GIS and GENERALIZATION

Methodology and Practice

GISDATA 1

Series Editors
Ian Masser and
François Salgé

EUROPEAN SCIENCE FOUNDATION
ESF

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

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

Copyright © Taylor & Francis Ltd 1995

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, electrostatic, magnetic tape, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

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'.

13

The importance of quantifying the effects of generalization

Elsa Maria João

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

13.1 Introduction

One of the most important characteristics of Geographical Information Systems (GIS) is their ability to analyse spatial data. This can, however, be jeopardized by data transformations which detrimentally affect the data quality (Rhind and Clark, 1988). One such transformation that can alter the results of GIS map manipulations is the generalization of geographical data. Generalization can be described as a process by which the presence of geographical features within a map is reduced or modified in terms of their size, shape or numbers (Balodis, 1988). The end-product of a generalization process is therefore a derived dataset with less complex properties than those of the original dataset. For certain GIS applications, the maintenance of one large, detailed database could partly solve the problems caused by generalization. However, the ability to generate less-complex data from a detailed source dataset is still fundamental to GIS.

Within a GIS, generalization needs to be performed for three main reasons. The first, most obvious purpose is for *display*. The plotting of a map is one of the most common and useful outputs from a GIS map manipulation. Depending on the output scale and the detail of the source map, it might be necessary to generalize the data to improve depiction following the same principles used by manual cartographers. A map can often help communicate the results of a complex analysis in a clearer way than a set of tables and graphs. This can play an important role, e.g. in environmental impact studies where the environmental scientists carry out the analysis but often it is politicians that take the final decisions. Generalization carried out for display purposes is called *cartographic generalization.*

The second reason for the need to generalize within a GIS is strictly for *data reduction.* Driven either by financial or technological constraints, it may be necessary to generalize data in order to reduce the amount of data storage or processing time. The third main reason for generalizing in a GIS is for *analysis*. This can involve the use of generalization for homogenizing datasets which have different resolution or accuracy levels (see Weibel, Chapter 5). Most importantly, generalization as an analytical tool can be used to help understand at which scale spatial processes occur (Müller, 1991). Generalization carried out either for data reduction or for analysis is called *model generalization*. Some results of model generalization can also be displayed but, unlike cartographic generalization, the generalization was not done for the sake of graphic clarity.

Generalization, therefore, poses a dilemma to the GIS user. On the one hand it is necessary to generalize in order to improve the display quality of a map at a scale smaller than the one it was compiled from, or to allow analysis with different degrees of detail. However, on the other hand, generalization can potentially cause unintended transformations of the data that can alter the topology of geographical phenomena, and affect subsequent statistical or geometrical calculations. Normally, GIS users would want to minimize, control and quantify the effects of generalization on their results. Knowledge about the type and magnitude of generalization effects embedded within spatial datasets should therefore be deemed essential by any GIS user. Despite this, almost all the published literature describe only qualitatively the consequences of generalization. Other researchers (such as Beard, 1988) have measured generalization effects of specific manually generalized features, but solely in order to evaluate automatic generalization procedures. Blakemore (1983) and Goodchild (1980) investigated generalization effects directly but did not compare maps across different scales. Moreover, the qualitative descriptions have been mostly restricted to the domain of cartographic generalization. Thapa and Bossler (1992, p. 838), for example, point out that 'substantial' shifts can occur in actual ground terms due to cartographic generalization, but no figures are given.

This chapter discusses the importance of the generalization effects caused by manual and automated generalization, drawing on some of the findings of a quantitative study carried out by the author. João (1994) compared different source scale maps for two different study areas in order to measure the generalization effects embedded in the smaller-scale maps. GIS map manipulations were also carried out, using the same features taken from the different source scale maps, in order to determine the consequences of generalization on the results. The maps used had been manually generalized but, in addition, the largest scale map for each study area was generalized using the Douglas–Peucker algorithm and the results compared with the manually generalized maps. This chapter interprets the results of the study by João (1994) within the context of cartographic and model generalization. In order to illustrate the ideas presented in this chapter, an additional overlay operation was also carried out. This chapter concludes with a proposal by which, in future, automated generalization could increase its scope by encompassing the quantification and control of generalization effects.

