

THE UNIVERSITY OF HULL

Cartographic Information Systems Research Group C.I.S.R.G.

Hon. Co-ordinator:
Dr. M. Visvalingam
Department of Computer Science
The University
HULL HU6 7RX

Tel. (0482) 465295/465951
Telex 592592 KHMAIL G,
f.a.o. HULIB375
Fax (0482) 466666

Discussion Paper Series Editors: Dr M Visvalingam
Dr M E Turner

C.I.S.R.G. DISCUSSION PAPER 10

Line Generalisation by Repeated Elimination
of the Smallest Area

by

M Visvalingam and J D Whyatt

Not to be quoted without the author's permission

© M Visvalingam and J D Whyatt
University of Hull

July 1992

Discussion Papers

1. Wade, P., Visvalingam, M. and Kirby, G.H. (1986)
From Line Geometry to Area Topology, 48 pp.
2. Visvalingam, M. (1987)
Problems in the Design and Implementation of a GKS-based
User Interface for a Graphical Information System, 18 pp.
3. Visvalingam, M. (1988)
Standardising Basic Spatial Units : Problems and Prospects, 19 pp.
4. Whyatt, J. D. and Wade, P. R. (1988)
The Douglas-Peucker Algorithm for Line Simplification: an
introduction with programs, 17 pp.
5. Visvalingam, M. (1988)
Cartography, Geographical Information Systems and Maps in
Perspective, 16 pp.
6. Visvalingam, M. and Whyatt, J. D. (1989)
The Douglas-Peucker Algorithm for Line Simplification: re-evaluation
through visualisation, 35 pp.
7. Visvalingam, M. (1989)
Trends and Concerns in Digital Cartography, 37 pp.
8. Visvalingam, M. and Whyatt, J. D. (1990)
Cartographic Algorithms : problems of implementation and evaluation
and the impact of digitising errors, 24 pp.
9. Visvalingam, M. (1991)
Visualisation, VISC and Scientific Insight, 21 pp
10. Visvalingam, M. and Whyatt, J. D. (1992)
Line Generalisation by Repeated Elimination of the Smallest Area, 16 pp

Special Issues

1. Visvalingam, M. (ed) (1986)
Research based on Ordnance Survey Small-Scales Digital Data
Proceedings of a meeting held as part of the 1986 Annual
Conference of the Institute of British Geographers (Reading,
Jan 1986), sponsored by the Ordnance Survey, 79 pp.
2. Visvalingam, M. and Kirby, G. H. (1987)
Directory of Research and Development based on Ordnance
Survey Small Scales Digital Data, sponsored by the Ordnance Survey,
38 pp.
3. Visvalingam, M. and Sekouris, N. M. (1989)
Management of Digital Map Data Using a Relational Database Model,
sponsored by the Ordnance Survey, 50 pp.

ABSTRACT

This paper presents a new approach to line generalisation which uses the concept of "effective area" for progressive simplification of a line by point elimination. Two coastlines are used to compare the performance of this, with that of the widely used Douglas-Peucker, algorithm. The results from the area-based algorithm compare favourably with manual generalisation of the same lines. It is capable of achieving both imperceptible minimal simplifications and caricatural generalisations. By careful selection of cut-off values, it is possible to use the same algorithm for both scale-dependent and scale-independent generalisations. More importantly, it offers scope for modelling cartographic lines as consisting of features within features so that their geometric manipulation may be modified by application- and/or user-defined rules and weights. The paper examines the merits and limitations of the algorithm and the opportunities it offers for further research and progress in the field of line generalisation.

C O N T E N T S

Abstract	
1. Introduction	1
2. Background	1
3. The area-based method	1
3.1 Elimination rather than selection	3
3.2 Effective area	3
3.3 The algorithm	3
3.4 Locally derived metric but holistic view of line	5
4. Data, results and Discussion	5
4.1 Data	5
4.2 Results	5
4.3 Discussion	10
5. Conclusion	13
Acknowledgements	
References	15

1. INTRODUCTION

The identification of the salient character or caricature of a line is central to the process of line generalisation. The caricatures produced by the Douglas-Peucker algorithm (Douglas and Peucker, 1973) are believed to be the most successful. The algorithm is reputed to select the critical, shape-preserving points on a line. However, several researchers have expressed reservations about this and other point-based methods (see Visvalingam and Whyatt, 1990, 1991a). These methods are believed to be suitable only for minimal simplification, and not for generalisation, especially of complex lines. Caricatural generalisations preserve only the major distinctive features and omit smaller, less significant ones. Results obtained using an area-based algorithm (specified by Visvalingam and first tested by Whyatt (1991)) suggest that it can filter out features within features progressively and thus achieve both minimal simplification and caricatural generalisation.

