National Center for Geographic Information and Analysis

**National Center for
Geographic Information and Analysis**

The Comparison of Complexity Measures for Cartographic Lines

by

Marcin J. Jasinski
National Center for Geographic Information & Analysis
Department of Geography
State University of New York at Buffalo
Buffalo, New York 14260

Technical Report 90-1

January 1990

(RE/ I-3)

PREFACE AND ACKNOWLEDGEMENTS

This report represents a project submitted to the Department of Geography of the State University of New York at Buffalo in partial fulfillment of the requirements for the degree of Master of Arts.

I would like to gratefully acknowledge the support and guidance of my advisor Dr. Barbara Buttenfield. I would also like to thank Dr. David Mark for his expertise and willingness in this effort. Thanks are also due to Dr. Keith Clarke who first introduced me to the world of computer cartography.

This project comprises a portion of the efforts of Research Initiative #3, "Multiple Representations", of the National Center for Geographic Information and Analysis, supported by a grant from the National Science Foundation (SES-88-10917); support by NSF is gratefully acknowledged.

ABSTRACT

The concept of 'line complexity' is often used in cartography in relation to geometric shape of linear features. The concept is particularly important in automated line generalization. Many measures have been applied over the years in an attempt to describe the shape of irregular one-dimensional features in a quantifiable manner. All of them are narrowly focused on single characteristics of complexity, and moreover they remain in an unknown relationship to one another. . This project reviews methods and techniques of capturing the complexity of lines. Further, it explores the existing relationships between eight measures of geometric shape of lines. The measures selected for the analysis involve such characteristics of shape as density of detail, length, angularity, and fractal dimension. The analysis performed on an especially designed test data set, comprising of both natural and cultural features, involves principal components analysis. The recommendations for future research on the topic are presented.

TABLE OF CONTENTS

LIST OF FIGURES

1. INTRODUCTION.

1.1 Introduction.

This project is a product of research conducted under the auspices of Initiative 3 "Multiple Representations" of the National Center for Geographic Information and Analysis. The goal of discussions and research sponsored by Initiative 3 is to determine the main impediments to automatically providing multiple views of the same cartographic database, and draw immediate and future research directions to overcome the obstacles. One of the main impediments involves automated generalization of spatial data. The question of how to redraw maps with change of scale or purpose of a map became a challenging problem in computer environment. The main stumbling block lies with the development of a rule base which would make it possible to translate the elusive steps of generalization into computer- understandable form. To achieve that, a better understanding is necessary of both the process of generalization and of the nature and geometry of cartographic features.

Work described here analyzes geometric properties of very irregular linear features at different levels of resolution. The initial hypothesis was that some of currently used measures of lines geometry are statistically redundant. By looking at what is common and what is different in the measures, it should be possible to determine a quantifiable definition of 'line complexity', and to discover its main components. A more extensive introduction to the problem is presented on the following pages. It is followed by a critical - review of research published on the subject, and the description of the work conducted for this study.

1.2 Problem Statement,

The advent of computer technology has revolutionized much of geography, and cartography in particular. It has changed the way maps are produced in terms of automation of drafting and plotting, and substantially affected cartographic methodology. The availability of new, very powerful means of handling and displaying spatial data on one hand raises hopes for complete automation of the map creation process, and on the other, made people realize that the new environment requires reformulation of cartographic rules of design. The pressure to automate the map-making process comes from two directions. One is Geographic Information Systems, for which the map is the front end and the main form of output (Weibel and Buttenfield, 1988). A second impetus comes from traditional map production environments. The spread of GIS packages and their use by people coming from various backgrounds, often with little or no geographic/cartographic training, raises concerns about the ease with which cartographic monstrosities may now be created. The need for automation in GIS has two sides: one is to bar a naive user from creating nonsensical maps, and the other, to facilitate an expert user in creating sophisticated output.

One of the main components of automated map design is map generalization. The problem of automating that process has currently become the most challenging issue in analytical and computer cartography. The process of generalization has to be performed either when we want to change the scale of the map, or when there is a need to change the purpose, and thus the content of the map. One can distinguish two general approaches to providing a user with multiple views at various levels of generalization of a digital database in a map form. One has some features of a 'brute force' type of approach, although with an important trait which is, that it works. This approach assumes either existence of separate databases for every scale or theme of representation or view or, if only one database exists, the elements that are to be retained at smaller scale representations have to be explicitly flagged in the existing database. This means much of redundant storage of information and also lack of flexibility, as the data can only be viewed at predefined scales at which it is represented in the database.

The alternative approach is more elegant and more difficult to operationalize. It involves automatic generation of required representations from a single database containing data at the largest scale available. The first approach may well be functional for institutions such as US Geological Survey which publish map series at only a few scales and are not interested in changing that in foreseeable future. For a GIS user, on the other hand, not being able to view the data at any suitable resolution for a given analysis, would be too much of an unnecessary limitation. Thus, the necessity of providing means for generating multiple scale views from a single database seems obvious.

In the realm of manual cartography, the process lies in the hands of a map designer who makes decisions regarding the sub-steps of generalization accordingly to experience, cartographic training and aesthetic taste. The process is holistic in nature, and its steps are interdependent. The translation and reformulation of generalization process into a machine understandable form has proven to be an extremely difficult task. The main thrust in cartographic research in automated generalization has gone into analytical dissection of the process in order to identify all possible components (operators) of generalization (Shea and McMaster, 1989). This has been done in hope that once all the steps are defined, they could be translated into algorithms, and their joint performance would simulate work of a manual cartographer. The assumption is that a formal rule base for cartographic generalization could be created in this way. But the process of generalization in the. manual environment is essentially intuitive and is governed by heuristics rather than strict rules. For instance, the steps of simplification, smoothing, displacement and merging may be made with one movement of cartographer's scribing pen. The problem of synthesizing that process in digital environments remains a challenge to be tackled.

Much of the research that has been conducted on cartographic generalization has been concerned with linear features. The reason is that lines can be used to describe most of the features which are present on maps. Points can be considered as degenerate lines, and areas can be described by their boundaries which are linear (interesting discussion of some exceptions to this assumption is provided by Mark and Csillag, 1989). There are two directions in research on line generalization. The more widespread assumes that cartographic lines are geometric constructs and that generalization parameters should be derived from the lines' geometric characteristics. A second approach says that one should not divorce a line from what it actually represents, so that a given piece of line should be only generalized as a specific type of feature e.g. fjord type of coast, meandering stream, interstate highway, etc. (Mark, 1989). Both these approaches are discussed in detail in the next chapter. Following the latter trend would require, on one hand, creating an extensive look-up table containing an exhaustive set of all- possible geographic features that a map pan depict with corresponding generalization parameters. On the other hand, it would require either developing methods of automated feature identification or flagging the features in the database as being of a specific type. Although it is beyond dispute that this approach would most closely simulate the process taking place in human mind while generalizing, the amount of effort required to create such an automated system would be gargantuan. For that reason most of the research effort concentrates on development of algorithms which treat lines as strings of coordinates and process them with regard to their geometric characteristics.

All of this is particularly pertinent to line simplification, the process that leaves out unnecessary details and preserves those points in a line which are considered most important. Simplification is considered successful if the simplified caricature of a line closely resembles the original version. A number of measures to evaluate simplification have been developed (McMaster, 1986; White, 1985). Quite a number of algorithms can be applied in simplifying a given line, and each approaches the problem from a different standpoint. Some look at intersegment angles in a line, and preserve those points where the angle of change is greater than some specified value. Others preserve the points which are most distant from the anchor line (the segment connecting the first and last points of a line). Others 'walk' the line looking at a few points at a time and select those where the line changes its direction most visibly. In general, all methods look for those places in a line where the change in information content occurs, which is compared against some pre-specified tolerance value.

The drawback of all such algorithms is that the tolerance value must be determined prior to the simplification process and remains constant during the run of an algorithm. Whether or not the shape of a line changes during its course, the tolerance parameter stays the same. It can only be changed interactively by stopping the process and tweaking the parameters. One can easily think of many cases, particularly with large data sets, where it would be desirable for the simplifying algorithm to be able to adjust itself to the changes in lines' shape.

Let's take for example a map of the Atlantic coast of the United States. To reduce the scale of a map, one of the necessary operations would be simplification of the coastline. Employing a single tolerance value would mean treating the Florida coast the same way as the rugged shores of Maine. Clearly, an algorithm which could recognize the parts of line which contain more intricate details and accordingly respond by lowering the tolerance parameter, and correspondingly adjusting it for other stretches, would be able to improve the product of simplification over a standard algorithm by carrying over to the coarser resolution the essential geometric characteristics of the original line. Such an approach should help further automate the generalization process of large data files containing lines of various levels of complexity.

One of the keys to further progress down that path (and one of the objectives of this research) lies in extending our knowledge and understanding of geometric shape of irregular cartographic lines. All of the briefly mentioned methods of automated simplification take one geometric characteristic of a line into account and go on with the process based on that single type information. Many researchers have called for a more comprehensive measure of the shape of lines, which could help to automatically distinguish between lines and compare, in a quantifiable way, lines of various complexity (Buttenfield, 1987; McMaster, 1986).

Existing measures of line shape are narrowly focused on, for instance, angular variation of a line, or its deviation from a straight line, or frequency of detail. Each provides a partial description of a line which is in a very fuzzy relation to other, similar descriptions, resulting from different measures. High angularity of a line does not tell anything about the frequency of turns, and vice versa. As an example of the effort to find such a. comprehensive measure of line complexity may serve the fact that many cartographers (Muller, 1987; Eastman, 1985; Shelberg at al., 1982; Carstensen, 1989) have attempted to' employ in that role the fractal dimension of lines. As a dimensionless measure relating the change in length to the scale of measurement, fractals have seemed to carry a lot of promise. One of the goals of this work is to compare fractal measures of linear features with traditional geometric measures.

Another objective of this project is to bring us closer to understanding what of complexity. The two terms, line geometric shape and line complexity seem to be used interchangeably (Dutton 1981; McMaster, 1986; Buttenfield, 1989). They came to be sort of catchall phrases with lack of clear definition of what they actually describe. The term complexity does not seem to be as neutral as the term geometric shape, and carries certain amount of preconception that the subjects' shape is not simple.

The current stage of knowledge of cartographic lines complexity resembles a situation where one is trying to peep into a closed room with many doors. The only means of seeing what is inside is by looking through the key holes, through which some shapes of what is in the room can be distinguished. But even the combination of all limited views from all of the key holes does not provide one with a better sense of what is really in there. The aim of this research is not perhaps to kick the doors wide open, but at least to drill bigger peeping holes and establish links between their views.

In the remainder of this paper, the focus of discussion will be on comparison of measures of lines' shape. The purpose is to search for commonalties and redundancy in measures applied to lines. Chapter 2 will review previous work relevant to the subject. Chapter 3 will contain the description of research design performed in the project. Chapter 4 will discuss the results of the comparative analysis on the measures of line complexity. The last chapter will draw the conclusions stemming from the analysis and present some future research recommendations.

2. REVIEW OF LITERATURE.

2.1 Overview.

Very little has been written about line complexity per se. In geographic literature the term is usually used as a synonym to 'shape of line' and is often mentioned with regard to topics tangential to the core of this project. The term complexity (but not line complexity) is often used in cartography with regard to visual effectiveness of a map, particularly in connection to the problem of spatial conflict stemming from crowding of a map content. There are four main areas in published research to which the concept of line complexity is relevant. The distinction between these research directions has been dictated by the wealth of existing literature around each of them, rather than by clear differences in area of interest between them. The first three can be encompassed by the broad term of generalization and they represent various approaches to, or steps in solving the problem. They are: 1. Line simplification and enhancement, 2. Feature identification, 3. Use of fractals in line generalization. The remaining area is: 4. Development of shape indices.

2.2 Line Simplification and Enhancement.

The objective of the line simplification is to reduce the number of points which describe the line. The process should be performed in such a way that the characteristics of the line are preserved and it is easily recognizable as a caricature of the original (Douglas and Peucker, 1973). The process may take the form of both statistical (reduction of volume of data) and cartographic (for display purposes) generalization (Brassel and Weibel, 1988). The former is applied when there is a need for reduction of the number of coordinates in a data file for storage or efficiency purposes, the latter when it is necessary to reduce the amount of detail in a line, as in scale reduction. The objective for appropriate simplification is to decide which points are to stay and which are disposable for a given level of generalization.

According to information theory, points carrying maximum information are those with which the largest amount of change is associated. Work by psychologist Frank Attneave (1954) introduced the idea of characteristic points, that is those carrying most information for preserving a recognizable shape. Research by Kelley (1977, as cited by Buttenfield, 1989) showed that the points considered by a number of subjects as critical for recognition of shape were characterized by high angular change.

Jenks (1989) points out that in the cartographic environment there are two types of characteristic points. There are points which carry purely geometric information, which relate to the shape of line, and there are points which have an important meaning (geomorphic, political) attached to them. Examples of the latter may be points constituting such features as San Francisco Bay, or Hel Peninsula on the Baltic. In 1979 Marino conducted an experiment using a group of subjects consisting of both cartographers and non-cartographers in order to find out whether both groups would consider the same points in a line as the most characteristic. The results showed very high positive correlation between the answers. Moreover, certain logical hierarchy in selection of points was observed. The points considered critical in very crude representation were also present, along with others, in less simplified stages. This fact allows one to conclude that application of a line simplification procedure which selects characteristic points will produce a recognizable caricature of the original line.

A review of line generalization algorithms published by Zycor corporation lists 24, grouped into 9 categories (Zycor, 1984). They are: selection, low-pass filtering, angle detection, DEM smoothing, tolerance bands, point relaxation, domain transformation, mathematical filtering and epsilon filtering. This rather complicated grouping was based on the predominant mathematical technique employed. A different classification was proposed by McMaster (1989), based on the extent of line that is taken into consideration by an algorithm at a time. The classification is included in Table 2.1.

The classification does not distinguish between the algorithms as far as what geometric characteristics of a line they are based on. The most simplistic algorithms, unfortunately still much used due to their computational efficiency, are nth or random point removal. The more sophisticated ones consider geometric properties most such as angular change (Jenks,1981; Deveau,1985), density of points, segment length (Johannsen,1974), perpendicular deviation from some line connecting other points (Douglas -Peucker, 1973; Lang,1969), or band width (Reumann-Witkam,1974; Opheim,1981). An algorithm- specific tolerance is based on such geometric properties, which when compared against during the run of an algorithm serves as the elimination criterion. A detailed description of individual algorithms can be found in McMaster (1986, 1989).

TABLE 2.1. A CLASSIFICATION OF ALGORITHMS USED TO SIMPLIFY DIGITIZED LINES (McMaster, 1989)

[CATEGORY 1]: INDEPENDENT POINT ALGORITHMS

Do not account for the mathematical relationships with the neighboring coordinate pairs; operate independent of topology.

Examples: nth point routine
random-selection of points

[CATEGORY 2]: LOCAL PROCESSING ROUTINES

Utilize the characteristics of the immediate neighboring points in determining selection/rejection.

Examples: distance between points
angular change between points
Jenks' algorithm

[CATEGORY 3: CONSTRAINED EXTENDED LOCAL PROCESSING ROUTINES

Search beyond 'immediate' coordinate neighbors and evaluate sections of the line. Extent of search depends on distance, angular, or number of points criterion.

Examples: Lang algorithm
Opheim algorithm
Johannsen algorithm
Deveau algorithm
Roberge algorithm

[CATEGORY 4]: UNCONSTRAINED EXTENDED LOCAL PROCESSING ROUTINES

Search beyond 'immediate' coordinate neighbors and evaluate sections of the line. Extent of search is constrained by geomorphological [sic] complexity of the line, not of algorithmic criterion.

Example: Reumann-Witkam algorithm

[CATEGORY 4]: GLOBAL ROUTINES

Considers the entire line, or specified line segment, in processing. Interatively [sic] selects critical points.

Example: Douglas algorithm

Many measures have been developed to assess how well a simplified version of a line approximates the original. More popular are vector displacement and areal displacement methods. The first looks at perpendicular distances (vectors) between the points eliminated and the simplified line, the second sums up the area of all 'spurious' polygons created when the simplified version and the original one are overlaid (McMaster, 1986). Research done by White (1985) and by McMaster (1986) indicate that the Douglas-Peucker algorithm has the best overall performance in terms of minimizing the above mentioned error measures. One of its important traits is that selection of points is highly correlated with human selection of characteristic points (White, 1985).

The step in the line generalization process that usually follows simplification is enhancement (Buttenfield, 1984; Dutton, 1981). The most widespread type of that is linear smoothing, which is frequently required to enhance the visual appearance of the simplified line. To use an image processing analogy, smoothing behaves as a low-pass filter by evening out the jags and sharp angles and creating a more continuously flowing line. A review of smoothing algorithms may be found in McMaster (1989). The alternative methods of smoothing a line in digital environments are moving averages techniques, epsilon filtering or mathematical approximation method s (McMaster, 1989). Moving averages look at n consecutive points at a time and adjust the position of the point considered accordingly to the relative position of its neighbors. The idea of epsilon filtering originated with the work of Julian Perkal (1958b). Based on that idea, although modified, was the first algorithmic implementation proposed by Brophy (1973). The idea is to 'roll' a circle of a given radius (epsilon) along the line. The points which are covered or touched in the process are retained, the other are eliminated. Thus, this technique may also serve as a simplification method. The mathematical approximation techniques in smoothing include spline functions and Bezier curves.

A very different approach to line enhancement was proposed by Dutton (1981) who attempted to add, instead of remove, detail to a line to improve the realism of its looks. This method involved measuring the fractal dimension of a line and then inserting detail into it in a way so that the new one would b e self-similar to the original. Dutton implemented self-similarity by standardizing intersegment angles along the extent of the coordinate files. It should be noted that his understanding of self-similarity is not identical to Mandelbrofs definition, which is discussed later in this chapter.

2.3 Feature Identification.

One of the still unsolved problems in line generalization is an objective and appropriate selection of tolerance values in line simplification and/or smoothing algorithms. As discussed in the previous chapter, existing algorithms require one to set an arbitrary tolerance prior to processing a line. Important work in direction of automating this aspect of generalization was reported by Buttenfield (1986, 1987). The aim of her research was to establish a methodology for automated identification of the graphic structure of a cartographic line. The proposed technique is based on breaking a line into its trend and features which bifurcate from it. The trend is described as "the smoothest possible approximation" (Buttenfield, 1987) of a line, and features include the remaining information in the line. Once the components of a line are isolated, geometric characteristics of features are measured and statistically categorized using clustering techniques. This procedure is used to create a look-up table of such structural measures. Any given line can be similarly 'dissected' and its structural measures compared against the look-up table to assign the line an appropriate "structural label" (Buttenfield, 1987), which guides the selection of tolerance value for generalization. Although the four test lines which were used in the project represented very different geomorphic features, the look-up table containing the set of line graphic structures cannot be yet considered complete and similar testing on a wider set of lines still needs to be done.

In the earlier study Buttenfield (1986) attempted to develop structure signatures for, again, a set of geomorphically distinct lines. This was achieved not by separating trend from features, but by examining the changes in line geometry with changing scale of graphic representation. Lines were put into a strip tree structure (Ballard, 1981), which for a given level contained a different representation in terms of simplification. For each level of the tree various geometric measures were computed, and they were the base from which structure signatures were constructed.

