CONCEPTUAL BASIS FOR GEOGRAPHIC LINE GENERALIZATION

David M. Mark
National Center for Geographic Information and Analysis
Department of Geography, SUNY at Buffalo
Buffalo NY 14260

BIOGRAPHICAL SKETCH

David M. Mark is a Professor in the Department of Geography, SUNY at Buffalo, where he has taught and conducted research since 1981. He holds a Ph.D. in Geography from Simon Fraser University (1977). Mark is immediate past Chair of the GIS Specialty group of the Association of American Geographers, and is on the editorial boards of The American Cartographer and Geographical Analysis. He also is a member of the NCGIA Scientific Policy Committee. Mark's current research interests include geographic information systems, analytical cartography, cognitive science, navigation and way-finding, artificial intelligence, and expert systems.

ABSTRACT

Line generalization is an important part of any automated map-making effort. Generalization is sometimes performed to reduce data volume while preserving positional accuracy. However, geographic generalization aims to preserve the recognizability of geographic features of the real world, and their interrelations. This essay discusses geographic generalization at a conceptual level.

INTRODUCTION

The digital cartographic line-processing techniques which commonly go under the term "line generalization" have developed primarily to achieve two practical and distinct purposes: to reduce data volume by eliminating or reducing data redundancy, and to modify geometry so that lines obtained from maps of one scale can be plotted clearly at smaller scales. Brassel and Weibel (in press) have termed these *statistical* and *cartographic* generalization, respectively. Research has been very successful in providing algorithms to achieve the former, and in evaluating them (cf. McMaster, 1986, 1987a, 1987b); however, very little has been achieved in the latter area.

In this essay, it is claimed that Brassel and Weibel's *carto*graphic generalization should be renamed *graphical* generalization, and should further be subdivided: *visual* generalization would refer to generalization procedures based on principles of computational

vision, and its principles would apply equally to generalizing a machine part, a cartoon character, a pollen grain outline, or a shoreline. On the other hand, *geographical* generalization would take into account knowledge of the geometric structure of the *geographic feature* or feature-class being generalized, and would be the geographical instance of what might be called *phenomenon-based* generalization. (If visual and geographic generalization do not need to be separated, then a mechanical draftsperson, a biological illustrator, and a cartographer all should be able to produce equally good reduced-scale drawings of a shoreline, a complicated machine part, or a flower, irrespectively; such an experiment should be conducted!)

This essay assumes the following:

> **geographical generalization must incorporate information about the geometric structure of geographic phenomena.**

It attempts to provide hints and directions for beginning to develop methods for automated geographical generalization by presenting an overview of some geographic phenomena which are commonly represented by lines on maps. The essay focuses on similarities and differences among geographic features and their underlying phenomena, and on geometric properties which must be taken into account in geographical generalization.

OBJECTIVES OF "LINE GENERALIZATION"

Recently, considerable attention has been paid to theoretical and conceptual principles for cartographic generalization, and for the entire process of map design. This is in part due to the recognition that such principles are a prerequisite to fully-automated systems for map design and map-making. Mark and Buttenfield (1988) discussed over-all design criteria for a cartographic expert system. They divided the map design process into three inter-related components: *generalization*, *symbolization*, and *production*. Generalization was characterized as a process which first *models* geographic phenomena, and then generalizes those models. Generalization was in turn subdivided into: *simplification* (including reduction, selection, and repositioning); *classification* (encompassing aggregation, partitioning, and overlay); and *enhancement* (including smoothing, interpolation, and reconstruction). (For definitions and further discussions, see Mark and Buttenfield, 1988.) Although Mark and Buttenfield's discussion of the modeling phase emphasized a phenomenon-based approach, they did not exclude statistical or other phenomenon-independent approaches. Weibel and Buttenfield

(1988) extended this discussion, providing much detail, and emphasizing the requirements for mapping in a geographic information systems (GIS) environment.

