HOW TO SURVIVE ON A SINGLE DETAILED DATABASE

M. Kate Beard
Department of Landscape Architecture, 25 Ag Hall
University of Wisconsin-Madison, Madison, WI 53706

ABSTRACT

This paper reports on an approach to a data base that assumes only a single detailed coverage is required in place of separately derived coverages at more generalized scales. Less detailed or smaller scale representations are produced by manipulation of the detailed coverage. The suggested approach requires scale changing and generalization tools to handle large reductions of detail. This approach is demonstrated using nautical chart data ranging from 1:10,000 to 1:40,000 scale. Some initial results are tested by comparison to an independently generalized product.

INTRODUCTION

It would be ideal for small scale maps to be derived from large or medium scale maps (Watson 1970). In such a case only a single survey would be required. In a digital environment, it should be less time consuming, less costly and potentially less error prone to convert, store, and maintain a single large scale coverage and derive all smaller scale coverages. In practice this approach has rarely been followed. The conventional order of national mapping priorities is to achieve complete coverage at small or medium scales first since these can be completed most rapidly. The USGS digital mapping effort epitomizes this priority scheme. Although their original intent was to start with the 1:24,000 series, the actual order of production was the conversion of the 1:2,000,000 scale series followed by conversion of the 1:000,000 scale sheets. Their future plans include digitizing the 1:24,000 scale series which will assure that a digital version of each scale will exist. Given the capabilities of digital mapping at the time this effort was started their order of production is perhaps understandable. Early systems were not well equipped to handle the bulk of large scale mapping nor did the Survey probably wish to risk a large investment on still experimental technology.

There are, however, sufficient disadvantages in conversion and maintenance costs, and map series inconsistencies that this approach should not be perpetuated. An alternative approach is to convert, and maintain a single detailed coverage. Once this coverage is complete it should be sufficient. All smaller scale coverages could then be derived from this source. Implementation of this approach, however, is not a simple matter. Detailed information is time consuming and expensive to collect, and once the data is collected the conversion to any smaller scale is not a trivial matter. A single coverage approach requires considerations far beyond line generalization algorithms and factor of 2 scale reductions. This paper begins with a discussion of the advantages and disadvantages of this approach then proposes a data base design and collection of automated generalization tools that would be required to carry it out.

ADVANTAGES OF A SINGLE COVERAGE APPROACH

Reliance on a single detailed data base has several advantages. It eliminates the time and cost of converting several smaller scale versions of the same coverage. It simplifies and reduces the cost of maintenance and updates since an update need only be applied to one coverage. Inconsistencies due to time lags in updating several different scale coverages are removed. Also inconsistencies due to application of different source material to different scale versions could be eliminated. If a range of scales are to be converted and stored the question will naturally arise as to how many versions are stored and which ones. At one extreme there could be as many scales as there are users or applications. The storage situation could easily get out of hand. Adoption of a single coverage avoids this situation.

DISADVANTAGES OF A SINGLE COVERAGE APPROACH

There are difficulties in implementing a single coverage approach a number of which are identified in the following quotations.

> Of fundamental importance for generalization processing is the issue of whether the spatial data are digitized and stored once or several times for different scales. Theoretically, this represents a trade-off between processing and storage requirements. In practice a single data base solution will not be feasible for several reasons; time delay for real time applications, lack of generalization procedures, scale-related variability of objects to be stored. (Brassel 1985 p.22)
>
> Small scale retrieval of lines stored at a single large scale involves incurring very large overheads, both in the quantity of data accessed and in processing by the generalization algorithm. (Jones and Abraham 1986 p.388)
>
> There is no overall program of large scale digital base mapping and as yet no suitable base is readily available. To be suitable such a base would have to exist at a number of scales since it would be too elaborate and expensive for users to derive a specific scale by generalization from a largest common denominator scale. (Tomlinson 1986 p.10)

The main disadvantages raised in the above comments, concern the size of files, excessive processing times, and inadequate treatment of scale-related variability within a file. These objections may be valid given current thinking and the status of automated generalization procedures in current production systems. A solution exists, however, in a rethinking of data storage arrangements and improvement of scale changing and generalization algorithms.