2.4 Use of Fractals in Line Generalization.

The idea of relating line length to the scale of measurement precedes fractal concepts of Mandelbrot. In 1894 Penck was wondering about lack of clear mathematical relationship between the progression of map scales and the length of the same linear feature measured on them (from Perkal, 1958a). Hugo Steinhaus published a paper in 1949 in which he observed a paradox of accuracy in cartometric measurements. The paradox is that the finer measurements of line's length are, the greater its length becomes. In view of that, the question becomes whether it makes any sense to talk about the true length of a river or a coastline. Whether the length of a feature was surveyed with 1 foot accuracy or 100 feet, the results are both correct, and dependent upon the units of measurement. It should be noted that this is not the problem of measurement error, which will be present at any level of accuracy. Steinhaus' ideas were elaborated by Perkal (1958b) who created an "epsilon longimeter", which provides a way of measuring the length of an empirical curve. It is achieved by rolling a circle with given radius (epsilon) along the empirical curve and calculating the length as the length of the band created around the curve. The length becomes the function of epsilon, which is the required accuracy level. Richardson (1961) came to similar conclusions as the Polish mathematicians on the difficulty of defining the length or area of geographical features, while studying the relationship between international political conflicts and location of countries and the shape and length of their borders. He observed that on the plots of the total length of a feature as a function of the 'measuring rod' in logarithmic space, a near-linear relationship can be found.

For Richardson, that finding was of little consequence, but it was used by Mandelbrot (1967, 1982) in his estimation of fractal dimension D as (1 - s), where s is the slope of the line from Richardson's plots. If we had only two measures of line's length with two different ruler sizes rl and r2, we may obtain the length of the line by multiplying the rulers by the number of steps n1 and n2. Then D can be derived as

$$D = \log (n2/n1) / \log(r1/r2)$$

Mandelbrot (1967) coined the term fractal to indicate that many spatial and temporal phenomena may have dimension in between the topological ones. Thus, fractal lines can have a dimension between 1 and 2 (1 -- straight line, 2 -- space filling curve). Fractal lines are not differentiable, as they can always be split into smaller parts showing more detail (Burrough, 1985). One of the basic properties of fractal lines is that they are self-similar, that is independent of scale: a subset is un-resolvable from the whole, as the shapes propagate across the ranges of resolution. As applied 1 to natural phenomena, Mandelbrot (1982) relaxed this constraint to statistical self- similarity, so that not exact shapes but statistical properties stay constant across the scales. Fractal geometry has received considerable attention in geography. In 1987 Goodchild arid Mark said:

"(...) the concepts of fractional dimension and dependence of measure on scale (...) may (...) offer the first effective tools for understanding the irregularity widely observed in the geometry of real phenomena. (...) the numerical value of D may be the most important single parameter of an irregular cartographic feature, just as the arithmetic mean and other measures of central tendency are often used as the most characteristic parameters of a sample."

Fractals aroused interest, on one hand as a new way of formally parametrising geometric shape of irregular features, and on the other as a simulation tool, which can be used for generating synthetic lines or landscapes, which can be easily controlled by a few parameters. A review of applications of fractals to geographical problems, as well as methods of estimating fractal dimension is given by Burrough (1985). One of the more popular methods of estimating fractal dimension is from Richardson plots. An algorithm for this, so called, 'walking divider' method was published by Shelberg et al. (1982). In the context of line generalization fractals were used by Muller (1987), Longley and Batty (1989a, 1989b), Carstensen (1989), and Dutton (1981). Muller proposed that "fractal dimension (...) [be used] as a guiding standard for the automated generalization of statistically self-similar geographic lines". A detailed discussion of I walking divider' method of estimating fractal dimension, as well as applicability of fractal measures in line generalization is provided in the following chapters.

It should be noted here that fractal concepts received also its share of criticism. Krantz (1989) wrote recently: "One notable difference between fractal geometry and calculus is that fractal geometry has not solved any problems. "(p. 14). He goes on to state: " (...) the hypotheses and conjectures that the fractal people generate are (like the objects which . they study) self-referential. One generates the pictures to learn more about the pictures, not to attain deeper understanding" (p.15). Although this critique may be harsh, there is some truth to what Krantz is suggesting. It seems that many people jumped on the 'fractal band wagon' without prior thinking about what is it that they really want to explain by using them.

There seem to be two different problems with the use of fractals in cartography. On one hand there is the problem of how good and reliable the current methods of estimating the fractal dimension of a line are. The other one is more fundamental in nature, and it has to do with the question of whether geographic features are at all self-similar, even in the statistical sense, and what kind of insights one may obtain from that. Buttenfield (1989) in her paper on scale dependence and self-similarity discusses that issue, and describes a method of determining the points at which lines geometry changes, and ranges of scale for which self-similarity becomes apparent.

2.5 Shape Indices.

There is a wealth of literature on shape analysis and on development of shape indices in urban and regional geography, as well as in spatial statistics. All of that work is exclusively concerned with two-dimensional shapes and thus is not directly applicable to the topic at hand. Some general ideas though which were developed in that area which are pertinent to this research. Moellering and Rayner (1981), while talking about development of shape indices, state what conditions ideal shape index should fulfill. Such indices should depict an object in such a way that it could be reconstructed from the indices only. Further, shape indices should not be affected by scale or coordinate system used. A thorough review of work done in that area is provided by Austin (1984).

An early example of a shape index is the radial vector measure proposed by Boyce and Clark (1964). The measure was based on the proportionate contribution of each of 16 or more radii drawn from the middle of the figure to its border, to the sum of all. That was then standardized by comparing to the radius of a circle circumscribed on the figure. The limitations of the measure lie in various interpretations of what should be considered a center of a figure (geometric or functional), as well as dependence on the number of radii selected, and their orientation. Examples of other such indices may be measures of oblongity, relative spread, convexity and relative dispersion of territories developed by Bachi (1973). Another measure which would look at compactness of a polygon was proposed by Bunge (1966). Some moment measures derived from weighted or un-weighted center of gravity were proposed by Massarn (1970, 1975). In more recent literature much attention has been given to harmonic analysis such as the Fourier transform (Moellering and Rayner, 1981, 1982).

Many shape measures, such as the Boyce-Clark index, elongation ratio, circularity ratio, compactness ratio, and other were used by Lo (1980) in a study which attempted to relate changes in the shapes of Chinese cities to the changes in their populations. A rigorous statistical examination of Lo's results, coupled with similar tests for Canadian cities, led Griffith et al. (1986) to the conclusion that application of shape indices does not lead to any substantive insights to the problems in urban and regional geography, and that they are "red herrings rather than useful measures".

2.6 Summary

This chapter has reviewed literature on topics either related to, or with interest in line complexity. In the broad spectrum of cartographic generalization, especially pertinent to the topic at hand are the areas of line simplification and enhancement, feature identification, application of the concept of fractal dimension. Also some ideas from a tangential field, which is shape analysis in cultural geography, were presented. It should be clear from this review that the need for better understanding of geometric properties of cartographic lines is very much present. The question of how to capture the essential characteristics of the shape of lines, or their complexity, in quantifiable terms still remains to be answered.

In the next chapter, the discussion of the logic behind the research design of the project is presented, which consists of the overview of the analysis part, creation of the test data set, selection of the appropriate measures of geometric shape of lines, and methodology of the analysis. The following chapters include data analysis and discussion of the results.

3. RESEARCH DESIGN.

3.1 Overview.

The goal of the project is to further knowledge and understanding of the complexity of cartographic line features. It was hoped that new insights into the measurement of line complexity could be gained through comparison of a number of existing measures on a set of test lines. The measures which have been used in the literature all look at a single aspect of line's shape, such as angularity or density of points. There is no simple way to make reasonable comparisons between those measures and to know how they are related to one another.

The analysis was performed on a data set consisting of a few basic lines representing various types of geometry which can be found on maps, and nine levels of progressively simplified caricatures of the original lines. Each simplified version can then be considered as a view of the original line at coarser resolution. It is difficult to make inferences about the measures themselves if they are tested on various irregular features, as the features are in unknown relation to each other. Having the progressions of simplification provides one with a set of lines with clear relationship which is controlled by parameters in line simplification algorithm. The intention is to make comparisons primarily between the measures within each of the groups containing representations of a basic line and look for any existing trends there. Only after that, the cross-group comparisons would be made. As the final stage of the analysis, the principal components analysis would be performed on each of the groups in order to examine the structure of correlations between the measures.

The two subsequent parts of this chapter provide description of two distinct stages in the design of the project, including creation of the test data set, and selection of the measures of line complexity. The analysis will be described in Chapter 4.

3.2 Creation of the Test Data Set.

A set of four lines was selected for the project. In order for the set of lines to be representative of a wide range of shapes, it was decided to include both - natural and cultural features. For obvious reasons natural features tend to be more irregular, and cultural linear features, such as roads, rail tracks, or political boundaries, tend to consist of straight segments and arcs of a more standardized geometry. Further, the lines must be roughly at the same scale to facilitate comparisons between them, and should have at least 1,000 coordinate pairs, so that the most simplified caricatures would still preserve the character of the original lines.

The two natural features selected include the coastline of Mount Desert Island in Maine (Figure 3.1), and a section of the Klamath River (Figure 3.2) in Northern California from its bifurcation with Scott river to its mouth. Mount Desert Island is a mountainous island whose relief was mainly sculptured by the last glaciation. The shoreline is very intricate, with numerous little coves and peninsulas, and also the island is almost cut in half by the only fjord on American Atlantic coast. Thus, a variety of both low and high level detail is present. The line was digitized from a map at a scale of 1:25,000 published by DeLorme Publishing Company. Klamath River was chosen as a typical young stream with characteristic high directional change. It was digitized from 1:62,500 USGS 15 minute quadrangle topographic map series. These lines therefore describe natural features running perpendicular to and parallel to their respective fall lines of the terrain.

The cultural features selected for the study represent two roads. One is a part of Federal Highway US 101 extending from the eastern town boundary of Port Orford, Oregon to 41000'N (Clam Beach) in Northern California (Figure 3.3). The road generally closely follows the coastline and in that respect shares some similarity with Mount Desert Island in its relationship to the local fall line (contour-like character).The line was also digitized from 1:62,500 USGS 15' quadsheets. The second road came from the Swiss Alps, from the region close to the French border, near towns of Martigny and Orsieres (Figure 3.4). The idea was to have an extremely convoluted road which would test the limits of complexity of man made features. As it proved impossible to find a road which would preserve its character long enough so that the lower threshold (1,000) of the number of coordinate pairs would be reached, it was decided to use three roads of visually similar complexity and piece them together. All of them came from the same Swiss 1:50,000 topographic map. The resulting road feature may be considered analogous to the Klamath River feature as they both follow the fall line instead of following contour lines in the terrain.

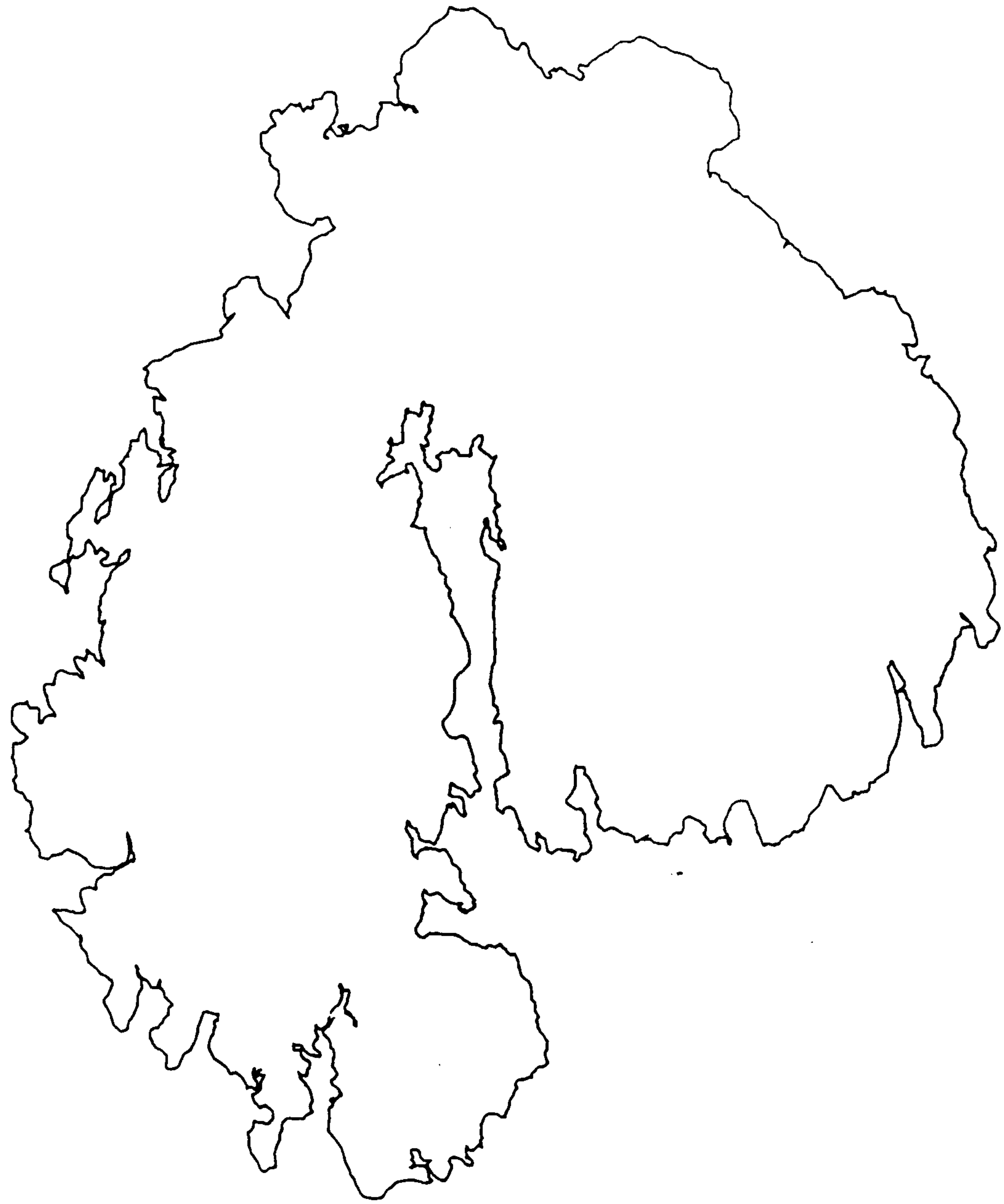


Figure 3.1 Mount Desert Island (2,257 points).
Original scale of representation 1:50,000.

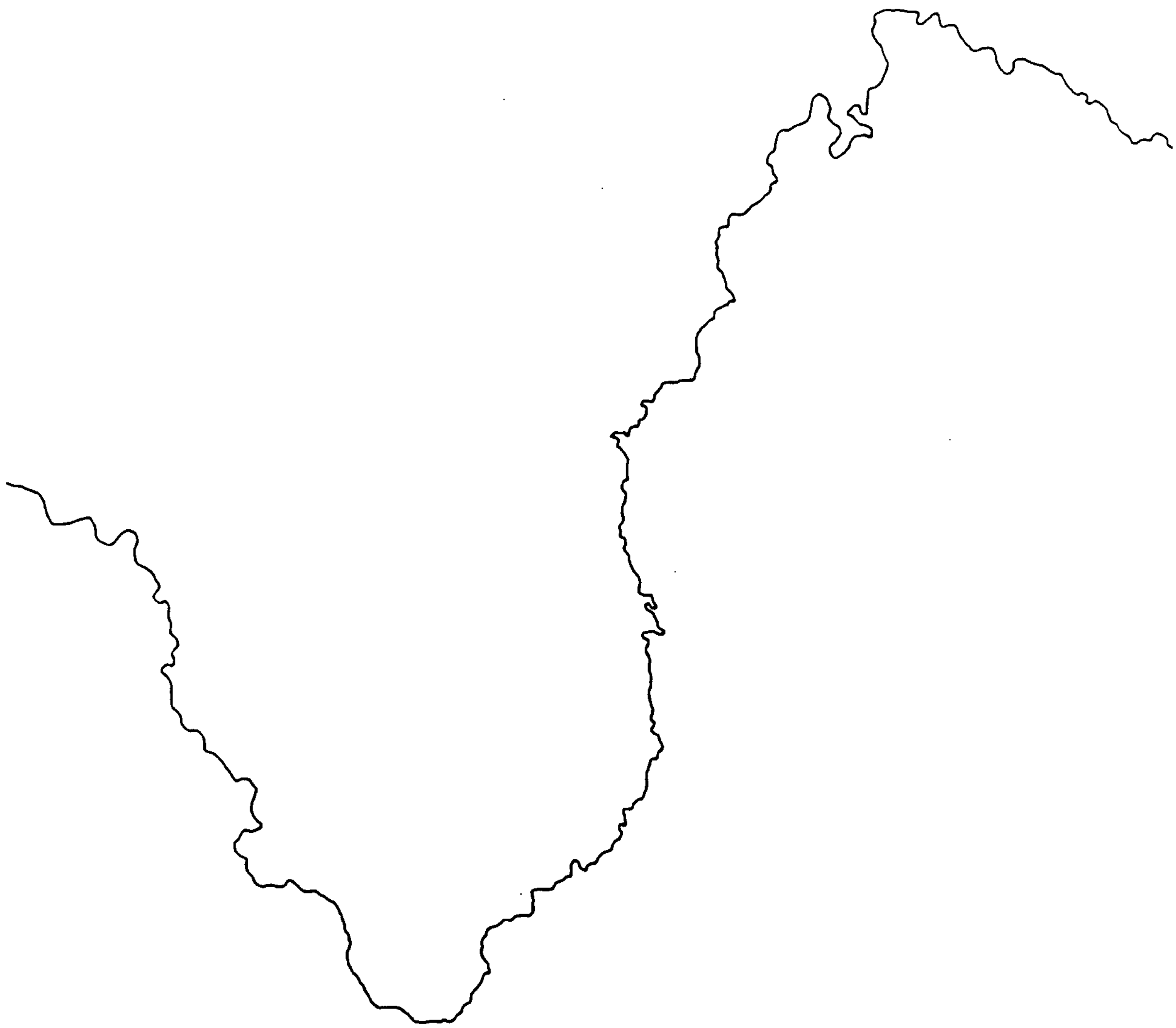


Figure 3.2 Klamath River (1,206 points).
Original scale of representation 1:62,500.

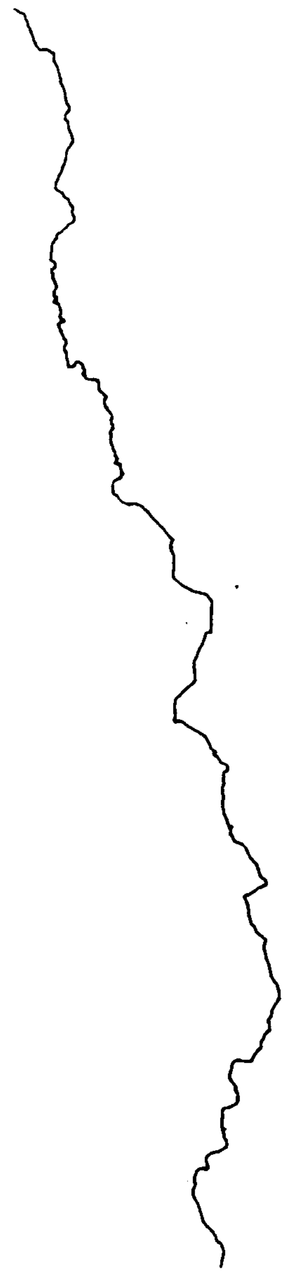


Figure 3.3 Highway US101 (1,218 points)
Original scale of representation 1:62,500.

