NOAA Technical Report NOS 112
Charting and Geodetic Services Series CGS 5

CARTOGRAPHIC GENERALIZATION: A REVIEW OF FEATURE SIMPLIFICATION AND SYSTEMATIC POINT ELIMINATION ALGORITHMS

Victoria H. Clayton
National Charting Research
and Development Laboratory

Rockville, MD
November 1985
Reprinted July 1986

U. S. DEPARTMENT OF COMMERCE
Malcolm Baldrige, Secretary

National Oceanic and Atmospheric Administration
Anthony J. Calio, Administrator

National Ocean Service
Paul M. Wolff, Assistant Administrator

Office of Charting and Geodetic Services
R. Adm. John D. Bossler, Director

NOAA Technical Report NOS 112
Charting and Geodetic Services Series CGS 5

CARTOGRAPHIC GENERALIZATION: A REVIEW OF FEATURE SIMPLIFICATION AND SYSTEMATIC POINT ELIMINATION ALGORITHMS

Victoria H. Clayton
National Charting Research
and Development Laboratory

Rockville, MD
November 1985
Reprinted July 1986

U. S. DEPARTMENT OF COMMERCE
Malcolm Baldrige, Secretary

National Oceanic and Atmospheric Administration
Anthony J. Calio, Administrator

National Ocean Service
Paul M. Wolff, Assistant Administrator

Office of Charting and Geodetic Services
R. Adm. John D. Bossler, Director

CONTENTS

FIGURES

CARTOGRAPHIC GENERALIZATION: A REVIEW OF FEATURE SIMPLIFICATION AND SYSTEMATIC POINT ELIMINATION ALGORITHMS

Victoria H. Clayton
National Charting Research and Development Laboratory
Charting and Geodetic Services
National Ocean Service, NOAA
Rockville, Maryland 20852

ABSTRACT: Numerous definitions and theories pertaining to cartographic generalization have been introduced by geographers and cartographers during the past century. In this paper, both manual and computer-assisted approaches to the process of linear and feature simplification are described and compared. Particular emphasis is placed on computer-assisted algorithms pertaining to vector data manipulation; specifically, systematic point elimination algorithms. In addition, several line smoothing algorithms are briefly discussed.

INTRODUCTION

One major difference between manual and computer-assisted cartography is the treatment of cartographic elements (point, line and area features) in map displays. The practice of manual cartography approaches this task subjectively, for the cartographers' individual bias affects the way in which map generalization is performed. The incorporation of computer assistance into cartography reduces individual bias by minimizing user intervention; permitting the user to specify certain standard criteria only in the initial stages of the simplification process.

The purpose of this paper is to present various theories, held by several prominent geographers and cartographers, which pertain to computer-assisted cartographic approaches to linear and feature simplification. In so doing, background material will be provided on manual methods of cartographic generalization, in order to establish a framework from which relatively recent developments in the field, specifically linear and feature simplification algorithms, will be discussed.

DEFINITIONS FOR CARTOGRAPHIC GENERALIZATION

Cartographic generalization has been discussed and analyzed by various geographers and cartographers since the early 1900's. In attempting to explain the process, each author has approached the topic from a different viewpoint. Some have methodically outlined what they perceive to be the proper steps for the cartographer to take when generalizing from large- to small-scale maps. Others have admitted their inability to describe accurately what the cartographer does when generalizing a map. The definitions, in this section illustrate the wide variety of viewpoints adopted by geographers and cartographers during the past century.

In 1908, Eckert asserted that "generalization depends on personal and subjective feelings," and therefore was "part of the 'art' that enters into the map making process" (Traylor 1979: 6). More recently, the "Multilingual Dictionary of Technical Terms in Cartography" prepared by the International Cartographic Association (ICA) defines cartographic generalization as "the selection and simplified representation of detail appropriate to the scale and/or purpose of the map" (Brophy 1973: 300). Brophy (1973: 300), however, maintains "generalization is an ambiguous process which lacks definite rules, guidelines or systematization." Keates (1973: 23), on the other hand, explains the outcome of the generalization process by describing it as "that which affects both location and meaning... (for) as the amount of space available for showing features on the map decreases in scale, less locational information can be given about features, both individually and collectively." Traylor (1979:24) contributed to the ICA definition by stating that generalization consists of "the selection and simplified representation of the phenomena being mapped, in order to reflect reality in its basic, typical aspects and characteristic peculiarities in accordance with the purpose, the subject matter, and the scale of the map." In addition, Koeman and Van der Weiden (1970) examined another aspect of the generalization process by considering the amount of information at the cartographer's disposal and the skill of the cartographer.

This report does not contain an exhaustive list of definitions, but it does indicate the prevailing attitudes of some of the most prominent theorists in geography and cartography. Terms such as "selection," "simplification," "map purpose," and "subjectivity" are often used in describing the generalization process. These terms will be discussed later in this paper. For, as there are many definitions of cartographic generalization, so are there many approaches to the nature of its practice.

THE PRACTICE OF MANUAL CARTOGRAPHIC GENERALIZATION

In 1908, Eckert divided the process of manual cartographic generalization into three stages: 1) the quantitative stage which involved the selection of the number of objects to be shown, 2) the technical stage which simplified the form, and 3) the qualitative stage where subjective evaluation of the elements for inclusion was performed (McMaster 1983: 30).

Lundquist (1959: 47-48) develops what he calls "basic rules for generalization." Under these rules the user must consider:

- Purpose of the map
- Reduction factor (scale factor)
- Objective evaluation using an examination of relevant data
- Local importance factor stressing a need for regional knowledge
- Attempt uniformity
- Awareness of the effects of personal prejudice

It was not until 1978 that Robinson et al. (1978: 150) introduced, in Elements of Cartography, the most comprehensive and systematic discussion of generalization. In this discussion, they define cartographic generalization as "the modification of specific data in order to increase the effectiveness of the communication by counteracting the undesirable consequences of reduction."

Robinson et al. begins by describing a "pregeneralization" step called selection. In this operation, the cartographer selects the information to be conveyed on the map. Robinson et al. (1978: 150) stress that selection is not a part of cartographic generalization because "selection is the intellectual process of deciding which information will be necessary to carry out the purpose of the map successfully." They classify the process of generalization into the following categories: 1) simplification (where "the cartographer determines the important characteristics of the data"), 2) classification (which "describes the ordering or scaling and grouping of the data"), 3) symbolization (the "graphic coding of the scaled and/or grouped essential characteristics"), and 4) induction (which refers to "the application in cartography of the logical process of inference").

Robinson et al. divides the concept of cartographic generalization into four separate and distinct elements where, for example, simplification is viewed as merely one aspect of the generalization process, and selection is considered outside of the process. This proves to be important when analyzing various computer-assisted generalization algorithms, as shown later in this report.

McMaster (1983: 10) summarizes the generalization scheme of Robinson et al. by stating, "linear generalization is a more comprehensive term which includes all facets of line manipulation: 1) simplification, 2) smoothing, and 3) possible feature displacement." Other views in cartographic generalization are discussed in the following paragraphs. However, they are neither as structured nor as thorough as Robinson's theory.

Jenks (1981: 8) expresses simplification as "a holistic process during which the cartographer simultaneously examines the naturally occurring line from a number of different contexts." "During one integrated activity points or features to be retained are selected, unwanted details are eliminated, and the new version of the line is drafted." In conclusion, Jenks (1981: 8) states "the quality of a simplified representation depends on an understanding of, and adherence to, good cartographic principles."

McMaster (1983: 1) claims that linear simplification is "part of the generalization process and is related to and dependent on scale change, (and) is used to solve the problem of clutter which arises with scale reduction." He further states "it is necessary to redraw a less cluttered 'generalized' line by selecting the lines' salient information and, at the same time, by eliminating the superfluous detail" (McMaster 1983: 1). McMaster then describes three generalization operations: selection of important characteristics, possible exaggeration of characteristics, and elimination of unwanted detail.

Pannekoek (1962: 56) views cartographic generalization as consisting of two processes: a selection of objects to be included on the map, and simplification of the shape to be given to the objects chosen for representation. "As a general rule, it should be said that the features that determine the essential character of the terrain should be stressed and nonessentials should remain subordinate to them or omitted altogether" (Pannekoek 1962: 56).

Keates (1973: 24) describes the process of generalization in the following fashion, "the first step is to select the individual features which are to be retained at the smaller scale, which at the same time will continue to represent the general characteristics of the...area. In addition, each individual feature has to be simplified in form by omitting minor irregularities and retaining only the major elements of shape."

Morrison (1974: 115) introduces an interesting approach to generalization, within a communication theory framework, in "A Theoretical Framework for Cartographic Generalization with Emphasis on the Process of Symbolization." Within this structure he describes the following items as important elements enabling communication to take place: 1) the cartographer and the map reader, 2) the channel of communication, or map, and 3) both cartographer's and map reader's conception of the physical universe and knowledge, i.e., sense of reality. Information flows through the map communication channel as follows (Morrison 1974: 116):

> The communication channel consists of one transformation of the sensory elements of the cartographer's reality to the physical elements on a map, and a second transformation from the map, consisting of the map reader's perception of the physical elements on a map to the sensory elements of the map reader's reality.

