University
De Boelelaan 1105
1081 HV Amsterdam
Netherlands

Dr John Smith
European Science Foundation
1 quai Lezay Marnesia
F67080 Strasbourg
France

Professor Esben Munk Sorensen
Department of Development and Planning
Aalborg University, Fibigerstraede 11
9220 Aalborg
Denmark

Dr Geir-Harald Strand
Norwegian Institute of Land Inventory
Box 115, N-1430 Ås
Norway

Dr Antonio Susanna
ENEA DISP-ARA
Via Vitaliano Brancati 48
00144 Roma
Italy

SECTION I

Introduction

1

Generalization: state of the art and issues

J.C. Müller,[a] R. Weibel,[b] J.P. Lagrange[c] and F. Salgé[c]

[a]*Geographisches Institut, Ruhr Universität Bochum, Universitätstraße 150, 44780 Bochum, Germany*
[b]*Department of Geography, University of Zurich, Winterthurerstraße 190, 8057 Zurich, Switzerland*
[c]*IGN, 2 Avenue Pasteur, 94160-Saint Mandé, France*

1.1 Introduction

This chapter intends to provoke discussions and reactions on a number of items relevant to GIS data visualization at multiple levels of scale (the ratio between the size of an object on the map and its real size on the ground) and resolution (the smallest object which can be represented on the map).

From a user point of view, visualization is the window of GIS and is essential for visual data exploration, interpretation and communication. Geographical processes are scale dependent and numerous applications in climate, water resources, agriculture, forestry, transportation, land and urban planning require changing degrees of detail and generalization when analysis and communication occur at the local or more global levels. Hence, there is a need for the modelling of geographic information at different levels of abstraction. Ideally, one should be able to view and analyse data at the level where geographical variance is maximized (Tobler and Mollering, 1972; Woodcock and Strahler, 1987) or where spatial processes are best understood.

From a data production point of view, the management and maintenance of spatial data are constrained by the requirements for accuracy, i.e. 'relationship between a measurement and the reality which it purports to represent' (Goodchild, 1991); precision, i.e. 'degree of detail in the reporting of a measurement' (Goodchild, 1991); and quality control. Requirements for the flexibility afforded by multiple scale production and update operations complicate the issues of accuracy, consistency and integrity.

The question, therefore, is not whether geographic information (in digital or analogue forms) should be made available at multiple levels of abstraction, but how it should be made available.

1.2 Pro- or contra-generalization?

Some authors argue that generalization — in the cartographic sense — is not a prerequisite to the delivery of geographic information at multiple levels of scale and resolution.

The ability of current GISs to zoom in and out of a given area, to break down a single multi-thematic layer into a series of mono-thematic layers, and concurrently to produce multiple windows of the corresponding zooming and layering operations, explains perhaps the historical lack of interest of the GIS community in cartographic generalization. Most GIS commercial firms have denied or ignored the cartographic generalization issue.

Other authors admit that generalization would be a useful tool in the GIS tool-kit but argue that automated generalization is either an 'NP-complete' problem (i.e. a computational solution cannot be devised) or the practical and economic benefits of a solution are dubious. The first view is strong among conventional cartographers. The latter view is shared by many national mapping agencies (NMAs) which store multiple scale versions of manually generalized data. In smaller countries like The Netherlands, where the size of the map series is rather limited, the inconvenience of storage overheads and duplication in updating efforts is perceived as a lesser evil compared to the potential processing cost of an automated generalization solution which has not yet arrived.

For larger countries like France, NMAs are still forced to store multiple scale versions, for a number of reasons: there is no production tool for generalization (on the market place) able to derive the required datasets; there is no tool to propagate updates through a series of derived datasets; the processes of regenerating datasets are expensive and require a long time (hence it is not profitable to carry those processes in an industrial context except once, which explains why the various datasets are maintained more or less separately). Finally, the smaller the scale, the shorter the update cycle. Hence, if a NMA wants to maintain only one scale version then it has to update it frequently, with the higher geometric accuracy, in order to respond to the update needs for all other smaller scale versions. This is a dilemma faced by NMAs which goes against the idea of one single database (the expression 'scaleless or scale-free database', which may lead to confusion, is purposefully avoided here; it seems that in the case of data coming from surveys or photogrammetry, it would be more appropriate to speak of precision, accuracy and resolution, not scale, since the notion of scale is meaningless in the absence of a mapping relation).

