COMPARISON OF MANUAL VERSUS DIGITAL LINE SIMPLIFICATION

Byron Nakos

Cartography Laboratory, Department of Rural and Surveying Engineering
National Technical University of Athens
9, Heroon Polytechniou STR., Zographos, GR-157 80, Greece
bnakos@central.ntua.gr

ABSTRACT

In the present study a comparison has been carried out between manually and digitally simplified lines in order to estimate appropriate algorithmic tolerances. For this purpose two algorithms have been used to simplify coastlines, cartographic features of high complexity, over a wide range of scale change. Digitally derived coastlines were compared with a reference data set representing the manual simplification procedure. The tolerances ensure small magnitude of displacement. This method may be followed for producing derived cartographic lines at a wide range of scales from a base map.

Introduction

The most commonly used line simplification algorithms have been designed to approach a result as close as possible to the manual simplification procedure. These algorithms are mainly based on geometric principles with the reduction rate affected by a predefined tolerance. However, there is a lack of information in the literature, regarding the choice of the appropriate algorithmic tolerances for simplifying cartographic lines during digital generalization tasks as stated by Cromley and Campbell (1992). Usually, the algorithmic tolerances can be indirectly estimated by applying the "Principles of Selection" (Töpfer and Pillewizer, 1966), a radical law which determines the retained number of objects for a given scale change and the number of objects of the source map. This approach is strongly affected by the number of vertices of the source map while is not explicitly related to the result of manual line simplification procedure. For the appropriate tolerance selection, Shea and McMaster (1989) argue that (p. 60): *The input parameter (tolerance) selection most probably results in more variation, in the final results than either the generalization operator or algorithm selection as discussed above*. Thus, the current way of handling this problem is generally based on trial and error.

In the present paper, an empirical study is presented referring to the comparison of manual versus digital line simplification. The study is a contribution to cover existing gaps (Li, 1993) on line generalization evaluation aspects in manually generalized version of existing maps. The comparison is based on:

- estimation of the critical number of points representing cartographic lines at different scales,
- various cartometric measures among manually and digitally simplified lines.

The digital line simplification was executed by applying two commonly used algorithms. The results of the comparison could be used for estimating appropriate algorithmic tolerances.

More specifically, coastlines that cover a central part of Greece were digitized. These cartographic lines were digitized on the same standards from maps of various scales produced by the national cartographic agency of Greece (Hellenic Geographic Army Service-HAGS).

Consequently, digitally derived lines were produced by applying two commonly used algorithms for line simplification using a wide range of tolerances. The two applied algorithms were:

- the Reumann-Witkam algorithm (Reumann and Witkam, 1974)
- the Douglas-Peucker algorithm (Douglas and Peucker, 1973)

A functional description of the above mentioned algorithms can be found in Weibel (1997).

The digitally simplified coastlines were compared to the reference data set produced by manual simplification. The results of the study may direct users to choose the most appropriate tolerances for specific scale changes during line simplification tasks in digital generalization procedures.

The reference data set

The reference data set consists of ten coastlines located at the central part of Greece (see Figure 1). These cartographic lines (Figure 2) are presented on the topographic maps produced by HGAS. The map series of HGAS cover the entire country with topographic maps at scales of 1:50,000, 1:100,000, 1:250,000, 1:500,000 and 1:1,000,000. The base map of these cartographic series is the map of scale 1:50,000. All the other map series are derived through successive manual generalizations. Thus, the selected coastlines are representative samples of typical manual line simplification, which transforms a base map to derived maps at different scales. This reference data set is assumed to be the base of the manual simplification procedure. Table 1 presents the number of map sheets of the map series that cover the region under study and Table 2 shows the names of the coastlines with their ID's.

Figure 1. The location of studied region.