A SOLUTION TO THE PROBLEM

One part of the solution would require changes in the institutional responsibility, creation, and format of a detailed database. The second part requires software development. A first task, however, is to define the term detailed coverage. This definition is then followed by discussion of data base changes and software improvements.

A Detailed Coverage

A coverage is a layer of one theme such as soils, hydrography, or land use. A coverage is based on a classification scheme and constructed by identifiying the spatial location of a class and delineating it from neighboring classes (Chrisman 1982). A detailed coverage implies detail in both the geometric and attribute components. Spatial detail (in a vector representation) will depend both on the number of points allowed to approximate boundary curves and the number of significant digits used to store the coordinates. For the attribute information greater detail is associated with a larger number of classes and finer discrimination between classes. Such a coverage should not be confused with a general purpose base map. The traditional base map attempts to anticipate the needs of several users by including all manner of information on a single map.

A detailed coverage cannot be tied to a specific scale, but will correspond generally to a large scale mapping. Detailed coverages may range from 1:200 to 1:50,000 scale depending on the nature of the resource or landscape and current knowledge of it. The level of detail or scale will depend basically on the size of the objects to be represented. A 1:50,000 scale map of bedrock in the Midwest might be considered detailed because of the size and homogeneity of the rock bodies. A detailed coverage of Rocky Mountain bedrock geology might require a larger mapping scale to depict the greater geological complexity. Similarly a detailed coverage of urban land use might mean a map scale of 1:500 while a detailed rural land use coverage might mean a 1:4,800 map scale.

Data base design changes

Distributed collection and maintenance of data. The first change is an institutional change in which responsibility for mapping would be delegated to responsible agencies at state, regional or local levels. Agencies at these levels are the prime users of detailed data and therefore should have a vested interest in its collection and maintenance. Detailed information is time consuming and costly to collect for large areas, but, if responsibilty for collection was dispersed the burden would be less for each contributor. The detailed data base could be built up incrementally by first focusing on selected areas such as urban, rapidly developing, or high environmental risk areas, then filling in the gaps as time and budgets allowed. State, regional, and local governments would be in a more logical position than the federal government to set such priorities. The data base could also be built from existing pieces such as detailed information collected for specific projects. As an example, detailed geologic information collected for the siting of a nuclear waste repository could be spliced into a more general coverage. Scale uniformity should not be a constraint if topological consistency is maintained. A larger scale or more detailed inset would simply mean a greater coordinate density for that area.

Substitution of detailed coverages for base maps A second change is to replace the general purpose base map with individual coverages of detailed information. A coverage of information, while detailed, should not create as large a storage and processing overhead as a general purpose base map with information on every conceivable object in the landscape. Maintenance of coverages also eliminates the problem of scale-related variation among different features since a coverage will include only one feature type. Hydrography treated independently should be simpler to generalize than a base map which includes roads, buildings, vegetation and contours in addition to hydrography.

Unfortunately the notion of a base map has become entrenched as a necessary foundation for all mapping activity (Bauer 1983). The base map provides a manual merger of several layers of information and has endured since the merger of digital layers has not been an efficient automated capability until recently. With the development of more sophisticated systems, and the availability of a geodetic reference framework and identifiable control points for each coverage, a base map may no longer be necessary (Chrisman and Niemann 1985).

Storage of data in practical analytical units There is no rule which requires a detailed coverage to exist as one large file. An alternative is to store data in manageable units based on obvious political or natural boundaries. The units could be counties, townships, watersheds, etc. depending on the information, its scale of variation, the level of detail currently available, or a legislative mandate. As the level of detail increases, a smaller areal unit could be used. If the size of the storage unit is well matched to the level of detail, processing times should not be unreasonable. Experience with the Dane County Land Records Project has shown that detailed (1:15,840 scale) soils and parcel (1:4,800 scale) data can be reasonably maintained as township coverages. These coverages are about 300K and 200K respectively. This storage approach is similar to the tiling scheme supported by the ARC/INFO map library (Aronson and Morehouse 1983). Such a structure has the potential to alleviate problems in handling regional scale-related variation within the same feature type. Dane County is bisected by a terminal moraine and soil geomorphologies across this boundary are quite different. Soil patterns within a township, however, are much more uniform, so generalization of this smaller unit is potentially simpler.