McMaster and Shea (1988) focussed on the generalization process. They organized the top level of their discussion around three questions: Why do we generalize? When do we generalize? How do we generalized? These can be stated more formally as intrinsic objectives, situation assessment, and spatial and attribute transformations, respectively (McMaster and Shea, 1988, p. 241). The rest of their paper concentrated on the first question; this essay will review such issues briefly, but is more concerned with their third objective.

Reduction of Data Volume

Many digital cartographic line-processing procedures have been developed to reduce data volumes. This process has at times been rather aptly termed "line reduction". In many cases, the goal is to eliminate redundant data while changing the geometry of the line as little as possible; this objective is termed "maintaining spatial accuracy" by McMaster and Shea (1988, p. 243). Redundant data commonly occur in cartographic line processing when digital lines are acquired from maps using "stream-mode" digitizing (points sampled at pseudo-constant intervals in x, y, distance, or time); similarly, the initial output from vectorization procedures applied to scan-digitized maps often is even more highly redundant.

One stringent test of a line reduction procedure might be: "can a computer-drafted version of the lines after processing be distinguished visually from the line before processing, or from the line on the original source document?" If the answer to both of these questions is "no", and yet the number of points in the line has been reduced, then the procedure has been successful. A quantitative measure of performance would be to determine the perpendicular distance to the reduced line from each point on the original digital line; for a particular number of points in the reduced line, the lower the root-mean-squared value of these distances, the better is the reduction. Since the performance of the algorithm can often be stated in terms of minimizing some statistical measure of "error", line reduction may be considered to be a kind of "statistical generalization", a term introduced by Brassel and Weibel (in press) to described minimum-change simplifications of digital elevation surfaces.

Preservation of Visual Appearance and Recognizability

As noted above, Brassel and Weibel (in press) distinguish statistical and cartographic generalization. "Cartographic generalization is used only for graphic display and therefore has to aim at visual effectiveness" (Brassel and Weibel, in press). A process with such an aim can only be evaluated through perceptual testing involving subjects representative of intended map users; few such studies have been conducted, and none (to my knowledge) using generalization procedures designed to preserve visual character rather than merely to simplify geometric form.

Preservation of Geographic Features and Relations

Pannekoek (1962) discussed cartographic generalization as an exercise in applied geography. He repeatedly emphasized that individual cartographic features should not be generalized in isolation or in the abstract. Rather, relations among the geographic features they represent must be established, and then should be preserved during scale reduction. A classic example, presented by Pannekoek, is the case of two roads and a railway running along the floor of a narrow mountain valley. At scales smaller than some threshold, the six lines (lowest contours on each wall of the valley; the two roads; the railway; and the river) cannot all be shown in their true positions without overlapping. If the theme of the maps requires all to be shown, then the other lines should be moved away from the river, in order to provide a distinct graphic image while preserving relative spatial relations (for example, the railway is between a particular road and the river). Pannekoek stressed the importance of showing the transportation lines as being on the valley floor. Thus the contours too must be moved, and higher contours as well; the valley floor must be widened to accommodate other map features (an element of cartographic license disturbing to this budding geomorphometer when J. Ross Mackay assigned the article in a graduate course in 1972!). Nickerson and Freeman (1986) discussed a program that included an element of such an adjustment.

A twisting mountain highway provides another kind of example. Recently, when driving north from San Francisco on highway 1, I was startled by the extreme sinuosity of the highway; maps my two major publishing houses gave little hint, showing the road as almost straight as it ran from just north of the Golden Gate bridge westward to the coast. The twists and turns of the road were too small to show at the map scale, and I have little doubt that positional accuracy was maximized by drawing a fairly straight line following the road's "meander axis". The solution used on some Swiss road maps seems better; winding mountain highways are represented by sinuous lines on the map. Again, I have no doubt that, on a 1:600,000 scale map,

the twists and turns in the cartographic line were of a far higher amplitude that the actual bends, and that the winding road symbols had fairly large positional errors. However, the character of the road is clearly communicated to a driver planning a route through the area. In effect, the road is categorized as a "winding mountain highway", and then represented by a "winding mountain highway symbol", namely a highway symbol drafted with a high sinuosity. Positional accuracy probably was sacrificed in order to communicate geographic character.