13.2 Why we should still be concerned with the effects of cartographic generalization

Because the purpose of cartographic generalization is for display (i.e. it is a visual-oriented process), it can be deduced that some of its products are probably ill-fitted for certain analyses. 'One can not expect a database which was generalised using procedures for cartographic generalisation to be a reliable source of predictable quality for analysis, because manipulation by cartographic generalisation will introduce unpredictable errors through processes such as feature displacement' (Brassel and Weibel, 1988, p. 236). Data quality should determine which datasets are used for analysis. However, at present, there is widespread use of cartographically generalized products for analysis, independent of their quality. This is mainly because the vast majority of existing spatial databases have been digitized from existing map series (such as the ones produced by national mapping agencies) which were originally generalized for display reasons (Fisher, 1991). The quality of these databases remains largely unknown.

Most of the spatial databases derived from analogue map series do not have associated accuracy values. One reason for this is the difficulty of associating quality measures with digital data (see Smith and Rhind, 1993). Even when values are given they are often vague, relating, for example, to the dataset as a whole rather than to individual features or feature types, and they usually fail to refer specifically to the error associated with generalization. It has actually been found much easier to quote error values for recent map products that are minimally generalized. This has been the case with the Land-Line93, an Ordnance Survey product, derived from very large-scale maps (Smith and Rhind, 1993). Generalization error is very difficult to quantify (Thapa and Bossler, 1992) because the amount of error introduced depends on the type of feature, its proximity to other features and *when* the feature was inserted in the database. As new features (e.g. a new road) are updated, not only can they be drawn using different techniques, but they are often forced to adapt to the existing features in the database (the cartographer maintaining relative accuracy at the expense of absolute accuracy). As a consequence of this, the same type of feature can have different positional accuracy in different parts of a map according to the density of features and when the different parts of the map were updated.

The quantification of the effects of generalization can be especially problematic in that most GIS users do not produce their own data. Externally supplied data, obtained either in a digital or analogue format, will already have generalization effects embedded in them — the type and magnitude of these effects are often unspecified. An exception to this is the Australian Survey which indicates average displacements between different feature types on their 1:250 000 scale topographic database. For example, a typical displacement of up to 200 m can be observed in situations in which one road and one railway are almost coincident and the road must be moved to ensure clarity (Australian Survey, 1992). However, it is not stated which features or portions of features suffered displacements of this magnitude, if any at all.

Given the present situation, GIS users will often ignore the lack of accuracy of their digital data due to generalization. Because of this, generalization has been pointed out as one of the main causes as to why the magnitude of errors in current GIS databases can be larger than the errors within their analogue counterparts. According to Openshaw (1989) the magnitude of errors in spatial databases can surpass those introduced by traditional cartographic manipulations of paper maps due to the way GIS perform operations on cartographic data 'which traditionally would not have been done, or else performed only under special circumstances, because of the problems of scale, complexity, and feature generalisation that might be involved' (Openshaw, 1989, p. 263).

Most GIS users are therefore put in the position in which the only available digital products are the ones derived from traditional analogue map series, often of an unpredictable quality. This lack of error information about cartographically generalized maps would be of less importance if they were not used for analysis. Unfortunately, in practice, these products are widely used to support spatial queries for GIS applications. In order to evaluate the consequences of cartographically generalized maps on GIS operations, João (1994) carried out an extensive quantitative study. This study found that cartographic generalization could strongly affect basic measurements such as length. For example, some features' length changed as much as 23 per cent between the scales 1:50 000 and 1:500 000. More importantly, features were especially affected in terms of positional accuracy. Displacements of well-defined points (such as intersections between roads and railways) as large as 994 m were found between the scales 1:50 000 and 1:625 000. The results also showed that the magnitude of the generalization effects could be exacerbated in a typical GIS analysis. This is because for most GIS map manipulation operations any generalization effect can be compounded when two or more features are considered simultaneously.