Morrison agrees with Robinson et al. (1978) as to the nature of the major elements of generalization: 1) simplification, 2) classification, 3) symbolization, and 4) induction. Morrison, however, adds two types of transformation: 1) one-to-one, and 2) onto. A function ($f: A \rightarrow B$) is said to be one-to-one if distinct elements in A have distinct elements in B, whereas a function ($f: A \rightarrow B$) is said to be onto if every $b \varepsilon B$ is the image of some $a \varepsilon A$, (where A = the reality and B = the map). "The process of simplification, therefore, becomes important on the overall generalization transformation...(and) the cartographer must decide which characteristics to portray on the map and which to ignore" (Morrison 1974: 120). The process of simplification, then, would be equivalent to a one-to-one transformation that was onto. Classification would be onto but not one-to-one. Symbolization would be one-to-one and "into," while induction would be neither one-to-one nor onto.

The major problems with manual generalization have been discussed by various geographers and cartographers. In 1974, Steward outlined six factors causing variations in manual simplification: 1) different human skills in drafting and checking map information, 2) nonuniformity of geographic knowledge among cartographers, 3) environmental working conditions, 4) urgency of production, 5) physical (muscular) control, and 6) mental well-being (McMaster 1983: 13).

Jenks (1981: 1), as well, notes that problems in manual generalization arise because "maps are conceived by people and so are subject to the psychological, physiological and logical limitations of the geo-cartograhers." In addition, McMaster (1983: 14) states "lack of consistency and repeatability amongst hand drawn simplifications of the same line is the primary disadvantage of manual simplification."

Steward accurately pinpointed the problem of manual generalization when he announced the "need to reduce individual bias by establishing impartial, universally acceptable criteria for line generalization" (Marino 1978: 4). Although the generalization scheme of Robinson et al. (1978) aids in developing an understanding of the thought processes undertaken by cartographers during manual line generalization, it does not specify which map features should be generalized. This problem is further compounded when making the transition from traditional or manual line generalization to computer-assisted generalization algorithms.

In manual cartography, the cartographer selects important characteristics, simplifies cartographic features, eliminates unwanted detail, and carries out feature displacement, all in one procedure. But, in computer-assisted cartography, three separate algorithms are required to complete the same tasks. For the "science of cartography" to progress, those tasks which are now performed manually must be automated. One important aspect that must be investigated when constructing an automated (or computer-assisted) cartographic system is the incorporation of an objective generalization program into the system design. By maintaining only critical points along a line, for example, the user reduces plotting time, storage, and storage costs (McMaster 1983: 19-20).

Topfer and Pillewizer (1966: 11) provide the first quantitative basis for conducting cartographic generalization. In "Principles of Selection," they introduce the following equation labeled "the radical law:"

$$nf = na \sqrt{Ma/Mf} \qquad (1)$$

where nf = the number of objects that can be shown at the derived scale
na = the number of objects shown on the source map
Ma = scale denominator of the source map
Mf = scale denominator of the derived map

Equation 1 expresses the relationship between the amount of detail shown on the source map and the amount of detail that can be shown on the generalized map. In other words, when "compiling from larger to smaller scales, the number of items that can be shown on the smaller scale will diminish according to the radical law" (Robinson 1978: 151). This is not an algorithm, but it does provide clues to cartographers on how much information could be transferred from a source map to a generalized map. The problem with "the radical law" is that it does not instruct the cartographer on which features should be retained and which should be eliminated. Therefore, it does not achieve the state of objectivity required in a computer-assisted generalization algorithm.

Srnka (1970: 54) introduces a mathematical equation that also could be used in the selection of linear features:

$$n(Poi)\,\% = e_{oi}\, n^{-f_{oi}}(Po)\, h^{g_{oi}}(Po) \qquad (2)$$

where: n(Poi) % = the percentage of the original number of lines represented within the area Poi in the I-th derived map
n(Po) = number of line elements within the reference area Po of the base map
h(Po) = length of the linear elements within the reference area Po of the base map
e_{oi} = total level of selection
f_{oi} = variable degree of selection at different numbers of linear elements in the base map
g_{oi} = variable degree of selection as a function of the length of the linear elements per unit area of the base map.

Equation 2 "takes into account the significance and density (character distribution) of the generalized phenomena" (White 1983: 8). Unfortunately, like "the radical law," it does not indicate which features are to be selected to remain on the generalized map.

Many algorithms have been developed that conduct cartographic generalization. These include algorithms for linear simplification, line smoothing, and mathematical curve fitting. The remainder of this paper focuses on reviewing algorithms pertinent to the manipulation of vector data, specifically, linear simplification algorithms. In addition, a brief discussion of several algorithms pertaining to line smoothing is included. Other aspects of cartographic generalization, such as mathematical curve fitting routines as well as approaches to classification, symbolization, and induction, will be left for future study.

FEATURE AND LINEAR SIMPLIFICATION ALGORITHMS

"The primary objective of most simplification algorithms is the selection of the major geomorphological characteristics along a line...these are often called critical or salient points" (McMaster 1983: 4). For discussion purposes, simplification algorithms will be classified as applying to either feature simplification, or systematic point elimination. Algorithms for systematically eliminating points will be further subdivided into two categories: sequential point elimination or global point elimination (McMaster 1983: 75). Furthermore, three basic methods of conducting sequential point elimination have been identified: N-th point elimination, establishing a tolerance limit, and creating a corridor/search area.

Feature Simplification

Essentially, the process of feature simplification deals with determining which cartographic features are to be retained when moving from larger to smaller scales. It is often necessary to perform feature simplification when "many small items of the same class are present in an area" (Morrison 1975: 102). When this process is performed manually, states Robinson (1978: 152), "the determination of which data elements to retain can be deduced from the purpose of the map and the place assigned the particular data distribution in the visual hierarchy specified in the map design." This determination depends on the cartographer's knowledge about the data being mapped and, therefore, it becomes subjective.

In computer-assisted cartography, both Robinson and Morrison suggest that feature simplification should be performed by assigning relative importance rankings to the various data elements, after they are input to the data file. The following standards may be used as ranking criteria: 1) size, 2) proximity, or 3) a combination of both size and proximity. For example, "the cartographer may specify the minimum size for retention based on the output scale and line width (Robinson et al. 1978: 160-161). In this way, less important features are suppressed to avoid clutter on the map. By establishing rank standards for all features, objectivity is achieved in conducting feature simplification by computer assistance.

Systematic Point Elimination Algorithms

In computer-assisted cartography, a line is seen as consisting of at least two (and up to a series of) individual X and Y coordinates that have been obtained through the digitizing process. Linear simplification consists of eliminating a certain amount of the X and Y coordinates along the digitized line, while maintaining the essential character of the line.

A number of algorithms have been designed to perform linear simplification. The two major categories are sequential point elimination and global point elimination. The difference between these two categories is the way they handle linear features. Sequential point elimination begins by looking at a small string of coordinate pairs. Point elimination is performed among those points, according to some specified criteria. The process progresses along the line, always working with a small string of coordinates at a time. Global point elimination looks at the entire line at one time and performs point elimination between the first and last coordinate pairs. Then, the algorithms work with consecutively smaller strings of coordinate pairs. Both sequential and global algorithms attempt to reduce the data file by eliminating redundant points. Many attempt to retain only those critical or salient points that characterize the linear feature.

Sequential Point Elimination

In 1966 Tobler introduced one of the first sequential point elimination algorithms, called the "N-th point routine." Beginning with the first coordinate pair on the line, this algorithm was designed to select every N-th coordinate pair to be retained on the generalized line segment (Rhind 1973: 54). The user specifies the value of "n." Hence, the larger the value of "n" the greater the simplification. (See fig. 1.)

This algorithm is conceptually simple, with results produced rapidly and cheaply, but it has three distinct disadvantages: the starting point of the line influences the end result, 2) straight lines are over represented, and safeguards are not established to ensure that critical points will be retained (Rhind 1973: 54). Even though this algorithm is not as subjective as manual techniques, more sophisticated simplification routines are needed.

The next category of sequential point elimination algorithms pertains to the concept of tolerancing, in which algorithms are designed to handle a triad of coordinate pairs at one time. With tolerancing routines, only coordinate pairs distant from the last plotted point by more than a predetermined distance are retained. This limit is established either through a specified line segment length or a particular angular distance.

Tobler, in 1965, and Hershey, in 1963, presented algorithms designed to eliminate superfluous X and Y coordinates through the use of a line segment tolerance limit. This limit would eliminate points whose distance apart was less than some function of the plotted line width (Marino 1978: 6, McMaster 1983: 45). Beginning with the first coordinate pair, the routines would judge whether the distance between it and the second point was eliminated and would then investigate subsequent coordinate pairs. If the point was at a distance greater than the width of the plotted line, that point would be retained. Unfortunately, these routines hold the same disadvantages as were identified in the "n" point routine, and so are not suitable for automated purposes.