2. BACKGROUND

Attneave (1954) used a caricature of a sleeping cat to illustrate the presence of "information loaded" critical points. He proposed that people perceived these points of high curvature along lines as perceptually important and high in information content. His concept of critical points has influenced research on line simplification in digital cartography. White (1983) observed that the now widely used Douglas-Peucker algorithm identified more of these critical points than some others she studied. McMaster repeatedly asserted that this algorithm was "mathematically and perceptually superior" to other algorithms he evaluated because it picked out more of these critical points and produced least displacement from the original line (see for instance McMaster, 1987).

Waugh (GIMMS, 1983) and Wade (Whyatt and Wade, 1988) proposed extensions to the algorithm; tag values were associated with each point as an indicator of its significance. These tag values may be used to rank points into a hierarchy of critical points. This concept of a fixed rank order of critical points is convenient since a tolerance parameter may be used to filter out the required points at run time. Visvalingam and Whyatt (1990, 1991a) used these tag values to study the assumptions underpinning its wide use, other properties of the method and their implications. They concluded that this algorithm does not necessarily select points which cartographers select from complex lines and reviewed other criticisms of this method. Even using relatively simple test data, White (1983) only found a 45% agreement between points selected by the algorithm and by respondents in her study. Furthermore, Ramer (1972) working in the field of Pattern Recognition had already noted that the algorithm selects some redundant points. Visvalingam and Whyatt (1991a & b) pointed out that shape distortions can occur as a result of selecting the furthest point from the anchor-floater point, since these extreme points can fall on spikes and minor insignificant features or be the outcome of tiny rounding and digitising errors. It is well known that the Douglas-Peucker method is only suitable for minimal simplification of lines in Jenk's (1981) class of imperceptible generalisation. Consequently, many believe that point-based methods are incapable of even approximating the cartographer's art of producing caricatural generalisations. This paper describes a point-based algorithm which is capable of achieving both minimal simplifications and caricatural generalisations.

3. THE AREA-BASED METHOD

Although this algorithm produces very encouraging results, we continue to view it as just one step towards a more intelligent system for line generalisation.

