# PIECE BY PIECE: A METHOD OF CARTOGRAPHIC LINE GENERALIZATION USING REGULAR HEXAGONAL TESSELLATION

P. Raposo

Department of Geography, The Pennsylvania State University. University Park, Pennsylvania - paulo.raposo@psu.edu

**KEY WORDS:** Cartographic Generalization, Line Simplification, Scale-Specificity, Tessellation, Hexagons

**ABSTRACT:**

Several methods of automated line simplification exist, but most involve parameters selected arbitrarily or heuristically, with little or no reference to the scale change between original data and generalized output. Also, while routines such as the Douglas-Peucker algorithm achieve simplified line correlates by retention of characteristic points from the input line, little analysis has been devoted to whether those points remain characteristic at the generalization target scale. A new algorithm is presented based on regular hexagonal tessellation. Mosaics of equilateral hexagons are used to sample lines, where the hexagon width relates directly to target scale. Inside each hexagon tessera, input line vertices are collapsed to a single vertex, and the resulting set of points constitute simplified correlate lines appropriate for the generalized map scale. Hexagonal width is derived in relation to target scale in two ways: by applying the Radical Law, and by selecting measures pursuant to Tobler's ideas on spatial resolution. Results yield a useful scale-specific method of line generalization.

## INTRODUCTION

Line simplification is arguably one of the most important generalization operators, since almost every map includes some form of lines. Yet the theory of this particular operator seems to exist in an unsatisfactory state in the literature, given the profusion and diversity of pieces written about it and the continued concern with unresolved geometric and practical issues. Not the least of reasons that may contribute to this is the relative lack of consensus regarding the definitions, requirements, and processes of the broader concept of generalization, but it is here suggested that most of the reason derives from the fact that line simplification is rarely considered in direct relation to map scale, which, in the author's opinion, is the most important factor driving the need to generalize at all. Rather, attention has gone to and significant advances in theory have been made in measures of geometric difference between original and generalized lines. This has resulted in importance being placed on characteristic points, without enough consideration of how these behave at smaller and smaller scales.

This paper is organized in two parts. The first part discusses and comments on current theory in line generalization, including aspects such as characteristic points, effects and measurements of simplification, and scale and resolution. The second introduces two algorithms developed by the author in response to outstanding issues in the line generalization literature, displays some preliminary results, and discusses present limitations and future development of the methods.

## 1. LINE GENERALIZATION THEORY

### 1.1 Characteristic Points

With the exception of those defined by functions such as Bezier curves, digital cartographic lines are ontologically different from lines drawn by manual cartographers, as they are composed of sequences of coordinates forming stations joined by straight line arcs. Essentially, a vertex in a cartographic line represents a chance for that line to change direction in any magnitude of degrees, and to continue as a straight segment for any distance to the next vertex. Collective measures of distance between vertices have been elegantly modeled by Peucker (1976) as frequency, a concept that allows for the simplification of a line by the reduction of that frequency according to a bandwidth, as occurs in the Douglas-Peucker algorithm (1973).

Placement of vertices in many vector data sets has been done by human digitization, and thus, whether guided by cartographic standards or not, reflects someone's approximation of what may be an infinitely complex natural line. This model is accepted on the assumption that selecting good points permits the capture of any line (Jenks, 1981). Research has shown that there is a tendency towards structure in the choices people make when asked to approximate complex figures with a finite number of points (Marino, 1979). The assertion of psychologist Fred Attneave (1954) that characteristic points exist in complex line drawings, and that these exist at those points of greatest directional change (salient points, apexes of curves, sudden angles, etc.) has been cited by many cartographic scholars concerned with vector generalization. Yet Attneave acknowledges nearly immediately after making this assertion that scale profoundly influences our perception of characteristic points: in describing his now famous 38-point approximation of a sleeping cat, he goes on to say that one could define a characteristic point at the tip of every strand of fur if one observed so closely. Characteristic points, then, are chosen as those that seem to compose an effective gestalt of the object at the viewing scale, and are thus scale-dependent.