A necessary prerequisite to geographic line generalization is the identification of the kind of line, or more correctly, the kind of phenomenon that the line represents (see Buttenfield, 1987). Once this is done, the line may in some cases be subdivided into component elements. Individual elements may be generalized, or replaced by prototypical exemplars of their kinds, or whole assemblages of sub-parts may be replaced by examples of their superordinate class. Thus is a rich area for future research.

GEOGRAPHICAL LINE GENERALIZATION

Geographic phenomena which are represented by lines on topographic and road maps are discussed in this section. (Lines on thematic maps, especially "categorical" or "area-class" boundaries, will almost certainly prove more difficult to model than the more concrete features represented by lines on topographic maps, and are not included in the current discussion.) One important principle is:

many geographic phenomena inherit components of their geometry from features of other kinds.

This seems to have been discussed little if at all in the cartographic literature. Because of these tendencies toward inheritance of geometric structure, the sequence of sub-sections here is not arbitrary, but places the more independent (fundamental) phenomena first, and more derived ones later.

Topographic surfaces (contours)

Principles for describing and explaining the form of the earth's surface are addressed in the science of geomorphology. Geomorphologists have identified a variety of terrain types, based on independent variables such as rock structure, climate, geomorphic process, tectonic effects, and stage of development. Although selected properties of topographic surfaces may be mimicked by statistical surfaces such as fractional Brownian models, a kind of fractal (see Goodchild and Mark 1987 for a review), detailed models

of the geometric character of such surfaces will require the application of knowledge of geomorphology. Brassel and Weibel (in press) clearly make the case that contour lines should never be generalized individually, since they are parts of surfaces; rather, digital elevation models must be constructed, generalized, and then re-contoured to achieve satisfactory results, either statistically or cartographically.

Streams

Geomorphologists divide streams into a number of categories. Channel patterns are either straight, meandering, or braided; there are sub-categories for each of these. Generally, streams run orthogonal to the contours, and on an idealized, smooth, single-valued surface, the stream lines and contours for duals of each other. The statistics of stream planform geometry have received much attention in the earth science literature, especially in the case of meandering channels (see O'Neill, 1987). Again, phenomenon-based knowledge should be used in line generalization procedures; in steep terrain, stream/valley generalization is an intimate part of topographic generalization (cf. Brassel and Weibel, in press).

Shorelines

In a geomorphological sense, shorelines might be considered to "originate" as contours, either submarine or terrestrial. A clear example is a reservoir: the shoreline for a fixed water level is just the contour equivalent to that water level. Any statistical difference between the shoreline of a new reservoir, as drawn on a map, and a nearby contour line on that same map is almost certainly due to different construction methods or to different cartographic generalization procedures used for shorelines and contours. Goodchild's (1982) analysis of lake shores and contours on Random Island, Newfoundland, suggests that, cartographically, shorelines tend to be presented in more detail (that is, are relatively less generalized), while contours on the same maps are smoothed to a greater degree. As sea level changes occur over geologic time, due to either oceanographic or tectonic effects, either there is a relative sea-level rise, in which case a terrestrial contour becomes the new shoreline, or a relative sea-level fall, to expose a submarine contour as the shoreline.

