G.F. Jenks, 1989
"Geographic logic in line generalization"
Cartographica, 26(1):27-42

GEOGRAPHIC LOGIC IN LINE GENERALIZATION

GEORGE F JENKS
Department of Geography, The University of Kansas
Lawrence, Kansas, United States

Maps are representations of the cultural and/or natural environmental characteristics of part or all of the surface of the earth. As such, the mapmaker must interpret the environment and apply geographic logic as he/she compiles a map. All maps are reduced scale representations and over time cartographers have developed a graphic vocabulary of conventional symbols logically related to the earth features that they represent. An artist may also interpret the environment but he/she is not restricted in the feature location, content and symbology used in the painting. The work dichotomy between the artist and the mapmaker means that the latter is constrained in creating his/her graphic and that spatial imperatives and geographic logic must be a guide in the compilation process.

Less than one hundred years ago cartographers created maps using hand methods for calculating projections, drafting lines, and the drafting of the lettering displayed on the body of a map. During this time period geographers were often cartographers who prepared illustrations for oral presentations and for publication. An abrupt change in the field of cartography began in the late 1930s and was accelerated by the desperate need for maps during World War II and by the dynamic expansion of mapping technology. There seems to be an inexplicable mystique which drives many to create maps and this phenomenon has been enhanced by the advent of the microcomputer. Manual skill is no longer a necessary prerequisite for the mapmaker who now can readily obtain the use of a computer as well as the software and data bases needed to create a map. It is now possible for very young children to follow a program menu and in a short time produce a map without understanding what is being displayed. This fact is of growing concern in the cartographic and geographic communities because many of those who use these new machines are creating maps that contain very serious logical errors (Jenks 1986).

Today, maps are created by persons in all walks of life utilizing a variety of data sources, materials and equipment. These mapmakers may work in the traditional or computer mode but in either case it is a rare occurrence when a map is created without linear symbols. These lines projected on the surface of the earth are creations of economic, political or social activities of man or the natural forces of the environment. Man-made linear features tend to be constructed in smoothly flowing curves and straight line segments, while naturally occurring linear features tend to be much more complex with many high intensity changes in direction. It is because of the highly complex nature of rivers, coastlines and drainage divides that this paper is focused on the problems of generalizing naturally occurring lines. The first part of the paper is composed of a number of brief statements used to develop geographically logical procedures for creating maps on computers. These sections include data acquisition, simplification and generalization, evaluation, comparisons, and how characteristic points can be used to enhance a linear display. The second half of the paper presents some of the author's recent research in attempting to understand how readers relate to lines displayed on topographic, atlas, textbook and media maps.

PROCEDURES

Linear data capture

The Fall River of Utah and Colorado used by Marino (1978) was selected as a typical naturally occurring line symbolized on maps. It is a 'young' fast moving stream with numerous meanders and rapids illustrative of the high intensity directional changes characteristic of mountain hydrography. Although much of the naturally occurring linear data acquired by large cartographic laboratories is captured with scanning technology a substantial amount, as in this case, is obtained using manually operated digitizers. The digitizer used for obtaining the original Fall River file, for instance, has a resolution of .005 inches which recorded 1096 coordinate pairs as shown in Figure 1, line A. As is often the case, manual digitization creates files with a number of errors due to operator inattention, muscular spasms and visual diffraction. Such errors, duplicate coordinates, spikes, switchbacks and knots illustrated by McMaster (1987) should be removed by a 'cleaning' process (Jenks 1981). Cleaning reduces the coordinate count, saving file space and plotting time, but more importantly it reduces plotter head chatter resulting in a truer linear representation. The cleaned file for line b of Figure 1 contains 875 coordinate pairs which is a reduction from line A of some twenty percent.

Some cartographers apply a smoothing operator to the cleaned data file to enhance the curvilinearity or 'flow' of a plot. The smoothing algorithms may be running averages or other mathematical procedures such as splines. The line labeled C in Figure 1 contains 875 coordinate pairs and has been created by such a gentle running average that one could describe it has having been modified in the imperceptible domain. Later, a discussion of more rigorous smoothing operators is presented, as related to the use of characteristic points in linear generalization. Many readers will be unable to perceive the differences between lines A,B, and C in Figure 1, but the latter two could logically be used to display Fall River or a simplification of it on a map.

Simplification

Previous experience has shown that mapreaders are unable to perceive the intricate detail that can be acquired by modern digitizers and drafted by modern plotters. As a result cartographers often resort to file reduction by a process