Lang (1969: 1) presents an algorithm which incorporates a Euclidean distance measure for point elimination. In this algorithm, coordinate pairs would be

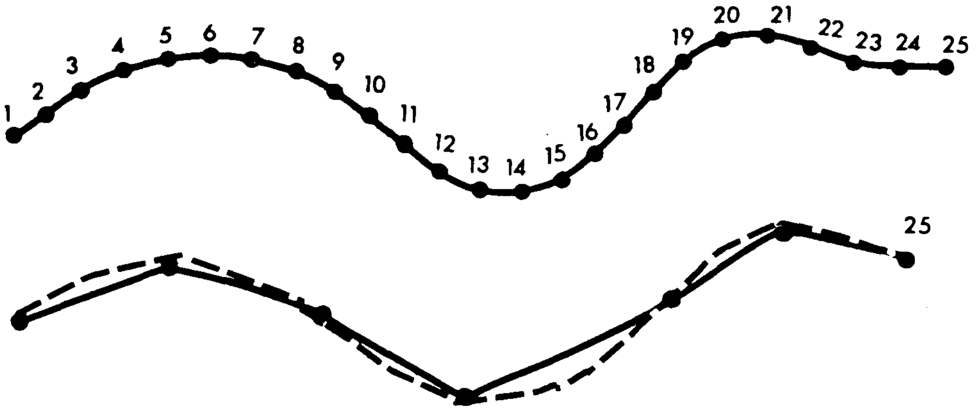
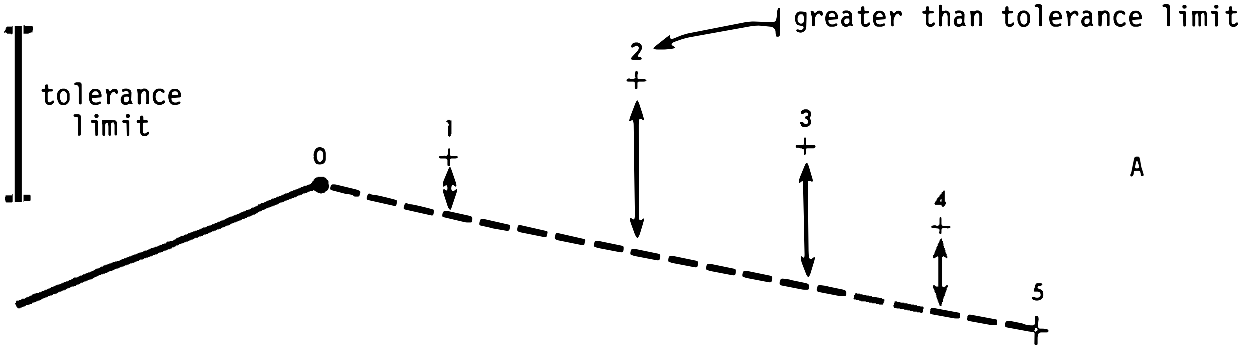
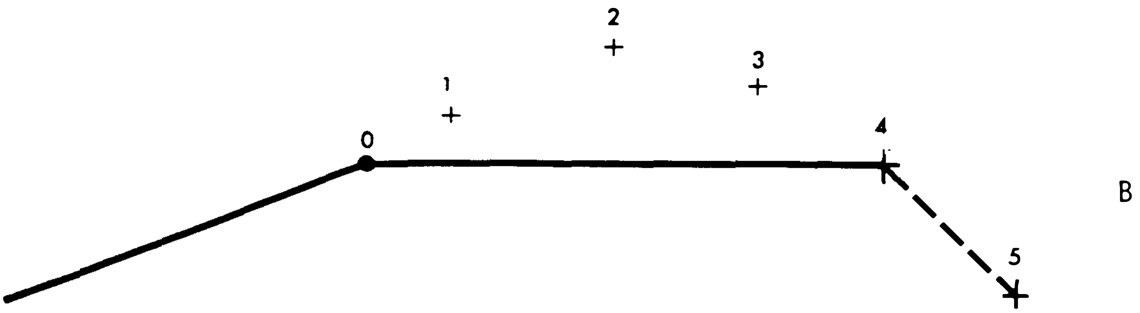


Figure 1. Tobler's "n^{th}" point elimination routine, where "n" equals 4.



A. The line is connected from point 0 to point 5. All intermediary points are evaluated as to whether their distance from the line (pt. 0 - pt. 5) is greater than the tolerance limit. Point 2 exceeds the tolerance limit.



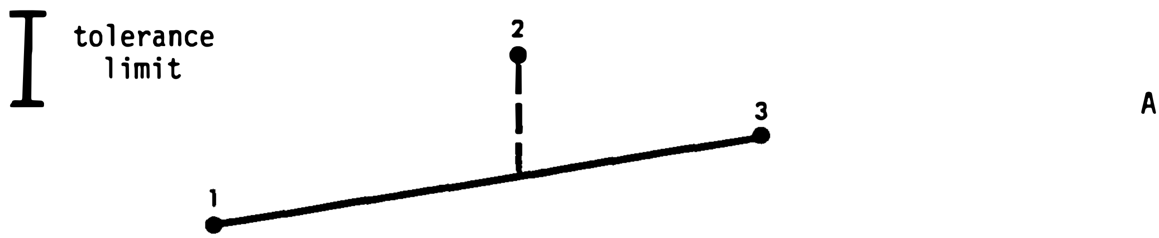
B. The line is redrawn to point 4 instead of point 5. The process is then repeated beginning with point 4.

Figure 2. Lang's tolerancing routine.

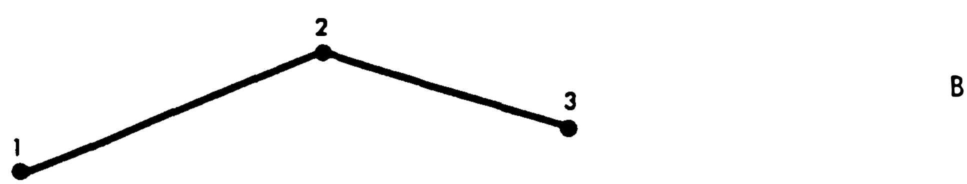
filtered out, "provided more of them lay further (sic) from the line...than some tolerancing accuracy." A tolerance limit (perpendicular line segment length) is first determined by the user. Next, a line is constructed connecting the first coordinate pair to each successive coordinate pair along the linear feature. Each time the line is connected to a new coordinate pair, perpendicular distances will be calculated from that line to all intermediary points. All intermediary points located at a perpendicular distance that is less than the tolerance limit are eliminated. However, intermediary points having a distance that exceeds the tolerance limit will be retained and the line is redrawn. (See fig. 2.)

Douglas and Peucker (1973: 116) state that while Lang's algorithm produced acceptable results on relatively smooth curves, it did not "detect the best representative points on sharp curves." They also note that it requires too much computer time for on-line processing systems. Lang also identified two disadvantages: pen movement is extremely slow except at large tolerances, and the total amount of drawing time is longer at smaller scales (Lang 1969: 1).

Jenks (1981) introduces another tolerancing algorithm based on a user-specified parameter. First, a line segment tolerance limit is set. Then, using a triad of coordinate pairs, a vector is calculated from the first to the third coordinate pair. The perpendicular distance between the second point and the vector connecting points one and three is then calculated. If the perpendicular distance between point two and the vector is larger than the tolerance limit, point two will be retained, because it is essential to maintain the character of the line (McMaster 1983: 75). If the distance between point two and the vector is less than the tolerance limit, point two will be eliminated and the triad will be advanced forward one step along the linear feature. (See fig. 3.)



A. Point 2 is above the tolerance limit.



B. Point 2 is retained as a salient/critical point.

Figure 3. Application of perpendicular tolerance algorithm as derived by Jenks.

McMaster (1983) describes a simplification algorithm that he obtained from Jenks through personal correspondence. This algorithm is based on working with a triad of coordinate pairs. First, a tolerance angle is established. Next, two vectors are constructed; the first vector connects point one with point two and the second vector connects point one with point three. Then, the angular change between the two vectors is calculated. If the calculated angle is greater than the tolerance angle, point two will be retained. This process is repeated, moving one step forward along the line.

Jenks (1983) cites an additional algorithm which he describes as a modification of the angular algorithm, using two parameter checks. In this algorithm, three tolerances are determined by the user. The first tolerance, MIN1, delimits the minimum allowable distance from point one to point two. The second tolerance, MIN2, determines the minimum allowable distance from point one to point three. Finally, the third tolerance, ANG, establishes the maximum allowable angle of change between the two vectors; point one-point two and point one-point three. (See fig. 4.) Jenks defines the modified angular algorithm (McMaster 1983: 47-48):

> If (1) the distance from one to two is less than MIN1, or (2), the distance from one to three is less than MIN2, point two is rejected. If both are larger, the angular check is calculated using ANG. An angle smaller than ANG will result in the removal of point two.