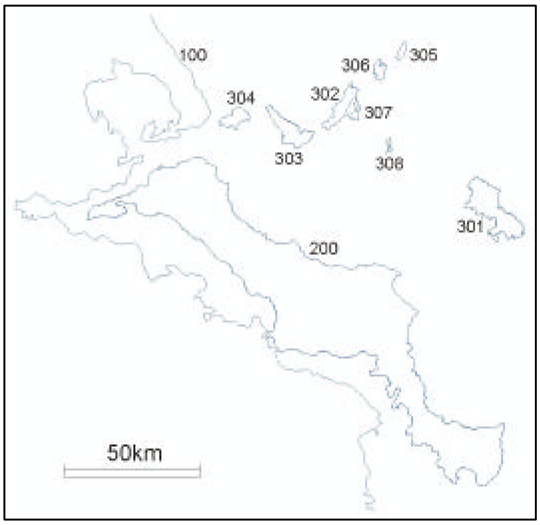

Figure 2. The coastlines under study.

The reference data were digitized using a Summagraphics MICROGRID II digitizer (of size A1) having a resolution of 1016 lpi. The coastlines were digitized in point mode using a magnifying glass on top of the cursor. The digitization procedure was performed using the Arcedit module of Arc/Info software platform. In order to suppress any bias, the lines were digitized following the same standards.

Table 1. The distribution of map sheets.

	Map scale	Map sheets
1	1:1,000,000	2
2	1:500,000	1
3	1:250,000	5
4	1:100,000	15
5	1:50,000	33
Total		**56**

Table 2. The studied coastlines.

	Coastline name	ID
1	Mainland	100
2	Isl. of Evia	200
3	Isl of Skyros	301
4	Isl. of Allonissos	302
5	Isl. of Skopelos	303
6	Isl. of Skiathos	304
7	Isl. of Gioura	305
8	Isl. of Kyra-Panagia	306
9	Isl. of Peristera	307
10	Isl. of Skantzoura	308

All co-ordinates of the reference data set were transformed to the Greek Geodetic Reference System (Transverse Mercator projection, ellipsoid GRS-80) with less than 0.2mm RMS error per sheet on the map. The data set was edited and cleaned in order to link the parts of coastlines that share various map sheets. Table 3 presents the average step and its standard deviation in mm on map for each coastline for all the scales. Analyzing Table 3, it could be stated that all lines were digitized with an overall average step of approximately 0.5mm.

Table 3. The digitized average step and standard deviation for each line.

Line ID	1:50,000		1:100,000		1:250,000		1:500,000		1:1,000,000	
	average step	std. dev.	average step	std. dev.	average step	std. dev.	average step	std. dev.	average step	std. dev.
100	0.5	±0.2	0.6	±0.3	0.7	±0.3	0.4	±0.2	0.6	±0.3
200	0.5	±0.2	0.6	±0.3	0.7	±0.3	0.5	±0.2	0.6	±0.3
301	0.5	±0.1	0.5	±0.2	0.7	±0.2	0.5	±0.2	0.4	±0.2
302	0.4	±0.1	0.5	±0.2	0.7	±0.3	0.4	±0.2	0.4	±0.2
303	0.4	±0.1	0.6	±0.2	0.8	±0.3	0.4	±0.2	0.4	±0.2
304	0.4	±0.1	0.6	±0.2	0.7	±0.3	0.4	±0.1	0.4	±0.1
305	0.4	±0.1	0.5	±0.2	0.5	±0.2	0.5	±0.2	0.4	±0.2
306	0.4	±0.2	0.6	±0.2	0.6	±0.2	0.5	±0.2	0.3	±0.2
307	0.4	±0.2	0.7	±0.3	0.8	±0.3	0.5	±0.2	0.3	±0.2
308	0.4	±0.2	0.5	±0.2	0.5	±0.2	0.5	±0.2	0.3	±0.1

The digitization process of the paper maps produces raw data representing the coastlines that include a number of redundant vertices. These raw data are inappropriate to be used for any comparison, so, they should be transformed to a reference data set that will form the basis of manual simplification by removing the redundant vertices; such a procedure suggested also by Jones and Abraham (1987). This means that the data set should be converged at a level that co-linear vertices do not exist. The transformation of the raw data to a set of critical points was performed by applying a data reduction algorithm to the lines of raw data. It is obvious that the removal of co-linear vertices along the lines does not affect their length. The length measure was used as the criterion of approaching the appropriate level of number of points representing each line. The angularity measure was not applied as a criterion since it is strongly influenced by the presence of spikes (Visvalingam and Whyatt, 1990). The raw data were reduced by applying Douglas-Peucker algorithm with small tolerances (0.002-0.05 mm on map), as it has been suggested by similar studies, (João, 1998).