Immediately upon the establishment of a water level, coastal geomorphic processes begin to act on the resulting shoreline; the speed of erosion depends on the shore materials, and on the wave, wind, and tidal environment. It is clear that coastal geomorphic processes are scale-dependent, and that the temporal and spatial scales of such processes are functionally linked. Wave refraction

tends to concentrate wave energy at headlands (convexities of the land), whereas energy per unit length of shoreline is below average in bays. Thus, net erosion tends to take place at headlands, whereas net deposition occurs in the bays. On an irregular shoreline, beaches and mudflats (areas of deposition) are found largely in the bays. The net effect of all this is that shorelines tend to straighten out over time. The effect will be evident most quickly at short spatial scales.

Geomorphologists have divided shorelines into a number of types or classes. Each of these types has a particular history and stage, and is composed of members from a discrete set of coastal landforms. Beaches, rocky headlands, and spits are important components. Most headlands which are erosional remnants are rugged, have rough or irregular shorelines, and otherwise have arbitrary shapes determined by initial forms, rock types and structures, wave directions, *et cetera*. Spits and beaches, however, have forms with a much more controlled (less variable) geometry. For example, the late Robert Packer of the University of Western Ontario found that many spits are closely approximated by logarithmic spirals (Packer, 1980).

Political and Land Survey Boundaries

Most political boundaries follow either physical features or lines of latitude or longitude. Both drainage divides (for example, the France-Spain border in the Pyrenees, or southern part of the British Columbia-Alberta in the Rocky Mountains) and streams (there are a great many many examples) are commonly used as boundaries. The fact that many rivers are dynamic in their planform geometry leads to interesting legal and/or cartographic problems. For example, the boundary between Mississippi and Louisiana is the midline of the Mississippi River when the border was legally established more than a century ago, and does not correspond with the current position of the river.

In areas which were surveyed before they were settled by Europeans, rectangular land survey is common. Then, survey boundaries may also be used as boundaries for minor or major political units. Arbitrary lines of latitude or longitude also often became boundaries as a result of negotiations between distant colonial powers, or between those powers and newly-independent former colonies. An example is the Canada - United States boundary in the west, which approximates the 49th parallel of latitude from Lake-of-the-Woods to the Pacific. Many state boundaries in the western United States are the result of the subdivision of larger territories by officials in Washington. Land survey boundaries are rather "organic" and irregular in the metes-and-bounds systems of most of the original 13 colonies of the United States, and in many other parts of the world. They are often much more rectangular in

the western United States, western Canada, Australia, and other "pre-surveyed" regions.

Roads

Most roads are constructed according to highway engineering codes, which limit the tightness of curves for roads of certain classes and speeds. These engineering requirements place smoothness constraints on the short-scale geometry of the roads; these constraints are especially evident on freeways and other high-speed roads, and should be determinable from the road type, which is included in the USGS DLG feature codes and other digital cartographic data schemes. However, the longer-scale geometry of these same roads is governed by quite different factors, and often is inherited from other geographic features.

Some roads are "organic", simply wandering across country, or perhaps following older walking, cattle, or game trails. However, many roads follow other types of geographic lines. Some roads "follow" rivers, and others "follow" shorelines. In the United States, Canada, Australia, and perhaps other countries which were surveyed before European settlement, many roads follow the survey lines; in the western United States and Canada, this amounts to a 1 by 1 mile grid (1.6 by 1.6 km) of section boundaries, some or all of which may have actual roads along them. Later, a high-speed, limited access highway may minimize land acquisition costs by following the older, survey-based roadways where practical, with transition segments where needed to provide sufficient smoothness (for example, Highway 401 in south-western Ontario).

A mountain highway also is an example of a road which often follows a geographic line, most of the time. In attempting to climb as quickly as possible, subject to a gradient constraint, the road crosses contours at a slight angle which can be calculated from the ratio of the road slope to the hill slope. [The sine of the angle of intersection (on the map) between the contour and the road is equal to the ratio of the road slope to the hill slope, where both slopes are expressed as tangents (gradients or percentages).] Whenever the steepness of the hill slope is much greater than the maximum allowable road gradient, most parts of the trace of the road will have a very similar longer-scale geometry to a contour line on that slope. Of course, on many mountain highways, such sections are connected by short, tightly-curved connectors of about 180 degrees of arc, when there is a "switch-back", and the hillside switches from the left to the right side of the road (or the opposite).