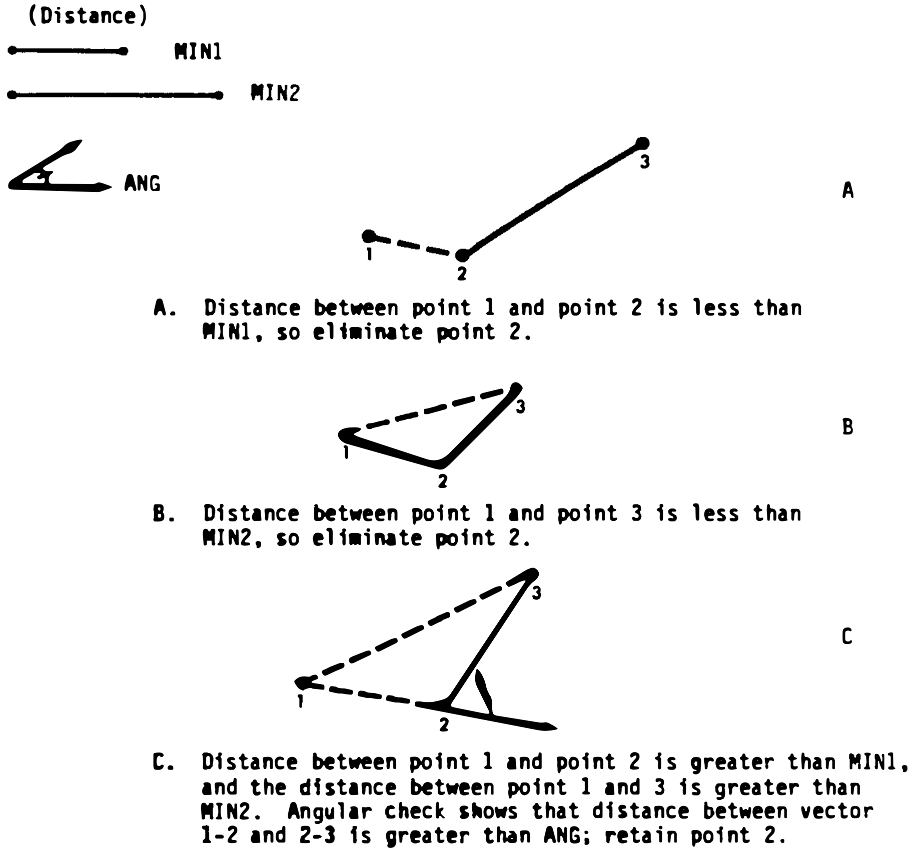


Figure 4. Jenks MIN1, MIN2, and ANG algorithm.

The third category of sequential point elimination algorithms pertains to the use of corridor/search areas. Basically, corridor/search areas are defined by either parallel line segments, rectangles, or circles which, as they are laid along the linear feature, eliminate points that fall within their predefined regions.

In 1974, Reuman and Witkam developed a corridor/search area using a point elimination algorithm entitled the "segmentation method." It is based on the use of two parallel lines whose distance apart is user specified. In essence, the two parallel lines are placed over the linear feature in "the direction of its initial tangent, until the end collides with the curve" (Opheim 1982: 34). The point where one of the parallel lines of the search area intersects with the linear feature is then retained as a characteristic point of the line. In addition, the intersection point becomes the starting point for the next search by laying the two parallel lines along the tangent of the remaining part of the linear feature. The entire process is then repeated until the end of the line is reached. (See fig. 5.)

Although this algorithm is considered one of the fastest methods for linear simplification, it does have disadvantages. First of all, Opheim (1982: 38) has found that it does not effectively handle sudden bends in the curvature of the line, thereby eliminating critical points along the line. Secondly, the choice of the direction tangent to the linear feature is not well calculated, where a straight line is drawn between the last two fix points and used to derive the direction. Although the method is simple, Opheim (1982: 38) states, "a natural choice could be the tangent of the curve through the last fix point as search area." Lastly, Opheim (1982: 34) argues that "two parallel lines to infinity usually gives bad results". He recommends that the distance between points retained should not exceed a predetermined distance (dmax) nor be closer than a specified minimum distance (dmin).

Rhind, in 1973, described a method defining a circular search area in place of two parallel lines (Opheim 1982: 35). The boundary of the circle would be specified by the user, including both a maximum radius and a minimum radius. When using the circle in the same procedure described by Reuman and Witkam, Opheim (1982: 35) found it to be a moderately fast method which deleted only those points, "which are of little interest; i.e., those along a median path." (See fig. 6.) The disadvantage of this algorithm is, that when applied to data collected by a "time-based" digitizer, it tended to suppress bends in the curvature of the line.

Opheim (1982: 35) recommends a routine for corridor/search area, point elimination which he calls a "blend of the search regions of Reuman-Witkam and Rhind." This method, originated by Skappel, uses both dmin, dmax and two parallel lines--it is as if one cut out a section from Rhind's circle. (See fig. 7.) Choosing the right parameters for dmin, dmax and the distance between the two parallel lines depends on, "the type of curve, the purpose of the data reduction,...and the user's own judgment" (Opheim 1982: 36).

When placed over a linear feature, as in the Reuman-Witkam routine, all points that lie within dmin are eliminated (too close to the initial point), and the last point within the search region is selected as the "critical" point to be retained. Unfortunately, the problem with this routine is that when the curvature of the line makes a sudden bend inside the search region, the "critical" points of the bend will be eliminated.

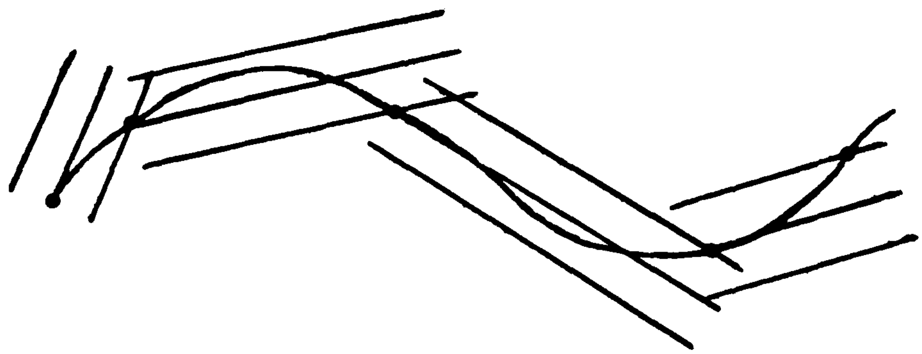


Figure 5. Reuman and Witkins corridor/search routine.

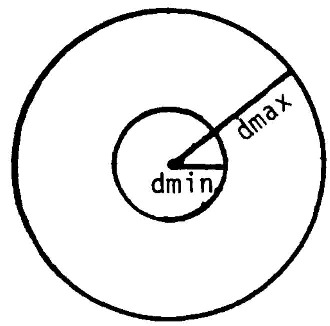


Figure 6. Rhind's circle-search area routine.

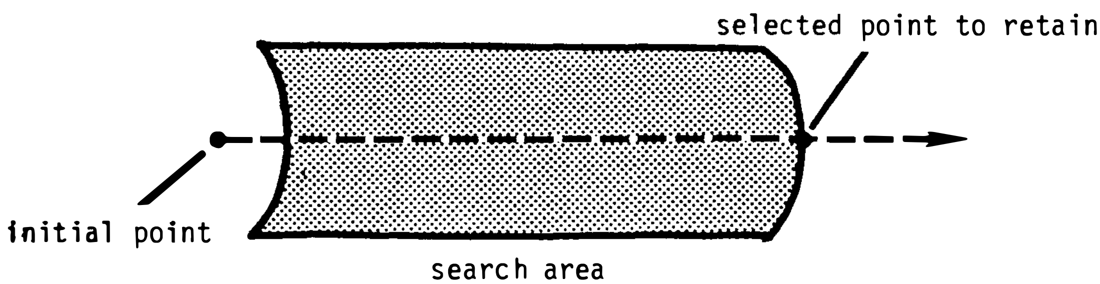


Figure 7. Opheim's corridor/search routine.

Global Point Elimination

The most important aspect of global point elimination is the fact that it is the method which comes the closest to imitating manual simplification techniques. This method looks at the whole linear feature at one time, rather than piece by piece (as in routines emphasizing sequential point elimination). Cartographers approach line simplification in the same fashion; they first consider the character of the entire line before drafting a generalized version. Three routines have been developed that perform global point elimination: method one, method two of Douglas-Peucker's (1973) point reduction algorithm, and Peucker's (1975) "band-width" algorithm.

Douglas-Peucker's algorithm begins by defining a maximum tolerance distance in the form of a line segment (Douglas-Peucker 1973: 116):

> Method one begins by defining the first point on the line as the anchor and the last as a floating point. These two points define a straight segment. The intervening points along the curved line are examined to find the one with the greatest perpendicular distance between it and the straight line deemed by the anchor and the floater. If this distance is less than the maximum tolerance distance, then the straight segment is deemed suitable to represent the whole line. In the case where the condition is not met, the point lying furthest (sic) away becomes the new floating point. As the cycle is repeated the floating point advances toward the anchor.
>
> When the maximum distance requirement is met, the anchor is moved to the floater and the last point on the line is reassigned as the new floating point. The repeat of this later operation comprises the outer cycle of the process. The points which had been assigned as anchor points comprise the generalized line.