Characteristic points in digital cartographic lines and their retention in line simplification have become a major aspect of the generalization literature. Jenks (1979) defines characteristic points as being of two types: those that are relevant to perceived form (e.g. curve apexes) and those that are given particular geographic importance (e.g. where a river passes under a bridge). This division is helpful, because it helps to clarify those processes of line simplification that are necessarily discretionary and those that may be effectively treated by the objective application of automated algorithms. However, many authors have not made this distinction while advocating strongly for the need to retain characteristic points in line simplifications. Emphasis on characteristic points is demonstrated by the manner in which it dovetails with the definition of line simplification given by many authors: the removal of vertices from a line to arrive at a representative subset of vertices (McMaster and Shea, 1992; McMaster and Veregin, 1997; Veregin, 1999; White, 1985). While many authors posit that simplification occurs properly when the subset of retained points is composed of characteristic points from the original line, only a few discuss characteristic points, before or after simplification, in explicit relation to scale change (Buttenfield, 1989; Cromley and Campbel, 1992; Dutton, 1999). Still others relate characteristic points to neighborhood-level line complexity and suggest ways in which such points can define simplification between them (Buttenfield and McMaster, 1991; Plazanet, 1995).

### 1.2 Simplification

Line simplification is an expressly spatial operator, concerned with the alteration of feature geometry for representation at smaller scales. Simplified lines will necessarily differ in geographic position at various places along their length from their original counterparts. McMaster's (1987) suite of geometric measures between original lines and their simplified correlates permits objective analysis and measurement of generalization degrees and positional error introduced by the simplification process. Simplification methods that retain a subset of the original line vertices are often advocated for on the grounds that they preserve some degree of positional accuracy. While many authors rightly observe the alteration of line position in absolute Euclidean space, and by corollary the positional errors introduced in line simplification, few consider whether given displacement values are acceptable or not at intended generalization scale, a matter best considered with regard to visual resolution at specific scales.

Perhaps a more serious concern is the manner in which a simplified line may acquire erroneous topological relationships with other map features (e.g. a river jumping to the opposite side of a point-feature city). Methods of dealing with this potential problem seem at present to necessitate human editing, but efforts by some have increased understanding of the problem toward efficient solutions (Saalfeld, 1999; Shi and Cheung, 2006). Here Jenks' distinction between characteristic points that are deemed geographically important and those that are geometrically descriptive may be useful in guiding edits, with greatest concern focused on the former.

Several authors have suggested the use of some or all of McMaster's displacement measures in parameterization or optimization of line simplification (Cromley and Campbell, 1992; Jenks, 1989; Veregin, 1999). One motivation for this is likely to be that input parameters determining degrees of generalization for most simplification algorithms, such as the Douglas-Peucker or Visvalingam-Whyatt (1993) methods, do not relate lucidly to displacement measures taken after simplification. Further, most of these input parameters do not relate to target map scale.

### 1.3 Scale and Resolution

With few exceptions (such as Cromley and Campbell, 1992; Dutton, 1999; Li, 1996), scale is little discussed in direct relation to other aspects of line simplification. Töpfer's Radical Law (Töpfer and Pillewizer, 1966) remains the most cogent treatment of generalization of any kind with direct relation to scale. The Law is a series of equations, each accompanied by certain constant and exponent values that tailor it to a particular map feature type. Each equation ingests the number of features of the relevant type at the starting map scale, and expresses how many of these features should be retained upon reduction to a specified scale. As several scholars have noted, the Law provides a rational guide to the quantity of features to retain, but does not address which features should be retained.