Resistence to storing a data base in separate units is due in part to past difficulties of merging them to create larger areal coverages. New software referred to as 'zipping' (Beard and Chrisman 1986) can overcome this difficulty. Given small areal coverage storage units as building blocks, the logical processing sequence is to generalize each unit to a desired level of detail. Then once the storage bulk is reduced the units can be "zipped" together quite rapidly to create smaller scale, larger areal coverages. To be workable, the combined generalization and "zipping" process for several units should be faster and less complex than generalization of one large detailed coverage.

Software development

The last part of the solution is to develop a flexible package of generalization tools. A change of scale is not available at the push of a button as some systems would promise. Reduction of detailed coverages to substantially coarser resolution requires more sophisticated processing than simple line generalization. Tomlinson and Boyle (1981) conducted a benchmark of nine geographic information systems in 1980 and reported that, "No sytem demonstrated a capacity to produce legible 1:250,000 scale maps from 1:12,500 scale source material". Automated scale changing capabilities have not improved substantially since then. An automated scale reduction process described by Leberl, Olson and Lichtner (1985) handles a reduction of 1:24,00 to 1:50,000 scale (about a factor of 2). Monmonier (1983) calls for operational algorithms that must handle scale reductions exceeding a factor of four. Reliance on a single detailed coverage could require scale reductions exceeding a factor of 25. Much of the current generalization effort is still focused on line generalization research (Zoraster *et al* 1984) which is becoming redundant and avoids the more difficult problems of scale reduction. To effectively handle large reductions a process should allow for variable reduction of line detail, feature simplification, and attribute reclassification and aggregation. The following section discusses automated generalization capabilities needed to carry out large scale reductions.

COMPONENTS OF AN AUTOMATED GENERALIZATION PACKAGE

Some people assume only three basic algorithms; simplification, smoothing and displacement are required, and that these should be designed to replicate the manual generalization process as closely as possible. Rhind (1973) was one of the first to recognize that automated generalization should not be a direct translation of manual techniques. Rhind (1973) identified the essential automated generalization functions as line sinuosity reduction, feature transposition, within and between category amalgamation, feature or category elimination, and graphic coding change. For 1973 this was an astute selection of functions which recognized the need to deal with attribute as well as spatial information. Graphic coding change is the only unnecessary function. Graphic representration is a consideration in the generalization process but it should remain a separate function. Brassel (1985) offered a model for automated generalization that consisted of objects and functions; the main objects being points, lines, and areas, with eight to twelve functions for each of these objects. His model also recognizes the need for more complex functions specifically related to digital representations. In an attempt to be exhaustive, however, his model is perhaps more complex than necessary. Five major functions would appear to be optimal and are described below.

SELECT: This routine allows a user to select features for elimination or to select or exclude a set of features for further processing. Selection criteria can be geometric data such as threshold lengths, widths, areas, distances, perimeters, etc. or attribute data such as names or geocodes In some situations a desired generalization could be accomplished by the selection process alone. The result of this selection process should be a new topologically consistent coverage.

AGGREGATE: This process condenses the attribute information by reducing the number of classes. The user specifies new classes and defines how they will be hierarchically reclassified from existing classes of the detailed coverage. As an example, residential, commercial and manufacturing land uses might be aggregated to a built-up or urban land use class. This routine eliminated lines and areas based on the reclassification. In the land use example, any lines separating residential and commercial land uses will disappear, and the combined area will appear as an urban land use area.

REDUCE: This is a routine to remove points from a line. The Douglas-Peucker routine is a logical choice for this function. It works on the principle that points of maximum deviation from a trend line connecting the end points of a line are retained while points less than a specified distance (tolerance) from the trend line are eliminated.. This routine has a well defined theoretical base (Peucker 1975, Marino, 1979, White, 1983), is computationally efficient and mimimizes positional displacement (McMasters 1986). Its utility has been proven by the test of time and the adoption by nearly everyone.