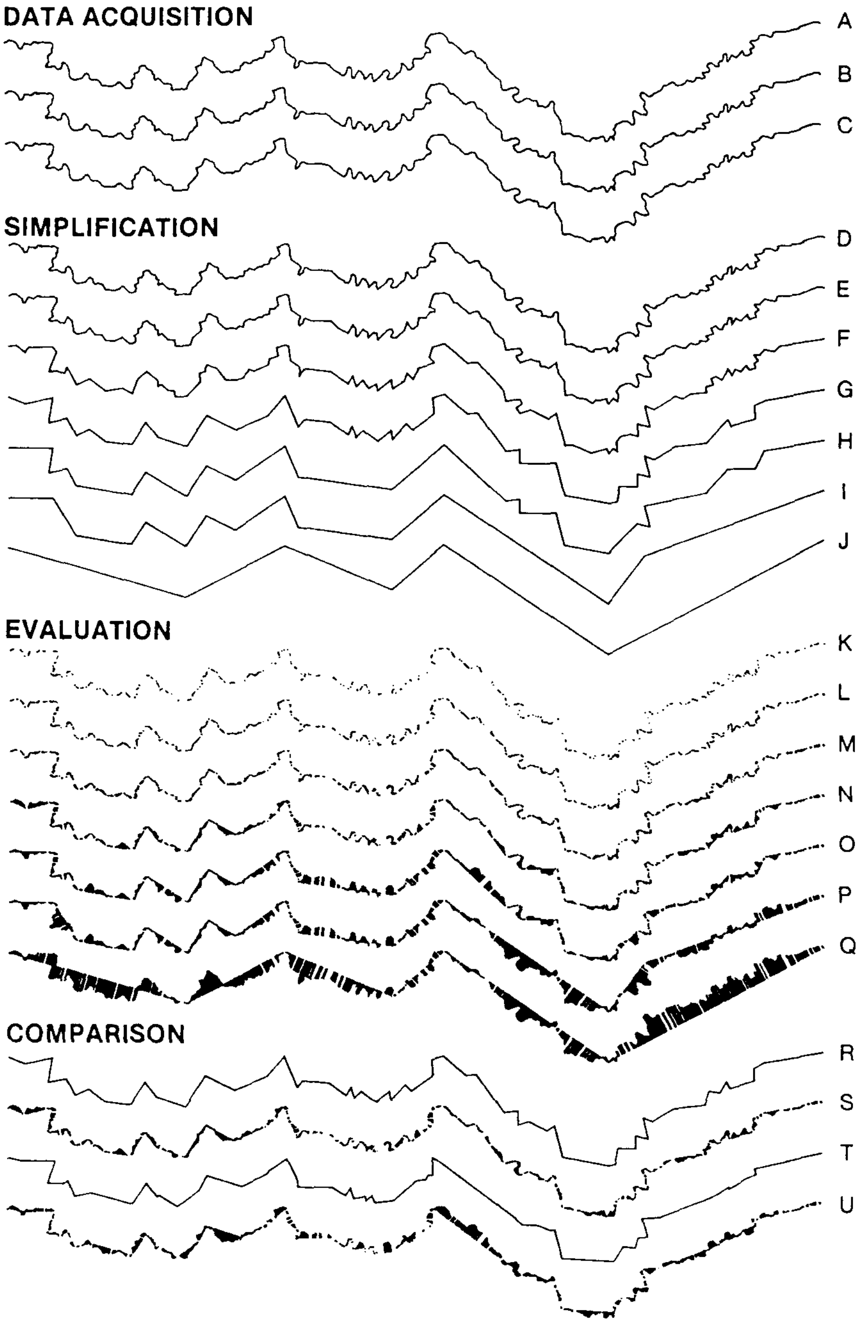


FIGURE 1 These lines are all different versions of Fall River of Utah and Colorado. Line A is a plot of the original digital file which was cleaned to create line B. Line C is a smoothed version of line B created using a gentle moving average operator. Few people can distinguish differences between lines B and C and thus either could be used on a map.

Lines D through J are DOUGLAS simplifications of line B. Lines K through Q are vector error plots of lines D through J. Line R is a DOUGLAS simplification of line B and line S contains the vector error plots of line R. Line T is a USGSB simplification of line B and line U contains the vector errors for line T.

commonly called 'thinning' or more specifically simplification. The process of simplification is applied to a file to reduce the number of coordinates considered to be superfluous. These may be coordinates crowded together, coordinates along nearly straight line segments or any other coordinates that are not considered necessary to maintain the character or flow of the line. A multitude of simplification algorithms have been developed, and new techniques are constantly being designed and tested (McMaster 1987).

For the purpose of this paper the simplification algorithm developed by David Douglas (Douglas and Peucker 1973) was selected. Many cartographers consider it to be the most accurate simplification algorithm available, while others think that it is too slow and costly in terms of computer processing time. The seven lines labeled D to J in Figure 1 were created using this algorithm. These simplified lines contain reduced coordinates sets which follow a downward geometric progression sometimes used in psychophysical research. The coordinate counts are: D = 439, E = 220, F = 112, G = 54, H = 28, I = 15 and J = 7.

Evaluation

Whenever lines are simplified they can be evaluated perceptually, but this is often an unacceptable method when two simplifications either appear to be identical or when the simplification algorithms have performed erratically. McMaster (1983a) investigated some thirty difference measures of cartographic lines and through various types of visual and statistical analysis reduced this number to seven. His measures are particularly appropriate for a study in which computer manipulation of linear files is necessary. Two families of measures are included: the first are attribute measures of an individual line, while the second involves displacement measures between the original line and a simplification of it. The attribute measures include the number of coordinates in the linear file, the average number of coordinates per inch of line and the standard deviation of the number of coordinates per inch of line, all of which provide information of particular concern to the computer cartographer. These are especially useful in evaluating the quality of the data in regard to the digitizing process. Two additional measures that add to our understanding include the angular change between contiguous vectors and the curvilinearity of the line. Anstey (1965), a geologist, calls the first of these a 'wiggliness' index which in reality is a simple angularity measure. A more complex measure counts the number of curvilinear segments along the line as a measure of the larger angular features or the sinuosity of the line. This latter measure was developed to separate meanders on a river from the smaller turns. A partial listing of the characteristics of the individual lines A through J is presented in Table 1 including the sum of angular changes, the average angular change, and the angular change per inch of line.

Alternatively, the comparative or displacement measures are an attempt to quantify the actual and perceptual movement of a line which takes place in the simplification process. One of these measures calculates the areal displacement between the original line and a derived simplification. The other calculates the vector distance between the eliminated coordinates on the original line and the straight vectors connecting coordinates on the simplification. A graphic presentation of the vector errors for lines D through J is presented in lines K

TABLE 1 CHARACTERISTICS OF THE VARIOUS VERSIONS OF THE FALL RIVER LINE

Line Ident.	Number Coords.	Length (inches)	Sum Ang. Ch.	Average Ang. Ch.	Ang. Ch. (per in.)
A	1096	14.15	605.6	.55	42.8
B	875	14.08	579.4	.66	41.1
C	875	12.96	268.6	.31	20.7
D	439	13.84	368.5	.84	26.6
E	220	13.38	228.9	1.05	17.1
F	112	12.68	148.1	1.35	11.7
G	54	11.56	71.6	1.38	6.2
H	28	10.54	34.2	1.31	3.2
I	15	9.76	16.0	1.23	1.6
J	7	9.0	5.2	1.04	.6

TABLE 2 VECTOR ERROR DISPLACEMENTS

Line Ident.	Number Coords.	Sum of Abs. Vector Error	Mean Abs. Vector Error
K	439	.87	.002
L	220	2.58	.004
M	112	6.05	.008
N	54	13.19	.016
O	28	21.27	.025
P	15	35.67	.041
Q	7	72.18	.083
R	58	12.09	.015
S	58	12.09	.015
T	59	21.31	.026
U	59	21.31	.026

through Q in Figure 1. Vector error measurements provide essential information on the amount of spatial displacement of a simplified line. This measure also makes it possible for a cartographer to compare the results of different algorithms as will be demonstrated in the second section of this paper. Displacement data for lines K through Q are shown in Table 2 including the number of coordinates, the sum of the absolute vector errors, and the mean absolute vector error.