As a result, most research efforts to resolve the problem of automated generalization, whether in the context of GIS or for the production of paper maps, have been confined to academia. Some academics have argued that the storage of a finite series of multiple scale cartographic databases provides major impediments from both a scientific and management point of view. From a scientific view point because 'what if' and 'if–then' scenarios in GIS require the possibility of navigating dynamically and continuously from any scale to any other scale automatically. From a management point of view because one cannot afford the duplication of efforts that occur in map series updating as well as the inconsistencies which may arise through this process. Solutions to these impediments require moving beyond the paradigm of traditional paper map series, without sacrificing support for the production of paper maps. That is to say, the generation of digital products can no longer be driven by paper map production, as the needs for spatial data have become much broader and complex. Generalization facilities must be provided by GISs to support the use of geographic information at multiple scales for multiple purposes and tasks. Note that besides GIS electronic displays, paper maps will continue to exist; paper maps are permanent, transportable documents, and they also offer a far better accuracy and can represent more data than the screen of a CRT. Furthermore, maps, as visual communication means, are still the easiest and quickest to read media for communicating geographical information to the reader.

Some mapping agencies and commercial firms are now investing resources to implement automated or semi-automated generalization facilities. Apart from the effect of individual leadership, this could mean that there is a growing belief among professionals that generalization could become an operational tool for the production of geographic information in the years to come. Professionals have also come to realize that the full potential of GIS can only be exploited if functions for automated generalization are available.

Assuming that the answer to the previous hypothesis is true — that is generalization in the context of GIS and automated map production is both desirable and feasible — we are now faced with the theoretical and practical issues of building systems for automated generalization. Some authors have used the term 'generalization machines' (e.g. João *et al.*, 1993) to describe such systems; the term has a strong mechanistic connotation, however, and we wonder whether we can use it in this context.

1.3 Generalization yes, but what are the issues?

We need to distinguish between the issues that are brought about by graphical representation from those which arise from modelling at different levels of spatial and semantic resolution. Generalization may be viewed as an interpretation process which leads to a higher level view of some phenomena — looking at them 'at a smaller scale'. This paradigm is always the first used in any generalization activity, whether spatial or statistical. Second, 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. These two categories have motivated research mainly in two areas: model-oriented generalization, with focus on the first stage above-mentioned, and cartographic generalization, which deals with graphic representation.

Issues relevant to graphical representations are well known to conventional cartographers. In geographical circles, people usually think of generalization as part of cartographic compilation whose purpose is to resolve legibility problems. An operation such as feature displacement is typically cartographic. Should we go beyond this and consider generalization in contexts which are not necessarily representational? The distinction, for example, between cartographic and statistically controlled generalizations was made before (Brassel and Weibel, 1988). Modelling reliability on statistical surfaces by polygonal filtering (Herzog, 1989) is not necessarily directed towards visualization but helps to understand data by providing higher levels of abstraction. In this case, the motivation as well as the solutions to bring about the necessary transformations are not the same as for cartographic generalization. Generalization in the sense of modelling is a requirement for spatial analysis and the tools (e.g. spatial districting and aggregation of spatial enumeration units, image classification, trend surface analysis, surface filtering, and kriging) have already been developed (see, in particular, Tobler, 1966; Tobler and Moellering, 1972). Do we need to consider this category of generalization in our research agenda? Is it relevant to the producers of geoinformation? Is there a need for future research? Or should we close the book on statistically oriented generalization instead?

The first part of this position paper deals with data abstraction, i.e. a reduction of spatial as well as semantic resolution, whether motivated by data analysis or cartographic representation. We will coin this kind of activity under the general term 'model-oriented generalization'.

1.4 Model-oriented generalization

The difference between the model view and the cartographic view of generalization is the possibility for database manipulation in the former case, independent of cartographic representation. Spatial objects may need to have multiple digital representations in which internal representations (models) should be distinguished from visualization (cartographic) representations. One reason for generalization at the modelling level is to facilitate data access in GIS. This need becomes urgent in view of the design of GISs in which the user interacts with the geo-objects without knowledge of their internal representation. Also, model generalization may be driven by analytical queries (Where are the trends?, What is the spatial average?, Where are the new classes to appear at this level of variance?, etc.) whereas cartographic generalization is mainly driven by communication requirements (legibility, graphical clarity, and understandability). But the two types are not independent, and one (model-oriented) can be a precursor to the other (graphics-oriented). The question is how much and what kind of model-oriented generalization support is required for the accomplishment of routine tasks in cartographic generalization?

In model-oriented generalization, methods are currently being developed to support insertions, deletions, updates and geometric queries at an arbitrary location for an arbitrary scale (Becker *et al.*, 1991). A generalization index may be applied to point data which, in turn, defines their priority for access or rescaling operations. Storage structures for seamless, 'scaleless' geographic databases have also been proposed (Oosterom, 1989). Hierarchical data structures, including quad trees and strip trees, are often used to subdivide and merge data for generalization purposes (Jones and Abraham, 1986). The working hypothesis put forward by database experts is that spatial proximity information must be implicitly available in order to favour access to local information and neighbourliness relationships.