COLLAPSE: This routine explicitly invokes a dimension change. Areas or polygons specifically selected can be collapsed to lines or points, i.e. a river represented as an area is collapsed to a single line, or a city represented as an area is collapsed to a point. Nodes are relocated on the center line or centroid. Neighboring areas are extended to occupy areas formerly covered by the collapsed feature.

COARSEN: This routine will simplify and also collapse features. The degree of simplification will depend on a specified distance, epsilon. This idea is based on the epsilon filter (Perkal, 1965, Chrisman 1983). The routine removes or modifies features by analyzing clusters of points which fall within epsilon of each other. As an example, if points defining an island or penisula are within this distance of each other, they will collapse to points and then be eliminated.

These functions can be invoked in different combinations and order to create the desired result. The number of steps and processing sequence depends on the final objective, scale and graphic output resolution. A large reduction to scale might require all of the functions or several iterations of a function. A number of these functions already exist in some form in such GIS as ODYSSEY and ARC/INFO (which implies that these require topology). With slight modification these functions could be adapted for use in an automated scale reduction package. The collapse routine does not yet exist but should not be extremely difficult to implement. The coarsen routine exists currently in ODYSSEY but needs refinement.

APPLICATION OF THE PROCEDURE

As a test of the single detailed database concept, this generalization procedure was applied to a detailed coverage to produce smaller scale versions. While not all of the generalization procedure is fully operational, the basic concepts were demonstrated on a simple data set. The data set was 1:10,000 to 1:40,000 scale digital coastline data provided by the National Ocean Service (NOS). The NOS data exemplifies the detailed coverage concept. Their database was constructed by digitizing only the largest scale, most detailed version of the coastline. Where 1:10,000 or 1:15,000 scale versions of the coastline were available these were substituted for the 1:40,000 scale coastline.

The success of the procedure was to be tested by comparison of the generalized results against an independently generalized product. The test data was provided by NOS and consisted of the detailed coastline data plotted, manually generalized to 1:250,000 scale and redigitized. To match the test data set, the detailed coastline data had to be reduced by a factor of 25, (a reasonably large scale reduction).

The NOS data was iteratively reduced and coarsened by using progressively larger tolerances. Table 1 summarizes this process. The final entry in this table is the manually generalized test data set.

	Number of polygons	Number of points
NOS detailed source data	65	2703
20 meter reduction	65	1362
20 meter coarsening	51	1291
50 meter coarsening	37	1013
50 meter reduction	37	597
80 meter coarsening	24	517
70 meter reduction	24	458
NOS manually generalized data	26	466

Table 1.

Maps 1 and 2 show the manually and automatically generalized versions respectively. As the number of points, polygons and graphic representations show, the automatically generalized version is a reasonable facsimile of the manually produced version. One of the main differences between the versions is the treatment of small islands. In the automated process these are eliminated, while in the manual generalization they are exaggerated. One advantage of the automated process is a reduction in error. Map 3 is an overlay of the manually generalized and original versions which shows the positional and attribute differences between the two. Map 4 is a similar overlay which shows the differences between the automatically generalized and original versions. Table 2 and a comparison of Maps 3 and 4 show that the both positional and attribute error in the automatically generalized result are reduced.

NOS Detailed Data Against NOS Generalized Data

		Hand Generalized Coastline	
		Land	Water
Detailed Coastline	Land	3928.18	313.25
	Water	101.73	4746.26

NOS Detailed Data Against Epsilon Generalized Data

		Automated Generalized Coastline	
		Land	Water
Detailed Coastline	Land	4144.93	90.76
	Water	109.25	4731.93