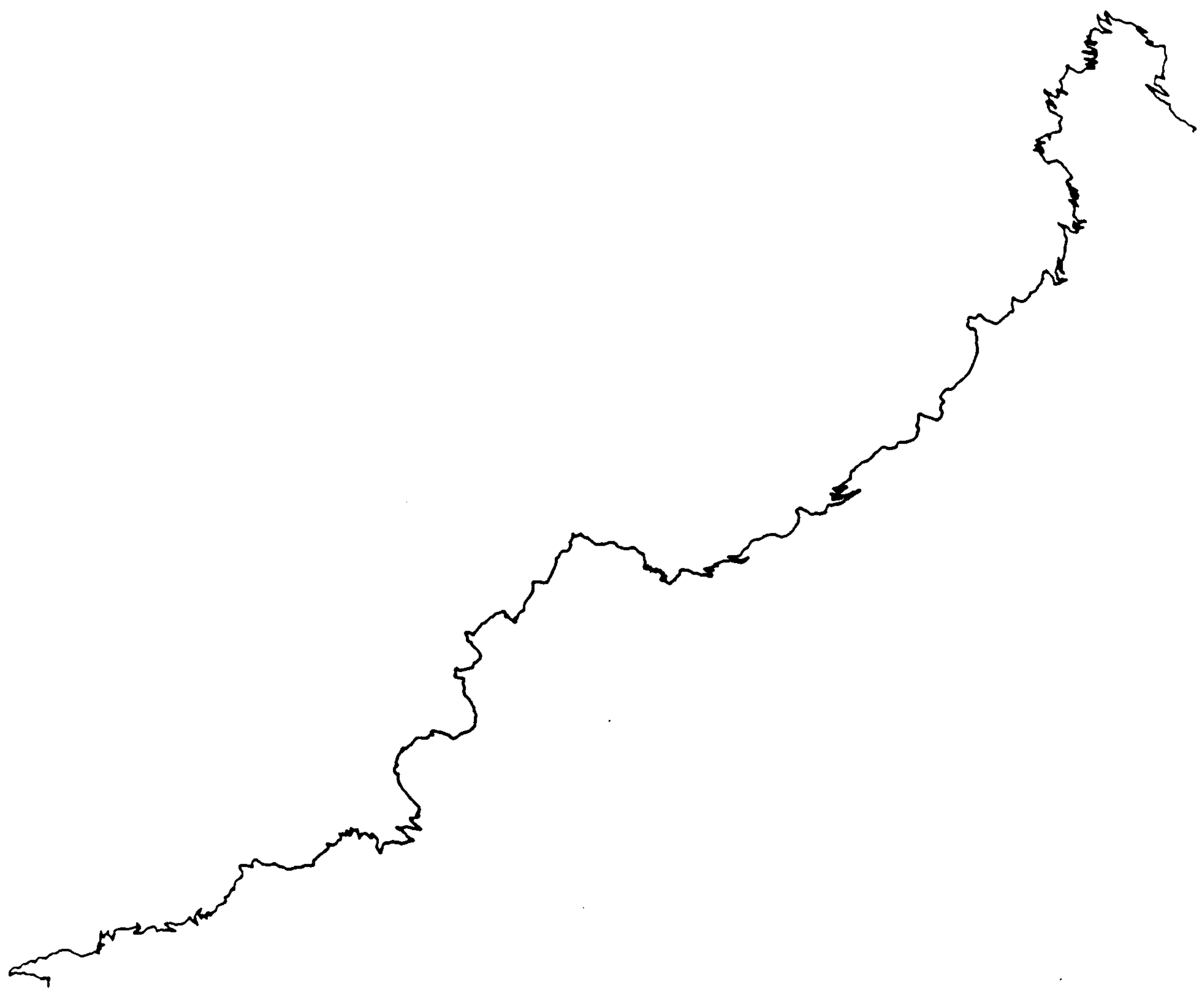


Figure 3.4 Swiss Road (1,268 points)
Original scale of representation 1:50,000.

All lines were digitized in point mode on an Altek digitizing tablet by the same person. Immediately following the digitizing the lines were 'cleaned' of spurious polygons and duplicate coordinate pairs. The former was achieved by simple viewing the blown up plots of the lines on a CRT terminal; the latter, by running an algorithm which looked for distances between consecutive pairs of coordinates, and eliminated one of them if the distance was smaller than the threshold. As Mount Desert Island was initially digitized from a larger scale map, it needed to be brought to par with the other lines. The line was simplified using Douglas-Peucker routine. The number of coordinates to be preserved in the new version were determined using the Radical Law (Topfer and Pillwizer, 1966). The law guides one on how much detail should be preserved on a map compiled from a source map, knowing the scales of both maps and the original number of items.

$$n_c = n_s \, (S_c \, / \, S_s)^{-1/2}$$

where: n_c number of items on compiled map with a scale fraction of S_c.
n_s number of items on compiled map with a scale fraction of S_s.

In the next stage, all four lines were subjected to multiple simplifications using the Douglas-Peucker algorithm (Douglas and Peucker, 1973). The choice of that algorithm was dictated by its general acclaim as the algorithm which minimizes displacement errors (Muller, 1987). At first an altered version of the algorithm was used in order to find out what tolerance values should be used to obtain a caricature containing the desired number of coordinate pairs. This was achieved by modifying the algorithm, so that instead of creating simplifications, it provided tabular data on the number of points retained in the simplified versions, given certain tolerance values (there were 200 values taken from 0.1 to 20.0 with the increment of 0.1).

The results of that can be visualized in Figure 3.5. When plotted in logarithmic space, the plots have an almost linear shape, particularly in the case of Mount Desert Island and the upper part of US101 (only Swiss road seems to have a rather curvilinear shape). This linearity of plots indicates the presence of full range of detail in the lines. As the tolerance value is incremented, the number of coordinates retained decreases at an almost constant rate. One might hypothesize that the linearity of plots -would be an indication of statistical self-similarity of the lines. The horizontal segment of the plot in the upper part of Mount Desert Island results from the fact that the line was already once simplified prior to this test. The minute detail has been removed and the few smallest tolerance values employed treat the line below the limits of its resolution.

Based on these data, a progression of ten levels of simplification was created for each line. Each consecutive level contains a smaller number of original coordinates. The progression is: 100% of points retained (the original line), and then 80%, 60%, 50%, 40%, 30%, 20%, 10%, 5%, and 2% of coordinates retained. This sequence is loosely based on perceptual logic. The levels 90% and 70% were skipped as they are not visually distinguishable from the neighboring ones. On the other hand, two very coarse representations, 5% and 2% were added. To illustrate this point Figure 3.6 was created which shows Klamath River in all the levels of simplification. As the original lines were digitized in point mode there are not many redundant points in these lines, which would not have been the case if the lines were digitized in stream mode. Thanks to that the 100% level lines are similar to subsequent simplified versions in the way that they also contain characteristic points only. During the simplification process special care was taken to ensure that each simplified version is a sample from all of the less simplified ones, that is, for example, all of the points in Klamath 10% can be found in Klamath 20% and up. Klamath 10% does not contain any coordinate pairs that Klamath 20% does not have and Klamath 30% does.

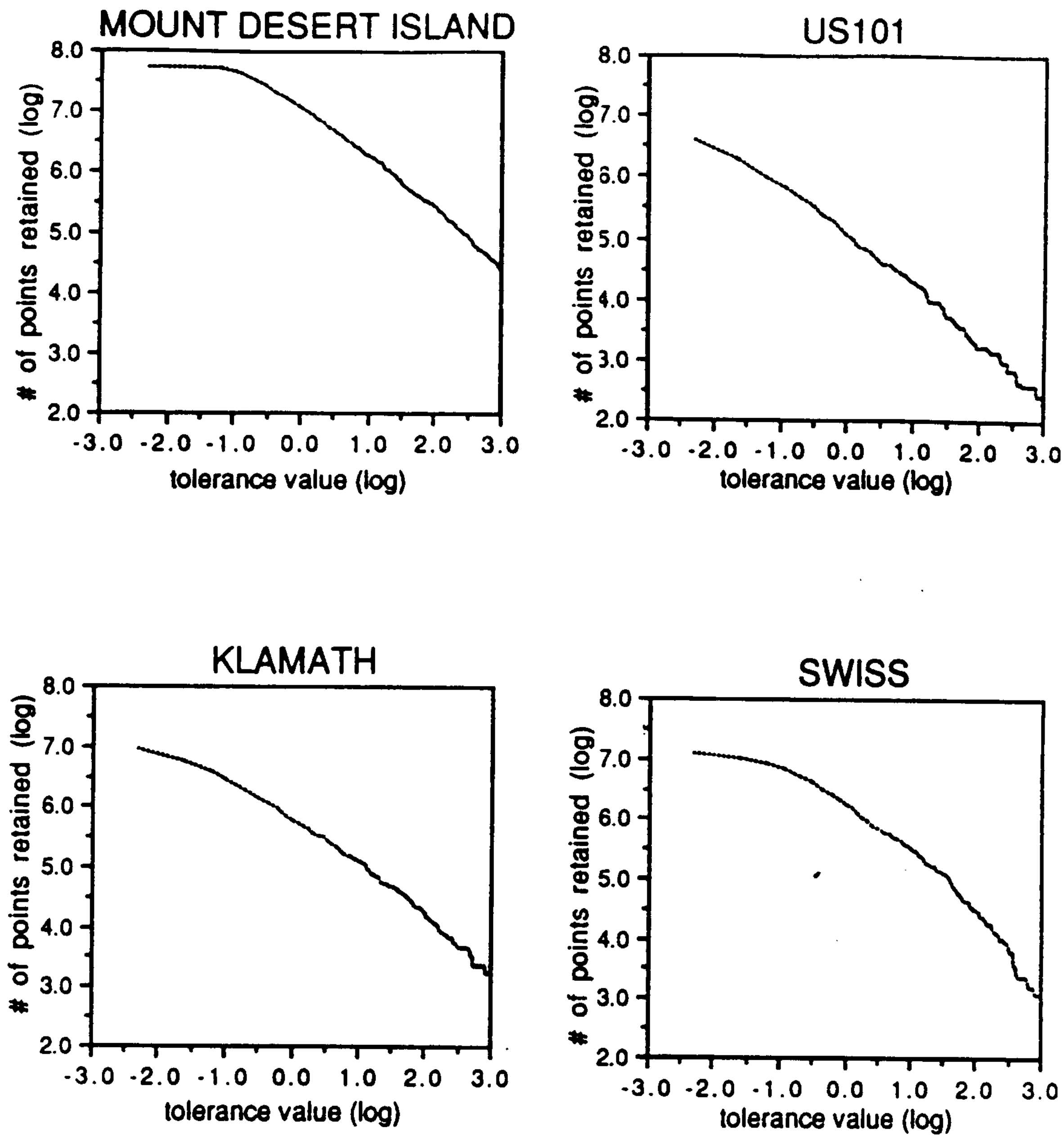


Figure 3.5 Number of Coordinates Retained with Changes in Tolerance Value in Douglas-Peucker Algorithm.

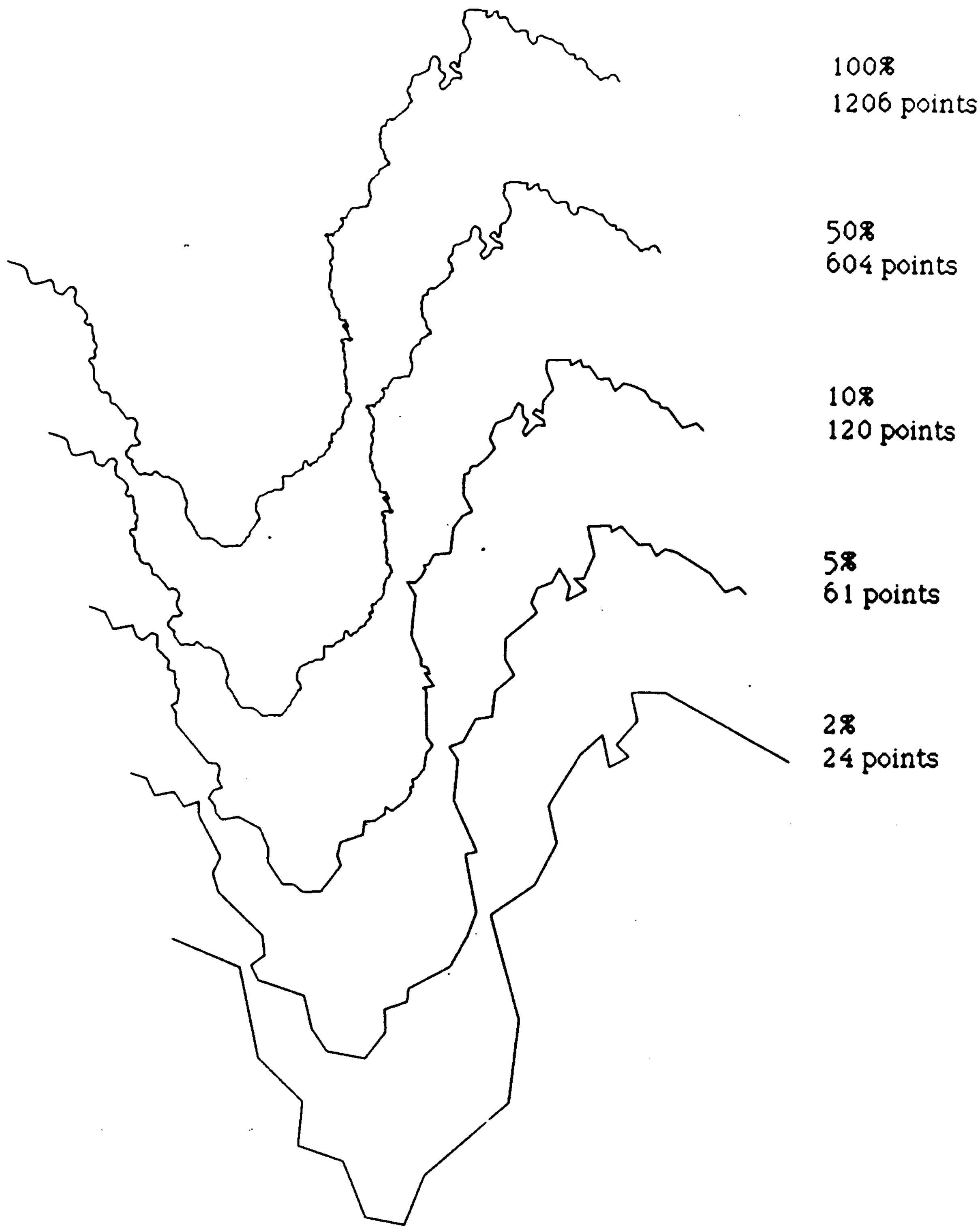


Figure 3.6 Perceptual Logic in Line Simplification Process.

3.3 Selection of Geometric Measures.

Comparison of a number of geometric measures may provide insights into the ways line complexity should be measured for cartographic purposes. Certain guidance for the selection of the measures was provided by the statement by Moellering and Rayner (1981) cited earlier, regarding ideal shape indices. It was not hoped that a set of indices would be created which would allow full reconstruction of original lines from the descriptions (as postulated by Moellering and Rayner, 1982). What was considered important, was that the measures should be general in nature, applicable to any kind of one-dimensional features, and that they were relative measures, not tied to any coordinate system or unit of measure.

Four aspects of line complexity were considered to be most important: angularity, relative length, density or frequency of detail, and fractal dimension. The choice of the first three was dictated by the existing literature (McMaster, 1986, Buttenfield, 1987). The fractal dimension could have been considered as a length measure, but was purposefully separated so that it could be compared against the other three groups. In the final analysis eight measures were used: 1. Average segment length, 2. Coefficient of variation of average segment length, 3. Error variance, 4. Coefficient of variation of error variance, 5. Average angularity, 6. Coefficient of variation of average angularity, 7. Curvilinearity ratio, 8. Fractal dimension. A few other measures were tried but for reasons described in the following chapter, these were the ones that entered the final analysis.

3.3.1 Average Segment Length.

This measure is an average of all segments between the points which constitute a digital line. It is expected that the values of the measure should increase as more simplified versions of the original line are measured, because fewer points are retained and segments automatically get longer. This measure can be considered as an approximation of the density of detail that exists in a line. An alternative measure, which was used by McMaster (1986), was the number of points per inch. It was decided not to use that measure as it is dependent on the length unit. In other words, if a line was plotted in a different scale, McMaster's measure would automatically change, although the actual geometry of the line would remain the same.

3.3.2 Coefficient of Variation of Average Segment Length.

The mean value of a distribution does not say anything about dispersion of values in the data set. Conversely, average segment length does not give much information whether most of the segments are of similar length or whether the line consists of some very short and some very long ones. In other words, we do not know how uniform is the detail in the line. This can be described by use of coefficient of variation, which gives one the relative variability of a distribution as a ratio of the standard deviation to the mean.

st. deviation $$s = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n - 1}}$$

coef. of variation $$CV = \frac{s}{\bar{X}}$$

It can be hypothesized that in the progression of resolution levels, for self-similar features this measure should be fairly constant, as a similar distribution of detail should be present in each version of the line.

3.3.3 Error Variance.

Error variance (Figure 3.7) is an average perpendicular displacement of every point in a line to the anchor line. The anchor line (Buttenfield, 1984) is defined as a straight segment linking the first and the last point of the coordinate string. Dutton (1981) refers to it as a "trend line". Error variance measures the deviation of the line from its straight line approximation. As the measure was used in only slightly different form by Buttenfield (1986), the name of the measure was retained in this project. A similar concept of looking at perpendicular distances between points and recursively smaller anchor lines constituted the basis for the Douglas-Peucker line simplification algorithm.

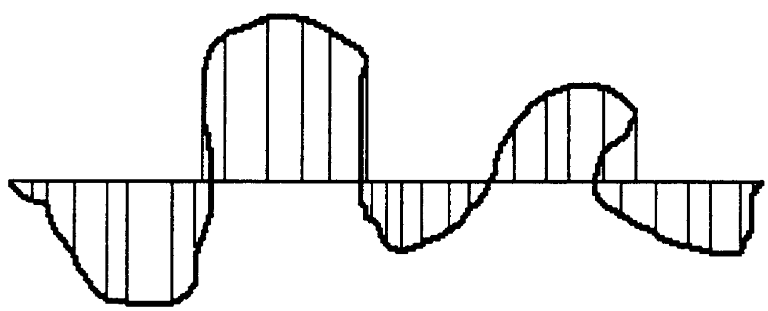


Figure 3.7 Error Variance.

3.3.4 Coefficient of Variation of Error Variance.

For the same reason for which coefficient of variation of average segment length is used, a standardized measure of dispersion is used for error variance. It is interesting to see the distribution of line's deviation from the anchor line and how it changes with progressive simplification.

3.3.5 Average Angularity.

As research done by Kelley (1977, as cited by Buttenfield, 1989) showed, angularity or 'wiggliness' of a line is one of its primary characteristics. Some line simplification algorithms are based on the variations in intersegment angles, e.g. Jenks' algorithm (1981). McMaster (1986) used the total angularity measure (sum of all intersegment angles) in his study. It was decided to use an average measure here, as the total angularity is clearly dependent on the number of points in a line. The average angularity variable is calculated according to the formula:

$$\text{Average Angularity} = \frac{\text{sum of } (1 - \sin \beta/2)}{n - 2}$$

The measure can take values ranging from 0 (straight line, intersegment angle is 0), to 1 (the line backs on itself, the angle is 180 degrees). The value of angularity is the same whether the angle of change is positive or negative.