A different perspective of the vector error distribution is shown in the histograms presented in Figure 2. The raw vector error is shown below each histogram so that the graphs can be tied to the data in Table 2. The first bar on the left side of each histogram represents the number of zero errors in the simplification. Each bar to the right of the first shows the number of errors in two-thousandths of an inch steps. Since it is impossible to simplify a naturally occurring line without error the histogram labeled B/O represents the fact that there are no errors on the cleaned line B. The histogram labeled K/.87 has a total of eighty-seven hundredths of an inch absolute vector error and represents the simplification with four hundred and thirty-nine coordinate pairs. The size and position of the bars in this histogram indicates the correct structure for a quality simplification. There are only two hundred and four errors in K/.87 and none of these exceed six thousandths of an inch. On the other hand, there are seven

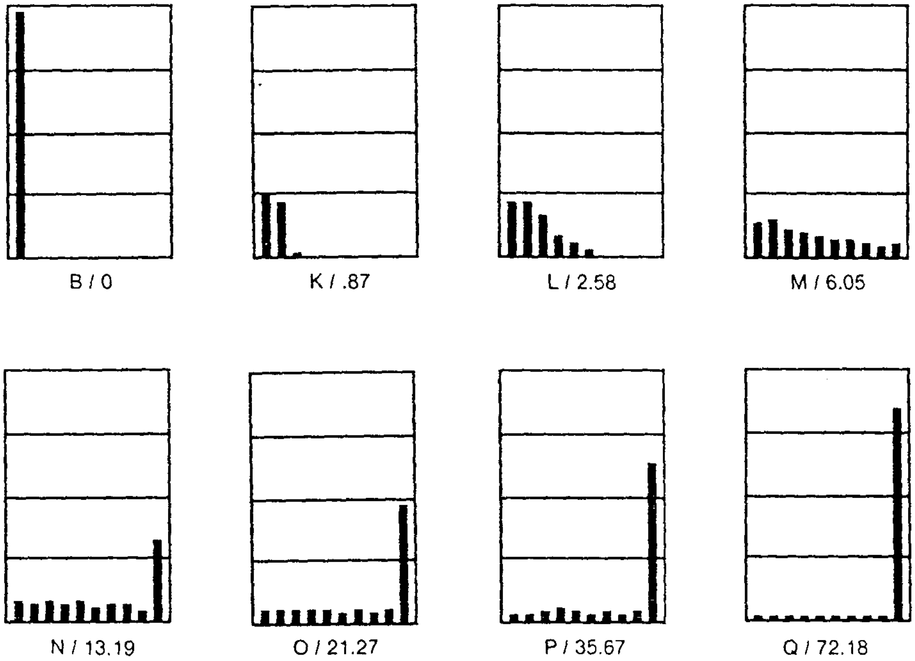


FIGURE 2 Vector error histograms for lines B and K through Q. The horizontal scale of the histograms has an interval of two thousandths of an inch. The vertical scale presents the percentage of coordinate pairs in each histogram category. The data printed below the histograms identifies the line and the sum of absolute errors for that line measured in inches.

hundred and fifty-two errors in the histogram for line Q/72.18 located in the lower right corner of the Figure 2. A similar set of histograms representing areal error is presented in McMaster (1983b).

Measure simplification algorithm consistency

Ideally a cartographer would like to have an algorithm which creates a simplification with a minimal sum of error vectors and a uniform or consistent distribution of the individual errors along the length of the line. A statistic for simplification consistency can be created in a three-step procedure. First, the simplified line is broken into a number of segments of equal length so that the total vector error sum can be subdivided into the same number of parts. For example, Fall River line B of Figure 1, created by the Douglas algorithm, was broken into twenty-seven segments of one-half inch each. Second, the absolute vector errors for each segment were summed and are displayed as vertical bars in Figure 3-A. The data for the displayed twenty-seven bars were changed to absolute values and then summed and averaged. This created a mean absolute vector error (MAVE) of .481 inches which is represented as a solid line on Figure 3-A. Third, the MAVE was subtracted from each of the segment vector error sums and these dispersion values are displayed as bars above or below the MAVE line in Figure 3-B. The simplification consistency statistic is calculated by averaging the absolute dispersion values to create a mean absolute deviation (MAD) for a value of .144 inches. The graphs and data displayed in Figures 3-A and 3-B represent