The basic categories of space found in the GIS literature, namely metric, topological, and structural categories, can be used to describe various levels of abstraction for spatial objects. The metric space describes distance relations and constitutes the lowest level of abstraction. The topological space, instead, 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). Abstraction of a road network using hypergraphs and graph theoretic concepts is an example of structural representation (Titeux, 1989; Salge *et al.*, 1990). The question is whether we can invent protocols to propagate changes (say through updating) from one level of abstraction to all others. This would go a long way towards detecting inconsistencies between representations.

Other models for data abstraction and data structuring are also available, but are still in the laboratories. For instance, what are the potentials of abstraction mechanisms known from semantic modelling (Smith and Smith, 1977; Hull and King, 1987), including classification, generalization, aggregation and association, in formalizing relations between spatial objects? There has been much excitement about the introduction of object-oriented programming in GIS. Apart from the confusion surrounding the idea of object orientation, most GIS vendors use the concept for advertising purposes. The object-oriented environment, where procedures (methods) are bound to the object, objects communicate with each other and inherit attributes and methods from others, seems to offer great potentials for implementing generalization procedures. The concept of 'delegation between objects', in particular, could be used to perform updates concurrently across all map-scale layers in the database. As with semantic modelling, the proposed models are attractive but have no proven records yet in the field of generalization.

Temporal abstraction is another type of data modelling which expresses changes occurring in spatial objects (and their attributes) at different intervals of time (Langran, 1992). Representations can either be snapshots of the real world, or they can express an average state over a certain interval of time. The subject has become increasingly relevant among custodians of ephemerous spatial databases (particularly in meteorology, forestry and navigation) who require consideration of the problems of object identity and changes not only in the spatial and attribute domains, but also in the temporal domain. The addition of the time dimension raises new problems in data structuring (time is topologically unidimensional) and representation. The tools to analyse, generalize and visualize temporal information are still in their infancy.

Model-oriented generalization research has been somewhat neglected in comparison with the efforts invested in graphics-oriented generalization. The traditional view of generalization in support of surveying and mapping organizations for multi-scale map production is overwhelming and has been much more studied. Busy implementing algorithms to perform the analogue of cartographic generalization tasks such as simplification, exaggeration, elimination and displacement, we have forgotten the intimate relationship between generalization at the modelling level and generalization at the 'surface' (e.g. graphical representation). Cartographic generalization requires (1) inside information regarding a spatial object (including spatial, semantic and perhaps temporal aspects), and (2) outside information regarding the relationship among objects and their contextual relevance. The resolution of conflicts, for instance, typifies the problem of generalization on the 'surface', but requires both types of information for its solution. As mentioned earlier, the way the data model is organized and can be generalized is likely to influence the performance of cartographic generalization.

1.5 Cartographic generalization

The tools currently available for automated cartographic generalization resemble those of manual generalization. In this sense, efforts in the automatic domain are oriented towards the manual domain. Furthermore, the quality of computer-produced maps is often tested by comparing the results with manually produced ones. The question is whether we should use manually generalized maps as a criterion of good performance for automated generalization. Should automatically produced maps look like manual ones? This is perhaps a dubious goal and probably unrealistic. Some authors have argued for methods whose results mimic the way people generalize by looking at objects from a distance (Li and Openshaw, 1993). But the fact remains that no new paradigms have emerged under the hat of automated generalization.

A prior attempt towards automated cartographic generalization was to provide a theoretical foundation by answering questions such as what, why, when, and how should we generalize, and providing a framework of objectives to attain, including philosophical, application, and computational ones (McMaster and Shea, 1988). A second step was to make an inventory of the tools available in order to attain those objectives. The list and the definition of those tools vary among generalization specialists, mainly because they fail to differentiate between the transformation applied to an object and the operators used to perform this transformation (Ruas *et al.*, 1993). For example, in the process of simplification, we can list various operators, including select-and-delete, aggregation, compression, smoothing, caricaturization, and collapse. Nevertheless, such inventories were useful since they were used as 'cahier des charges' by commercial firms to set up their development agenda. For instance, a partial catalogue of generalization operators has been already implemented or is intended to be developed by INTERGRAPH, including selection/elimination, simplification, typification, aggregation, collapse, classification, symbolization, exaggeration, displacement, and aesthetic refinement (Lee, 1993). A third attempt was to model the generalization process by suggesting sequential and recursive scheduling scenarios of the generalization steps involving different operators. Those could be different depending on the map subject (Lichtner, 1979; Müller, 1991; Lee, 1992; Müller and Wang, 1992).

One can essentially distinguish between two approaches for the implementation of the working tools in automated generalization. One is batch while the other is interactive.