In method two, the operation is exactly the same as in method one. However, all points assigned as floaters are recorded by stacking them in a vector. "After the anchor point is moved to the floating point, the new floating point is selected from the top of the stack, thereby avoiding the necessity of re-examining all the points between the floater and the end of the line" (Douglas-Peucker 1973: 117). (See fig. 8.)

The advantage of method two over method one is that it takes only 5 percent of the computing time required for method one, and, "is thought to produce better caricatures" (Douglas-Peucker 1973: 117). However, it does result in the selection of a greater number of points than method one.

Peucker (1975: 511) devises another line simplification algorithm initially conceived as an afterthought to the Douglas-Peucker algorithm. In this algorithm he introduces the concept of "band-widths," wherein given a certain general direction of the line, the band will become a bounding rectangle. "...the sides and ends of the band are parallel, and perpendicular, respectively, to the general direction (of the line), totally enclosing it" (Peucker 1975: 511). Point elimination is conducted by partitioning the line into subsets, "until each subset is a band with a width less than a predetermined threshold." (Peucker 1975: 511). "At each step, the partitioning process is performed by selecting

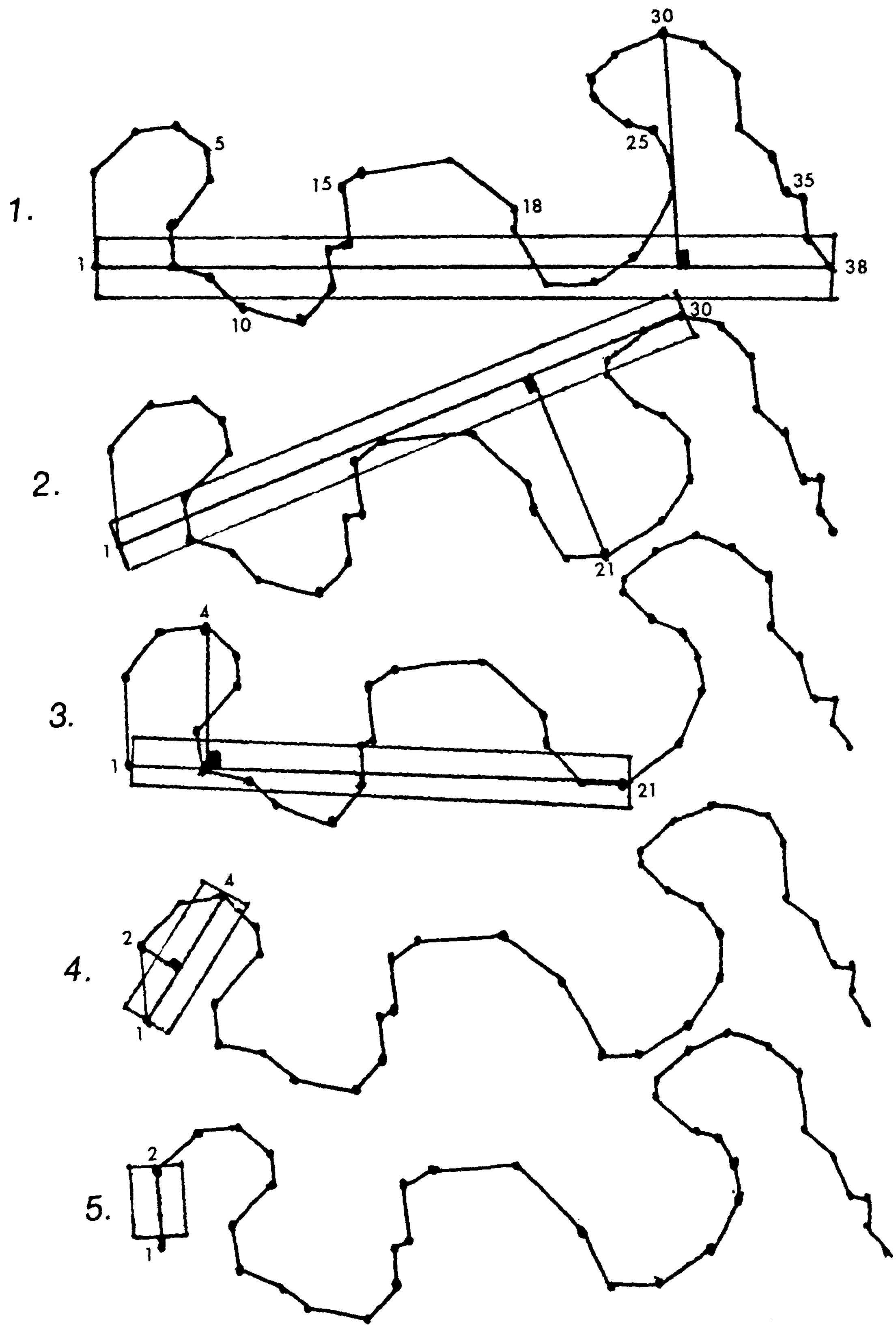
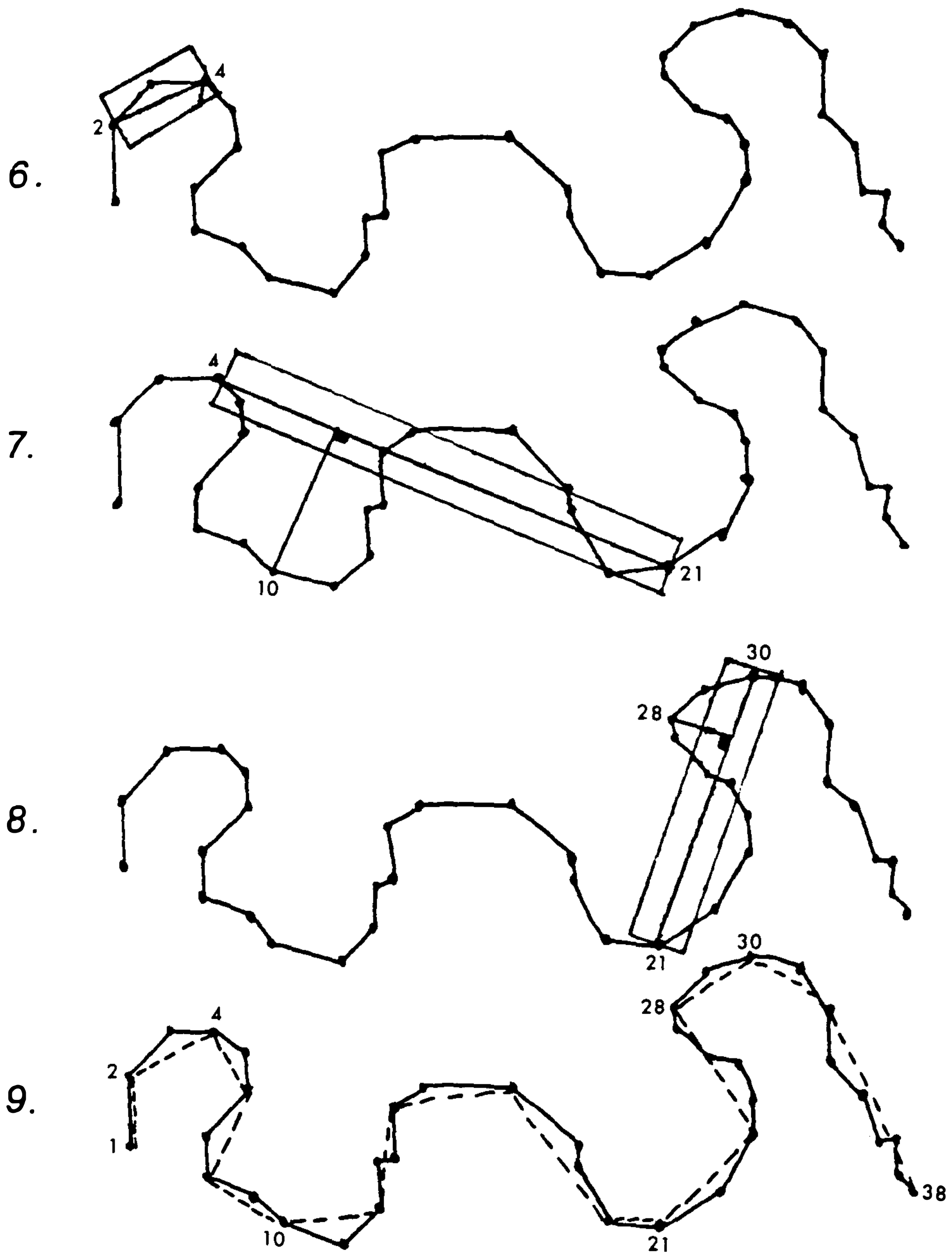


Figure 8. Douglas-Peucker's point reduction routine (McMaster 1983: 78-79).