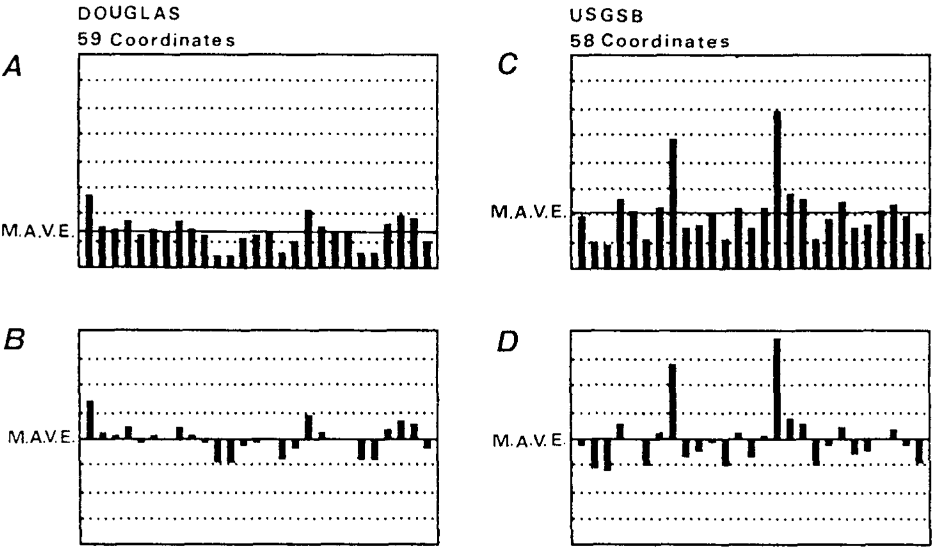


FIGURE 3 Consistency graphs for lines R and T.
Graph A. Each bar represents the sum of the vector errors for the 27 linear segments of the DOUGLAS simplification of Fall River. The solid horizontal line across the graph is the average of the 27 vector error sums – the mean absolute vector error MAVE equals .481 inches.
Graph B. The black bars represent the 27 dispersion values on either side of MAVE. The mean absolute deviation of these bars is 0.144 inches–MAD is a measure of consistency for this simplification.
Graph C. The bars on this histogram represent the 27 vector error sums for line T of the Fall River simplification created by the USGSB algorithm.
Graph D. The mean absolute deviation for line T–MAD is 0.261 inches indicating that the DOUGLAS algorithm performed more consistently than USGSB.

measures for the Douglas simplification shown as line R in Figure 1. Line S of Figure 1 shows the individual vector errors for the simplification R (Davis 1986).[1]

The consistency measure MAD is not particularly useful by itself but it becomes a very useful tool for comparing the simplification of one algorithm with the simplification of another. Line T in Figure 1 was created using the USGSB algorithm which was used in the 1960s by some cartographers in the U.S. Geological Survey (Jenks 196?). Line U of Figure 1 is a display of the individual errors for the simplification of line T. Note that lines R and T are visually comparable and that both contain nearly identical numbers of coordinates as shown in Table 2 and Figure 3. The graphs C and D of Figure 3 are graphic displays of the information used to create a new MAVE and MAD for the USGSB simplification T of Figure 1. The values for graphs C and D in Figure 3 are MAVE of .781 inches and for MAD of .261 inches. These values for line T when compared with those obtained for line R confirm visual impressions obtained from studying Figure 3. Clearly the Douglas algorithm performed its simplification task more consistently than the USGSB algorithm.

Characteristic Points

In 1954 psychologist Fred Attneave illustrated how a limited number of points connected by straight line vectors could convey the essential information about a sleeping cat (Attneave 1954). Members of other disciplines – for example cartoonists and other artists – often connect points with straight lines to emphasize facial features. Geographers and cartographers use similar points as basic anchors when they draft or use computers to represent naturally occurring lines on maps. Few cartographers compiling a map of the coast of California would eliminate San Francisco Bay or omit Cape Cod from the coastline of Massachusetts. These characteristic or critical points can be used to maintain the spatial integrity of lines, but they can also be used to represent the position of economic, political or social features.

Characteristic points along cartographic lines do not all have the same informative value in a mapping situation: for instance, the center of attention may require either political points or geomorphic points on a river (Thapa 1988). The selection of characteristic points also changes with map scale. Large scale maps usually carry more detailed environmental information than small scale maps and thus require many more critical points. This scalar requirement has recently been given great attention by computer cartographers seeking to develop hierarchical data bases. It is also interesting that practitioners of this new methodology have found the Douglas algorithm useful. When the simplification parameters of the Douglas algorithm are modified, the points selected in the simplification process are themselves in a hierarchy. The number of points selected using the most stringent parameter are repeated whenever less binding parameters are utilized.

In 1978, Jill Skoog Marino (1978) created an experiment to determine whether cartographers and non-cartographers would select a series of similar characteristic points on naturally occurring lines. The percentage of similar choices by the two groups of participants were surprisingly consistent generating a correlation of $r = .9$ at the .001 significance level.

Fall River was one of the lines that Marino (1978) used in her experiments dealing with the selection of characteristic points by cartographers and non-cartographers. A series of twenty characteristic points selected by cartographers was chosen for use in the demonstration which follows. The simplified line labeled AC of Figure 4 contains fifty-seven coordinate pairs. A copy of that line with the twenty characteristic points selected in the Marino experiment is labeled BC. Line CC is a smoothed version of line BC in which the anchoring characteristic points were utilized to control spatial displacement as shown in line DC. The algorithm used to reduce spatial movement of lines in the simplification process is a controlled running average operator. The data to be used must have one of two numerical identifiers preceding each coordinate pair.[2] As the operator approaches a coordinate pair with a characteristic point identifier the program passes control to a subroutine which triplicates this coordinate pair forcing the line through the selected point. Line EC is a rather severe smoothing of line AC without characteristic point control and the displacement resulting from this smoothing is shown in line FC of Figure 4. This illustration emphasizes that cartographers should not generate smoothings of naturally occurring lines without understanding the logic behind what they are doing. In the most severe instances, smoothing without characteristic point control can cause a stream to be moved away from river port or moved so that the city appears on the opposite bank of the stream.

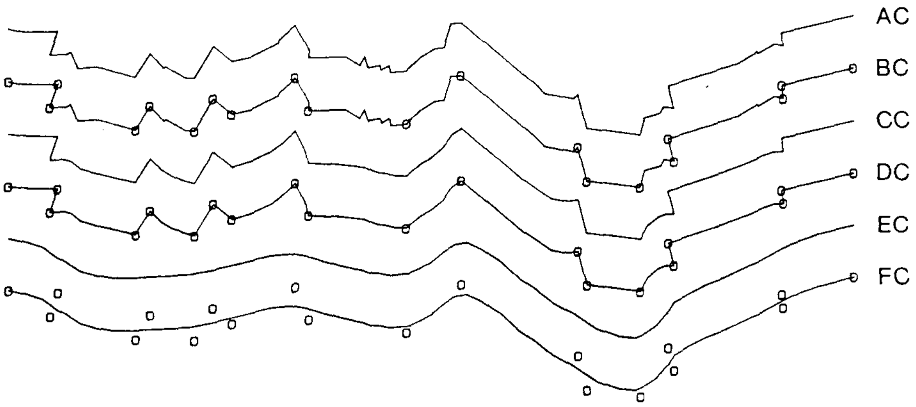


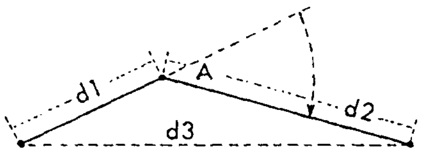
FIGURE 4 The small circles in this illustration are twenty characteristic points selected by cartographers in one of Marino's experiments. These points can be used to minimize spatial displacement as can be seen in plots of lines CC and DC. When characteristic points are not utilized when smoothing lines the end product may have severe displacement as shown in lines EC and FC.