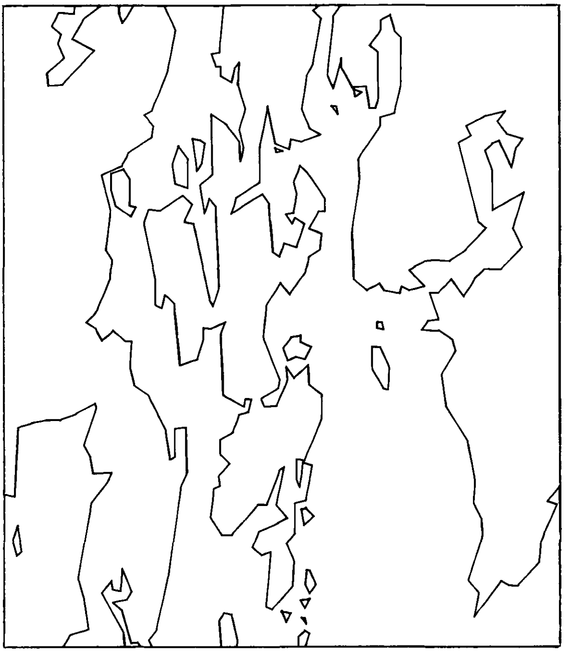
Table 2.

Another aspect of an automatically generalized product which cartographers have been striving for is a more objective result. The handling of features in this case was entirely objective. The only subjective input is a tolerance for the reduction and an epsilon distance for the coarsening.

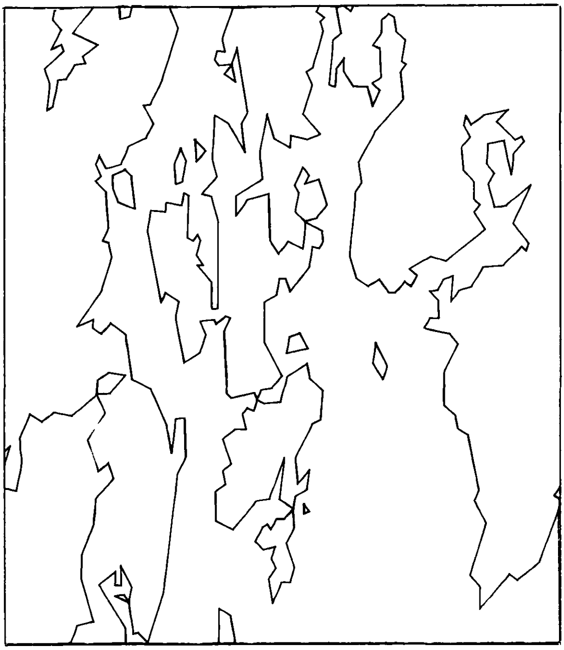
This test data involved the simple case of just two attribute classes; land and water, but the results are encouraging. The results show very reasonable reduction of the geometric data and handling of the range of different scales in the source material. With additional development and refinement of the functions, the process should be able to handle coverages with more complex attribute data.