Finally, the reference data set was formed by those reduced lines, which were preserving their length and were reduced with a tolerance of approximately 0.01mm on the map. This filtering procedure did not affect significantly the average steps of the digitized coastlines presented in Table 3.

Comparison between manual and digital simplification

The first stage of the comparison deals with a test of the reference data set against the "Principles of Selection". The lines of the reference data set were tested on whether or not they follow the "Principles of Selection". Since the coastlines are represented by linear symbols having the same width at all scales the "Principles of Selection" can be expressed as follows (Jones and Abraham, 1987):

$$n_d = n_s \frac{m_d}{m_s},$$

where: n_d and n_s are the numbers of segments at the derived and source scales, and: m_d and m_s are the derived and source scales, respectively. According to the above equation, the expected value of the segments ratio was: 0.5 for three scale changes (1:50K-1:100K, 1:250K-1:500K and 1:500K-1:1M) and: 0.4 for one scale change (1:100K-1:250K). The estimated values of segment ratio varied from a minimum of 0.32 to a maximum of 0.85, when the "Principles of Selection" give a value of 0.5. In the case that the "Principles of Selection" give a value of 0.4, the estimated values of segments ratio varied from a minimum of 0.27 to a maximum of 0.38, respectively. The test showed a significant divergence with "Principles of Selection" referring to the number of segments representing the ten lines at the four scale changes. The divergence of the estimated segment ratios values can be mainly explained by the differense in line complexity, existing between the coastlines. The results of the test showed that the "Principles of Selection" might not be used for estimating the number of vertices of the derived line, a conclusion that has been also stated by Visvalingam and Whyatt (1993).

Consequently, the reference data set was digitally simplified by applying the two algorithms using a predefined range of tolerances. The output of this procedure was a set of digitally simplified derived lines for each scale change, which are appropriate to be compared with the lines of the reference data. The processing was done using MicroStation 95 as software platform. The needed software tools for applying the two algorithms was developed using MicroStation Development Language (MDL).

The comparison between manual and digital line simplification is based on the overlay of two lines, a typical GIS function, and the estimation of different cartometric measures referring to the generated polygons (sliver polygons). This concept has been used in the past for evaluating line simplification algorithms (White, 1985; McMaster, 1986; Muller, 1987; Jenks, 1989). Among the different cartometric measures the one most commonly used is the "area displacement measure". During the evaluation of line simplification algorithms, lines produced from the same basic line, are compared. Since most of the line simplification algorithms are data reduction procedures, their application does not produce such magnitudes of displacement as the human hand does. For this reason cartometric measures such as the area displacement, as a global measure, are useful for evaluating algorithms, but is ambiguous for comparing manually and digitally simplified lines (João, 1998). When manually and digitally simplified lines are compared, because they belong to different cartographic series, a critical parameter is how close the two lines actually are.

Consequently, the cartometric measures are useful for controlling the displacement during line simplification, in such a way that no significant generalization error is produced. In order to ensure their reliability, the ratio of "inner" area versus "outer" area (area ratio) of the generated sliver polygons was estimated. When it is close to the value of 1, then the two lines are wiggling on both sides equally, which actually means that the estimated cartometric measures are more reliable.

The cartometric measures applied were:

- the mean area displacement,
- the total number of (sliver) polygons per unit length, and
- the mean polygon area.

The mean area displacement was calculated from the sum of the area of all "inner" and "outer" sliver polygons, divided by the length of the reference line. The mean area displacement measure can be expressed in m^2/m on the ground or mm^2/mm on the map. The total number of polygons per unit length was calculated by dividing the number of all sliver polygons with the length of the reference line expressed in km. The mean polygon area measure was calculated from the sum of the area of all "inner" and "outer" polygons divided by the total number of all polygons, expressed in m^2 on the ground or mm^2 on the map. The described way of expressing these three cartometric measures allows their values to be compared at all the scales. The process of calculating those measures was performed using several software tools that developed in MDL. Table 4 illustrates an example of the comparison results for the case of Peristera Isl. (ID=307).