Tobler (1987) defines average spatial resolution as "the content of the geometric domain of observation divided by the number of observations, all raised to the power one over the spatial dimension", where the domain is a length, area or volume in one, two or three dimensions, respectively. Working with the measure of the smallest mark that can be made on a map as approximately a half-millimeter, Tobler relates resolution to map scale with a simple rule: "divide the denominator of the map scale by 1,000 to get the detectable size [of features drawn to scale] in meters. The resolution is one half of this amount". Resolution is half since, from sampling theory, an object can only be certainly detected if the sampling frequency is half its width, thus ensuring the object can't pass undetected between samples. Tobler further suggests that a good rule of thumb is to use a sampling rate one-fifth the size of the smallest features one wishes to detect. (For example, at a scale 1:25,000, resolution is 12.5 m, detectable object width is 25 m, and rule-of-thumb detectable object width of 62.5 m.) Other authors have also described resolvable units: McMaster and Shea (1992) note that 0.02 mm at a viewing distance of 30 cm is about the smallest size of object the human eye can resolve, and they wisely suggest no cartographer should make marks so extremely small.

Li and Openshaw (1993) have suggested that generalization be carried out according to a "natural principal," found in the effect of resolution change in human vision as viewing distance increases. They assert that generalization should be undertaken primarily in response to reduction in map area as a consequence of scale change, and that scale change provides an "objective criterion that can be used in analytical algorithms to automate the generalization process" (1993). They declare that "the remaining problem is to determine how the scale change can best be linked to the degree of generalization needed to retain legibility".

## 2. GENERALIZATION USING HEXAGONAL TESSELLATIONS

### 2.1 Algorithm Description

The remainder of the paper describes two line generalization algorithms developed by the author. Motivation behind the development of these originated from the author's belief that map scale plays a defining role in determining the appropriateness of levels of generalization, and thus should be directly referenced in the generalization process itself, rather than used *a posteriori* to evaluate generalizations produced without careful scale specificity. Further motivation originated from the desire to devise algorithms with input parameters directly and objectively related to target scale and/or scale change.

The algorithms do not attempt to locate and retain characteristic points. On the belief that these are scale-specific, and therefore not necessarily appropriate for representing a given figure at smaller scales, the author instead adopts a position similar to that described in Buttenfield (1985), wherein all vertices along a cartographic line are considered equiprobable in position. Rather than rely on the retention of a critical point at the apex of a curve, for example, the algorithms presented below will retain the curve only if it remains large enough on the map after scale reduction to visually register at the map's calculated resolution.

The methods presented here are predicated on the notion that regular tessellations of equilateral hexagons at various resolutions (i.e. various hexagon widths, being the perpendicular distances from one side to the side opposite) can be used in line sampling strategies that capture the essential form of a line at varying levels of spatial detail. Essentially, the idea is to sample a line with uniform aerial frequency directly scaled to target map scale, and to collapse all input vertices in the neighborhood of a sample locale to a single vertex. If the distance between sampling locales is proportioned according to expectable visual resolution in the reduced map space at target scale, the collapsed vertices should constitute a generalized line with visually resolvable detail that maintains geometric form to the degree at which a viewer may expect, at target scale, to perceive it. Equilateral hexagons are used in a honeycomb configuration because of the radial symmetry inherent to hexagonal tessellations (i.e. distances from the centroid of one hexagon to any of its six topologically-adjacent neighbors are equal), which creates a spatially-uniform sampling strategy. Also, tessellations of equilateral hexagons, if considered along with the centroids of the

Figure 1. Generalized approximation. Input line vertices are in grey. Output vertices are located in each hexagon at the mean x-y coordinate pair calculated from the input vertices in that hexagon. These vertices, as well as the resulting line, are in black.