Phase 9 indicates the final simplification.

those points which touch the sides of the bands as starts and ends of the subsets. In order to partition the line, a link is drawn between the beginning and ending points of the line. This link describes the general direction of the line. "For every point, the vertical distance form the link is computed and the point with the maximum absolute value is retained as the new point that divides the line into two portions which are subsequently treated independently the same way" (Peucker 1975: 513). The process is repeated until the maximum distance between the original line and the general direction line is less than a predetermined value. (See fig. 9.) Marino (1978: 9) argues that the advantage of this algorithm is that it, "provides a quantitative expression for the characteristics of the elements of a line, as well as objective rules for their selection."

MEASURING SIMPLIFICATION ERROR

McMaster (1983: 268) formulates two methods of evaluating the efficiency and accuracy of several simplification algorithms. These methods were designed to measure the displacement between a digitized line and its simplification. Displacement is defined as the "result of the simplified line no longer lying in the same geographical position as those sections of the original line, or a shift in the true geographic location." The most accurate simplification algorithms will produce as few displacements as possible, and "space these displacement errors evenly along the line."

Areal displacement and vector displacement are the two measures used by McMaster to judge the effectiveness of four types of simplification algorithms: N-th point (Tobler), angular tolerance (Jenks), perpendicular distance tolerance (Lang, Jenks), and the Douglas-Peucker point reduction algorithm. Areal displacement is the polygonal (areal) difference between the original line and its simplification. (See fig. 10.) This difference is observed in several ways: positive differences (polygons to the left side of the original line), and negative differences (polygons to the right side of the original line).

Vector displacement is measured as the "perpendicular distance from the eliminated coordinate on the base line (original line) to the new vector on the simplified line" (McMaster 1982: 268). The results of McMaster's measurements indicates that in spite of the degree of simplification, the Douglas-Peucker algorithm has less areal and vector displacement. On the other hand, the angular tolerance algorithm was found to have the most areal displacement at all levels and the most vector displacement at most levels. Results of measuring areal and vector displacements of the N-th point algorithm and the perpendicular distance algorithms were difficult to conclude. It was found that the perpendicular distance algorithms were superior to a level of 60 to 70 percent of the coordinates eliminated. However, at a more "rigorous simplification," the lines for both perpendicular distance algorithms and the N-th point algorithm had equivalent amounts of displacement. In conclusion, McMaster ranks the algorithms as follows: 1) Douglas-Peucker, 2) perpendicular tolerance algorithms, 3) N-th point, and 4) angular distance algorithms.

Opheim (1982: 39) argues that the Reuman-Witkam algorithm for the corridor/search area is "a faster method than the Douglas-Peucker." However, he acknowledges that the Reuman-Witkam algorithm is not as careful as the Douglas-Peucker method when retaining salient points along a curve. He also notes that costs escalate for the Douglas-Peucker method when the routine

"A" shows the concept; "B" the actual practice.

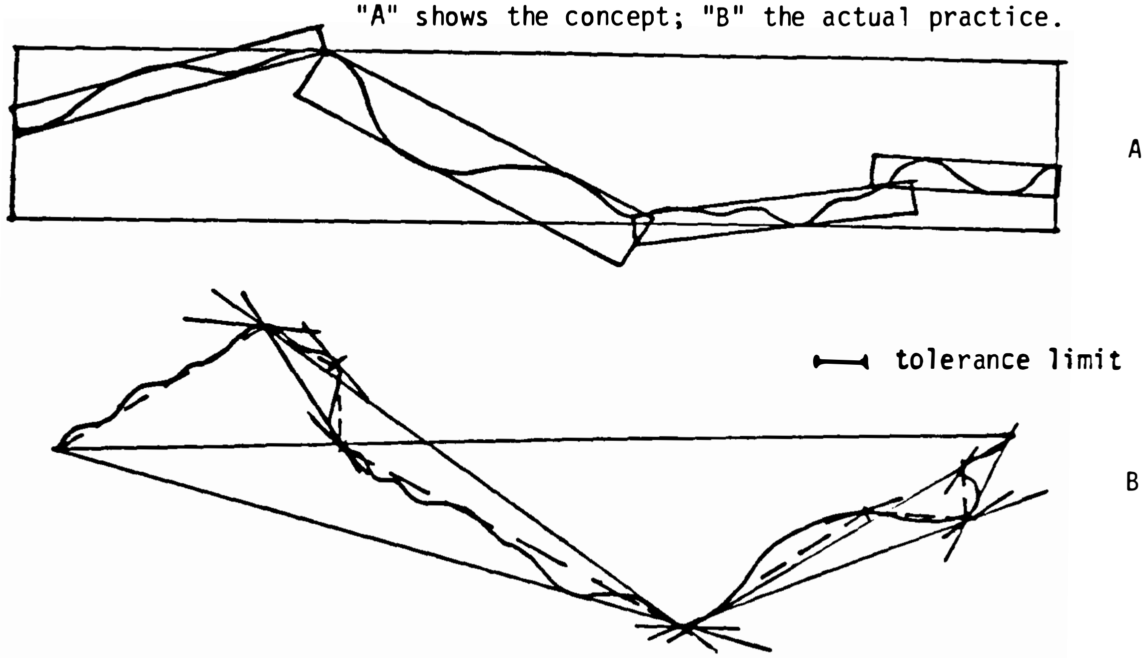


Figure 9. Peucker's "band-width" algorithm.

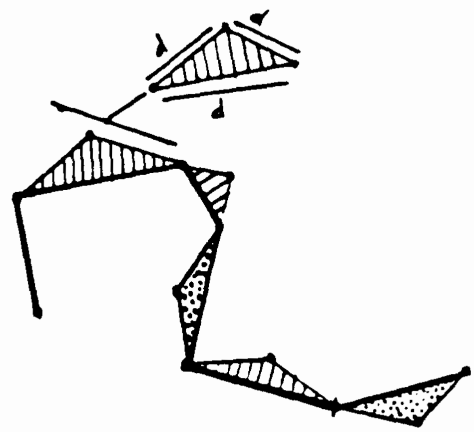


Figure 10. McMaster's areal displacement.

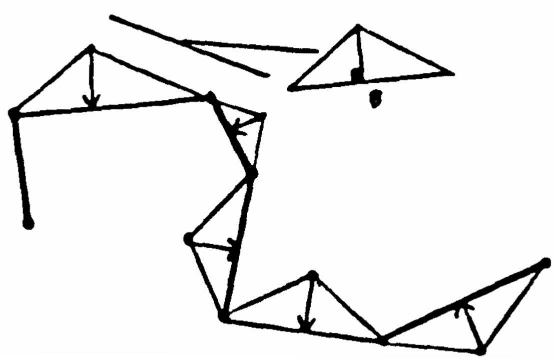


Figure 11. McMaster's vector displacement.

attempts to handle too many points at one time. In such a case, Opheim recommends that the linear data set be divided into several small sets.

Opheim (1982: 33) then suggests that a combination of his corridor/search area with that of Douglas-Peucker, "will result in a powerful tool for data reduction." In this fashion, the combination of algorithms will provide the capability of retaining salient points along curved lines and will "perform as well or even better than the Douglas-Peucker method (alone) and at a lower cost." Unfortunately, dividing the linear data set into smaller sets, in order to run the Douglas-Peucker algorithm at a lower cost, defeats the purpose of the global routine. It then becomes just another sequential point elimination routine which does not follow the same approach as manual generalization techniques. Although Opheim's recommendation may produce a cost-effective solution to point eliminination, it does not bridge the gap between computer-assisted cartography and traditional cartographic methods.

AN INTERACTIVE, LINEAR SIMPLIFICATION ALGORITHM

Brophy (1973: 300) introduces an algorithm that allows for direct control by the cartographer as well as the capability of performing both feature elimination and systematic point elimination. In this algorithm, he examines both generalization theory and mathematical theory. In addition, the algorithm utilizes both objective and subjective approaches. It is objective in the sense that mathematical concepts are used along with procedural techniques, and subjective because the routine relies upon user-specified input parameters such as scale, line, width, and feature size.

Brophy maintains that "cartographic generalization cannot be completely removed from the cartographer's control" (1973: 303). Therefore, his algorithm ALIGEN is made of six components that allow the cartographer to select from a variety of generalization options. Component one provides for the elimination of smaller features. Here the selection of features to be removed along a line is on the basis of a specified minimum portrayal size. This component determines the spacing of coordinate pairs along the line according to "a multiplicative function of the scale change between the source map and the generalized map, the line weight change between these maps, and the level of generalization selected for the generalized map...all specified by the cartographer." Component two redefines the new curve created in component one by connecting the points "as a series of tangent points of finite width equal to the line weight of the line on the generalized map" (Brophy 1973: 304). Component three is optional and allows the cartographer to specify control points along the linear feature, thereby assuring that those points are not eliminated on the generalized line. Component four is optional, and permits the cartographer to select larger features that should be eliminated, resulting in further simplification of the line. Component five provides for systematic line smoothing and exaggeration of particular features. Finally, component six plots each line after generalization.