This concludes the discussion which has focused upon geographic logic and the use of various types of linear evaluation procedures. The simplification characteristics of mappable lines may be evaluated singly or comparisons between lines can be used to determine which have the best quality. The impetus behind this discussion is an attempt to draw attention to procedures which often result in the production of lower quality maps.

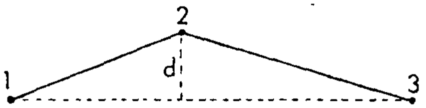
PERCEPTION RESEARCH

The discussion will now turn to a research project dealing with linear simplification which included three basic elements. The first involved the search for guidelines that can be used by mapmakers who want to simplify digital files used to plot naturally occurring lines. Here, questions such as: What percentage of the coordinates are superfluous? Does an individual algorithm simplify different types of geomorphic lines equally well? Is there a perceptual difference between lines created using different algorithms? The second research topic revolved around comparisons of the perceptual similarities and differences between three groups of map users. For example, do professors and graduate students in geography, university freshmen, and sixth graders in elementary school make similar judgments about simplified lines? Thirdly, are different levels of simplification appropriate for topographic maps, atlas and textbook maps, and maps for the media (Jenks 1985)?

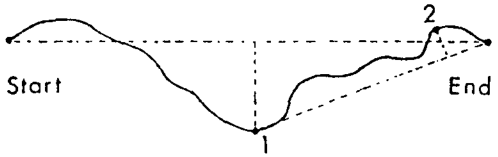
Research psychologists have found over- and under-estimation of the length of vectors, and angular changes between vectors in their work with lines. Such findings indicated that an experimental design involving scaling ought to be selected where multiple stimuli are used. Eleven different geomorphic lines used by Robert McMaster (1983a) in his dissertation, including three rivers, four coastlines and four contours, were selected and each was simplified by three different algorithms at eight levels of coordinate content. Creation of comparably similar (DOUGLAS, SIMPLIFY, USGSB) simplifications using three mathe-



SIMPLIFY



USGSB



DOUGLAS

FIGURE 5 The SIMPLIFY simplification algorithm created by Jenks uses four parameters, three distances and one angle to determine whether the middle coordinate in the triad is to be kept or discarded.

The USGSB algorithm was created by unknown cartographers in the United States Geological Survey. One distance parameter is utilized to decide whether to keep or discard the center coordinate of the triad.

The DOUGLAS algorithm was created by David Douglas and is discussed in detail in Douglas and Peucker 1973. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. (*The Canadian Cartographer* 10/2: 112–122).

matically different algorithms posed problems which were solved by an iterative process. The nature of the three algorithms are displayed graphically in Figure 5. A geometric progression, similar to ones used by psychologists in psychophysical research, was first used to determine the desired eight levels of coordinate count.[3] A part of the coastline of Chesapeake Bay was selected to illustrate the levels of generalization for each of the algorithms. Each of the simplified lines was evaluated using a program to generate a vector error value which could later be used in arraying the lines in rank order. The number of coordinate pairs (italic lettering) and the vector error values (block lettering) are shown with the Chesapeake simplifications in Figure 6.

Administration of the perceptual experiment

Cartographic colleagues in ten different universities in the southwestern part of the United States cooperated in the research effort by making themselves and their students available (Jenks 1984). Five hundred individuals participated in the evaluation of one of the graphic data sets after having received a brief training session to apprise them of the terminology used in the experiment. The experiment involved a graphic scaling task in which the participants were given a randomly arranged set of simplified lines composed of twenty-four simplifica-