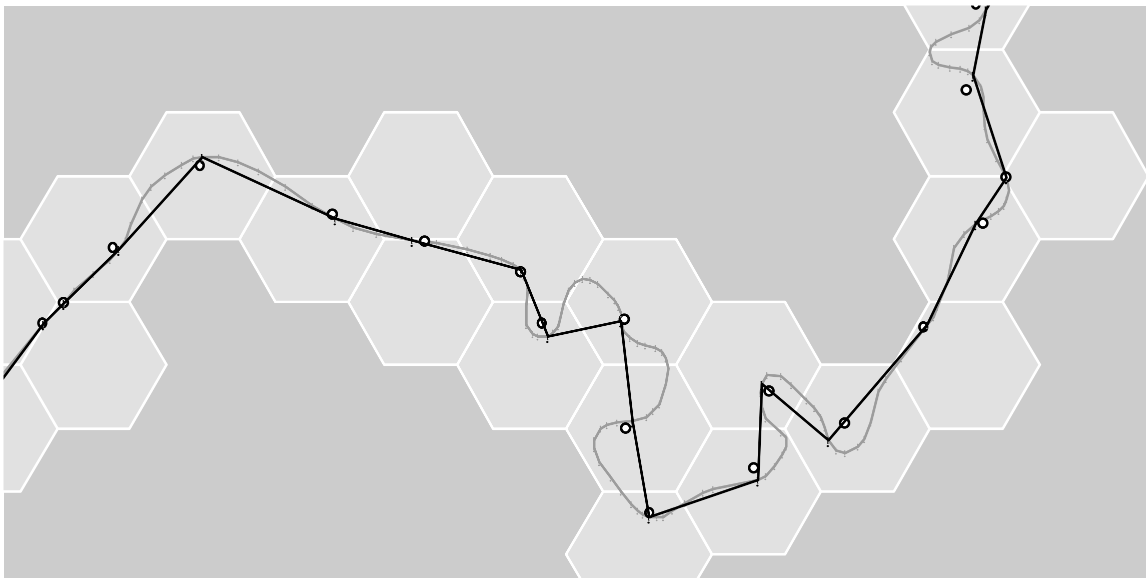

Figure 2. Simplification. Input line and vertices are in grey. Mean x-y coordinates in each hexagon calculated from the input vertices in that hexagon, drawn with white circles. The vertex among the input vertices in each hexagon closest to the mean x-y coordinate is selected; these, as well as the resulting line, are in black.

hexagons, also qualify as Dirichlet or Voronoi tessellations, meaning that any point in any given hexagon is closer to the centroid of that hexagon than to the centroid of any other hexagon.

The algorithms, in using scalable aerial units to perform local generalization, bear resemblances to those of Dutton (1999), Li and Openshaw (1993), and Perkal (1966). Differences to the work of these authors lie in the non-fixed positioning of the tessellation, and the ways in which tesserae width is related to target scale.

The first algorithm is a method of creating a generalized approximation of the input line, and is illustrated in Figure 1. The products of the first algorithm are considered approximations because they do not create a line from a subset of the input line vertices, but rather use new, derived points in each tessera. The second algorithm is a method of creating a simplified correlate of the input line, and is illustrated in Figure 2. (In both figures, only those hexagons from a continuous tessellation which intersect the input line are drawn for visual clarity.) The second algorithm does create a line from a subset of the input line vertices. The algorithms are essentially alike, except in the decision of how and where to collapse input points in each hexagon to a single vertex. Both algorithms calculate a spatial mean inside each hexagon for the set of input line vertices that fall within it: the first uses the x-y coordinate of that spatial mean as the single vertex representative of the input vertices in that hexagon, while the second selects the input vertex closest to the spatial mean as the single vertex to be retained in that hexagon.

The algorithms presented here can be categorized using schemes developed by other authors. McMaster and Shea (1992) identify five classes of line simplification algorithms, namely: 1) independent point algorithms, which do not consider line geometry or topological relationships, 2) local processing routines, which use neighboring points to select generalized points, 3) constrained extended local processing routines, which consider geometry beyond neighboring points in nearby line segments, 4) unconstrained extended local processing routines, which consider geometry beyond nearby line segments and are defined by geomorphic characteristics of the line, and 5) global routines, which consider whole lines and work iteratively. The algorithms presented here belong best in their second category. Also, in their present implementation, the algorithms operate irrespective of any other map features with which the lines will be drawn, and thus constitute an example of *in vacuo* generalization (Saalfeld, 1999); this is distinct from *en masse* or *en suite* generalization, which take topological relationships of the line and all other features into consideration, and which take topological relationships of the line and nearby features into consideration, respectively. It can be seen therefore that generalizations made using these algorithms may be subject to topological inconsistencies with other map features. While it is true that generalized lines may pass outside the set of hexagons which intersect the input line, this seems to be rare and to occur with small aerial displacement measures as compared to the area of one hexagon. While further testing on multiple input lines will better define this phenomenon, it is presently suggested that so long as other map features lay outside of the hexagons which intersect the input line, there is negligibly low chance they will acquire erroneous topological relationships with the line upon its generalization.