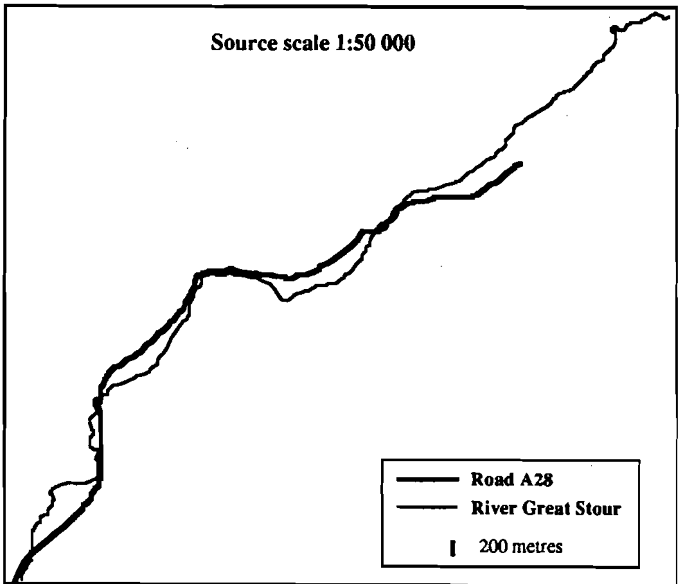
In addition to the fact that cartographic generalization can strongly affect the quality of digital data, the unpredictable nature of this process makes it a more serious and difficult problem to tackle. The study by João (1994) found that the results of cartographic generalization were very variable according to feature classes but were even more unpredictable across individual features. It was often difficult to predict beforehand which of the feature classes would be most affected by generalization and for which of the different scales. For example, the feature class that presented the highest lateral shift at one scale was not necessarily the same as that having the highest value at another scale. This made it impossible to generate hard and fast cartographic generalization rules for individual feature classes. The variability and associated unpredictability of cartographic generalization is encapsulated in the notion that often manual generalization is governed by *best depiction* rather than by a *particular rule* (Geoff Johnson, Ordnance Survey, 1993, personal communication). The fact that cartographic generalization does not always produce predictable results, due to the extreme variability of the distribution of geographical phenomena, is a major difference between cartographic and model generalization.

The unpredictability of the effects of cartographic generalization reflects itself more seriously in the results of GIS map manipulations. In order to illustrate further the effects of generalization on a typical GIS map manipulation, a straightforward overlay operation was carried out to find out the length of a certain road which lay, respectively, within 100, 200 and 300 m of a river. Figure 13.1 shows the river (the Great Stour, that passes through the city of Canterbury in the south of England) and the road (the A28) used in the overlay operation.

The overlay was repeated using the same river and same road taken out of three different source scales: 1:50 000, 1:250 000 and 1:625 000. The data were obtained in a digital format from the Ordnance Survey of Great Britain. The 1:50 000 data were part of a trial digital product, but data from the other two scales are current digital cartographic data products sold by the Ordnance Survey. Table 13.1 shows the results of the overlay operation for the three distances when the features for the three different scales were used. It can be seen, for example, that the length of the road that lies within 100 m of the river Great Stour at the scale 1:250 000 is only 35 per cent of its length at the scale 1:50 000, while at the scale 1:625 000 it is only 52 per cent of its original length.

The results presented in Table 13.1 illustrate the issues mentioned above in connection with the effects of cartographic generalization. The fact that in the case of the 100 m buffer at the scale 1:250 000 only 35 per cent of the original road length remains, reinforces how large the effect of cartographic generalization can be. The reason for such a large effect is partly due to the geographic position of the two features — as the road and the river run parallel to each other (see Figure 13.1), even small displacements can cause a large impact on the results. The extent of cartographic generalization (such as the simplification, displacement and elimination of selected detail) is determined by the cartographer's aim to achieve a satisfactory representation of the landscape and this, in turn, is affected by the original landscape itself.

Moreover, the generalization effects do not necessarily increase progressively with the decrease of scale. In the case of the 100 m buffer, the overlay generated the smallest road



N.B. The road ends at the city boundary of Canterbury.

Figure 13.1 The river and the road used in the overlay operation.