Railways

Railways have an even more constrained geometry than roads, since tight bends are never constructed, and gradients must be very low. Such smoothness should be preserved during generalization, even if curves must be exaggerated in order to achieve this.

Summary

The purpose of this essay has not been to criticize past and current research on computerized cartographic line generalization. Nor has it been an attempt to define how research in this area should be conducted in the future. Rather, it has been an attempt to move one (small) step toward a truly "geographic" approach to line generalization for mapping. It is a bold assertion on my part to state that, in order to successfully generalize a cartographic line, one must take into account the geometric nature of the real-world phenomenon which that cartographic line represents, but nevertheless I assert just that. My main purpose here is to foster research to achieve that end, or to debate on the validity or utility of my assertions.

Acknowledgements

I wish to thank Bob McMaster for the discussions in Sydney that convinced me that it was time for me to write this essay, Babs Buttenfield for many discussions of this material over recent years, and Rob Weibel and Mark Monmonier for their comments on earlier drafts of the material presented here; the fact that each of them would dispute parts of this essay does not diminish my gratitude to them. The essay was written partly as a contribution to Research Initiative #3 of the National Center for Geographic Information and Analysis, supported by a grant from the National Science Foundation (SES-88-10917); support by NSF is gratefully acknowledged. Parts of the essay were written while Mark was a Visiting Scientist with the CSIRO Centre for Spatial Information Systems, Canberra, Australia.

References

Brassel, K. E., and Weibel, R., in press. A review and framework of automated map generalization. *International Journal of Geographical Information Systems*, forthcoming.

Buttenfield, B. P., 1987. Automating the identification of cartographic lines. *The American Cartographer* 14: 7-20.

Goodchild, M. F., 1982. The fractional Brownian process as a terrain simulation model. *Modeling and Simulation* 13: 1122-1137. Proceedings, 13th Annual Pittsburgh Conference on Modeling and Simulation.

Goodchild, M. F., and Mark, D. M., 1987. The fractal nature of geographic phenomena. *Annals of the Association of American Geographers* 77: 265-278.

Mandelbrot, B. B., 1967. How long is the coast of Britain? Statistical self-similarity and fractional dimension. *Science* 156: 636-638.

Mark, D. M., and Buttenfield, B. P., 1988. Design criteria for a cartographic expert system. Proceedings, 8th International Workshop on Expert Systems and Their Applications, vol. 2, pp. 413-425.

McMaster, R. B., 1986. A statistical analysis of mathematical measures for linear simplification. *The American Cartographer* 13: 103-116.

McMaster, R. B., 1987a. Automated line generalization. *Cartographica* 24: 74-111.

McMaster, R. B., 1987b. The geometric properties of numerical generalization. *Geographical Analysis* 19: 330-346.

McMaster, R. B., and Shea, K. S., 1988. Cartographic generalization in digital a environment: A framework for implementation in a geographic information system. *Proceedings, GIS/LIS '88*, vol. 1, pp. 240-249.

O'Neill, M. P., 1987. Meandering channel patterns-- analysis and interpretation. Unpublished PhD dissertation, State University of New York at Buffalo.

Packer, R. W., 1980. The logarithmic spiral and the shape of drumlins. Paper presented at the Joint Meeting of the Canadian Association of Geographers, Ontario Division, and the East Lakes Division of the Association of American Geographers, London, Ontario, November 1980.

Pannekoek, A. J., 1962. Generalization of coastlines and contours. *International Yearbook of Cartography* 2: 55-74.

Weibel, R., and Buttenfield, B. P., 1988. Map design for geographic information systems. *Proceedings, GIS/LIS '88*, vol. 1, pp. 350-359.