ACKNOWLEDGEMENTS

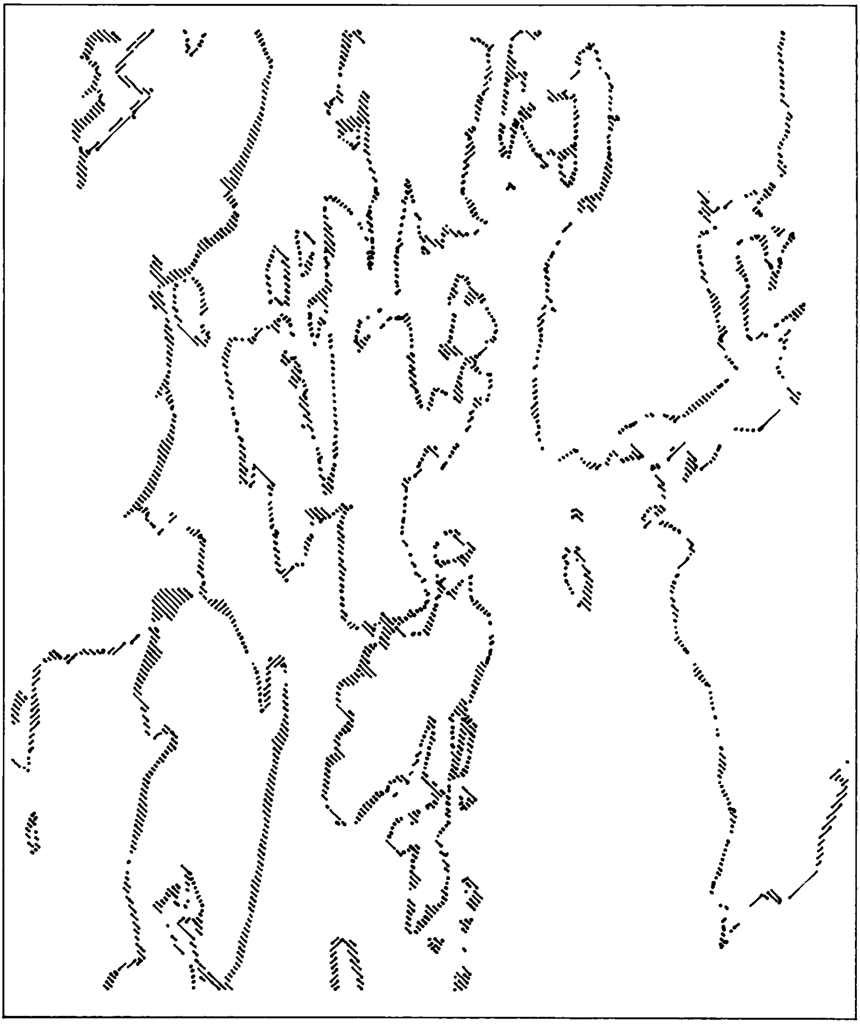
This research was funded in part by the Nautical Charting Branch of the National Ocean Service. The author thanks NOS for providing the digital coastline data.



Map 1



Map 2

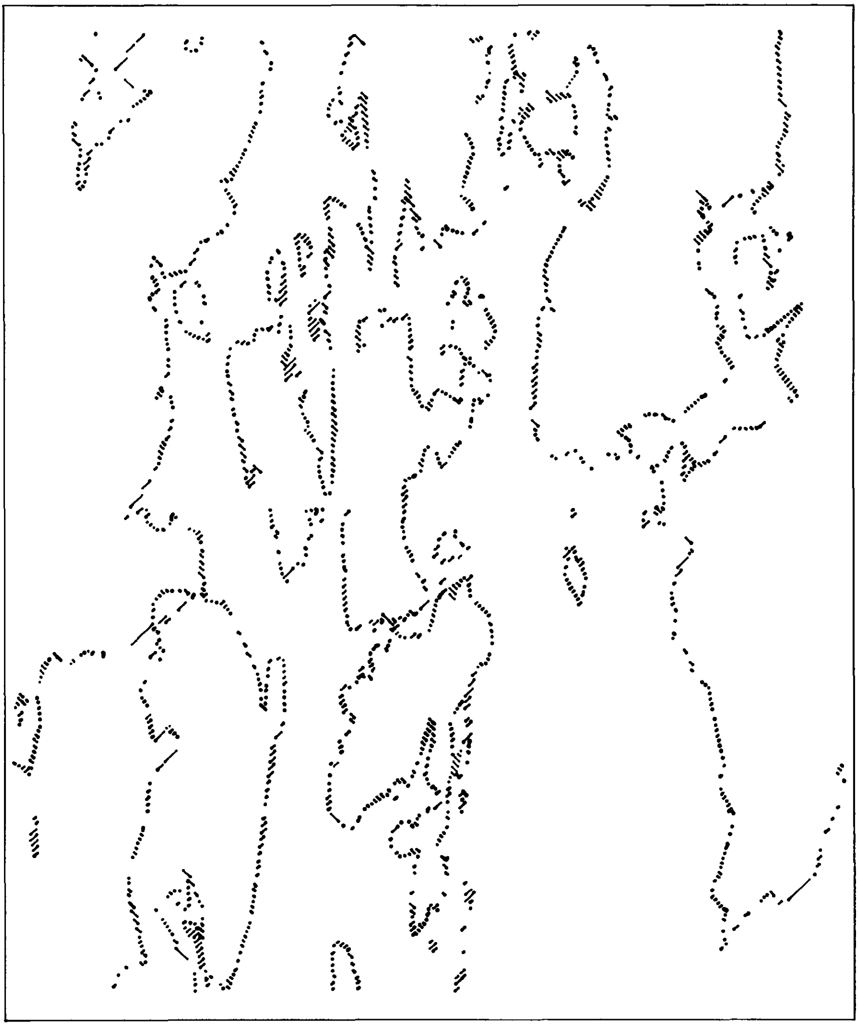


Map 3

Differences Between Large and
Small Scale Charts
Generalized Coast Manually
Derived from the Detailed Data