Two methods of determining hexagon sizing in relation to target map scale have thus far been pursued. The first is an application of the Radical Law. By this determination, hexagon width is defined such that the frequency of vertices in the output line reflects a linear decrease in the frequency of vertices in the input line, proportional to the difference in scale between input and target lines. The mean distance between each node in the input line is calculated, and the hexagon width is set as this value increased by the target and input data scale quotient:

$$w = \frac{\sum l}{n} \times \frac{t}{d} \tag{1}$$

where $w$ = hexagon width
$\Sigma\, l$= the sum of all arc lengths between vertices in the input line
$n$ = the number of all arcs between vertices in the input line
$t$ = the denominator of the target scale
$d$ = the denominator of the input data scale.

The second determination is based on Tobler's (1987) rule for spatial resolution. In applying the rule to the perception of successive vertices in a cartographic line, the author has chosen to select a resolution one half the rule-of-thumb detectable size. This value is taken for the hexagon width, which defines the sampling resolution of the resulting tessellation. Thus, for example, at a map scale of 1:250,000, detection width is 250 m, resolution is 125 m, rule-of-thumb detection width is 625 m, and the hexagon width for the algorithms described here is 312.5 m.

### 2.2 Preliminary Results and Future Work

Results from both algorithms, using both methods of hexagon width determination discussed above, are presented as a series of four in Figure 3. The input line is an arbitrary section of the Black River in Orleans County, Vermont, taken from the USGS National Hydrography Dataset (NHD), High resolution. This portion, measuring approximately 12.9 km in length, was chosen upon visual inspection for its sinuosity and the presence of at least two distinguishable levels of physical form (i.e. tight sinuous sections needing generalization at smaller scales, and larger trends in shape). The author hoped such complexity would test the abilities of the algorithms to reproduce both general form as well as local detail. The maps are projected in UTM. In all cases input scale was 1:24,000, and the lines were generalized to a target scale of 1:100,000. Hexagon widths, being the only input parameters, were set at 57.24 m for the examples pursuant to the Radical Law (A & B), and at 125 m for the examples pursuant to Tobler's resolution theory (C & D).

Results are presented here for visual comparison and demonstration; no attempt has yet been made to analyze the results quantitatively, though future work will be devoted to this. While at present evaluations are exclusively subjective, the approximation in line C, generated with a hexagon resolution persuant to Tobler's resolution theory and utilizing the mean x-y coordinates calculated in each hexagon, seems to generate the most aesthetically pleasing and geometrically lucid line.

Future work will involve testing the algorithms on many more cartographic lines representing features with broadly varying geomorphologies, such as rivers, shorelines and roads, as well as polygon boundaries such as state borders. Particular attention will go to testing the algorithms on input lines with widely-varying component arc lengths. Quantitative analysis and measurement of the products of the algorithms will also be performed and compared to the products of other, existing routines. It is presumed that the algorithms developed here will be best suited to lines representing naturally sinuous, complex features such as watercourses. Variations in the placement and orientation of the hexagonal tessellation are also possible, and work will explore the consequences of these.