Table 13.1 Results of the overlay using manually generalized features

	Length of the road A28 within certain distance of the river Great Stour (m)		
Scale	Within 100 m of the river	Within 200 m of the river	Within 300 m of the river
1:50 000	4548	7287	9658
1:250 000	1584 (35%)	5433 (75%)	8946 (93%)
1:625 000	2385 (52%)	3135 (43%)	4973 (52%)

length at the middle scale rather than at the smallest scale. Because of the unpredictability of the generalization effects, the way generalization influences GIS map manipulations is also not straightforward. Deriving data from a larger scale does not necessarily mean a reduction of generalization effects (cf. where model generalization can actually be used to increase accuracy — see Müller, 1991). With cartographic generalization some generalization rules used by cartographers might actually cause larger generalization effects at larger scales. Such is sometimes the case when a house is positioned between a road and a river. If, at a particular scale, a rule ensures that the house should be included, then this house might force the river and the road to move apart to allow room for the clear depiction of the house. However, if at a smaller scale the house no longer needs to be shown, then the river and the road might no longer need to be displaced.

The findings of this overlay operation corroborate the results of João (1994). They show both that errors associated with cartographic generalization can be large and unpredictable — especially in the case of more complex GIS map manipulations that use a large variety of features. As a consequence, it is particularly important to measure the impact that these errors are causing on GIS map manipulations, as long as cartographically generalized data continues to be used for analysis purposes. In the future — as digital data specifically tailored for analysis become increasingly accessible — the use of cartographically generalized products for analysis purposes will almost certainly diminish. Only when digital data generalized by model generalization (i.e. for analysis) are commonly available, might there be a less pressing need to attach accuracy values to cartographically generalized maps. The control of the data quality of *model* generalization will instead become increasingly important.

13.3 Model generalization and data quality

The importance of data quality can be encapsulated in the notion that information of unreliable quality can be worse than no information at all. For many GIS applications, users will only be able to evaluate the fitness for purpose of their data if they have access to specific quantitative measures of accuracy (Chrisman, 1991). Data can be considered as the most valuable asset of a GIS (Rhind, 1991) and so this concept of fitness for purpose within a GIS relates to the *multiple* applications which the data have the potential to be used for. In the situations in which data might be needed for purposes that had not been foreseen when the data were created, it is essential to have information on data quality to determine the suitability of the data for those new applications. The importance of enabling users to judge the fitness of data for their own use has been recognized by the US Spatial Data Transfer Standard (SDTS) which requires a data quality report coupled with every data transfer (Fegeas *et al.*, 1992).

Model generalization is usually done for analytical reasons and therefore, by definition, needs to keep close control of the impact of generalization on data quality. According to Brassel and Weibel (1988), model generalization aims at minimizing error, such as minimum average displacement, and so has to be done under parametric control. As more tools for model generalization start being developed (see Grünreich, Chapter 4), there will be a pressing need for a systematic analysis of the effects of model-oriented generalization.

In order to contrast the effects caused by model generalization with those caused by cartographic generalization, João (1994) generalized features with the commonly used Douglas–Peucker algorithm (Douglas and Peucker, 1973) and compared them with the same features generalized by manual cartographic methods. The Douglas–Peucker algorithm is an example of an algorithm which is specific to neither analysis nor display — it depends why and how it is being used. Because the algorithm was used with very small tolerances, no potential topological errors occurred (such as lines crossing back on themselves — see Müller, 1990; Visvalingam and Whyatt, 1990). This meant that, for the tolerances used, it conformed with model-oriented quality objectives (see Weibel, Chapter 5) and so can be considered a model-generalization tool.

João (1994) found that generalization effects were typically greater in manually generalized topographic maps than in those produced by the Douglas–Peucker algorithm. Automatic generalization retained length and angularity very well, and most importantly, displaced features much less. In other words, it caused much less distortion than manual generalization. This is because the Douglas–Peucker algorithm only filters the high-frequency components, causing a reduction in local detail of the lines without the more global displacement that the manually generalized lines often suffered. This is related to the way most line simplification algorithms work — they do not displace a line along its entire length, as even if the line loses the majority of its points, its first and last point will remain constant. These findings support the results of Beard (1987) who found that, for a thematic map depicting land and water areas, both positional and attribute errors were reduced by automatic generalization.

Model and cartographic generalization share the use of some common generalization tools, such as selection, simplification and smoothing. However, the tools exclusive to the use of cartographic generalization (such as enhancement, feature displacement and shape change) are the ones that are particularly responsible for the displacement and distortion of mapped features. João (1994) also found that because cartographically generalized features usually suffered more displacement than model generalized features (e.g. in terms of areal displacement as defined by McMaster, 1987), the results of GIS map manipulations were less affected by model generalization.