3.3.6 Coefficient of Variation of Average Angularity.

The rationale behind selecting this measure was analogical to that in the case of previous coefficients of variation. It was hypothesized that with the progression of simplification the values of the average angularity measure should grow, whereas the values of the coefficient of variation of average angularity should decrease. The reason is that more and more critical points only will be retained and those should have higher angular change associated with them, and so the mean value should increase and the dispersion value should decline.

3.3.7 Curvilinearity Ratio.

Curvilinearity is defined as the number of inflection points in a line, meaning the number of times a line switches its direction of turn from positive to negative (or right to left) or vice versa (Figure 3.8). This measure is a variation on the theme of a curvilinearity measure used by McMaster (1986). Although the name 'wiggliness' would seem more appropriate for this measure, it was decided to retain the name as it was used by McMaster (1986). Again, in his study, the sum of all inflection points was used as a variable. In this project the measure is used in the form of a ratio of total curvilinearity to the number of all turns. The measure was selected in order to complement the average angularity by accounting not for the magnitude of angular - changes but their direction. If a line flows smoothly turning in none or one direction only (straight line, circle) the value of the variable will be 0, but it will climb drastically if the turns constantly switching from left to right and back.

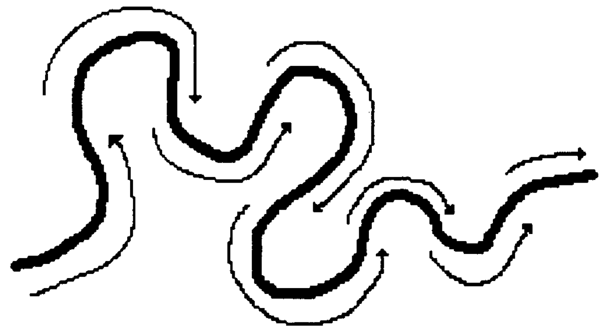


Figure 3.8 Curvilinearity.

3.3.8 Fractal Dimension.

Three different algorithms were initially tested for estimating the fractal dimension: the walking divider method as published by Shelberg et al. (1982), the hybrid walk method (Clark, 1986) and Eastman's (1985) single-pass angular computation. A detailed discussion of the results of these tests is included in the following chapter. The walking divider method was selected to be used in the final analysis. The algorithm is conceptually based on the original Richardson's manual walking divider method of comparing changes in the length of a line depending on the scale of the measurement. It was implemented by Shelberg et al. (1982) in an algorithmic form, which simulates traversing the extent of the line by a series of divider's walks with various spacings. One has to specify the initial length of the divider's step, the progression of subsequent step sizes (decreasing or increasing), and some sort of cut off point. That can be either a hard-wired number of times the line is going to be traversed, or some tolerance value, which if reached, stops the algorithm, e.g. continue increasing step size and 'walking' the line until not less than five full steps can be fit in the line.

The implemented algorithm in this research is an interpolating one, meaning that in the 'walking' process, the exact intersection of the line with the current divider length is found for each step. That point is then interpolated from the two neighboring coordinate pairs. Very often the last step of the divider is incomplete as the span of the divider is greater than the distance between the last interpolated point and the last point of the line. That distance is then added as a partial step to the total length (calculated as number of divider steps times the step length). After the termination of all walks, the results which are step lengths with corresponding total line length values, are regressed in the log-log space. The fractal dimension D is then estimated as the (1 - s), where s is the slope of the regression line.

3.4 Summary

This chapter presents the logic behind the design of the research of this project. First, the test data set consisting of four basic lines (two natural and two cultural features) and nine levels of progressive simplifications was created. Then, eight measures of geometric shape of lines were selected. It was important to include measures which would account for the main characteristics of geometric shape of one-dimensional features. With this in mind the indices chosen measure angular variation (average angularity and its coefficient of variation, curvilinearity ratio), density of detail and length variations (average segment length, error variance and their respective coefficients of variation) and fractal dimension. All of these were run across the levels of simplification for all four test lines.

The results of those tests are presented in the following chapter. Also, comparative trials of three methods of estimating fractal dimension are reported. The final chapter, which follows, contains general conclusions stemming from this research and recommendations for directions of future studies on the topic.

4. DISCUSSION OF DATA AND ANALYSIS.

4.1 Overview.

The initial part of the analysis is based on plotting out the values of all measures against the progressions of resolution for each of the four original lines. The interpretation and discussion of the graphs follows the same sequence in which the measures were described in the previous chapter. The four basic lines, Mount Desert Island, Klamath river, highway US101, and Swiss road, will be referred to as MDI, Klamath, US101, and Swiss, respectively.

4.2 Measures of Length.

Figure 4.1 show what is happening to the average segment length when the line is more and more simplified. As it could be expected, the mean segment length is increasing. This trend is gradual for the levels with large percentage of points retained, and rapidly gains momentum when less than 20% of points are preserved. What is interesting, is that the variable behaves almost identically for all four lines. The values are the highest for Mount Desert Island, which is understood, as this feature is overall the longest of the four.

Interpretation of the standardized measure of dispersion of mean segment length seems to be much more difficult (Figure 4.2). For MDI, Klamath and Swiss the measure behaves similarly in a way that oscillates roughly between the values of 0.5 and 0.7. One could look for some signs of periodicity here and it may also be indication of statistical self-similarity within this range of resolutions. That would seem to confirm the earlier hypothesis about potential use of this measure to determine the self-similarity of a line. Interestingly, the variable monotonically decreases for US101, which means that each level of simplification has segments of more uniform lengths. That would indicate an increasing level of regularity in terms of the detail present through the whole extent of the line.

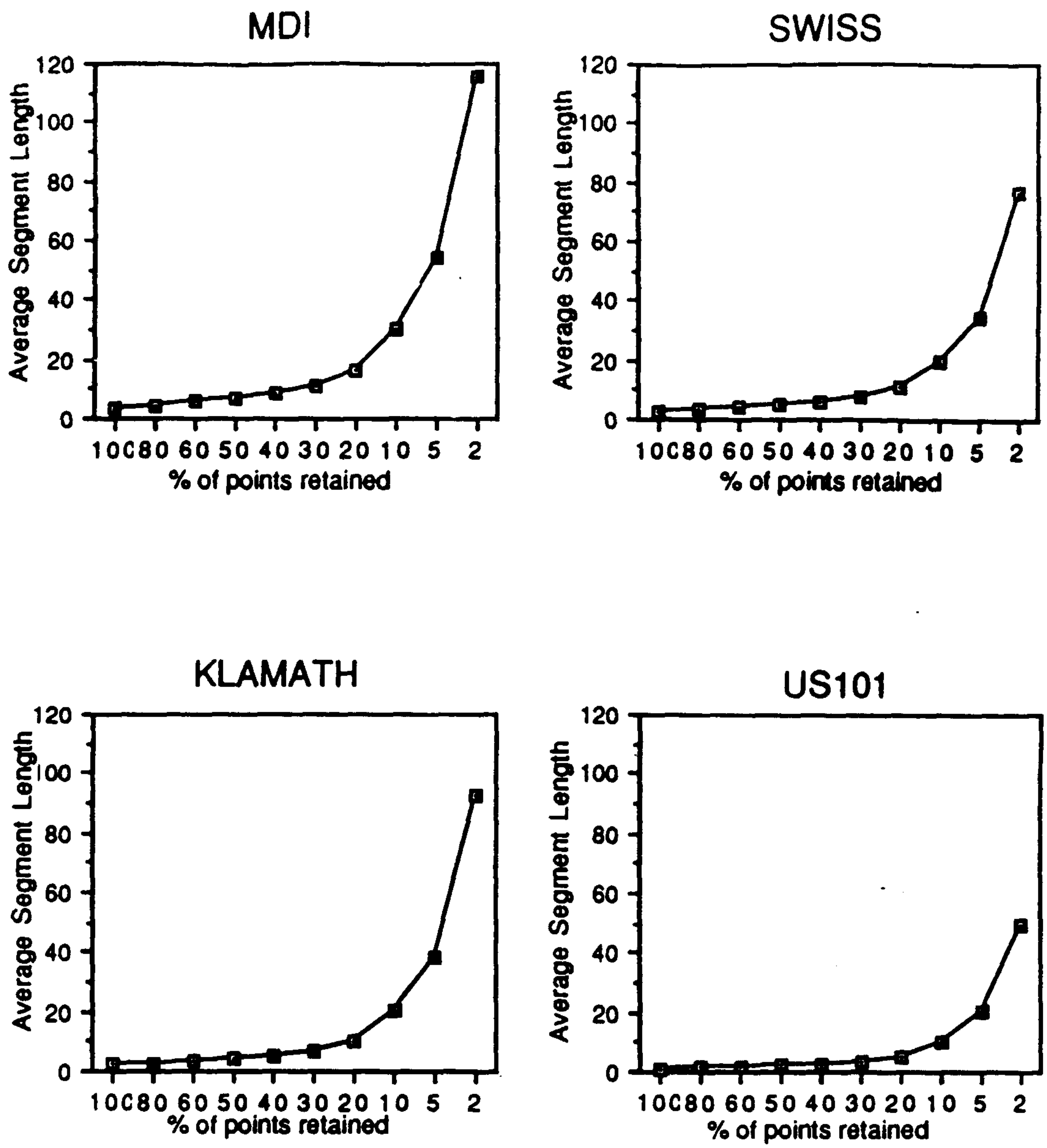


Figure 4.1 Average Segment Length.

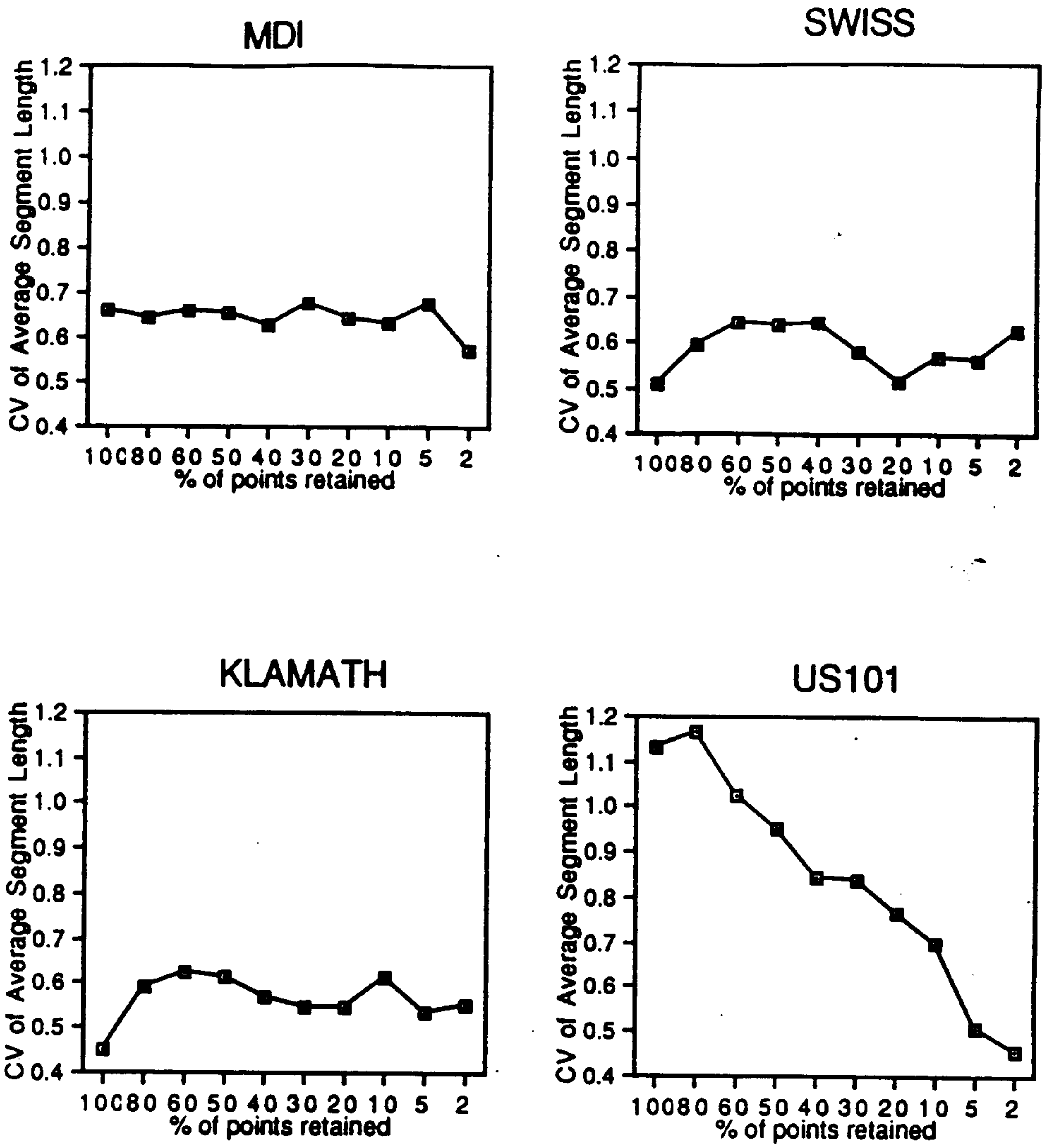


Figure 4.2 Coefficient of Variation of Average Segment Length.

In the case of error variance (Figure 4.3), it seems that some different trends can be noticed for natural and for man made features. Both roads have a generally increasing trend, the greater the simplification. This is logical, as with progressive simplification by Douglas-Peucker routine characteristic points are retained which should tend to have high error variance. As they both do not have much high level detail, the range of values that error variance takes in absolute terms is much lower in comparison to both Klamath and MDI. The interpretation of the error variance for MDI is much tougher as there is no obvious anchor line for that feature. It was decided to use the first and last points in the file as limits of the anchor line so it is extremely short.

The coefficient of variation (Figure 4.4) shows that for all lines the dispersion of values of error variance is pretty stable. Similarly to its dispersion measure of average segment length, US101 exhibits a decreasing trend which indicates that the points preserved in subsequent simplifications are more regularly spaced in terms of error variance from the anchor line. The very stable behavior of the measure in the case of MDI could have been expected, as MDI's anchor line is degenerate, so that it is in a different relation to that line than in the other three cases.

4.3 Measures of Angularity.

The next measure is average angularity (Figure 4.5). The highest values can be observed for Swiss and MDI and lower for US101 and Klamath. The first two features are full of high frequency, sharp turns, when the latter two have more of a meandering stream like character. The values for all lines exhibit a very clear increasing trend. It is interesting that for US101, Swiss, and Klamath, that trend is clearly stopped and even reversed at either 10% (US101 and Swiss) or 5% level (Klamath). One may expect that the same would have happened to MDI should the simplification process continued. One should remember here that MDI has almost twice as many points as any of the remaining lines at any of the simplification levels. Thus, certain threshold in terms of angular variation, which the other three lines reach at that level, is delayed for MDI. The increasing trend confirms the findings of earlier cited research by Kelley (1977, as cited by Buttenfield, 1989), who found that characteristic points are located at places demonstrating angular change. It thus seems that the peak of average angularity could be interpreted as a threshold of maximum reasonable simplification, beyond which further reduction of the number of points will lead to loss of geometric characteristics of the original line. Using the terminology of Buttenfield (1989), the line would preserve its structural signature up that point in simplification process. Beyond that resolution, details in the digital form of the line would no longer provide an appropriate representation of the original feature.

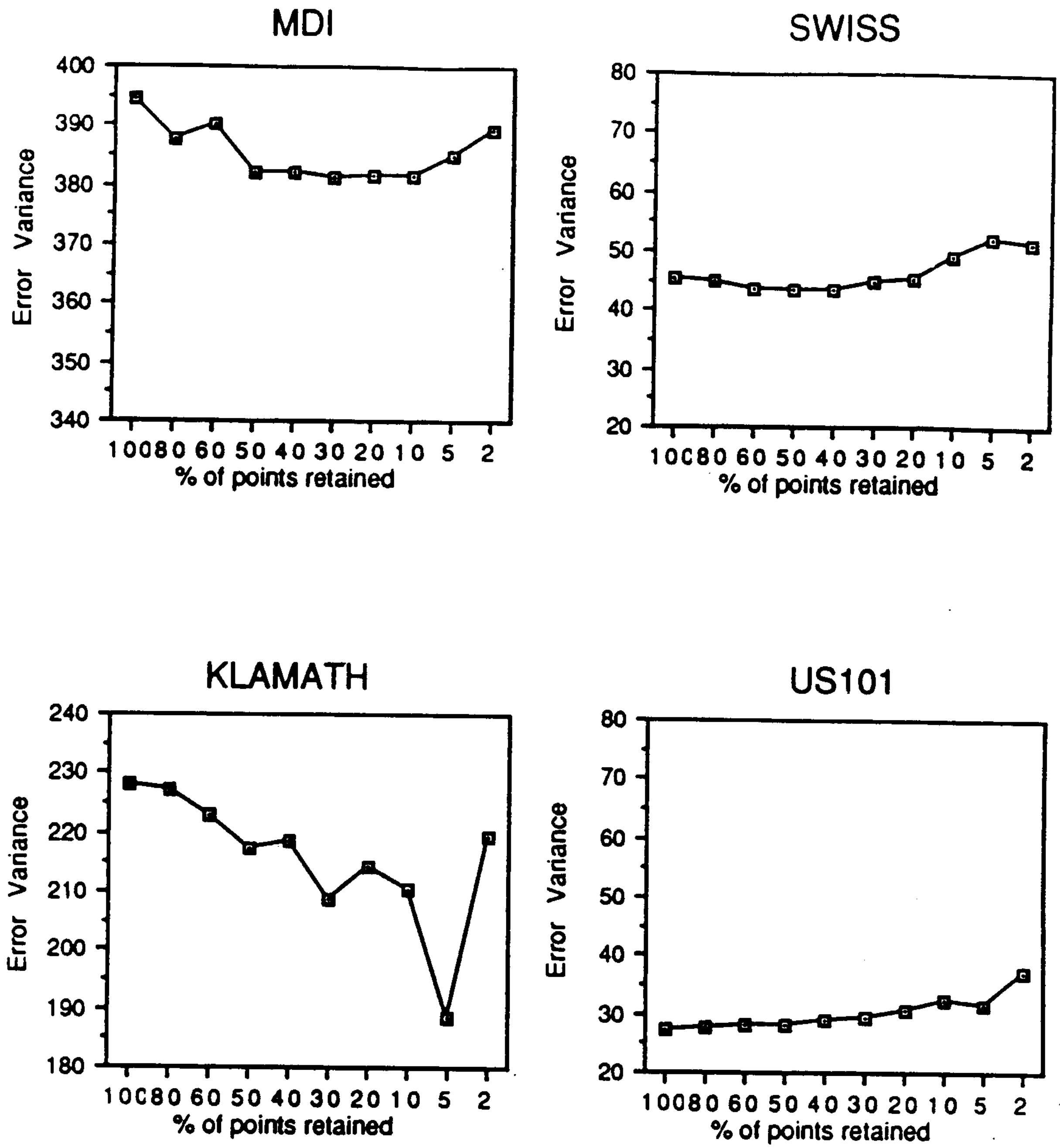


Figure 4.3 Error Variance.