In analyzing Brophy's algorithm, Morrison (1975: 106) writes, "although it is complex, it runs rather efficiently and gives reasonable results." Further, Morrison states that in choosing between Douglas-Peucker's algorithm and Brophy's algorithm, one must consider the fact that the Douglas-Peucker algorithm does not remove features along a line (it results in dark, heavy lines). It tends to retain a higher coordinate density around curves; whereas Brophy's algorithm will remove unwanted linear features. In addition, Brophy's algorithm "appears to have a firmer base in cartographic theory" (Morrison 1975: 106).

Therefore, the advantages and disadvantages of choosing between the two algorithms must be weighed. The user must realize that Brophy's algorithm is an "interactive" process, whereas the Douglas-Peucker method has the user select the tolerance limit only once (thereby narrowing the user's input to one action in the beginning of the process). Also, Brophy's algorithm does not work with the entire linear feature at one time (it is an algorithm that eliminates points). Ultimately, the question of which is the better algorithm relies on the cost of operating the program at the specified facility, and, therefore, lies outside of the realm of the algorithm itself.

LINEAR FEATURE SMOOTHING

Smoothing is the process of reproducing linear features by removing a series of points along a digitized line and replacing them with new coordinate pairs. Basically, smoothing algorithms "simplify lines by diminishing variations in direction and reducing angles" (White 1983: 26). In other words, smoothing algorithms are used to give a digitized line a "smoother" look.

What distinguishes smoothing algorithms from simplification algorithms is that smoothing algorithms take original (digitized) coordinate pairs and transform them into new coordinate pairs using mathematical functions, thereby changing the original data set by moving point locations. Simplification algorithms on the other hand eliminate coordinate pairs from the original data set, but do not change the location of those coordinates selected to remain on the simplified line.

Smoothing algorithms are most often employed to minimize digitizing irregularities, particularly when the scale of the map is reduced. However, "smoothing techniques appear to be most appropriate for data sets already reduced in size, in order to provide a more 'natural' looking line" (White 1983: 26). Four smoothing algorithms will be discussed next: Koeman and Van der Weiden's "moving means," Tobler's "weighted means," Boyle's "forward look," and Perkal's "circle" algorithms.

Koeman and Van der Weiden (1970: 47) developed a smoothing routine based on the concept of structural generalization. They define structural generalization as "the change in linear map elements by reduction in scale, maintaining the original form of the lines" (1970: 47). In essence, their algorithm incorporates the use of moving means, whereby the average value of a series of coordinate pairs is assigned to the center point of the string of coordinate pairs. Thus, one coordinate pair (the average) replaces the entire series of coordinate pairs. For example, values of the first two coordinate pairs along a line are averaged. These pairs are then replaced by their average value and the routine moves to the next two coordinate pairs along the line, etc.

Koeman and Van der Weiden (1970: 48) conducted their smoothing process four times on a single line, using a series of 2, 3, 5, and 12 successive coordinate pairs. They concluded that the technique gives good results when "the ratio between the distances between points recorded on straight or smoothly curved lines, and the distance between points on very irregular lines is not too large." The process tended to lose curves when the averaging was conducted between relatively large distances along smooth curves and small distances along rough curves. The advantages of this routine are: rapid production rates, line quality that could not be maintained manually, and all uses of original digitized data.

The following disadvantages were noted: the greater the number of points the more costly and complex the operation, and the algorithm was "inclined to depress the effect of extreme points, often the very points which gave character to the line" (Marino 1978: 7).

In 1966 Tobler devised a smoothing algorithm that used weighted, moving means to smooth the line. The equation he described pertained to three dimensional data sets, but it could be adjusted to handle two dimensional data sets as well. Unfortunately, Tobler did not provide further information on how the two-dimensional equation should be formed. However, Tobler described the process of weighted, moving averages as that which "assigns particular 'weights' to each point in the calculation, in order to increase (or decrease) its influence on the final point position" (White 1983: 27). The process is similar to moving averages, but has the additional property of applying weights to each coordinate pair. For instance, if large weights are assigned evenly to each neighboring coordinate pair, the amount of simplification would be small. In such a case, the center point would not move too far and the character of the line would not be severely affected. If, however, large weights are given to extreme points then they will have the effect of "pulling" the center point in their direction. In addition, Tobler describes two advantages of this method: large-scale features are retained, and small-scale features are filtered out.

Boyle (1970 :91) introduced a different approach to smoothing linear features which he calls "forward look linear interpolation." He believes that when using a digital computer the problem is describing the line, "to the required accuracy in a compact numerical manner and then...regenerate it in the original form... (where the) drawn output has to be smoothed in a manner suitable to the type of line and acceptable to the eye."

Forward look linear interpolation, then, is conducted along a line by aiming at a point that lies either two or more points ahead. Boyle (1970: 94) describes the following example, "each time one-quarter of the distance of a four-point look ahead, or one-tenth of a ten-point look ahead, is completed, a new aiming point is created on a logical basis." (See fig. 12.) One advantage of this process is that the output is a "series of vectors that vary only by a small angle from one to the next." However, according to Boyle, the major attribute of this routine is that its visual appearance is pleasing to the eye and it appears to work well with continually varying cartographic line data. Coastlines may be sufficiently reproduced using a four-point look ahead. Water contours are better portrayed using a ten-point look ahead (for sharp contours will not appear in the smoothed output).

Perkal 1966 introduces the most comprehensive and objective approach to linear feature smoothing. The degree of generalization for this method is established by the assignment of a real number ε to represent the diameter of a circle. (See fig. 13.) The object of this routine is to generalize "a region (D) by placing the circle (of diameter ε) inside the region and rotate it in such a manner that the circle always lies completely inside the area (D), it is never outside the area (D)." The term ε-generalization applies to the set of all points (p) that have the property that they are contained within the circle of diameter ε which can be completely included in the region (D). The boundary formed by ε-generalization set of points is then the smoothed boundary for D. Those points that are not covered by the edge of the circle are eliminated (q).

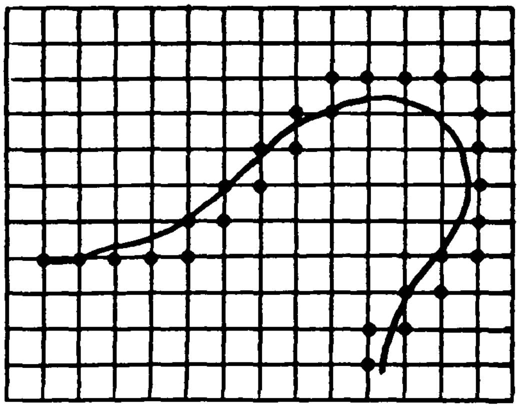


Figure 12. Boyle's "forward look" routine. This example portrays a four-point ahead aim.

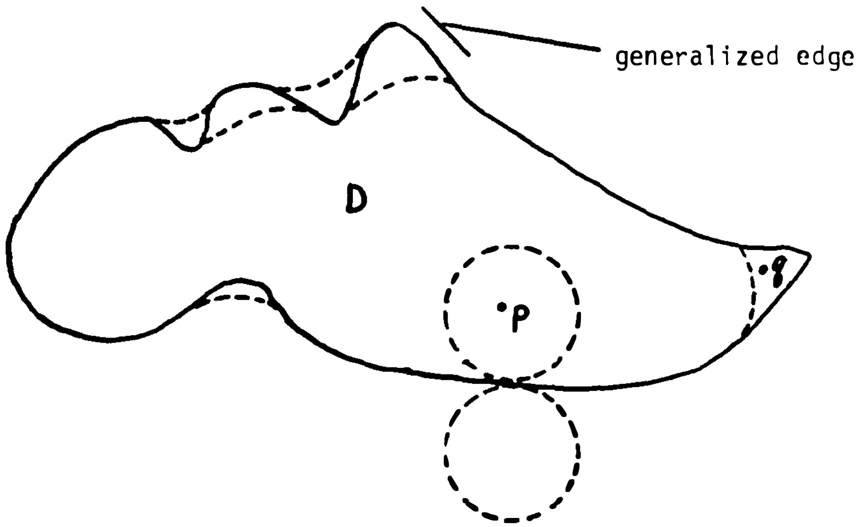


Figure 13. Perkal's "circle" routine.

In addition to the "inner" line generalization boundary of area D, an outer line generalization boundary is created of the "complement of area D" (D'). This complement (D') refers to the area, "of the region that remains after deletion of D" (Perkal 1966: 5). Perkal's smoothing process has two phases: moving the circle (of diameter ε) inside the region D, and moving the circle outside the region D. The generalized edge will be the region of divergence, i.e., the shaded region between D and D'. By using a large circle, a greater area of divergence can be created and, therefore, a greater generalization.