Table 13.2 illustrates how model generalization can affect the results of a typical GIS map manipulation to a lesser degree than cartographic generalization (compare with the results presented in Table 13.1). It shows the results of the same overlay operation as described in section 13.2, but this time using automatically generalized features. The road and the river at the scale 1:50 000 were generalized using the Douglas–Peucker algorithm. The level of generalization was determined by the number of crucial points (i.e. spurious points which did not add any extra detail were discounted) used to represent the road and the river at the scales 1:250 000 and 1:625 000.

It can be seen from Table 13.2 that although, in general, there is a constant increase in the generalization effects, these effects are much smaller than those for the manually generalized lines (cf. Table 13.1). However, there still remains some degree of unpredictability. In the case of the 300 m buffer, the result of the overlay at the smallest scale generated a slightly larger road length than at the largest scale. As in the case of the lines generalized manually, this was due to a combined effect of the reduction of length and the change of the relative position of the road and the river. The reduction by the algorithm of the number of points used to represent the lines caused a decrease of the road length but at the same time caused the sideways shift of sections of the road or the river (vector displacement as described by McMaster, 1987). It was this sideways shift which caused sections of the two lines to lie closer together, and therefore a lengthening of the road within 300 m of the river at scale 1:625 000.

Table 13.2 Results of the overlay using automatically generalized features

	Length of the road A28 within certain distance of the river Great Stour (m)		
Scale	Within 100 m of the river	Within 200 m of the river	Within 300 m of the river
1:50 000	4548	7287	9658
Equivalent to 1:250 000 (automatically generalized)	4488 (99%)	7281 (100%)	9626 (100%)
Equivalent to 1:625 000 (automatically generalized)	4461 (98%)	7122 (98%)	9822 (101%)

Despite the fact that it has been found that automated generalization can cause less generalization effects than manual generalization (Beard, 1988; João, 1994), the data quality control of algorithms such as the Douglas–Peucker is still very rudimentary. At present, most generalization algorithms lack mechanisms for the control of the quality of their output. Ideally, a GIS user would require help in the choice of the algorithm's tolerances so as to avoid, for example, generating lines which were excessively spiky or lines crossing back on themselves. So far, to a large extent, the control of the consequences of generalization on data quality has been invariably dissociated from the research into its automation. This is despite the fact that the automation of generalization would benefit from simultaneously scrutinizing the quality of its generalized products.

13.4 Increasing the scope of automated generalization by controlling its effects on data quality

Within a GIS, not only are the consequences of uncontrolled generalization more serious, but at the same time the quantification and control of generalization can be made easier by automation. Despite this, most of the effort in automating generalization has been dissociated from the control of generalization effects. As a consequence, most of the existing automated generalization tools cannot control the distortion they cause on the data, except for the elementary setting of the generalization tolerance (João, 1991). In particular, as model generalization advances, it is important that a system includes the minimization and quantification of unwanted generalization effects as an integral part of its model generalization tools.

To overcome this problem, a three-stage control process of generalization effects based on quality measures is proposed here. The first stage in this process would start by evaluating the need for generalizing. The control of generalization effects could then start even before the generalization process begins. GIS functions could give advice to the user about whether generalization is required or advisable — for example, by identifying conflicts (e.g. caused by overlapping or imperceptible features) using the approach suggested by Beard and Mackaness (1991). It is also important to define the limits of what is possible and sensible to attempt, and to educate the user about these limits. For example, there may be situations where it is inadvisable to combine a dataset at a particular scale with data from a very different scale, without generalizing the larger-scale data.

If it was confirmed that generalization was needed, then the second stage could control how, and how strongly, features were generalized, based on quality parameters. This could be carried out using the system suggested by João *et al.* (1993) in which the selection of the best available generalization procedure (in terms of algorithms and respective tolerances) was based on purpose, scale, data types and, most importantly, *quality requirements*. This improved automated generalization system would select generalization processes on the basis of the minimization of generalization effects as specified by the user. For example, the extent of generalization effects could be minimized by avoiding over-simplification of a line by the user specifying the maximum allowed length change or displacement.