▨ Land in Large Scale, Water in Small

▧ Water in Large Scale, Land in Small



Map 4

Differences Between Large and Small Scale Charts
Coast Automatically Generalized from the Detailed Data

▨ Land in Large Scale, Water in Small

▧ Water in Large Scale, Land in Small

REFERENCES

Aronson P. and Morehouse, S., 1983. "The ARC/INFO Map Library: A Design for a Digital Geographic Database", **Proceedings AUTO-CARTO 6**, 1, 372-382.

Bauer, K., 1983. "Public Planning: the Role of Maps and the Geodetic Base", IES Report 123. University of Wisconsin-Madison. 130-140.

Beard, M. K., and Chrisman, N. R., 1986. "Zipping: New Software for Merging Map Sheets", **Proceeding of American Congress on Surveying and Mapping 46th Annual Meeting,** 1, 153-161.

Brassel, Kurt, 1985. "Strategies and Data Models for Computer-Aided Generalization", **International Yearbook of Cartography**, 25, 11-30.

Chrisman, N.R., 1982. Methods of Spatial Analysis Based on Error in Categorical Maps. unpublished Ph.D. thesis, University of Bristol.

Chrisman, N. R. 1983, "Epsilon Filtering: a Concept for Automated Scale Changing", **Proceedings of American Congress on Surveying and Mapping, 43rd Annual Meeting**, 322-331.

Chrisman, N.R. and Niemann, B. J., 1985, "Alternate Routes to a Multipurpose Cadastre: Merging Institutional and Technical Reasoning", **Proceedings AUTO-CARTO 7**, 84-94.

Jones, C.B. and Abraham, I. M., 1986, "Design Considerations for a Scale-Independent Cartographic Database", **Proceedings: 2nd International Symposium on Spatial Data Handling**, Seattle, WA 384-398.

Leberl, F., Olson, D., and Lichtner, W., 1985. "ASTRA: A System for Automated Scale Transition", **Technical Papers 51st Annual Meeting ASP**, 1, 1-13.

Marino, J., 1979. "Identification of Characteristic Points Along Naturally Occuring Lines: an empirical study", **The Canadian Cartographer**, 16, 1, 70-80.

McMasters, R.B., 1986, "A Statistical Analysis of Mathematical Measures for Linear Simplification", **The American Cartographer**, 13, 2, 103-116.

Perkal, J., 1966. "An Attempt at Objective Generalization', **Discussion Paper 10**, Ann Arbor, Michigan, Inter-University Community of Mathematical Geographers.

Peucker, T. K., 1975. "A Theory of the Cartographic Line", **Proceedings AUTO-CARTO 2**, 508-518.

Rhind, D.W., 1973, "Generalization and Realism Within Automated Cartography", **The Canadian Cartographer**, 10, 51-62.

Tomlinson, R., and Boyle, R., 1981. "The State of Development of Systems for Handling Natural Resource Data", **Cartographica**, 18, 4, 65-95.

Tomlinson Associates. 1986, Review of North American Experience of Current and Potential Uses of Geographic Information Systems, Tomlinson Associates, Ontario, Canada.

Watson, W., 1970. "A Study of the Generalization of a Small Scale Map Series", **International Yearbook of Cartography**, 16, 24-32.

White, E.R., 1983. "Assessment of Line Generalization Algorithms Using Characteristic Points", **The American Cartographer**, 12, 1, 17-27.

Zoraster, S., Davis, D., Hugus, M., 1984. **Manual and Automated Line Generalization and Feature Displacement** ETL-0359. Fort Belvoir, VA.