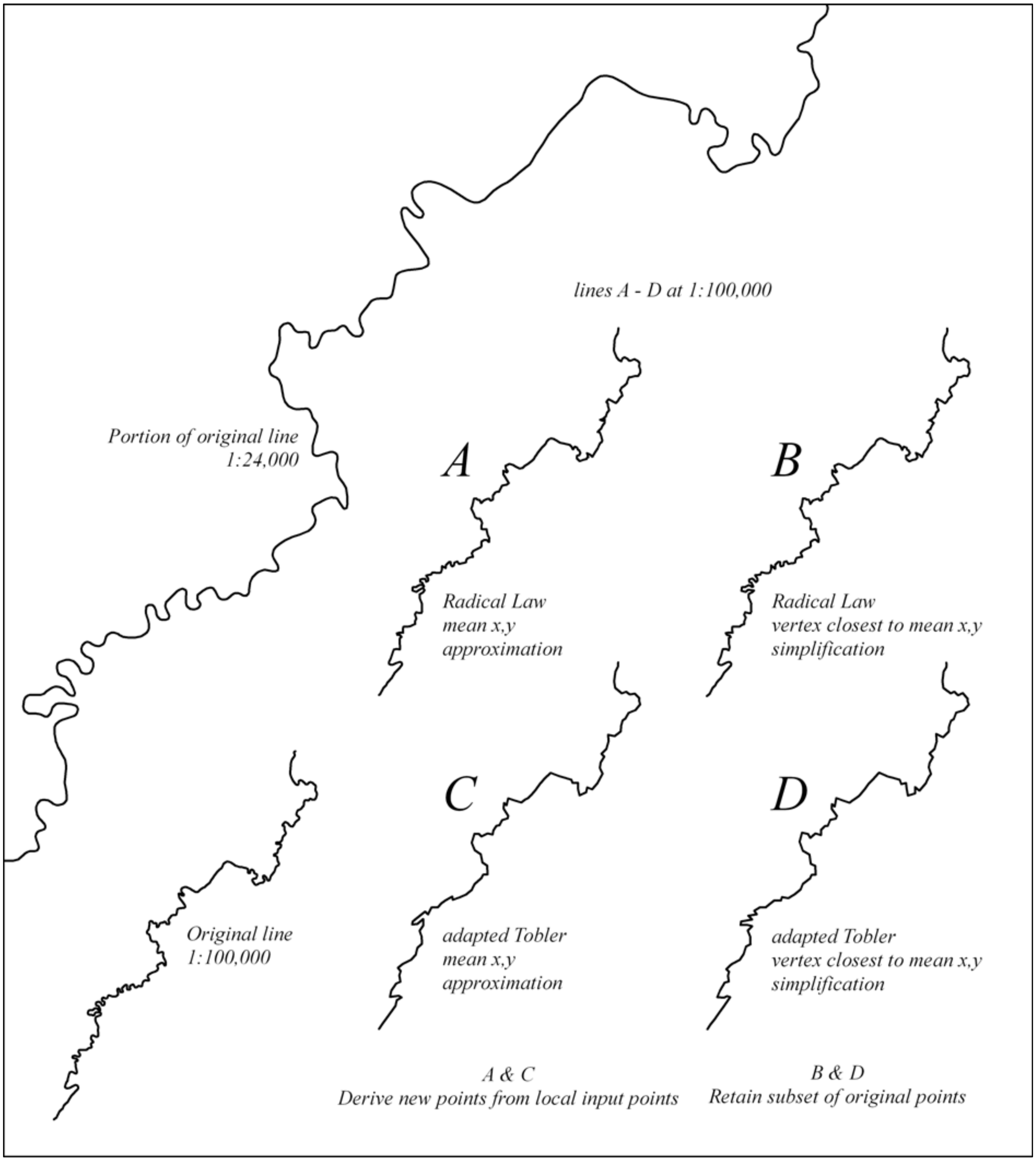


Figure 3. Results from application of both algorithms using both tessellation resolution (i.e. hexagon width) definition methods.

**REFERENCES**

Attneave, F., 1954. Some informational aspects of visual perception. *Psychological review*, 61(3), pp. 183-193.

Buttenfield, B. P., 1985. Treatment of the cartographic line. Cartographica: *The International Journal for Geographic Information and Geovisualization*, 22(2), pp. 1-26.