The final stage of generalization control would take place after the generalization process was completed. This would involve a quantification and storage of the transformations caused by generalization. The quantification of generalization effects can be done at two different levels of detail. The simplest would entail storing only the quality constraints used in the generalization procedure (e.g. the maximum threshold value of allowed deviation between models — see Weibel, Chapter 5). A more detailed quantification of generalization effects would entail actually measuring what had happened. These *quantified* generalization effects could then be used to tag the data (so as to avoid misuse) or could be taken into account in GIS map manipulations based on error propagation models (Goodchild and Gopal, 1989; Openshaw *et al.*, 1991).

13.5 Conclusions

It is often argued that a dataset generalized for display purposes should not be used for analysis. However, this ignores the fact that, at present, they often are. If cartographically generalized products are the only type of digital data available, then GIS users have no other alternative than to use them. At the same time, if these cartographically generalized products lack quality values associated with them, it will be very difficult (or even impossible) for users to evaluate the acceptability of these datasets for analysis purposes.

This chapter has discussed how cartographic generalization effects are usually more serious than the effects caused by model generalization. The implications of this are twofold. First, it emphasizes the importance at the moment of controlling and quantifying the effects of cartographic generalization on GIS map manipulations, while a large proportion of datasets used for analysis continue to be derived from cartographically generalized analogue maps. Second, it points out the pressing need for the development of more model generalization tools for generating datasets specifically suited to GIS analysis. As datasets derived from model generalization become increasingly available, it will, in turn, become a priority to evaluate and control their generalization effects.

Generalization is important to GIS, not only because of the drive towards more and better-automated tools but also due to the way generalization affects data quality. The development of automated tools and control of generalization effects have so far been dissociated to a large degree. In future, however, the control of generalization effects should be an extra parameter to be taken into account during the automation process. Research into the control of the quality of the products produced by generalization does not need therefore to be separated from the research into the automation process.

Independently of the development of a more intelligent system that would control generalization effects according to the preferences of users (João *et al.*, 1993), there is still scope for improving the information given to users about the type and magnitude of generalization contained in spatial datasets. If quality requirements are used to guide generalization, it should be possible to store these quality requirements as indicators of the quality of the generalized product. The information given by the Australian Survey (1992) in terms of the expected displacements for different feature types according to the number of features under conflict, sets the example that other mapping agencies should follow and enhance.

In addition, all digital datasets should be accompanied by a statement on data quality statistics that would indicate to the user the accuracy of the different features in different parts of the map. SDTS suggests three methods of quality reporting: textual narration, defined-quality attributes, and quality overlays (Fegeas *et al.*, 1992). For example, in order to help the user visualize generalization error in the database, 'grey boxes' could be drawn around heavily cartographically generalized areas to warn about excessive generalization effects and possibly motivate the user to obtain larger-scale maps (or to investigate the possibility of using data derived from model generalization).

In future, it might be the case that mapping agencies will commonly supply different datasets at the same source scale (say 1:250 000), but one which resulted from cartographic generalization and another that resulted from model generalization. These two distinct datasets would have different accuracy levels, bearing in mind the different types of uses that the data would have. However, the concept of fitness for purpose can only be evaluated and judged by the users themselves and not by the producer of the data (Chrisman, 1986). For most applications, this requires access to specific quantitative measures of accuracy rather than general qualitative statements. It is therefore fundamental that when automating generalization there is an associated quantification and control of the transformations suffered by the data.

Acknowledgements

Comments and suggestions from anonymous reviewers are gratefully acknowledged.

References

Australian Survey, 1992, *TOPO-250k Data User Guide*, Belconnen: Australian Surveying and Land Information Group.

Balodis, M., 1988, Generalisation, in Anson, R. (Ed.) *Basic Cartography for Students and Technicians*, 2, pp. 71–84, London: Elsevier, on behalf of the International Cartographic Association.

Beard, K., 1987, How to survive on a single detailed database, in *Proceedings of AUTO-CARTO 8*, Baltimore, MD, 30 March–2 April, pp. 211–20.

Beard, K., 1988, 'Multiple representations from a detailed database: a scheme for automated generalization', unpublished PhD thesis, The University of Wisconsin-Madison.

Beard, K. and Mackaness, W., 1991, Generalization operations and supporting structures, in *Proccedings of AUTO-CARTO 10*, Baltimore, MD, Vol. 6, pp. 29–45.