Table 4. Comparison results of Peristera Isl. coastline for scale change 1:100K-1:250K

Toler.	Vertices		Area Ratio		Mean Area Displ. (mm^2/mm on map)		Pol./Length (1/km)		Area/Pol. (mm^2 on map)	
	D-P	R-W	D-P	R-W	D-P	R-W	D-P	R-W	D-P	R-W
10m	265	349	1.08	1.09	0.10	0.10	3.68	3.55	0.11	0.11
20m	170	244	1.08	1.03	0.10	0.10	3.28	3.34	0.12	0.12
30m	139	198	1.01	0.97	0.10	0.10	3.09	3.25	0.13	0.12
40m	120	165	0.93	1.06	0.10	0.10	3.09	3.31	0.13	0.12
50m	98	146	0.85	0.89	0.10	0.10	3.09	3.55	0.13	0.11
60m	91	125	0.74	1.16	0.11	0.10	3.15	3.12	0.13	0.13
70m	85	115	0.78	0.95	0.11	0.11	3.03	3.03	0.14	0.14
80m	79	104	0.71	0.83	0.11	0.12	3.15	2.97	0.14	0.16
90m	69	103	0.79	0.79	0.13	0.12	2.84	2.78	0.18	0.17
Reference Lines			1.10		0.10		3.62		0.11	

Coastlines are among the cartographic features with the least displacement, when they are manually simplified. In this case small displacement arises when small peninsulas or bays have to be eliminated. For this reason, the estimated values of these cartometric measures referring to the two reference lines of scale 1:100K and 1: 250K, are lower than those estimated in a similar study (João, 1998). In that context features such as roads have been simplified accepting higher level of displacement. In addition, Table 4 shows that there is no significant variation between the estimated cartometric measures of the derived and the reference lines. Actually, this observation could be explained by the fact that line simplification algorithms are data reduction procedures and cause small magnitudes of displacement.

Following the concept of generating the reference data set, the number of critical points representing the coastline at scale of 1:250K has been estimated to be 166. Table 4 indicates that the suggested tolerances of Douglas-Peucker and Reumann-Witkam algorithms are 20m and 40m respectively. With similar way, the suggested tolerances of all other lines and scale changes can be estimated, and are presented in Table 5. Averaging the tolerances of all lines for the different scale changes of both algorithms the graph presented Figure 3 has been produced. It could be seen that the line connecting the tolerance values approaches a straight line as it has been proposed by Jones and Abraham (1987).

Table 5. Suggested tolerances for both algorithms and each line.

Line ID	1:50K – 1:100K		1:100K – 1-250K		1:250K – 1:500K		1:500K – 1:1M	
	D-P	R-W	D-P	R-W	D-P	R-W	D-P	R-W
100	4m	8 m	20m	35 m	20m	40 m	75m	125 m
200	4.5m	8 m	20m	35 m	20m	40 m	50m	95 m
301	4.5 m	8 m	25m	40 m	35m	60 m	25m	50 m
302	5 m	8 m	30m	50 m	50m	40 m	60m	115 m
303	5 m	12 m	25m	45 m	25m	30 m	50m	80 m
304	5 m	12 m	25m	45 m	30m	50 m	100m	100 m
305	5 m	12 m	30m	45 m	40m	75 m	50m	70 m
306	5 m	8 m	20m	45 m	35m	65 m	40m	50 m
307	7.5 m	13 m	20m	40 m	25m	50 m	50m	100 m
308	5 m	8 m	20m	35 m	40m	75 m	20m	50 m