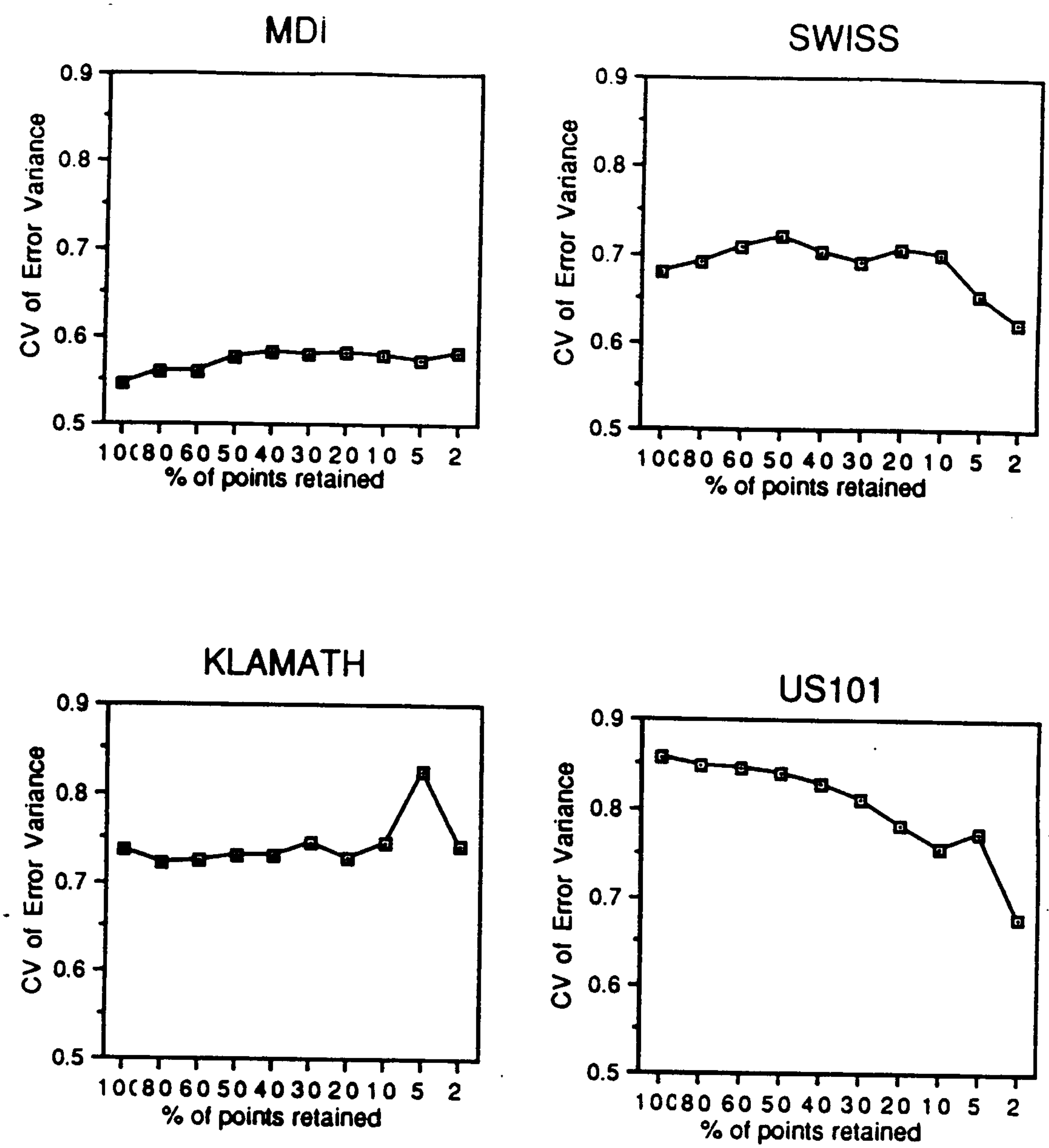


Figure 4.4 Coefficient of Variation of Error Variance.

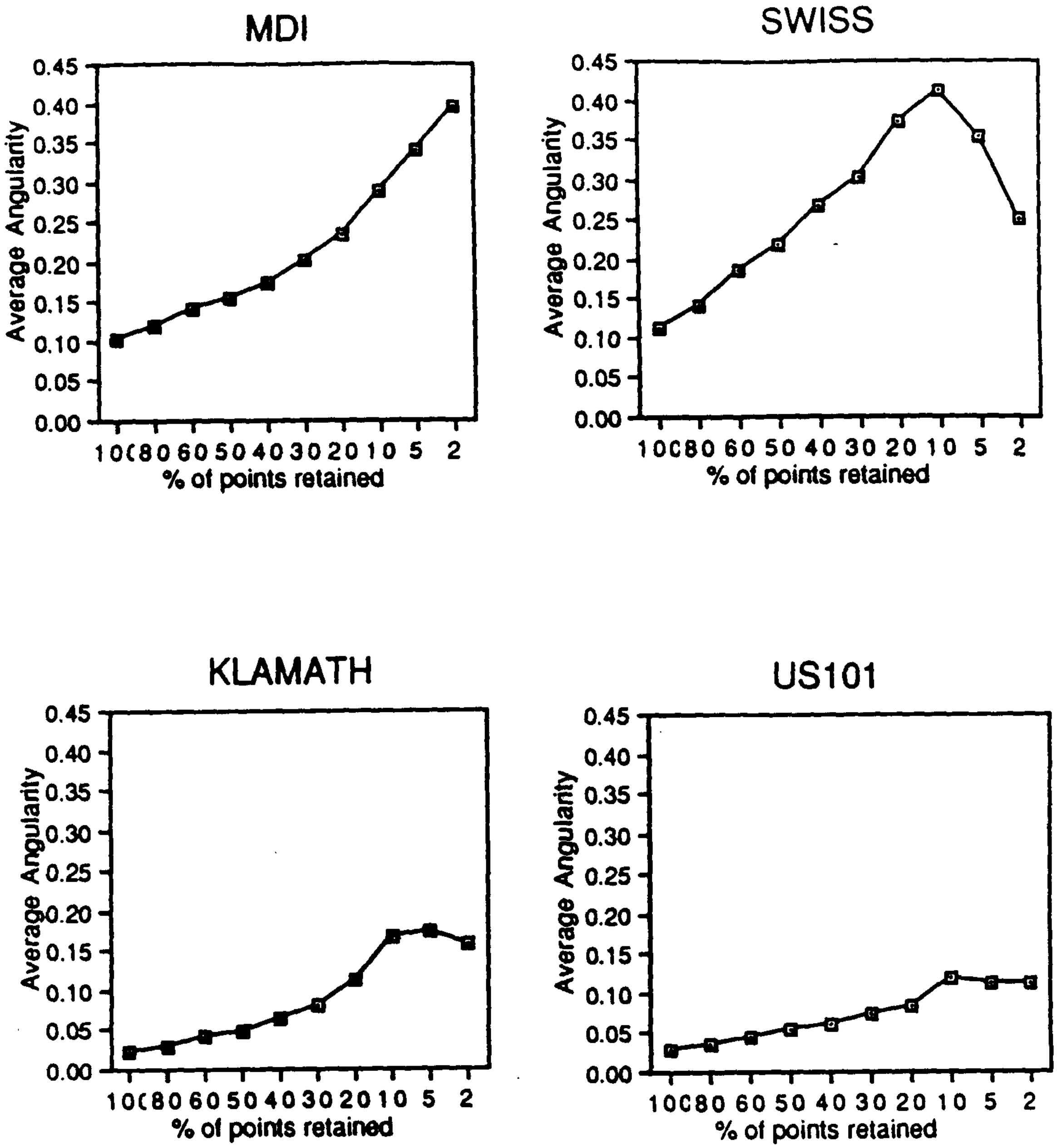


Figure 4.5 Average Angularity.

The coefficient of variation for average angularity (Figure 4.6), confirms what was stated above about average angularity. The values are quite stable (particularly for MDI) until the level of 10% is reached, and then exhibit seemingly random variations. The exception is an unexplainable 'jump' at level 40% for US101. Although at the design stage it was thought that the measure would have a decreasing trend, there is no evidence of that.

The curvilinearity ratio (Figure 4.7) exhibits similar pattern to average angularity, that is a very clear positive trend across the range of simplifications. In the case of Klamath and Swiss, the trends' cusps are at the same level as for average angularity. The increase in curvilinearity ratio with coarser resolution comes as no surprise. The characteristic points which are retained by Douglas-Peucker are those with high angular change, which should coincide with the change in curvilinearity. This measure seems to give one a very good sense of a line's character. For example, the values for Klamath are generally low, starting at 0.37, which means that the direction of the line's trend changes for 37% of coordinates. If only every third turn is an inflection point, then the line has to have a meandering stream like flow, with wide, smooth turns. In the case of Swiss, though, the values start with 0.58, which means that almost two-thirds of turns involve change from left to right or the opposite. This suggests a more rugged line, with higher frequency small detail.

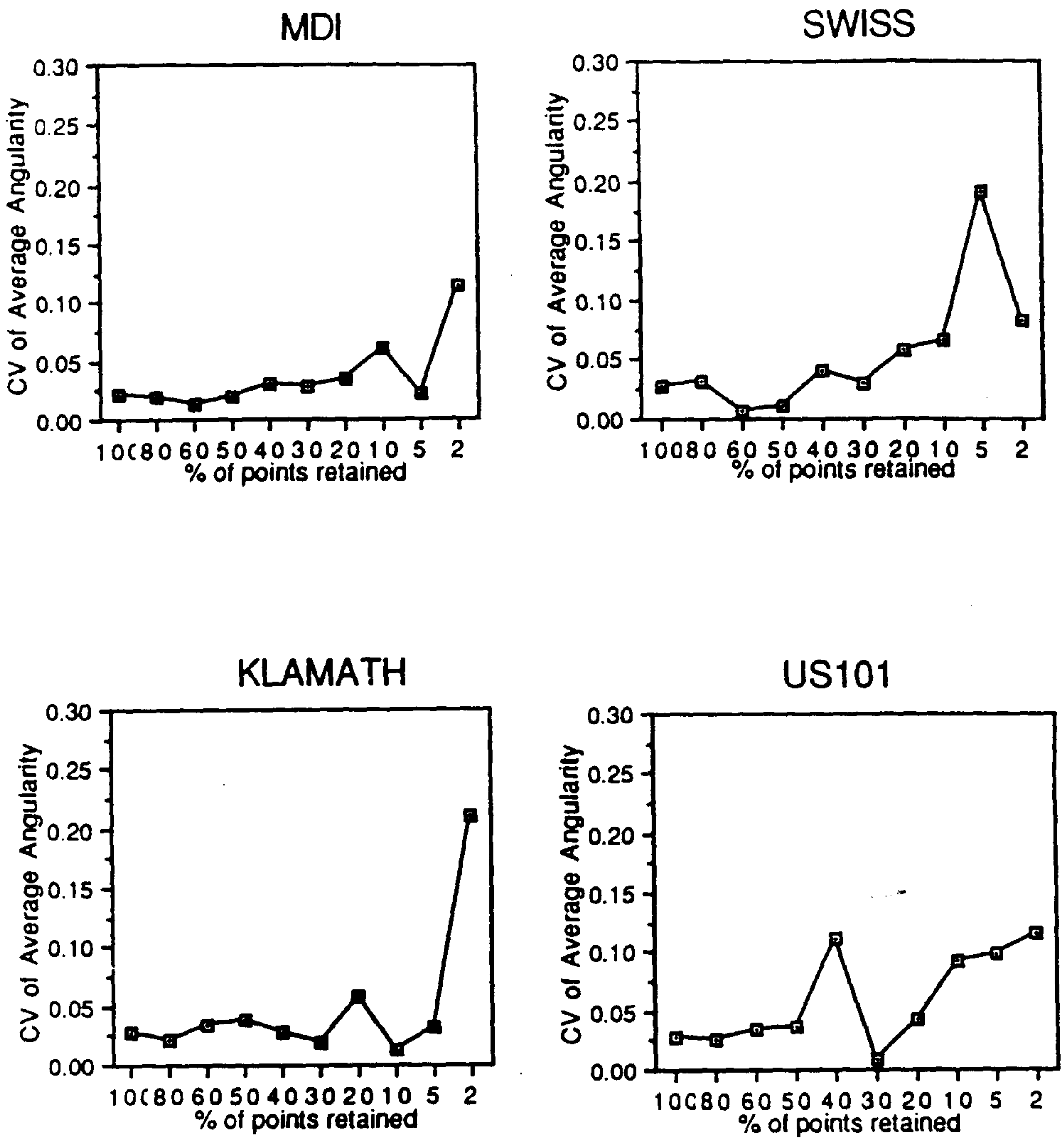


Figure 4.6 Coefficient of Variation of Average Angularity.

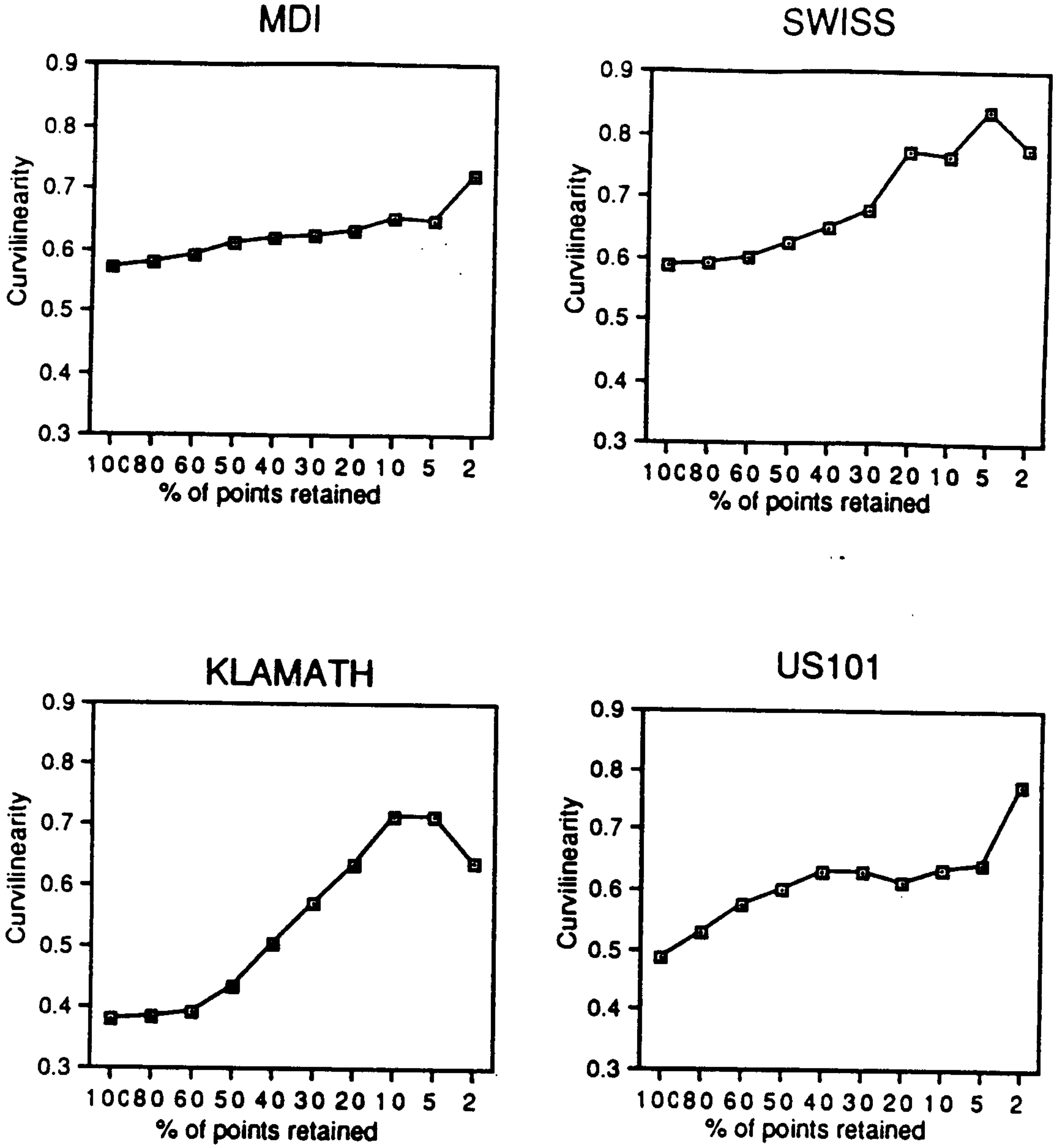


Figure 4.7 Curvilinearity Ratio.

4.4 Measures of Fractal Dimension.

Initially, it was intended to include at least two different methods of estimating fractal dimension in the final set of measures. The methods examined were the walking divider (Shelberg et al., 1982), described in the previous chapter, the hybrid walk (Clark, 1986), Eastman's single-pass measure (Eastman, 1985), and Richardson's plots as utilized by Buttenfield (1989).

In the walking divider method in the algorithmic form, one simulates traversing a line manually with a pair of open dividers. Three parameters have to be set before the run. They are: the initial opening of the dividers, the progression with which the dividers' step size is going to change, and the threshold parameter which will terminate running of the algorithm. There seems to be widespread agreement on using geometric progression as a rate of increase in the size of dividers opening (Shelberg et al., 1982; Muller, 1987). The advantage of that progression is that it results in even spacing of dividers step size when plotted later against the line's length in logarithmic scale.

Determining the values of the other two parameters seems to be governed by rule of thumb. Shelberg et al. (1982) suggest that the initial step length should be equal to half of the average segment length because "(...) the sampling theorem states one should sample at 1/2 the wavelength so that no significant variation escapes", and the say that "(...) the form of the curve often dictates this manner". Unfortunately they do not say how the form of a line dictates the use of that length. Theoretically, there is no lower bound which should determine the initial spacing of the dividers. In practice, the digital lines have well defined resolution limit which is given by the shortest segment length and that could be used as the initial step length. Walking a line with the step size smaller than that would not make any sense as no differences in line's length could be observed.

The third parameter governs the point at which doubling of the step size should terminate. In Shelberg's et al (1982) study, the line was walked at least five times, but the authors do not say what kind of upper bound was used. They conclude that five to eight solution steps are sufficient to determine the slope of the regression line and thus fractality", and that the choice of the ideal initial step size should be based on observation of the scatterplots (step length vs. line length in logarithmic space), r2 values, and the number of walks of line. The upper bound for number of walks used by Longley and Batty (1989a) was that the line had to fit at least seven steps of dividers for any dividers' length.

In this study a number of different initial divider openings were used as well as various cutoff values, to determine the influence on the estimate of fractal dimension. The results of those trials are presented in Tables 4.1 through 4.5. Figure 4.8 shows the scatterplot for MDI at 100% level. The initial step size was set to half the average segment length and the cutoff point was set so that at least five full steps had to fit in any walk. Given those parameters the estimate of fractal dimension is D =1.187 with r^2 value of 0.997. If the initial step size is set equal to the average segment length, fractal dimension equals D = 1.201 with r^2 = 0.987. Conversely we may decide to set the upper limit of the number of times the line is walked to, say five times. Then the estimate of D = 1. 13 1, with $r^2 = 0.955$.