Blakemore, M., 1983, Generalization and error in spatial data bases, in *Proceedings AUTO-CARTO 6*, Falls Church, VA, 1983, pp. 313–22.

Brassel, K. and Weibel, R., 1988, A review and framework of automated map generalization, *International Journal of Geographical Information Systems*, **2**(3), pp. 229–44.

Chrisman, N., 1986, Obtaining information on quality of digital data, in *Proceedings of AUTO-CARTO London*, London, Vol. 1, pp. 350–8.

Chrisman, N., 1991, The error component in spatial data, in Maguire, D., Goodchild, M. and Rhind, D. (Eds) *Geographical Information Systems: Principles and Applications*, pp. 165–74. Harlow: Longman.

Douglas, D. and Peucker, T., 1973, Algorithms for the reduction of the number of points required to represent a digitised line or its caricature, *Canadian Cartographer*, **10**, 112–22.

Fegeas, R., Cascio, J. and Lazar, R., 1992, An overview of FIPS 173, the Spatial Data Transfer Standard, *Cartography and Geographic Information Systems*, **19**(5), 278–93.

Fisher, P., 1991, Spatial data sources and data problems, in Maguire, D., Goodchild, M. and Rhind, D. (Eds) *Geographical Information Systems: Principles and Applications*, pp. 175–89, Harlow: Longman.

Goodchild, M., 1980, The effects of generalisation in geographical data encoding, in Freeman, H. and Pieroni, G. (Eds) *Map Data Processing*, pp. 191–205, New York: Academic Press.

Goodchild, M. and Gopal, S. (Eds), 1989, *Accuracy of Spatial Databases*, London: Taylor & Francis.

João, E., 1991, The role of the user in generalisation within GIS, in Mark, D. and Frank, A. (Eds) *Cognitive and Linguistic Aspects of Geographic Space*, pp. 493–506, Dordrecht: Kluwer Academic Publishers.

João, E., 1994, 'Causes and consequences of map generalisation', unpublished PhD thesis, Birkbeck College, University of London.

João, E., Herbert, G., Rhind, D., Openshaw, S. and Raper, J., 1993, Towards a generalisation machine to minimise generalisation effects within a GIS, in Mather, P. (Ed.) *Geographical Information Handling — Research and Applications*, pp. 63–78, Sussex: John Wiley.

McMaster, R., 1987, The geometric properties of numerical generalisation, *Geographical Analysis*, **19**(4), 330–46.

Müller, J.-C., 1990, The removal of spatial conflicts in line generalisation, *Cartography and Geographic Information Systems*, **17**(2), 141–9.

Müller, J.-C., 1991, Generalization of spatial databases, in Maguire, D., Goodchild, M. and Rhind, D. (Eds) *Geographical Information Systems: Principles and Applications*, pp. 457–75, Harlow: Longman.

Openshaw, S., 1989, Learning to live with errors in spatial databases, in Goodchild, M. and Gopal, S. (Eds) *Accuracy of Spatial Databases*, pp. 263–76, London: Taylor & Francis.

Openshaw, S., Charlton, M. and Carver, S., 1991, Error propagation: a Monte Carlo simulation, in Masser, I. and Blakemore, M. (Eds) *Handling Geographical Information: Methodology and Potential Applications*, pp. 78–101, Harlow: Longman.

Rhind, D., 1991, Data access, charging and copyright, and their implications for GIS, in *Proceedings of the Second European Conference on GIS*, Brussels, Vol. 2, pp. 929–45.

Rhind, D. and Clark, P., 1988, Cartographic data inputs to global databases, in Mounsey, H. and Tomlinson, R. (Eds) *Building Databases for Global Science*, London: Taylor & Francis, pp. 79–104.

Smith, N. and Rhind, D., 1993, Defining the quality of spatial data: a discussion document, in *Proceedings of the Land Information Management and GIS Conference*, University of New South Wales, July, pp. 199–207.

Thapa, K. and Bossler, J., 1992, Accuracy of spatial data used in Geographic Information Systems, *Photogrammetric Engineering and Remote Sensing*, **58**(6), 835–41.

Visvalingam, M. and Whyatt, J., 1990, The Douglas–Peucker algorithm for line simplification: re-evaluation through visualization, *Computer Graphics Forum*, **9**(3), 213–28.