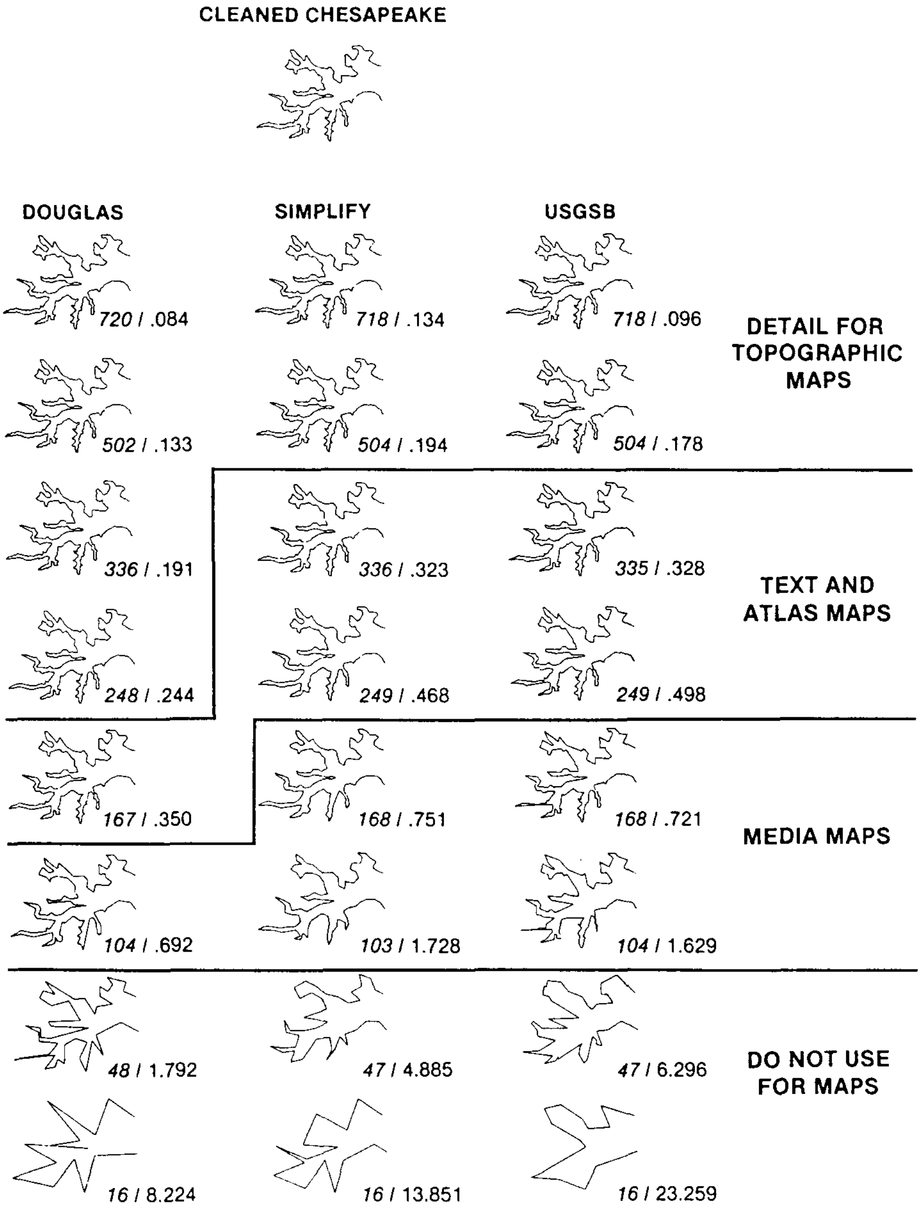


FIGURE 6 A portion of the coastline of Chesapeake Bay simplified by three different algorithms–DOUGLAS, SIMPLIFY and USGSB. The level of simplification indicates the practical use of the modified lines shown by the heavy type on the right-hand side of the figure.

tions – as well as two copies of the original line. The simplifications and one copy of the original line were identified by a single letter or number. The second copy of the original line was labeled with the word 'original'. The directions were to remove the maps from the envelope and to locate the map labeled 'original' and place it on the upper left-hand corner of the desk. They were then asked to locate the one of the remaining lines which they considered to be an exact duplicate of the one labeled original and place it on the lower left-hand corner of the desk. The participants were then asked to locate the one map that was most unlike the original and place it on the lower right-hand corner of the desk. The remaining maps were to be arranged in perceptual order – in terms of complexity – between the two extremes. No time limit was set and many participants spent a rather long time arranging and rearranging the maps. When satisfied, the subjects copied the identifying letters or digits onto a prepared answer sheet. The final task was to draw three vertical lines on the answer sheet between the map identification letters or numbers so that they created four groupings. The first three blocks were to be labeled as 1 suitable for topographic maps, 2 maps for atlases or text books, and 3 maps for media levels of simplification. The fourth block contained identification letters or numbers for maps which were considered so over simplified that they should not be used for any communicative purpose. These four blocks for the Chesapeake series of simplifications are shown in Figure 6. At least thirty persons evaluated each of the eleven data sets.

Data Analysis

In his early research in perception the psychologist Fred Attneave developed the notion of psychological distance between two visual stimuli (Attneave 1954).[4] Seven error measures were tested and vector error was considered to be the most appropriate surrogate for psychological distance. As a result vector error was used to rank order the twenty-five Chesapeake lines displayed in Figure 6 (Jenks 1985). These ranks were utilized to make comparisons with the perceived order that was assigned by the participants. Prior to data collection an assumption was made that many participants might array the ranked lines in perfect vector error order. Such was not the case for the Chesapeake simplifications because none of the thirty-five participants ranked all twenty-five lines in correct order. In fact, thirteen lines correctly ordered was the maximum and the average number was only eight. A thirty-five row (for the number of participants) and twenty-five column array (for the perceived ranks) was created. There are 875 possible positions in this array and only 278 were correctly used for a 32 percent correct response. The response accuracy for the other ten lines was similar. As a rule, the excessively simplified lines were perceptually different enough that the respondents could array them correctly. However, the simplifications with a large number of coordinate pairs were so perceptually similar that most participants were unable to perceive differences between them, resulting in ranks which were often in error.

The three-dimensional graph shown in Figure 7 illustrates the response set in a twenty-five by twenty-five matrix. Had the assumption of great accuracy proven correct all of the polygon positions along the axis from the front to the rear of the graph would have been very tall with little dispersion. Instead, the broad spread in the foreground demonstrates participants' confusion in their attempts to determine differences between the less drastic simplifications which appear to be very similar. In contrast, the tallest prism in the rear of the graph represents thirty-three correct rankings out of a possible thirty-five. The prism heights at the front of the graph are short with the prism at the apex being six