One may argue that the differences between the estimates of fractal dimension are minor and that they only show that MDI is self-similar. The point is that those differences nevertheless exist and they may very well be much more pronounced for other lines. What is even more important, is that even at first glance at Figure 4.8 one can see that the actual distribution of points is not linear but rather curvilinear. Fitting a regression line may provide an inappropriate approximation (Buttenfield, 1989) and one may start wondering whether it really is methodologically correct. In research done by Andrle and Abrahams (1989), scatterplots of step size versus line length were analyzed using a goodness of fit test and the Durbin-Watson statistic. Both the tests indicated that the linear model is inappropriate, and the former revealed a possible curvature in the plots. The concave shape of the plots was detected in all cases, including those where linear fit had r2 = 0.99.

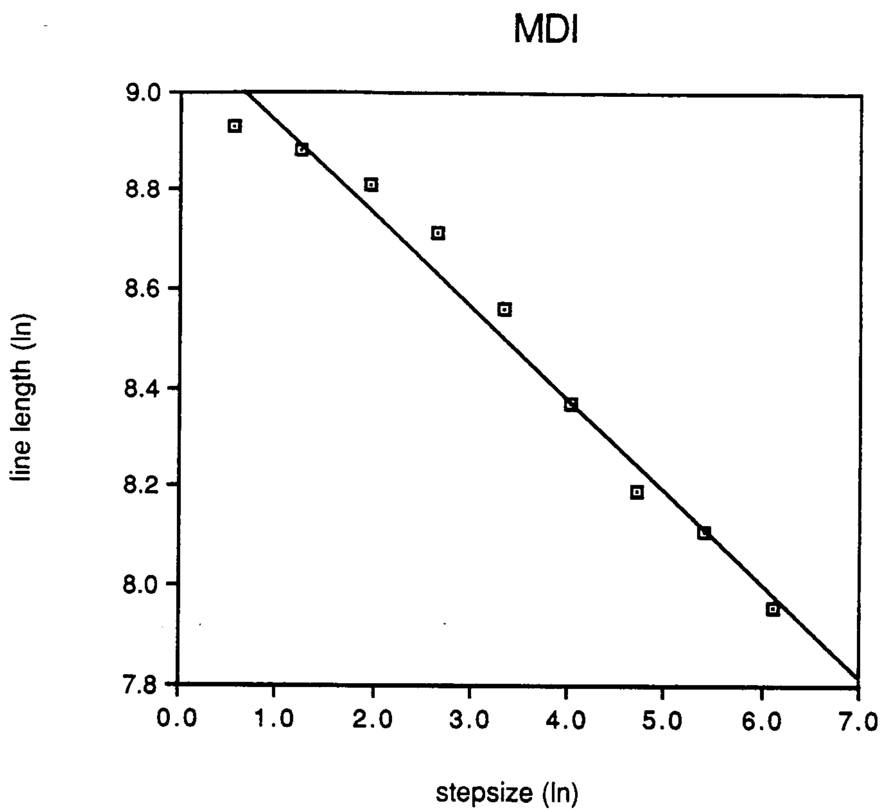


Figure 4.8 Richardson's Plot for Mount Desert Island

TABLE 4.1. ESTIMATES OF FRACTAL DIMENSION USING WALKING DIVIDERS METHOD.

The initial step size: equal to minimum segment length.
The terminating condition: until at least 3 full steps can be fitted in the line.

LINES % of points	KLAMATH D (r-sq)	MDI D (r-sq)	SWISS D (r-sq)	US101 D (r-sq)
100	1.067 (0.89)	1.187 (0.89)	1.095 (0.97)	1.029 (0.91)
80	1.067 (0.89)	1.175 (0.95)	1.103 (0.97)	1.029 (0.91)
60	1.064 (0.91)	1.189 (0.93)	1.107 (0.98)	1.040 (0.80)
50	1.073 (0.87)	1.218 (0.91)	1.105 (0.98)	1.040 (0.80)
40	1.061 (0.96)	1.178 (0.97)	1.115 (0.99)	1.033 (0.93)
30	1.083 (0.88)	1.237 (0.93)	1.100 (0.97)	1.046 (0.76)
20	1.071 (0.95)	1.244 (0.91)	1.103 (0.98)	1.040 (0.83)
10	1.078 (0.98)	1.259 (0.91)	1.093 (0.97)	1.047 (0.84)
5	1.067 (0.91)	1.301 (0.95)	1.119 (0.85)	1.027 (0.86)
2	1.115 (0.92)	1.211 (0.96)	1.111 (0.95)	1.066 (0.89)

TABLE 4.2. ESTIMATES OF FRACTAL DIMENSION USING WALKING DIVIDERS METHOD.

The initial step size: equal to the average segment length.
The terminating condition: until at least 3 full steps can be fitted in the line.

LINES	KLAMATH	MDI	SWISS	US101
% of points	D (r-sq)	D (r-sq)	D (r-sq)	D (r-sq)
100	1.076 (0.97)	1.214 (0.99)	1.100 (0.95)	1.056 (0.85)
80	1.094 (0.93)	1.293 (0.94)	1.102 (0.97)	1.055 (0.78)
60	1.090 (0.94)	1.232 (0.99)	1.131 (0.95)	1.041 (0.86)
50	1.081 (0.96)	1.222 (0.99)	1.096 (0.95)	1.064 (0.84)
40	1.102 (0.92)	1.281 (0.89)	1.101 (0.94)	1.027 (0.90)
30	1.100 (0.93)	1.238 (0.99)	1.092 (0.92)	1.047 (0.85)
20	1.119 (0.89)	1.313 (0.89)	1.105 (0.97)	1.023 (0.68)
10	1.180 (0.93)	1.277 (0.98)	1.058 (0.80)	1.028 (0.62)
5	1.082 (0.99)			1.048 (0.82)
2				

* Missing values mean that even 3 full steps of the dividers could not be fitted.

TABLE 4.3. ESTIMATES OF FRACTAL DIMENSION USING WALKING DIVIDERS METHOD.

The initial step size: equal to the half of average segment length.
The terminating condition: until at least 3 full steps can be fitted in the line.

LINES	KLAMATH	MDI	SWISS	US101
% of points	D (r-sq)	D (r-sq)	D (r-sq)	D (r-sq)
100	1.072 (0.97)	1.202 (0.99)	1.105 (0.96)	1.050 (0.85)
80	1.086 (0.92)	1.269 (0.94)	1.109 (0.98)	1.048 (0.78)
60	1.084 (0.95)	1.222 (0.99)	1.129 (0.97)	1.039 (0.86)
50	1.077 (0.97)	1.218 (0.99)	1.103 (0.97)	1.057 (0.83)
40	1.095 (0.93)	1.263 (0.91)	1.106 (0.96)	1.028 (0.93)
30	1.092 (0.94)	1.233 (0.99)	1.105 (0.95)	1.043 (0.87)
20	1.105 (0.89)	1.289 (0.92)	1.107 (0.98)	1.026 (0.80)
10	1.139 (0.87)	1.266 (0.99)	1.065 (0.90)	1.030 (0.76)
5	1.085 (0.99)	1.211 (0.99)	1.132 (0.95)	1.040 (0.82)
2				

* Missing values mean that even 3 full steps of the dividers could not be fitted.

TABLE 4.4. ESTIMATES OF FRACTAL DIMENSION USING WALKING DIVIDERS METHOD.

The initial step size: equal to the half of average segment length.
The terminating condition: until at least 5 full steps can be fitted in the line.

LINES % of points	KLAMATH D (r-sq)	MDI D (r-sq)	SWISS D (r-sq)	US101 D (r-sq)
100	1.066 (0.96)	1.187 (0.98)	1.116 (0.99)	1.034 (0.94)
80	1.064 (0.97)	1.204 (0.97)	1.118 (0.99)	1.043 (0.78)
60	1.065 (0.97)	1.198 (0.98)	1.114 (0.99)	1.028 (0.96)
50	1.072 (0.97)	1.205 (0.99)	1.119 (0.99)	1.036 (0.93)
40	1.071 (0.99)	1.196 (0.97)	1.118 (0.98)	1.027 (0.95)
30	1.071 (0.98)	1.214 (0.99)	1.110 (0.97)	1.031 (0.94)
20	1.072 (0.99)	1.208 (0.99)	1.105 (0.99)	1.033 (0.93)
10	1.077 (0.90)	1.236 (0.99)	1.095 (0.98)	1.038 (0.92)
5	1.080 (0.99)	1.215 (0.99)	1.104 (0.91)	1.041 (0.84)
2				1.033 (0.78)

* Missing values mean that even 3 full steps of the dividers could not be fitted.

TABLE 4.5. ESTIMATES OF FRACTAL DIMENSION USING HYBRID WALK METHOD.

The initial step size: equal to the average segment length.
The terminating condition: until at least 3 full steps can be fitted in the line.

LINES % of points	KLAMATH D (r-sq)	MDI D (r-sq)	SWISS D (r-sq)	US101 D (r-sq)
100	1.233 (0.54)	1.405 (0.64)	1.169 (0.98)	1.026 (0.83)
80	1.076 (0.89)	1.198 (0.97)	1.194 (0.98)	1.498 (0.36)
60	1.079 (0.89)	1.220 (0.97)	1.160 (0.99)	1.027 (0.86)
50	1.237 (0.57)	1.456 (0.67)	1.167 (0.98)	1.025 (0.64)
40	1.078 (0.89)	1.212 (0.98)	1.168 (0.99)	1.573 (0.40)
30	1.080 (0.86)	1.223 (0.98)	1.154 (0.98)	1.024 (0.72)
20	1.073 (0.89)	1.217 (0.98)	1.137 (0.99)	1.424 (0.54)
10	1.062 (0.98)	1.220 (0.99)	1.074 (0.82)	1.822 (0.50)
5		1.611 (0.78)		1.483 (0.57)
2				

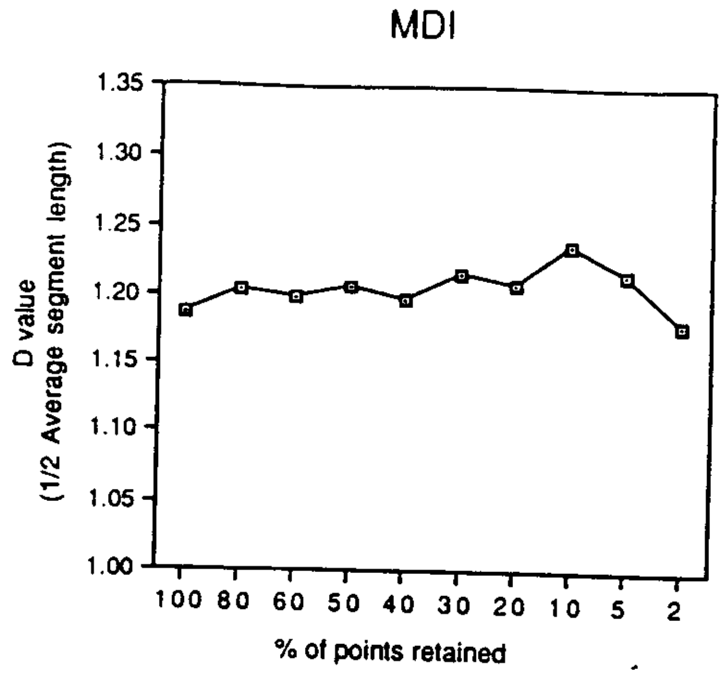
* Missing values mean that even 3 full steps of the dividers could not be fitted.

One could believe that the linear fit of r2 = 0.997 is an evidence that MDI is self-similar. If that is the case, then all of the levels of simplification should also exhibit the same property. Figure 4.9 shows the estimates of fractal dimension for all the levels of resolution For the upper graph the initial dividers step size was set to half of the average segment length, and the cutoff point was set so that no less than five full steps would fit in each walk. For the lower graph the initial opening of the dividers was set to the average segment length, and the respective cutoff was set to three steps. The value for the level 2% is missing because the line does not contain enough points to walk the dividers at least three times. If the line was self-similar then the estimates of fractal dimension could differ slightly between the graphs, because of tweaking of the parameters, but should be quite stable on each of the graphs. This does not seem to be the case for MDI, although is certainly more true for Klamath (Figure 4.10) and Swiss (Figure 4.11).- Whether it is also the case for US101 (Figure 4.12) is hard to say as in half of the levels the line fit is not valid.

The hybrid walk method (Clark, 1986; Longley and Batty, 1989a, 1989b) works similarly to the walking divider (called "structured walk" by Longley and Batty, 1989a). The difference is that no points are interpolated. During the walk, for each step it is found between which two points the step would fall, and the length of the step is adjusted to the closer one of those points. This way no new points are introduced and only those already existing in the line are used. It can be assumed that the effects of stretching some steps and shortening other even out and that the total length calculated by multiplying the step size by the --number of steps should be close to the one that could be obtained by using the interpolating algorithm. The advantage of using this algorithm lies in its computational efficiency as compared to the structured walk. Longley and Batty (1989) reported that the selection of starting point from which the dividers are walked may bias the estimate of fractal dimension.

For that reason the algorithm used was adjusted, so that it could start traversing a line from any point, going towards both ends and summing the result afterwards. Robert Andrle (personal communication, 1989) indicated that use of 50 random starting points and taking an average of obtained results should be sufficient to exclude any bias. Longley and Batty (1989a) used every point in their lines as a starting one. Initial runs of the algorithm in this study were based on 50 random starting points, but the results obtained did not seem to be stable.

Finally, every point in the test lines was used as a starting one, but still the results (presented in Table 4.5) were not satisfactory. For surprisingly high number of lines, as compared to the results from the structured walk method, the linefit was worse than 0.9, and thus treated with little confidence. The reason why this method worked poorly in this study may be the distribution of points in the lines. In Andfle's and Abraham's (1989) study the lines were stream digitized with high density of points. Even the initial opening of the dividers was many-fold larger than average segment length. In this study, with the spacing between coordinate pairs increasing with the progression in levels of simplification, in many cases opening of the dividers was smaller than some segment lengths. The error resulting from not interpolating could accumulate instead of compensating.



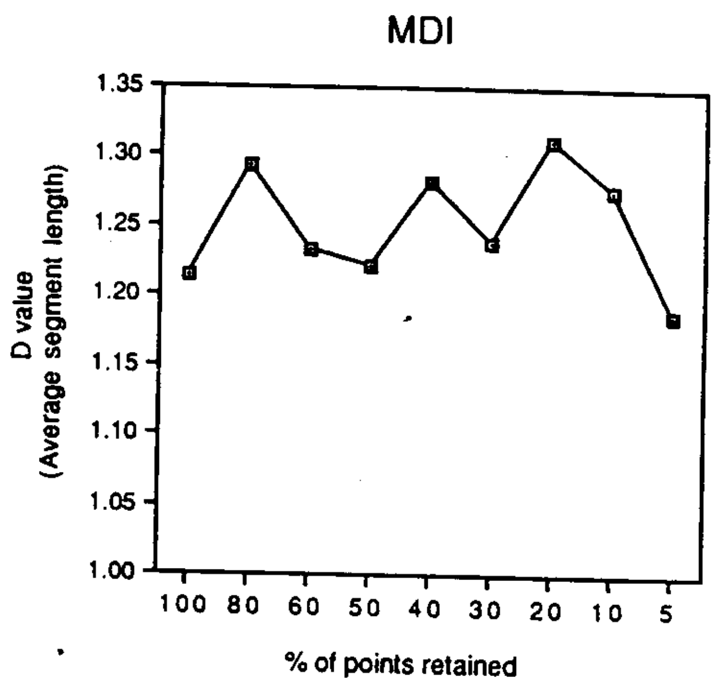


Figure 4.9 Stability of Fractal Dimension -- MDI.

Upper graph: The initial step size: half of average segment length. The terminating condition: until at least 5 full steps can be fitted in the line.

Lower graph: The initial stepsize: average segment length. The terminating condition: until at least 3 full steps can be fitted in the line.

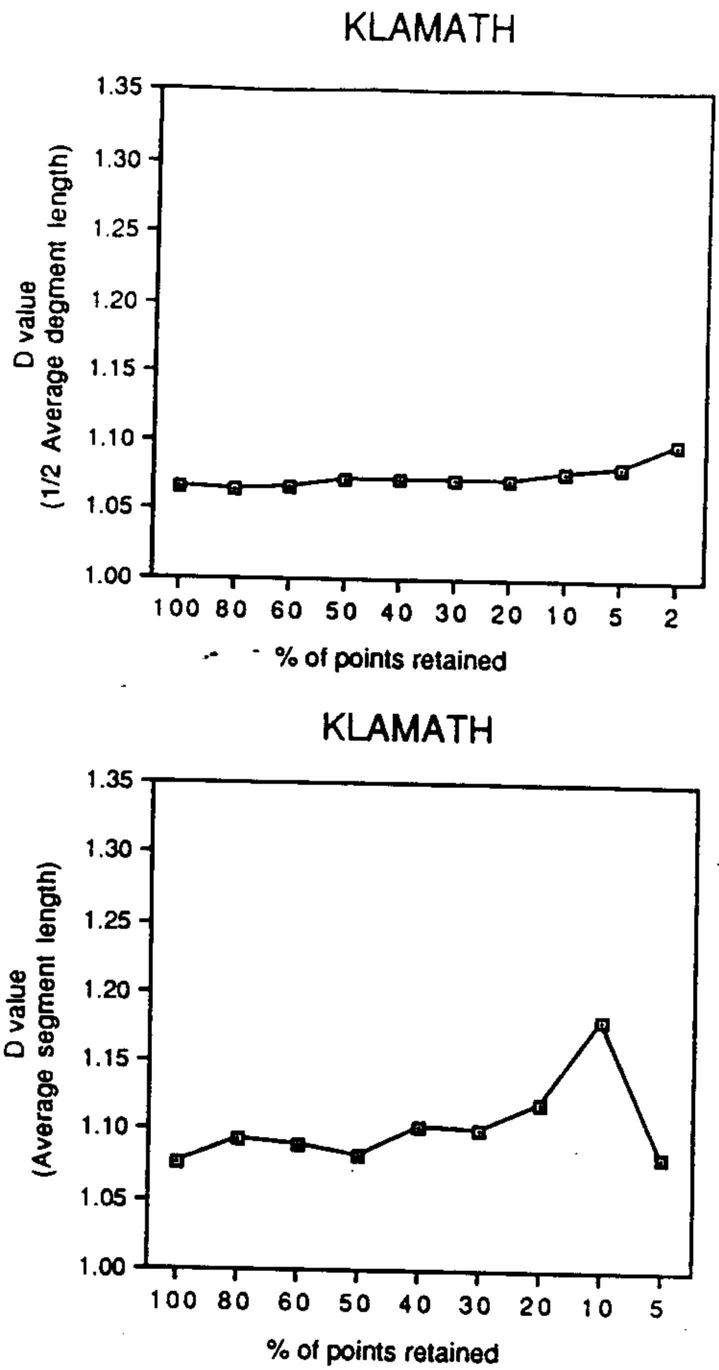
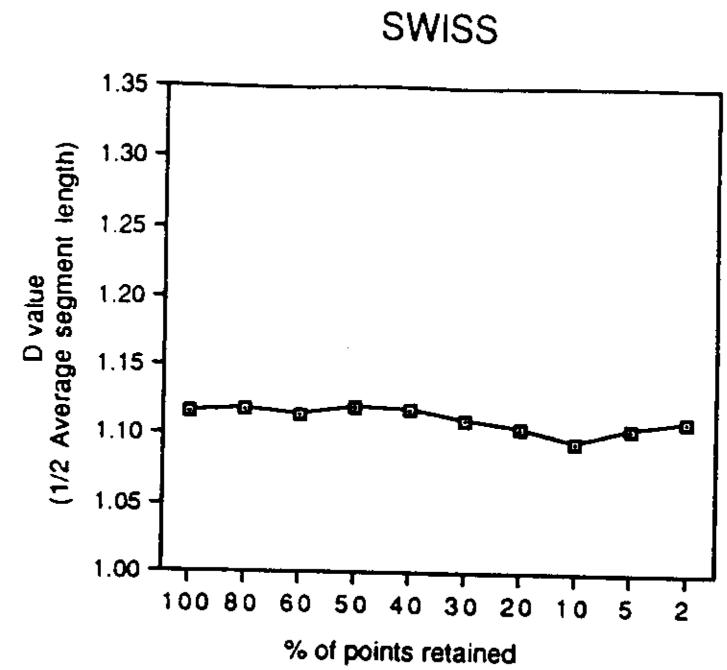


Figure 4.10 Stability of Fractal Dimension -- Klamath.