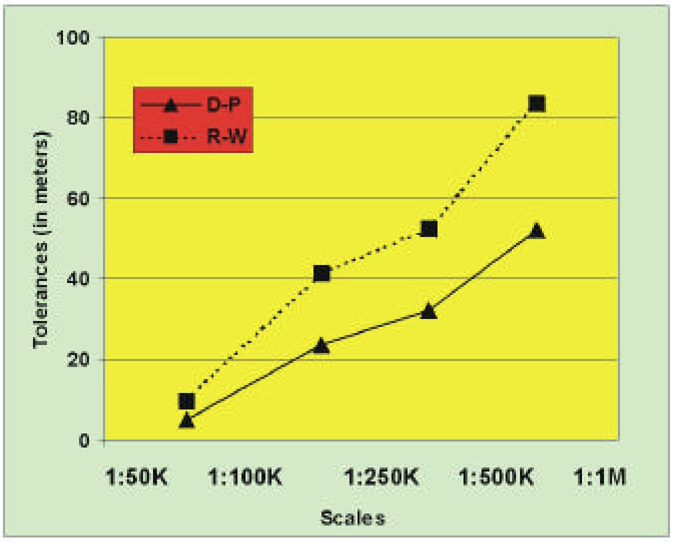

Figure 3. Suggested tolerances for both algorithms

Conclusions

The main aim of this empirical study is the introduction of a method for selecting the appropriate tolerances for a wide range of scale changes during line simplification while performing digital generalization. These tolerances select the same number of points when digital simplification is carried out with that of critical points representing manually simplified lines. Based on the comparison results, there is evidence, that the two algorithms using the suggested tolerances do not produce significant generalization error (displacement).

The coastlines have been chosen as subjects of study, since they are considered as having a rather high level of complexity. Their complexity can be expressed by the fractal dimension that previous research estimated to be 1.23 (Nakos, 1996). This is equivalent to the fractal dimension of the coastline of west Britain estimated to be 1.25 and defined as a characteristic example of high level of complexity (Mandelbrot, 1967).

However, the research must be extended including various complex cartographic lines in order to reach a general approval. Additionally, this aim could be supplemented by studying various kinds of linear cartographic features (i.e. roads, rivers, boundaries etc.) as well.

Acknowledgments

The author would like to thank the Hellenic Geographic Army Service for providing the paper maps used in this study and Spyros Katsareas, undergraduate student of Rural & Surveying Engineering Department of NTUA, for carrying out time consuming data processing tasks.

References

Cromley, R.G., and Campbell, G.M. (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. The Canadian Cartographer, 10(2), pp. 112-122.

João, E.M. (1998). Causes and Consequences of Map Generalization. Research Monographs in Geographical Information Systems. Taylor & Francis, London, p. 266.

Jenks, G.F. (1989). Geographic Logic in Line Generalization. Cartographica, 26(1), pp. 27-42.

Jones C.B., and Abraham, I.M. (1987). Line Generalisation in a Global Cartographic Database. Cartographica, 24(3), pp. 32-45.

Li, Z. (1993). Some observations on the issue of line generalisation. The Cartographic Journal, 3(1), pp. 68-71.

Mandelbrot, B.B. (1967). How long is the Coast of Britain? Statistical Self-similarity and Fractional Dimension. Science, 156(3775), pp. 636-638.

McMaster R.B. (1986). A Statistical Analysis of Mathematical Measures for Linear Simplification. The American Cartographer, 13(2), pp. 103-116.

Muller, J.-C. (1987). Fractal and automatic line generalisation. The Cartographic Journal, 24(1), pp. 27-34.

Nakos, B. (1996). The Use of Fractal Geometry Theory for Performing Digital Generalization Tasks. Proceedings of the 2nd National Cartographic Conference, Hellenic Cartographic Society, pp. 293-301. (In Greek)

Reumann, K., and Witkam, A.P.M. (1974). Optimizing Curve Segmentation in Computer Graphics. International Computing Symposium. Amsterdam, North Holland, pp. 467-472.

Shea, K.S., and McMaster, R.B. (1989). Cartographic Generalization in a Digital Environment: When and How to Generalize. Proceedings of 9th International Symposium on Computer–Assisted Cartography, Baltimore, Maryland, pp. 56-67.

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

Visvalingam, M., and Whyatt, J.D. (1990). The Douglas-Peucker algorithm for line simplification: Reevaluation through visualisation. Computer Graphics Forum, 9(3), pp. 213-228.

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

Weibel, R. (1997). Generalization of Spatial Data: Principles and Selected Algorithms. In Algorithmic Foundations of Geographical Information Systems (eds. van Kreveld et al.) Springer-Verlag, Berlin, pp. 99-152.

White, E.R. (1985). Assessment of Line-Generalization Algorithms Using Characteristic Points. The American Cartographer, 12(1), pp. 17-27.