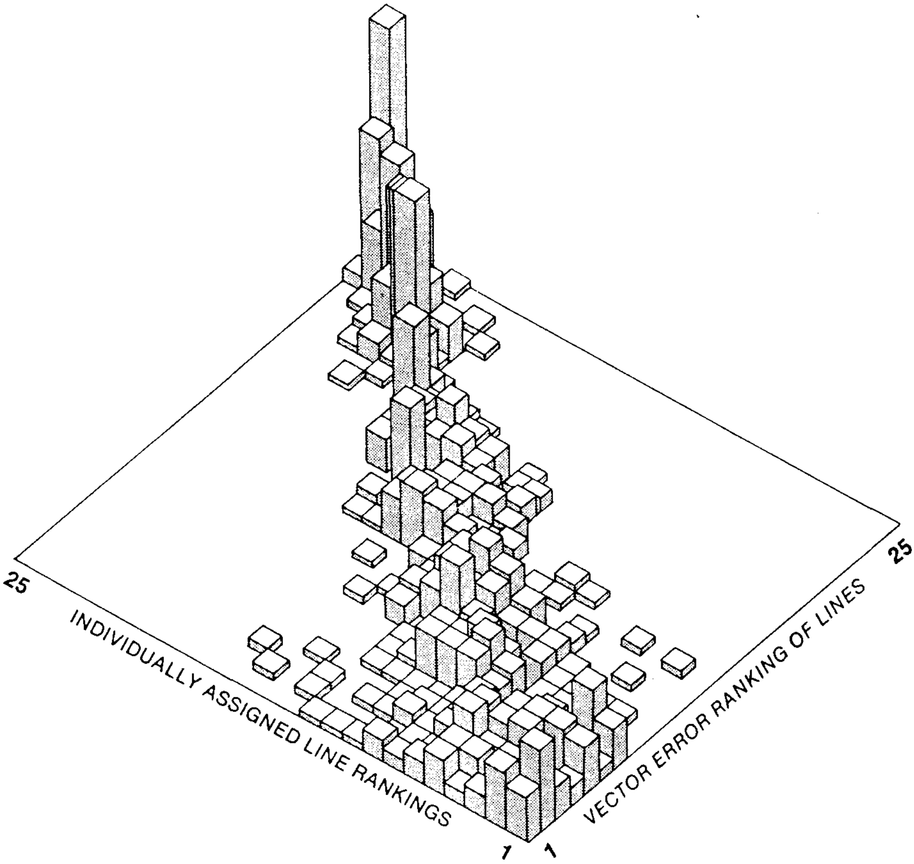


FIGURE 7 A twenty-five by twenty-five 3D histogram of responses in the scaling experiment. The broad area of short prisms in the foreground indicates participant confusion in ordering the Chesapeake lines with minimal amounts of vector error. The taller prisms at the rear of the histogram show that the Chesapeake lines with larger vector errors were scaled correctly.

units high, indicating that twenty-nine respondents erred in the ranking of the copy of the original line. Interestingly one individual, in arraying the twenty-five maps, thought that the line with a vector error ranking of eleven was identical to the copy of the original with a vector error ranking of one.

The unexpectedly broad scattering of responses became a major cause of concern about the experimental design. One could expect some scattering of responses due to perceptual error or inattention of the participants, but such scattering should have taken place equally along the whole central axis of the graph. A careful analysis of the preliminary results revealed the cause of this distribution. The experimental design for testing the three algorithms was based upon triads of nearly equal coordinate counts for each level of simplification and, in a number of instances, this resulted in vector error sums that were very similar. A close inspection of Figure 6 indicates nearly identical pairs of vector error sums of .133 and .134, .191 and .194, and .323 and .328. An inspection of the individual responses indicates that these pairs and a few other close couples were indeed misranked and should have been assigned ties. This finding posed an unsolvable dilemma, because if the triads had been created on the basis of similar vector error the coordinate count, and therefore the levels of simplification, would have been highly varied. Note, for example, the pair of simplifications with vector errors of .133 and .134 which have a difference of two hundred and sixteen coordinate pairs. Further inspection of the data in Figure 7 shows that as the differences in vector error increase near the rear of the graph, respondents were able to more correctly array the lines. In addition, the narrowing width (less dispersion) of the response spread along the central axis toward the rear of the figure corresponds with the increasing differences in vector error measures.

Earlier in this section a number of questions were posed. Answers to some of these questions are to be found in the lines and data presented in Figures 6 and 7. For example, the lines drawn across Figure 6 from left to right delineate the accepted levels in linear simplification for different kinds of maps. The first two boundaries were correctly anticipated but the experimental participants evidently want more detail in media maps than is found today in newspapers and in news reports on television. Another interesting finding is the similarity of the responses of professors and graduate students in geography, university freshmen, and sixth grade students. Several pieces of evidence, such as the broad scatter in Figure 7 and the nearly identical vector error pairings and the location of the border between topographic and atlas/textbook maps in Figure 6 indicate that there is substantially more detail in non-simplified naturally occurring lines than mapusers can assimilate. One could therefore take the position that two lines are identical if the viewer cannot perceive differences between them. Based upon this concept the line resulting from the Douglas algorithm simplification with a vector error of .191 and a coordinate count of 336 pairs could be used for topographic maps instead of other simplifications with greater numbers of coordinate pairs. This would mean that seventy-five percent of the 1293 coordinates of the original line are superfluous. Likewise, sixty percent of the coordinates on the original line (line A in Figure 6) are superflous if one elects to use either the SIMPLIFY or the USGSB algorithm. An answer to the question of whether an algorithm can create quality simplifications of all geomorphic lines is still unanswered. The great diversity of both algorithms and naturally occurring lines must be tested further using many different lines and thinning algorithms to provide a positive response.

AFTER-THOUGHTS

Maps are scalar representations of some aspect of the environment, and as such they should be made by environmentalists or cartographers with environment training. Too often members of the academic community and others interested in making maps have circumvented training in cartography or graphic display. This is particularly true since the advent of the microcomputer. This paper is an attempt to insert some cartographic and geographic logic into the minds of all who wish to create a map. Compilation of a map is an integrated process which involves concepts, acquisition and manipulation of data, and the ability to generate clear concise map displays. This requires education, experience, and geographic thinking in cartographic decision making.

NOTES

[1] There are two extremes that could be utilized in an attempt to understand the consistency of algorithmic simplifications. The first is to compare the absolute sum of all vector errors on the line. The second is to evaluate each individual vector error and then obtain the mean absolute deviation for the total series. Experience gained by studying the production of a variety of algorithms indicates that neither of these methods is particularly useful.

Naturally occurring lines tend to have a mixture of linear features in terms of the level of curvilinearity. One usually finds that there are a number of high intensity directional changes interspersed with smoother gentle curves. Some algorithms work very well with one of these forms but not with others. What is needed is an algorithm which handles all curvilinear types. If one studies lines K through Q in Figure 1 the variation in vector error patterns becomes clear.

Dividing cleaned files into equal linear sections is an attempt to isolate parts of the simplified line into segments which have uniform characteristics. The use of twenty seven segments in the Fall River line was arbitrary and other cartographers might very well use another number.