Upper graph: The initial step size: half of average segment length.
The terminating condition: until at least 5 full steps can be fitted in the line.

Lower graph: The initial stepsize: average segment length.
The terminating condition: until at least 3 full steps can be fitted in the line.



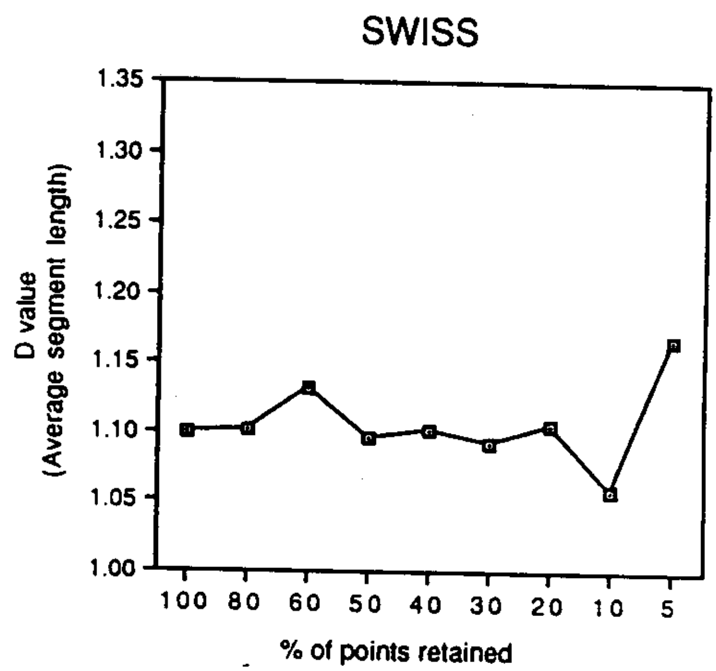
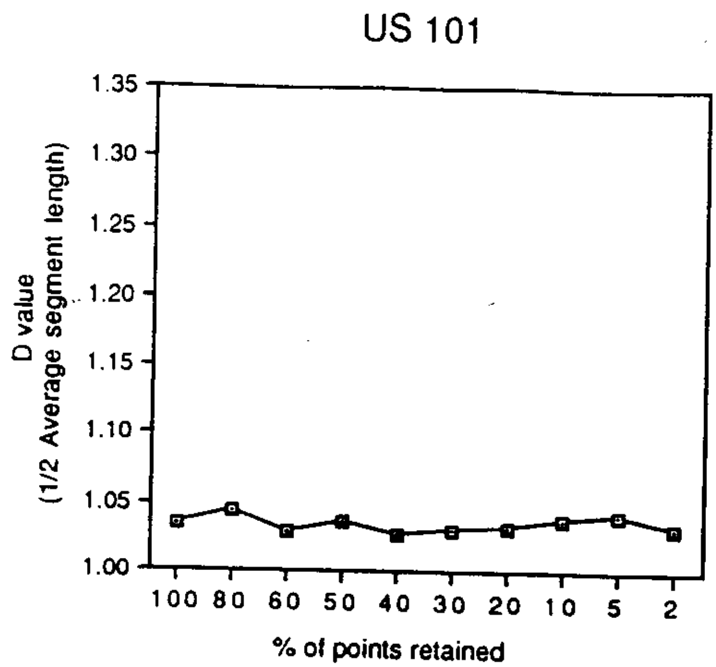


Figure 4.11 Stability of Fractal Dimension -- Swiss.

Upper graph: The initial step size: half of average segment length. The terminating condition: until at least 5 full steps can be fitted in the line.

Lower graph: The initial stepsize: average segment length. The terminating condition: until at least 3 full steps can be fitted in the line.



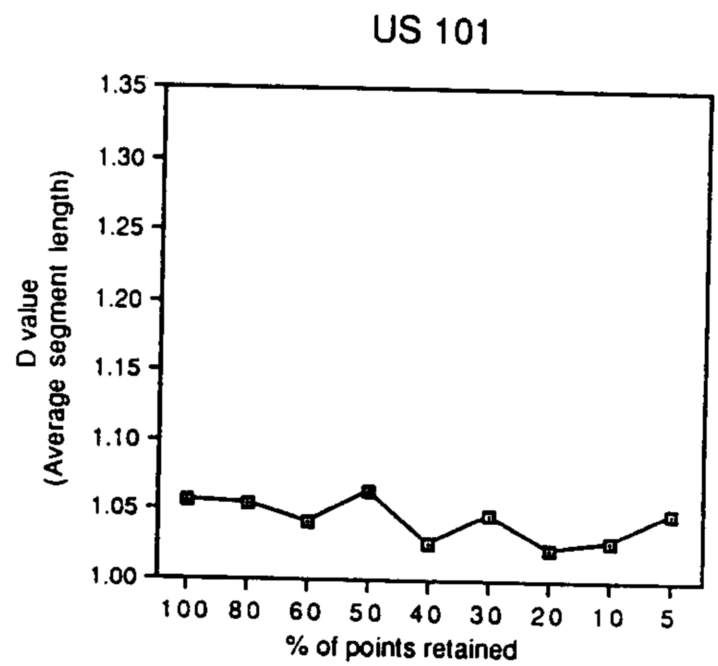


Figure 4.12 Stability of Fractal Dimension -- US101.

Upper graph: The initial step size: half of average segment length. The terminating condition: until at least 5 full steps can be fitted in the line.

Lower graph: The initial stepsize: average segment length. The terminating condition: until at least 3 full steps can be fitted in the line.

Eastman (1985) proposed a method of estimating fractal dimension of a line in a single pass by looking at the intersegment angles in isolation. His argument was that the angles between the triplets of points determine the sinuosity of a line, and that the sinuosity is "(...) to be directly related to fractal dimension under the assumption that random fractals arise from the statistical replication of a 'characteristic angle"' (Eastman, 1985). In the absence of any established mathematical relationship between Eastman's method and any other 'classical' one, as evidence of the fact that his method does measure fractal dimension, empirical data were presented that showed strong correlation between the results obtained from walking divider method and single-pass angular measure. The following formula for estimating fractal dimension was provided by Eastman:

$$D = \frac{\log(2)}{\log(2) + \log(\sqrt{K})}$$

$$K = \sum \frac{\left[\frac{c}{a+b}\right]^2}{n}$$

where a -- distance from point i to i+1,
b -- distance from point i+1 to i+2,
c -- distance from point i to i+2,
n -- number of intersegment angles.

Eastman's measure was employed with high hopes that it would provide a link between angular and fractal measures. Figure (4.13) shows the results obtained. When compared with Figure (4.5) with the values for the average angularity, it is surprising how highly these two measures are correlated. What this indicates is that Eastman's method does measure the average angular variation of a line, but the conclusion that by the same token it also estimates the fractal dimension seems unwarranted. There may be many measures of line geometry which may be forced to reach values that would fall between 1 and 2 (incidentally, Eastman's measure for MDI 2% is almost 2.2), and have their values within the range of some estimates of fractal dimension. But that does not prove any logical connection between the two. What fractal dimension is based on is the concept of self-similarity which is checked for by either looking at a feature in various scales, or by changing the scale of measurement on one feature under examination. Eastman's measure ignores the problem of self-similarity completely. As the results of Eastman's measure are almost identical with the average angularity, they were not included in the statistical analysis.

The reservations voiced about the sensibility of explaining the relationship between the step size and the total length by fitting a regression line can only be reinforced when plots in Figure 4.14 are considered. They are in essence Richardson's plots, where instead of the dividers step size, the average segment length is used. Buttenfield (1989) argued the validity of this method. For all the lines the curvilinear shape of the Richardson's plot is very well pronounced.

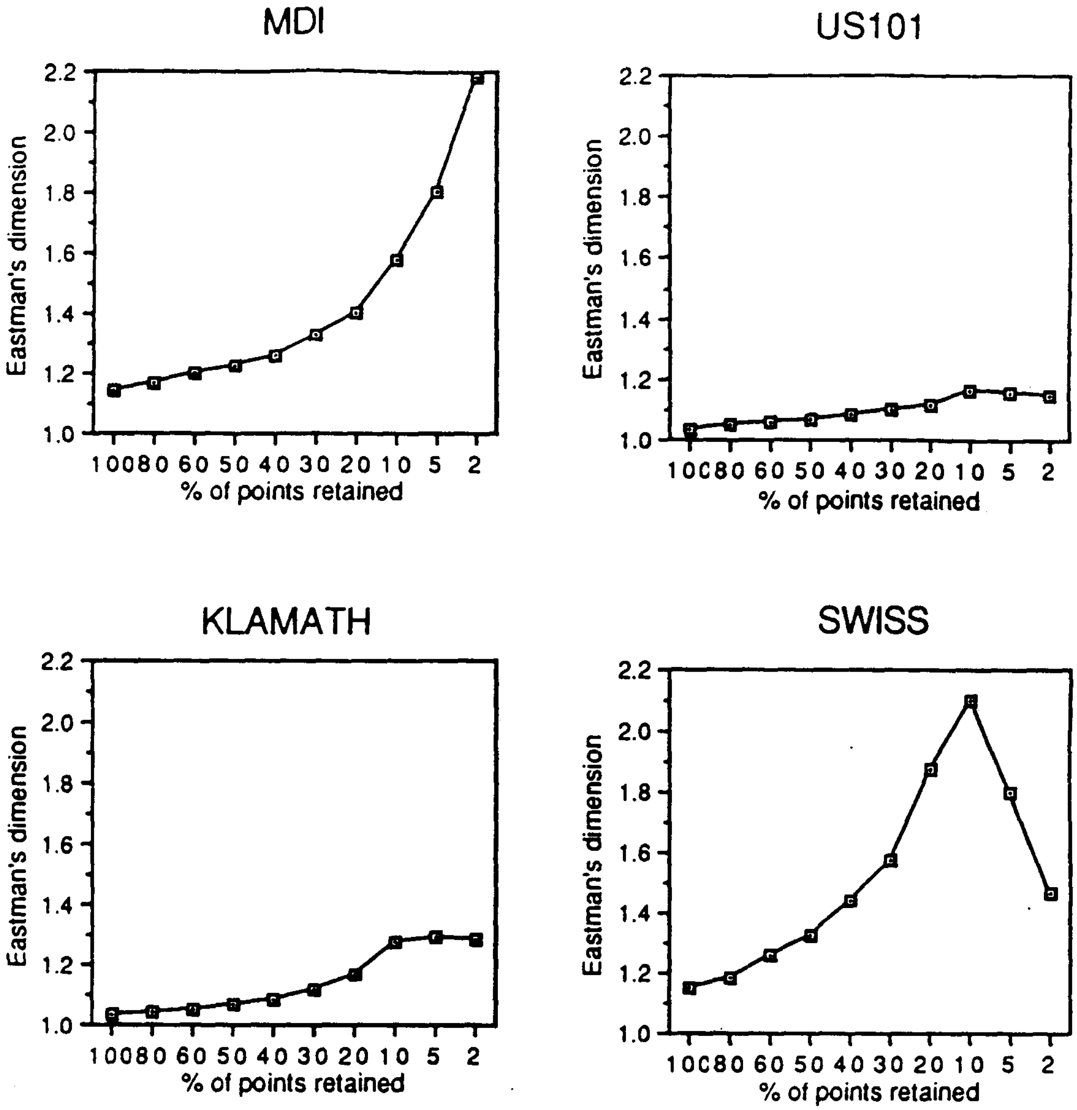


Figure 4.13 Eastman's Dimension.

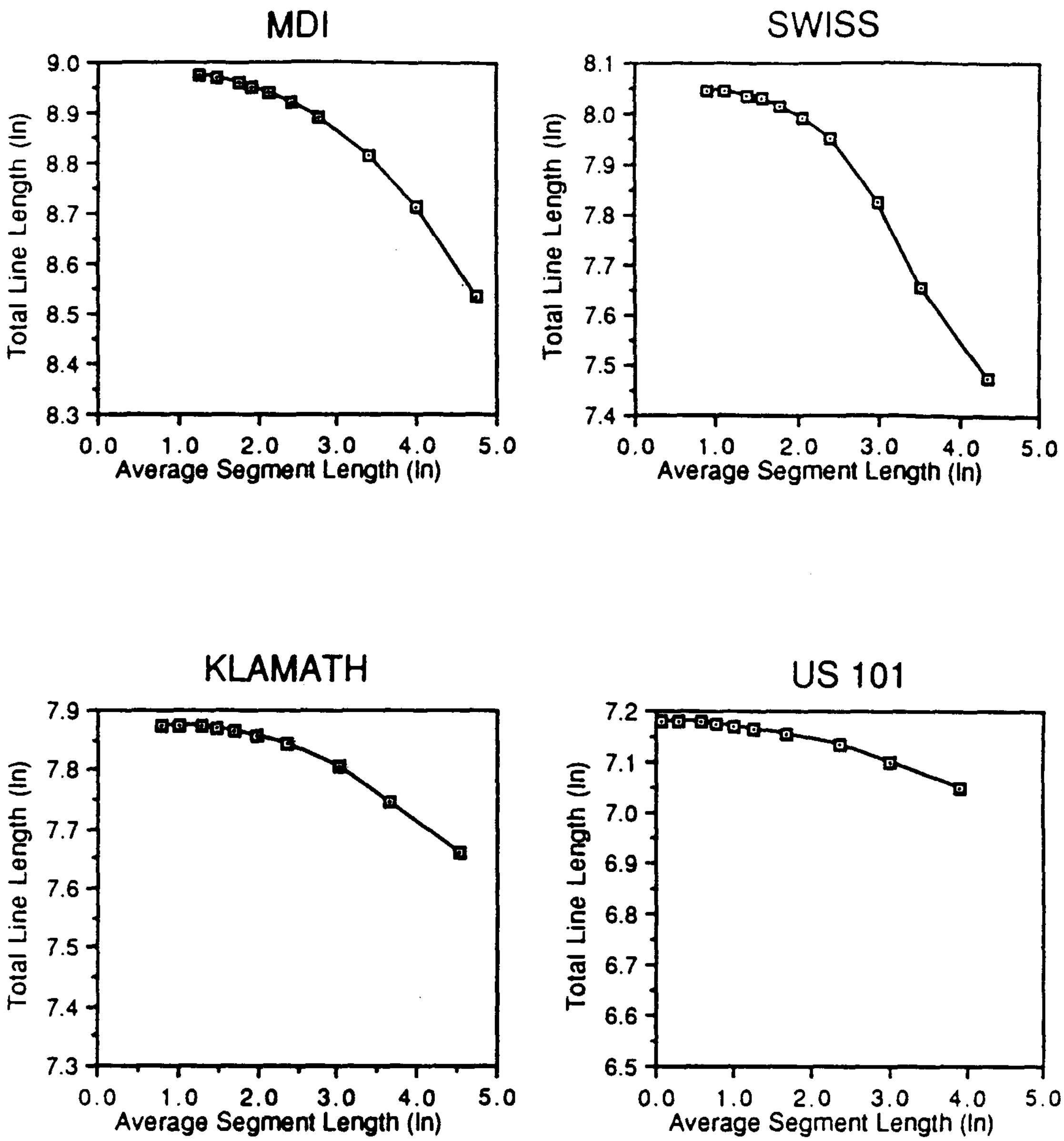


Figure 4.14 Rate of Decrease in Length.

TABLE 4.6. PRINCIPAL COMPONENTS ANALYSIS -- KLAMATH

EIGENVALUES OF THE CORRELATION MATRIX: TOTAL = 8 AVERAGE = 1

	1	2	3	4
EIGENVALUE	4.487	1.947	1.078	0.377
DIFFERENCE	2.540	0.869	0.701	0.306
PROPORTION	0.561	0.243	0.135	0.047
CUMULATIVE	0.561	0.804	0.939	0.986

	5	6	7	8
EIGENVALUE	0.071	0.030	0.007	0.002
DIFFERENCE	0.041	0.023	0.004	
PROPORTION	0.009	0.004	0.001	0.000
CUMULATIVE	0.993	0.999	0.999	1.000

FACTOR PATTERN

	FACTOR1	FACTOR2	FACTOR3	FACTOR4
AVL	0.852	0.491	-0.041	0.140
CVAVL	-0.125	0.037	0.974	0.182
EV	-0.692	0.687	-0.066	-0.118
CVEV	0.679	-0.585	-0.246	0.357
AVA	0.943	-0.141	0.189	-0.182
CVAVA	0.579	0.791	-0.082	0.098
CURV	0.875	-0.269	0.133	-0.373
FRAC	0.899	0.411	0.033	0.022

	FACTOR5	FACTOR6	FACTOR7	FACTOR8
AVL	0.080	0.006	0.049	-0.023
CVAVL	0.008	0.001	0.005	0.007
EV	0.170	-0.011	0.001	0.022
CVEV	0.082	0.004	0.000	0.024
AVA	0.120	0.057	-0.045	-0.012
CVAVA	-0.111	0.089	-0.013	0.016
CURV	-0.030	-0.001	0.038	0.021
FRAC	-0.029	-0.135	-0.026	0.003

VARIANCE EXPLAINED BY EACH FACTOR

FACTOR1	FACTOR2	FACTOR3	FACTOR4
4.487	1.947	1.078	0.377

FACTOR5	FACTOR6	FACTOR7	FACTOR8
0.071	0.030	0.007	0.003

TABLE 4.7. PRINCIPAL COMPONENTS ANALYSIS -- MDI

EIGENVALUES OF THE CORRELATION MATRIX: TOTAL = 8 AVERAGE = 1

	1	2	3	4
EIGENVALUE	4.597	2.347	0.676	0.321
DIFFERENCE	2.250	1.671	0.356	0.283
PROPORTION	0.575	0.293	0.085	0.040
CUMULATIVE	0.575	0.868	0.953	0.993

	5	6	7	8
EIGENVALUE	0.038	0.014	0.003	0.001
DIFFERENCE	0.024	0.009	0.005	
PROPORTION	0.005	0.002	0.001	0.000
CUMULATIVE	0.998	0.999	0.999	1.000

FACTOR PATTERN

	FACTOR1	FACTOR2	FACTOR3	FACTOR4
AVL	0.920	-0.262	0.255	0.104
CVAVL	-0.785	0.344	0.421	0.273
EV	-0.172	-0.944	0.263	-0.073
CVEV	0.650	0.675	-0.302	0.165
AVA	0.907	0.150	0.372	0.089
CVAVA	0.934	-0.230	-0.067	-0.218
CURV	0.984	0.093	0.107	0.077
FRAC	-0.121	0.852	0.344	-0.374