Buttenfield, B. P., 1989. Scale-dependence and self-similarity in cartographic lines. *Cartographica: The*

*International Journal for Geographic Information and Geovisualization*, 26(1), pp. 79-100.

Buttenfield, B. P. and McMaster, R. B., 1991. *Map generalization: making rules for knowledge representation*. Longman Group UK, London.

Cromley, R. and Campbell, G., 1992. Integrating quantitative and qualitative aspects of digital line simplification. *The Cartographic Journal*, 29(1), pp. 25-30.

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. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 10(2), pp. 112-122.

Dutton, G., 1999. Scale, Sinuosity, and Point Selection in Digital Line Generalization. *Cartography and Geographic Information Science*, 26(1), pp. 33-53.

Jenks, G. F., 1979. Thoughts on line generalization. AutoCarto. *Am. Congress on Surveying and Mapping*, pp. 209-220.

Jenks, G. F., 1981. Lines, Computers, and Human Frailties. *Annals of the Association of American Geographers*, 71(1), pp. 1-10.

Jenks, G. F., 1989. Geographic logic in line generalization. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 26(1), pp. 27-42.

Li, Z., 1996. Transformation of spatial representation in scale dimension: A new paradigm for digital generalization of spatial data. *International Archives of Photogrammetry and Remote Sensing*, 31(pp. 453-458.

Li, Z. and Openshaw, S., 1993. A natural principle for the objective generalization of digital maps. *Cartography and Geographic Information Science*, 20(1), pp. 19-29.

Marino, J., 1979. Identification of characteristic points along naturally occurring lines/an empirical study. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 16(1), pp. 70-80.

McMaster, R. B., 1987. Automated line generalization. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 24(2), pp. 74-111.

McMaster, R. B. and Shea, K. S., 1992. *Generalization in digital cartography*. Association of American Geographers, Washington, D.C.

McMaster, R. B. and Veregin, H., 1997. Visualizing Cartographic Generalization. AutoCarto. 13, pp. 174–183.

Perkal, J., 1966. An attempt at objective generalization. Michigan Inter-University Community of Mathematical Geographers, Discussion Paper, 10.

Peucker, T., 1976. A theory of the cartographic line. *International yearbook of cartography*, 16, pp. 134-143.

Plazanet, C., 1995. Measurement, characterization and classification for automated line feature generalization. AutoCarto. *American Congress on Surveying and Mapping & American Society for Photogrammetry and Remote Sensing*, pp. 59-68.

Saalfeld, A., 1999. Topologically Consistent Line Simplification with the Douglas-Peucker Algorithm. *Cartography and Geographic Information Science*, 26(1), pp. 7-18.

Shi, W. and Cheung, C., 2006. Performance evaluation of line simplification algorithms for vector generalization. *The Cartographic Journal*, 43(1), pp. 27-44.

Tobler, W. R., 1987. Measuring spatial resolution. International Workshop On Geographic Information Systems. *International Geographical Union (IGU); Geographical Society of China; National Natural Science Foundation of China; National Remote Sensing Center of China; International Development Research Center, Canada*, Beijing, China, 48, pp. 42-48.

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

Veregin, H., 1999. Line simplification, geometric distortion, and positional error. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 36(1), pp. 25-39.

Visvalingam, M. and Whyatt, J., 1993. Line generalisation by repeated elimination of points. *The Cartographic Journal*, 30(1), pp. 46-51.

White, E., 1985. Assessment of line-generalization algorithms using characteristic points. *Cartography and Geographic Information Science*, 12(1), pp. 17-28.

## ACKNOWLEDGEMENTS

I'd like to acknowledge Cindy Brewer for her invaluable guidance on this and other projects. I have used the Repeating Shapes v1.5.121 tool offered at the Jenness Enterprises website to generate hexagonal tessellations in ArcGIS, and would like to thank Jeff Jenness and Lois Engelman for that great piece of software. Also, conversations with the following persons have helped me develop these ideas, and I thank them each: Alan Saalfeld, Charlie Frye, and Donna Peuquet.