The distinguishing feature between Perkal's smoothing routine and that of Boyle's routines is that it is used for enclosed features. The method is capable of identifying regions of the map that should be eliminated due to their limited size. Most cartographers maintain "in reduction to scale of small details, a cartographer should comprehend all of the peculiarities necessary to describe the type of line and place them on a map at even the smallest scale (Perkal 1966: 3). However, this contributes to map clutter and complexity. It is more beneficial for the cartographer, when working at different scales, to have the ability to identify areas that are too small to portray clearly on the map. Perkal provides this option by allowing the user to change the diameter of the rotating circle, where "the disappearance of an inner area at a given level of generalization can be taken as a reason for omitting the given element from a map..." (Perkal 1966: 9). In such a case, the epsilon diameter of the circle is so large that the circle cannot rotate within the boundary of the region. While Perkal does not specify appropriate circle diameters to use at particular scales, his approach to feature smoothing brings computer-assisted cartography one step closer to complete automation. Once the cartographer determines the diameter of the generalizing circle, the remainder of the program does not require his intervention.

SUMMARY AND CONCLUSIONS

The linear simplification algorithms discussed in this report were evaluated upon two criteria: 1) were they objective approaches to simplification, and 2) did they retain critical/salient points along the line? These were considered to be the most important criteria to address.

Objectivity is necessary in order to "automate" the linear simplification process. The biggest difference between computer-assisted and manual cartography lies in their treatment of cartographic elements. Manual cartography is affected by the influence of individual bias in the performance of generalization since every cartographer has a certain way of doing things. However, to have computer assistance in performing simplification it is necessary to reduce bias "by systematically searching for impartial criteria that can be universally applied" (McMaster 1983: 32). This can only be attained by minimizing the amount of user intervention, e.g., allowing the user to specify certain standardized criteria only in the initial stages of the simplification process. An objective approach, then, is required in any linear simplification algorithm.

Retention of critical points assures that the true portrayal of the line's "character" is provided as the cartographer conducts linear simplification. When conducting manual simplification cartographers choose similar "critical" points to characterize the line (Marino 1978). It is important to transfer this quality over to computer-assisted simplification algorithms. However, the difficulty is that most simplification algorithms do not look at the entire line at one time. Algorithms that only look at small sections of the line at a time have difficulty in identifying abrupt curvature changes which often signify critical points. Therefore, the simplification algorithms recommended by most cartographers and geographers are those that look at the whole line at one time, i.e., global simplification algorithms.

Peucker's "band-width" algorithm fits the above mentioned criteria because it closely resembles the Douglas-Peucker point reduction algorithm, method one. However, Douglas-Peucker's method two algorithm (where floating points are stacked in vectors) better suits the requirements of objectivity and retention of critical points. While it selects more points than method one, it works faster and often produces better caricatures. This algorithm then would be a good choice for "automating" linear simplification. The user must weigh the attributes and disadvantages to select the appropriate linear simplification algorithm to suit the specific purpose.

Finally, a distinction needs to be made regarding feature simplification. Feature simplification in this report is equivalent to the phrase "feature elimination," for the process is aimed at selecting cartographic features that will remain on the generalized map. The process of "smoothing" boundary lines of features is then a linear smoothing process and not a simplification process.

Robinson and Morrison's "ranking" of data elements, or Perkal's "circle" algorithm, are suitable feature simplification algorithms. Perkal's algorithm, though, is the more straightforward approach. It allows for the elimination of features by specifying an ε-diameter circle measure that is not small enough to rotate within the boundary lines of the feature. Once the diameter is specified user intervention is no longer required, whereas, assigning "rankings" to data elements on a map can prove to be tedious and may involve frequent user intervention. Although the concept is simple, the practice of assigning "rankings" is complicated. Ranking is an arbitrary process, which means that an element of subjectivity is involved. Therefore, if a standard diameter can be derived, so that for each scale an appropriate measure can be chosen, feature simplification would best be performed by using Perkal's "circle" routine.

BIBLIOGRAPHY

Boyle, A. R., 1970: The quantised line. The Cartographic Journal, 7(2), 91-94.

Brophy, David Michael, 1973: An automated methodology for linear generalization in thematic cartography. Proceedings of the ACSM, 33rd Annual Meeting, Washington, D.C., 300-314.

Douglas, David H. and Peucker, Thomas 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.

Eckert, Max, 1908: On the nature of maps and map logic. Bulletin of American Geographic Society, 40(6), 344-351.

Hershey, A. V., 1963: The plotting of maps on CRT printer. Naval Weapons Laboratory Report 1844, U. S. Navy, Dahlgren, Virginia.

International Cartographic Association, 1973: Multilingual Dictionary of Technical Terms in Cartography, Franz Steiner, Verlag GMBH (Wiesbaden, Germany).

Jenks, George F., 1979: Thoughts on line generalization. Proceedings of the International Symposium on Cartography and Computing (Auto-Carto IV), American Congress on Surveying and Mapping and American Society of Photogrammetry, 209-220.

Jenks, George F., 1981: Lines, computers and human frailties. Annals of the Association of American Geographers, 71(1), 1-10.

Jenks, George F., (Department of Geography, University of Kansas, Lawrence) 1983: (personal communication with Robert McMaster).

Keates, J. S., 1973: Cartographic Design and Production. John Wiley and Sons, New York, N.Y.

Koeman, C. and Van der Weiden, F. L. T., 1970: The application of computation and automatic drawing instruments to structural generalization. The Cartographic Journal, 7(1), 47-49.

Lang, T., 1969: Rules for robot draughtsman. Geographical Magazine, XLII (1), 50-51.

Lundquist, Gosta, 1959: Generalization--a preliminary survey of an important subject. The Canadian Surveyor, 14(10), 466-470.

Marino, Jill S., (University of Kansas, Lawrence) 1983: Characteristic points and their significance in cartographic line generalization. M.S. thesis (unpublished).

McMaster, Robert, 1983: A mathematical evaluation of simplification algorithms. Proceedings of the Sixth International Symposium on Automated Cartography (Auto-Carto VI), 11, 267-276.

McMaster, Robert, (University of Kansas, Lawrence) 1983: Mathematical measures for the evaluation of simplified lines on maps. Ph.D. dissertation (unpublished).

Morrison, Joel L., 1974: A theoretical framework for cartographic generalization with emphasis on the process of simplification. International Yearbook of Cartography, vol. 13, 59-67.

Morrison, Joel L., 1975: Map generalization: theory, practice and economics. Proceedings of the International Symposium on Computer-Assisted Cartography (Auto-Carto II), 21-25 September 1975, U.S. Department of Commerce, Bureau of the Census and American Congress on Surveying and Mapping, 99-112.

Opheim, Harold, 1982: Fast data reduction of a digitized curve. Geo-Processing, 2, 33-40.

Pannekoek, A. J., 1962: Generalization of coastlines and contours. In Imhof, Edward (ed.), International Yearbook of Cartography, vol. 2, 55-75.

Perkal, Julian, 1966: An attempt at objective generalization. Translated by W. Jackowski from Julian Perkal, "Proba obiektywnej generalizacj," Geodezja i Kartografia, tom VII, zeszyt 2 (1958). In Nystuen, John (ed.), Michigan Inter-University Community of Mathematical Geographers, Discussion Paper No. 10. University of Michigan, Ann Arbor, MI.

Peucker, Thomas K., 1975: A theory of the cartographic line. Proceedings of the International Symposium on Computer-Assisted Cartography (Auto-Carto II), 21-25 September 1975, U.S. Department of Commerce, Bureau of the Census, and the American Congress on Surveying and Mapping, 508-518.

Reuman, K. and Witkam, A.P.M., 1974: Optimizing curve segmentation in computer graphics. International Computing Symposium, North-Holland, Amsterdam, 467-472.

Rhind, D. W., 1973: Generalization and realism within automated cartographic systems. The Canadian Cartographer, 10(1), 51-62.

Robinson, Arthur H., Sale, Randall, and Morrison, Joel L., 1978: Elements of Cartography, (4th edition). John Wiley and Sons, New York, N.Y.

Srnka, Erhart, 1970: The analytical solution of regular generalization in cartography. In Frenzel, Konrad (ed.), International Yearbook of Cartography, vol. 10, 48-62.

Steward, H. J., 1974: Cartographic generalization: some concepts and explanations. Cartographica Monograph, No. 10, University of Toronto Press, Toronto, Canada.

Tobler, Waldo, 1965: Automation in the preparation of thematic maps. The Cartographic Journal, 2(1), 32-38.

Tobler, Waldo, 1966: Numerical map generalization. In Nystuen, John (ed.), Michigan Inter-University Community of Mathematical Geographers, Discussion Paper No. 8. University of Michigan, Ann Arbor, MI.

Topfer, F. and Pillewizer, W., 1966: The principles of selection. The Cartographic Journal, 3(1), 10-16.

Traylor, Charles T., (University of Kansas, Lawrence) 1979: The evaluation of a methodology to measure manual digitization error in cartographic data bases. Ph.D. dissertation (unpublished).

White E. R., (University of Oklahoma, Norman) 1983: Perceptual evaluation of line generalization algorithms. M.S. thesis (unpublished).