	FACTOR5	FACTOR6	FACTOR7	FACTOR8
AVL	-0.031	-0.086	-0.000	0.007
CVAVL	0.108	-0.002	-0.004	0.002
EV	-0.003	0.050	0.017	0.012
CVEV	0.005	0.031	0.003	0.016
AVA	-0.063	0.032	-0.036	-0.005
CVAVA	0.140	0.000	-0.025	0.000
CURV	0.035	0.017	0.056	-0.009
FRAC	-0.016	-0.005	0.012	0.005

VARIANCE EXPLAINED BY EACH FACTOR

FACTOR1	FACTOR2	FACTOR3	FACTOR4
4.597	2.347	0.676	0.321

FACTOR5	FACTOR6	FACTOR7	FACTOR8
0.038	0.014	0.005	0.001

TABLE 4.8. PRINCIPAL COMPONENTS ANALYSIS -- SWISS

EIGENVALUES OF THE CORRELATION MATRIX: TOTAL = 8 AVERAGE = 1

	1	2	3	4
EIGENVALUE	5.493	1.332	0.855	0.227
DIFFERENCE	4.161	0.478	0.628	0.155
PROPORTION	0.687	0.167	0.107	0.028
CUMULATIVE	0.687	0.853	0.960	0.988

	5	6	7	8
EIGENVALUE	0.072	0.015	0.005	0.001
DIFFERENCE	0.057	0.010	0.004	
PROPORTION	0.009	0.002	0.001	0.000
CUMULATIVE	0.997	0.999	0.999	1.000

FACTOR PATTERN

	FACTOR1	FACTOR2	FACTOR3	FACTOR4
AVL	0.961	0.003	0.257	-0.010
CVAVL	-0.497	0.302	0.802	-0.118
EV	0.947	-0.194	0.078	-0.209
CVEV	-0.646	0.735	-0.074	0.037
AVA	0.766	0.615	-0.023	0.112
CVAVA	0.912	-0.257	0.263	0.158
CURV	0.941	0.264	0.016	0.202
FRAC	-0.834	-0.384	0.250	0.297

	FACTOR5	FACTOR6	FACTOR7	FACTOR8
AVL	0.073	-0.023	0.050	-0.008
CVAVL	-0.048	-0.008	-0.007	0.003
EV	0.090	0.078	-0.006	0.007
CVEV	0.185	0.009	-0.007	-0.002
AVA	-0.131	0.060	0.001	-0.007
CVAVA	0.057	-0.018	-0.043	-0.010
CURV	0.013	-0.032	0.002	0.018
FRAC	0.032	0.058	0.016	0.002

VARIANCE EXPLAINED BY EACH FACTOR

FACTOR1	FACTOR2	FACTOR3	FACTOR4
5.493	1.332	0.855	0.227

FACTOR5	FACTOR6	FACTOR7	FACTOR8
0.072	0.015	0.005	0.001

TABLE 4.9. PRINCIPAL COMPONENTS ANALYSIS -- US101

EIGENVALUES OF THE CORRELATION MATRIX: TOTAL = 8 AVERAGE = 1

	1	2	3	4
EIGENVALUE	5.671	1.355	0.668	0.252
DIFFERENCE	4.316	0.687	0.416	0.213
PROPORTION	0.709	0.169	0.084	0.032
CUMULATIVE	0.709	0.878	0.962	0.993

	5	6	7	8
EIGENVALUE	0.039	0.014	0.001	0.000
DIFFERENCE	0.025	0.012	0.001	
PROPORTION	0.005	0.002	0.000	0.000
CUMULATIVE	0.998	0.999	1.000	0.000

FACTOR PATTERN

	FACTOR1	FACTOR2	FACTOR3	FACTOR4
AVL	0.936	0.287	0.082	-0.141
CVAVL	-0.962	0.197	0.123	-0.039
EV	0.970	0.188	-0.042	-0.129
CVEV	-0.963	-0.208	0.093	0.087
AVA	0.990	0.109	-0.068	-0.008
CVAVA	0.366	-0.347	0.745	0.040
CURV	0.829	-0.357	-0.200	0.374
FRAC	-0.083	0.945	0.184	0.253

	FACTOR5	FACTOR6	FACTOR7
AVL	0.106	-0.043	-0.016
CVAVL	0.116	0.071	-0.004
EV	0.012	0.048	0.023
CVEV	0.088	-0.056	0.020
AVA	0.043	-0.019	0.011
CVAVA	-0.014	0.009	0.000
CURV	0.057	0.025	-0.004
FRAC	-0.028	-0.008	0.002

VARIANCE EXPLAINED BY EACH FACTOR

FACTOR1	FACTOR2	FACTOR3	FACTOR4
5.671	1.355	0.668	0.252

FACTOR5	FACTOR6	FACTOR7
0.039	0.014	0.001

4.5 Comparison of the Measures.

The final step of the analysis was to perform the principal components analysis on all four groups of lines in order to find out what is the structure of correlations between the complexity measures. Each group had ten observations (levels of simplification) and eight variables (measures). The fact that the observations are not independent was accepted, as the goal of the analysis was not parameter estimates, but descriptive factor loadings (Fotheringham, personal communication, 1989).

The results of the principal components analysis, performed in SAS, are included in Tables 4.6 through 4.9. For Klamath the first factor accounts for 56% of variation in the data, but the first three factors together account for 94%. The variables which load most strongly are average angularity, fractal dimension, curvilinearity ratio and average segment length onto the first factor, and error variance and its coefficient of variation onto the second one. The third factor is dominated by the coefficient of variation of average segment length. In the case of MDI the results are similar in terms of the overall percentage of variation explained by the factors -- 57.5% by the first one, and 95% by the first three. The loadings of the factors are different though. All three angular measures and average segment length load on the first one. Error variance and fractal dimension are most prominent in the second factor.

In both cultural features the first factor accounts for more variation in the data than in the case of natural features. For Swiss the first factor accounts for 69% (the first three for 96%) and for US101 the first for 71% (the first three also 96%). Patterns of factor loadings are not quite the same. In Swiss it is average segment length, error variance, curvilinearity ratio, coefficient of variation of average angularity and fractal dimension which load high on the first factor, with coefficient of variation of error variance on the second, and coefficient of variation of average segment length on the third. In the case of US101 all measures but fractal dimension and coefficient of variation of average angularity load up on the first factor, with the former dominating the second factor, and the latter the third factor.

Overall, the principal components results do not give clear picture of relationships between the measures, especially between fractal dimension and the others. Three measures: average segment length, average angularity and curvilinearity ratio load highly on the first factor in all four cases. Error variance loads highly on the second factor for both natural features. So some degree of statistical redundancy may be observed between those measures.

4.6 Summary.

This chapter provided discussion of the analysis performed on the test data set using eight measures of geometric shape of lines. In the first part the results were interpreted from the plots of the measures against the progression of simplifications for each of four sets of lines. Next, the discussion of tests of three methods of estimating fractal dimension was given. From the methods tested, the walking divider method (structured walk) was selected to be incorporated in further analysis. The other two methods, the hybrid walk and Eastman's angular measure, were found inappropriate for the purposes of the study. The final part contained the discussion of principal components analysis, which was performed on the results of the eight measures to find out the structure of correlations between them. More general conclusions of the results presented above can be found in the following, final chapter. Also some recommendations for future research directions involving measurements of shape cartographic linear features are proposed there.

5. CONCLUSIONS.

The purpose of the work presented in the previous four chapters was to examine the relationships between a number of measures of geometric shape of lines on a test data set constructed especially for the project. The measures selected for the analysis covered the main characteristics of line shape, such as angularity, density of detail and relative length, as well as fractal dimension. Finally, the structure of correlations between all measures was examined using principal components analysis.

The interpretation of behavior of individual measures over the set of test lines revealed some interesting characteristics of the angular measures. In the case of all lines the average angularity demonstrated an increasing trend which at some point had an abrupt reverse. That point could serve as a threshold, guiding the user on what the maximum reasonable level of simplification would be for a given feature. The other angular measure, the curvilinearity ratio, seems to give one a good feel of the shape of line in terms of what kind of sinuosity is prevailing (e.g. sharp turns, smooth meanders).

The results of the principal components analysis revealed two things. A strong correlation was observed between one length measure (average segment length) and two angular measures (average angularity and curvilinearity ratio), which in all cases loaded highly on the first factor. All three measures exhibit similar increasing trend with progressive simplification. The trend is obvious in the case of the average segment length, but it was interesting to observe it do clearly in the case of the other two. This finding confirms the notion that the characteristic points preserved by Douglas-Peucker algorithm have high angular change associated with them, and at the same time the line tends to charge the direction of its trend line at those points.

A second finding of the principal components analysis is the lack of a visible relationship between fractal dimension and the remaining measures. For the lines following contours of the terrain (Klamath and Swiss) fractal dimension loaded highly onto the first factor), and for the lines following the contour lines (Mount Desert Island and US101) it loaded onto the second one. It would seem in this case that the configuration of the terrain has more influence on the measure than the origins of the features (natural or cultural).

Initially it was hoped that some link between the angular measures and fractal dimension would be provided by Eastman’s single-pass angular measure, which was supposed to give estimates of fractal dimension. The method was found to actually measure average angularity of lines, and its possible correlation to fractal dimension (not found in the tests conducted in this project) was considered to be accidental.

Tests comparing two different ways of estimating fractal dimension using a walking divider approach showed that the use of the hybrid walk for lines with sparse distribution of points, relative to the step size, is inappropriate. The tests with the other method, the interpolating walking dividers (structured walk), showed that the estimates of fractal dimension are sensitive to the choice of the parameters controlling the measurement (initial step size, terminating condition, acceptable r2 value in linear regression). With the lack of established and well founded mathematically rules which govern the selection of those parameters, the question is as to how confident can we be about the results of that method.

More general doubts about usefulness of fractal concept in cartographic generalization may also be drawn. The discussion to what extent natural phenomena are self-similar or scale-dependent has been alive* for quite some time now (Mandelbrot, 1967; Carpenter, 1980; Buttenfield, 1989). It has been established that natural features are not self-similar, even in the statistical sense, in their full extent, but that they might be self-similar in some limited ranges of scale.

Having that in mind, one may start wondering about the applicability of fractal dimension as a measure of complexity in the context of cartographic generalization. What good is a measure which is applicable only to some features, or parts of some features? How will application of such a measure facilitate further automation of line generalization process? It seems that instead of looking for self-similar fragments of features, for the purposes of cartographic processing of lines, it would be much safer to assume full scale-dependence at all scales. The fractal concept may perhaps be well applied in other fields where, by definition, the area of interest is confined to a very small scale range e.g. powder technology, (Clark, 1986). In the context of looking for a comprehensive measure of line complexity fractal model clearly can not be applied as it is not a universal measure.

There is a definite need for studies using larger data sets to find further relationships between various shape measures, and establish which are of use and which are either trivial or redundant. Once this is established, the combination of the measures could serve in line generalization (not only simplification) algorithms as 'sensors' of changes in geometric shape of a line, which would lead to automatic adjustment of appropriate tolerance values.

LITERATURE CITED

Andrle, R. and Abrahams, A.D., 1989. Fractal techniques and the surface roughness of talus slopes. Earth Surface Process and Landforms, vol.14, pp.197-209.

Attneave, F., 1954, Some informational aspects of visual perception. Psychological Review, vol.61, pp.183-193.

Austin, R.F., 1984, Measuring and comparing two-dimensional shapes. in Gale & Willmott (eds) Spatial Statistics and Models, Holland: D.Reidel Publishing Co.

Bachi, R., 1973, Geostatistical analysis of territories. Proceedings, 39th Session, Vienna, Bulletin of the International Staistical Institute, vol.4, pp.121-132.

Ballard, D.H., 1981, Strip trees: A hierarchical representation for curves. Communications of the Association for Computing Machinery, vol.24, No.5, pp.310-321.

Boyce, R.and Clark, W.A.V., 1964, The concept of shape in geography. Geographical Review, vol.54, pp.561-572.

Brassel, K.E. and Weibel, R., 1988, A review and framework of automated map generalization. International Journal of GIS, vol.2, No.3, pp.310-321.

Brophy, D.M., 1973, An automated methodology for linear generalization in thematic cartography. Proceedings, ACSM Conference, 33rd Annual Meeting, Washington, DC, pp.300-314.

Bunge, W., 1966, Theoretical Geography. Lund Studies in Geography, Series C, General and Mathematical Geography No.1, Gleerup, Lund.

Burrough, P.A., 1985. Fakes, facsimiles and facts: fractal models of geophysical phenomena. Science and Uncertainty, ed. S. Nash, IBM Science Reviews Ltd.

Buttenfield, B.P., 1984, Line structure in graphic and geographic space. Unpublished Ph.D. Dissertation, University of Washington.

Buttenfield, B.P., 1986. Digital definitions of scale-dependent line structure. Proceedings-AUTO- CARTO London, vol.1, pp-497-506.

Buttenfield, B.P., 1987. Automating the identification of cartographic lines. The American Cartographer, vol.14, No.1, pp.7-20.

Buttenfield, B.P., 1989. Scale-dependence and self-similarity in cartographic lines, Cartographica, vol. 26, No.1, pp.79-100.

Carpenter, L.C., 1981, Computer rendering of fractal curves and surfaces. Proceedings, ACM SIGGRAPH Conference, Seattle, Washington.

Carstensen, W., 1989, Fractal self-similarity in cartographic generalization. The American Cartographer, vol.16, No.3, pp. 181-189.

Clark, N.N., 1986, Three techniques for implementing digital fractal analysis of particle shape. Powder Technology, vol.46, No.1, pp.45-52.

Deveau, T.J., 1985, Reducing the number of points in a plane curve representation. Proceedings, AUTO-CARTO 7, pp.152-160.

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, vol.10, No.2, pp. 112-122.

Dutton, G., 1981. Fractal enhancement of cartographic line detail. The American Cartographer, vol.8, No.1, pp.23-40.

Eastman, J.R., 1985, Single-pass measurement of the fractal dimension of digitized cartographic curves. Paper presented at the Annual Meeting of the Canadian Cartographic Association, New Brunswick.

Goodchild, M.F. and Mark, D.M., 1987. The fractal nature of natural phenomena. Annals of the Association of American Geographers, vol.77, No.2, pp. 265-278.

Griffith, D.A.,O'Neill, M.P., O'Neill, W.A., Leifer, L.A., Mooney, R.G., 1986, Shape indices: useful measures or red herrings? Professional Geographer, vol.38, No.3, pp.263-270.

Jenks, G.F., 1981, Lines, computers, and human frailties. Annals of the Association of American Geographers, vol.71, No.1, pp.1-10.

Jenks, G.F., 1989, Geographic logic in generalization. Cartographica, vol. 26, No.1, pp.27-42.

Johannsen, T.M., 1974, A program for editing and for generalizing operations (For derivation of a small scale map from digitized data in 1:50,000). in Erno Csati ed. Automation: The New Trend in Cartography, Budapest, Hungary, The Geocartotechnical Research Department, pp.131-138.

Krantz, S.G., 1989, Fractal geometry. The Mathematical Intelligencer, vol.11, No.4, Springer-Verlag, New York.

Lang, T., 1969, Rules for robot draughtsmen. Geographical Magazine, vol.62, No.1, pp.50-51.

Lo, C.P., 1980, Changes in the shape of Chinese cities, 1934-1974. Professional Geographer, vol.32, pp.173-183.

Longley, P.A. and Batty, M., 1989a. Fractal measurement and line generalization. Computers & Geosciences, vol.15, No.2, pp.167-183.

Longley, P.A. and Batty, M., 1989b. On the fractal measurement of geographical boundaries. Geographical Analysis, vol.21, No. 1, pp.47-67.

Mandelbrot, B.B., 1967, How long is the coastline of Britain? Statistical self-similarity and fractal dimension. Science, vol.154, pp.636-638.

Mandelbrot, B.B., 1982, The Fractal Geometry of Nature. San Francisco, W.H. Freeman and Co.

Marino, J.S., 1979, Identification of characteristic points along naturally occuring lines: An Empirical Study. The Canadian Cartographer, vol.16. No.1, pp.70-80.

Mark, D.M., 1989, Conceptual basis for geographical line generalization. Proceedings, AUTO-CARTO 9, vol.1, pp-68-77.

Mark, D.M. and Csillag, F., 1989, The nature of boundaries on 'area-class' maps. Cartographica, vol. 26, No.1, pp.65-78.

Massam, B.H., 1970, A note on shapes. Professional Geographer, vol.22, No.4, pp.197-199.

Massam, B.H., 1975, Location and Space in Social Administration. John Wiley, New York.

McMaster, R.B., 1986, A statistical analysis of mathematical measures for linear simplification. The American Cartographer, vol.13, No.2, pp.103-117.

McMaster, R.B., 1987, Automated line generalization. Cartographica, vol. 24, No.2, pp.74-111.

McMaster, R.B., 1989, Integration of algorithms in line generalization. Cartographica, vol. 26, No.1, pp.101-121.

Moellering, H. and Rayner, J.N., 1981, The harmonic analysis of spatial shapes using dual axis Fourier shape analysis (DAFSA). Geographical Analysis, vol.13, pp.64-77.

Moellering, H. and Rayner, J.N., 1982, The dual axis Fourier shape analysis of closed cartographic forms. The Cartographic Journal, vol.19, pp.53-59.

Muller, J.-C., 1986, Fractal and automated line generalization. The Cartographic Journal, vol.24, No.1, pp.27-34.

Opheim, H., 1981, Smoothing a digitized curveby data reduction methods. inJ.L.Encamacao ed. Eurographics '81, Amsterdam: North-Holland Publishing Company, pp.127-135.

Pierkal, J., 1958a, 0 dlugosci krzywych empirycznych. Zastosowania Matematyki, vol.3, No.3-4, pp.257-286.

Perkal, J., 1958b, Proba obiektywnej generalizacji. Geodezja i Kartografia, vol. 7, No.2, pp.130-142.

Reumann, K. and Witkam, A.P.M., 1974, Optimizing curve segmentation in computer graphics. International Computing Symposium, 1973, Amsterdam: North-Holland Publishing Company, pp.467-473.

Richardson, L.F., 1961, The problem of contiguity: An appendix to the statistics of deadly quarrels. General Systems Yearbook, vol.6, pp.139-187.

Shea, K.S. and McMaster, R.B., 1989, Cartographic generalization in a digital environment: When and how to generalize. Proceedings, AUTO-CARTO 9, pp.56-67.

Shelberg, M.C., Moellering, H. and Lam, N., 1982, Measuring the fractal dimensios of empirical curves. Proceedings, AUTO-CARTO 5, pp.481-490.

Steinhaus, H., 1949, 0 dlugosci krzywych empirycznych i jej porniarze, zwlaszcza w geografii. Sprawozdania Wroclawskiego Towarzystwa Naukowego 4, Dodatek 5.

Topfer, R. and Pillewizer, W., 1966, The principle of selection. Cartographic Journal, vol.3, No.1, pp.10-16.

Weibel, R. and Buttenfield, B.P., 1988, Map design for geographic information systems. Proceedings, GIS/LIS Conference, San Antonio, TX.

White, E.R., 1985, Assessment of line generalization algorithms using characteristic points. The American Cartographer, vol.12, No.1, pp.17-28.

Zycor, Inc., 1984, Manual and Automated Feature Displacement. Repert fot the U.S. Army Engineer Topographic Laboratories, Fort Belvoir, Virginia, 204 pages.