1.6 Batch generalization

At the most basic level, we have a batch approach where individual algorithms are used to execute various tasks (elimination, simplification, etc.) applied to various kinds of objects. Line generalization has been the most thoroughly studied subject in academic circles (for over 20 years). As with map projections, new algorithms for line generalization keep popping up in the literature. This is no coincidence. Eighty per cent of the cartographic objects are perceived to be lines (in fact, many of them are polygons). Furthermore, single lines viewed in isolation are easier to handle than complex objects like a building or a polygon nesting. Can we now claim that we have reached the state of the art in line generalization? Probably not, especially in view of a lack of theory as to which algorithm is the most appropriate for which line object (river, contour, road, census boundary). Perhaps we need to concentrate more on the application of existing algorithms than on the invention of new ones (Weibel, 1991b). Besides line simplification, we now dispose of algorithms to aggregate and simplify polygons, to exaggerate object size, to collapse complex objects into simpler ones, and to classify and to symbolize cartographic features. But we need an inventory of the performance and the applicability of the different algorithms currently available at universities, national mapping agencies or in private industry. Nobody has a really clear view of what is exploitable. In the generalization tool-kit, however, displacement is not well represented. This is without doubt the most difficult operator to implement, and although some solutions are available (Lichtner, 1979; Nickerson, 1988; Jäger, 1990), they are not comprehensive enough to cover the entire range of possible conflicts. Displacement has become a priority item on the research agenda.

Going one step further, individual batch solutions may be bundled into one 'total' comprehensive batch solution that can be applied for the generalization of an entire map composed of many different objects. Issues such as scheduling management and object interaction have then to be resolved.

A program such as CHANGE, developed at the University of Hannover, is a combination of procedural steps which comes close to the idea of a 'total' solution (Powitz and Schmidt, 1992; Gruenreich, 1993). To develop effectively such a program, one has to define clear objectives. In the case of CHANGE, for instance, the goal was to provide the automated generalization of some feature classes of German topographic maps for a limited range of scales, going from 1:5 000 to 1:25 000. Even in this case, however, the program is suboptimal in the sense that it performs only 50 or 60 per cent of the work. At the end, the user is still required to intervene to perform the necessary quality control and corrections required by operations that could not be entirely automated, such as displacement of conflicting objects in complicated surroundings. As a further example, Nickerson (1988) developed a system for automated generalization of topologically structured cartographic line data. The system is capable of handling feature elimination, feature simplification, and interference detection and resolution. The system is implemented in Fortran, but uses English-like rules for the user to specify generalization options. The intended scale range is 1:24 000 to 1:250 000.

The question is whether a 100 per cent batch solution in generalization will ever be attainable (or desirable). Performance in batch solutions is more likely to follow the economics of 'diminishing returns'. The landscape of geographical features portrayed on topographic maps, for example, can vary almost to infinity. This great variation creates generalization problems which cannot all be foreseen and the research required to cover all cases is so complex and so demanding that it would not be economical. The situation may improve in the future, however, when our methods will be derived from the 'deep' structure (semantic and topology) rather than from the surface level (form and size), and, therefore, will be less sensitive to the variation of individual objects.

1.7 Interactive generalization

The difficulties of providing a batch solution and the disappointment over the progress of the formalization of generalization knowledge (see below) have led some researchers to put their efforts towards the exploitation of interactive techniques. In this case, low-level tasks are performed by the software, but high-levels tasks, such as the choice of an object to be generalized or a particular routine or parameter, are performed or controlled by humans. In other words, the computer implements some tasks (usually execution) which it is good at solving but relies on the user for control and knowledge. Such an approach was suggested by Weibel (1991a) and was termed the 'amplified intelligence approach'. Furthermore, batch technology reflects a line of thought more appropriate to the 60s and 70s than to the 90s. The present trend is to use the interactive environment made available through work stations, PCs and powerful interfaces. So one might say that the dichotomy between batch and interactive generalization is rather artificial and will vanish in the future.

Interactive solutions are based on a user-friendly interface (including multi-window displays, pull-down menus, tool palettes, and menu shortcuts) which allows the user to navigate easily through the system's options and select the objects to be generalized as well as the tools used for generalization. Weibel (1991a, 1991b) gives a detailed list of components required for an 'amplified intelligence' system (mentioned above). Among these are facilities that support the user in making correct generalization decisions (e.g. measures giving data statistics or indicating object complexity; query and highlight functions, etc.) as well as functions for logging of interactions and scripting facilities required. For an interaction approach to be successful, it is essential that it does not just replace the cartographer's pen, but really enables the user to make decisions about generalization on a high level, that is, the system must be capable of amplifying human intelligence. The approach of interactive systems could also be regarded as an equivalent to decision support systems which are frequently used in business and planning applications (Sprague and Carlson, 1982).