[2] Characteristic points for a naturally occurring line can be selected by a digitizer operator, by an algorithm, or by an experimental procedure such as that developed by Marino (1978). In any case, the characteristic coordinate pairs must be labeled to signify that they are different from normal coordinate pairs. The labeling process might follow the form P, X, Y where P is an integer index. For example, a P of 7 might be used for characteristic points while a P of 3 could be used to identify common coordinate pairs.

[3] The basic premise of the experimental design used in this paper was to evaluate the quality of simplifications created by different algorithms. This notion was difficult to carry out because each of the three algorithms used was created using different mathematical bases.

Psychologists often apply logarithmic or geometric progressions to subdivide their data sets into clearly distinct categories. A geometric progression was applied to the 1293 coordinate pairs for the Chesapeake line. The coordinate sets produced are as follows: 718, 504, 336, 248, 168, 104, 47, and 16. These, or nearly identical values are shown below each simplification in Figure 6. The variation in the values was caused by the algorithm formula limitations.

An iterative procedure for arriving at the desired number of coordinate pairs was utilized. In the case of the USGSB algorithm, which has a single decision parameter, the operator selected a parametric value and ran the algorithm with the Chesapeake data set. If the result was not as desired a new parametric value was chosen and the procedure repeated until satisfactory results were achieved.

[4] Attneave (1954) developed the concept of psychological distance between graphic figures so that viewers might visually separate them and be able to rank them in order. Seven error measurements were applied to the experimental data previously discussed. The rankings for each measure were evaluated and the vector error measure provided values that could be ranked in logical order.

REFERENCES

ANSTEY, N.A. 1965. Wiggles, Jnl. *Canadian Society of Exploration Geologists* 1/1: 13–43.

ATTNEAVE, F. 1954. Some informational aspects of visual perception. *Psychological Review* 61/3: 183–193.

DAVIS, J. 1986. *Statistics and data analysis in geology*, New York: John Wiley and Sons.

DOUGLAS, D.H. and PEUCKER, T.K. 1973. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *The Canadian Cartographer* 10/2: 112–122.

JENKS, G.F. 196? Personal communication.

——— 1981. Lines, computers, and human frailties. *Annals of the Association of American Geographers* 71/1: 1–10.

——— 1984. I am particularly indebted to Professors Dulio Peruzzi, Eugene Turner, Richard Wright, Dennis Fitzsimons, and Hans Meihoefer.

——— 1985. Linear simplification: how far can we go? Paper presented to the *Tenth Annual Meeting, Canadian Cartographic Association,* Fredrickton, NB.

——— 1986. Computer cartography–problems and potential. Paper presented to the *6th Australian Cartographic Conference Astra-Carto Two,* Melbourne, Oct. 26–30, 1986.

MARINO, J.S. 1978. Characteristic points and their significance in cartographic line generalization. Unpublished Master's thesis, Department of Geography-Meteorology, University of Kansas.

——— 1979. Identification of characteristic points along naturally occurring lines: An empirical study. *The Canadian Cartographer* 16/1: 70–80.

MCMASTER, R.B. 1983a. Mathematical measures for the evaluation of simplified lines on maps. Unpublished Ph.D. Dissertation, Department of Geography-Meteorology, University of Kansas.

——— 1983b. A mathematical evaluation of simplification algorithms. *Proceedings of the Sixth International Conference on Automated Cartography* AUTO-CARTO 6 267–276.

——— 1986. A statistical analysis of mathematical measures for linear simplification. *The American Cartographer* 13/2: 103–117.

——— 1987. Automated line generalization. *Cartographica* 24/2, Monograph 37.

THAPA, K. 1988. Critical points detection and automated line generalization in raster data using zero crossings. *The Cartographic Journal* 25/1: 58–68.

ABSTRACT The logical problems faced by a geographer who generalizes linear symbols on a map are the same whether traditional or computer methods are used. One must recognize that there are a great many lines on the earth's surface, some of which were created by natural processes while others result from the economic, political or social activities of man. Geographic logic requires that these earth lines, when plotted on a map, should be recognizable representations in terms of symbolization and shape. Furthermore, geographers must understand that the generalization process shortens the length of linear symbols and moves them in geographic space. Critical or characteristic points should be selected so that the flow and position of the line is maintained.

RÉSUMÉ Le géographe qui généralise les symboles linéaires d'une carte affronte des problèmes logiques identiques, que les méthods, utilisées soient traditionnelles ou automatisées. Nombreuses sont ces lignes qui parcourent la surface du globe: certaines sont issues de processus naturels alors que d'autres proviennent des activités économiques, politiques ou sociales de l'humanité. Selon la logique du géographe, ces lignes au sol doivent être représentées fidèlement sur la carte en termes de symbolisation et de forme. De plus, le géographe doit comprendre que le processus de généralisation raccourcit l'étendue des symboles linéaires et les déplace dans l'espace géographique. Il convient donc de choisir des points critiques ou caractéristiques afin que la position et l'enchaînement d'une ligne soient préservés.

ZUSAMMENFASSUNG Ein Geograph, der lineare Kartenzeichen generalisiert, steht dendelben Problemen gegenüber, ob er nun herkömmliche oder rechnergestützte Verfahren benutzt. Man muβ erkennen, daβ es zahllose Linien auf der Erdoberfläche gibt, die entweder aus Naturabläufen entstanden sind oder von wirtschaftlichen, politischen oder gesellschaftlichen Tätigkeiten der Menschheit herrühren. Wenn man diese Linien auf einer Karte darstellt, erfordert die geographische Logik, daβ sie erkennbare Wiedergaben hinsichtlich Symbolisierung und Gestalt sind. Ferner muβ der Geograph verstehen, daβ der Generalisierungsprozeβ die Länge der linearen Zeichen verkürzt und sie im geographischen Raum verschiebt. Kritische oder charakteristische Punkte müssen ausgewählt werden, damit Verlauf und Lage der Linie bewahrt werden.

GEORGE F JENKS is Professor Emeritus in the Department of Geography, University of Kansas, Lawrence, Kansas, MS *submitted August 1988.*