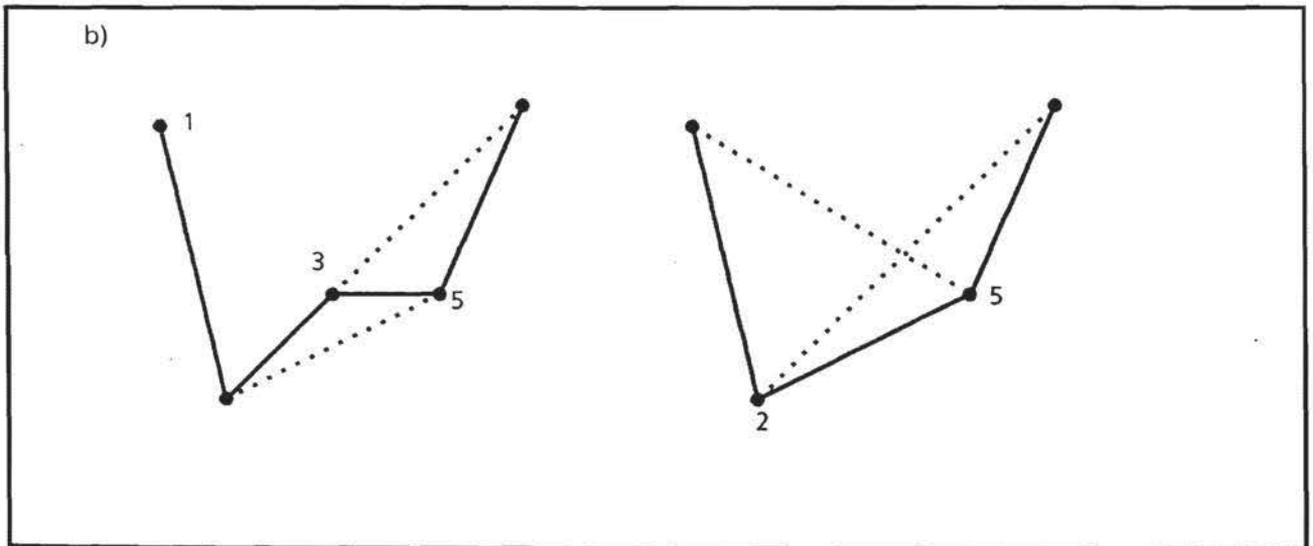
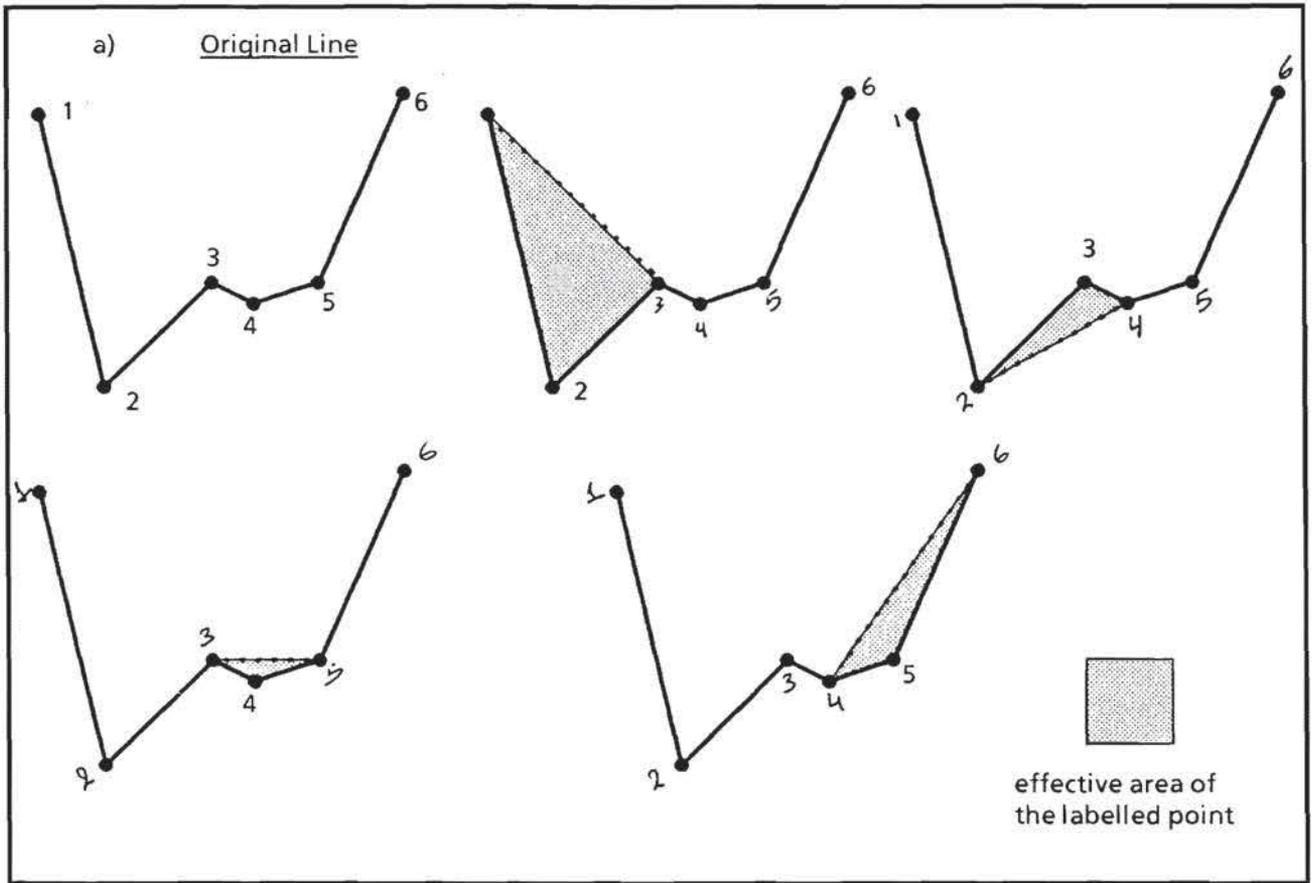


Figure 1 : Area-based Line Generalisation a) effective areas of points
 b) generalisation by repeated elimination of the smallest area.

The basic idea underpinning this algorithm is iteratively to drop the point which results in least areal displacement from the current part-simplified line. This results in the progressive elimination of geometric features, from the smallest to the largest. Area was chosen because other metrics, such as shape, only start to have an impact when the size of a feature exceeds a perceptual limit. This initial study set out to evaluate the impact of geometric size on line generalisation.

3.1 Elimination rather than Selection

The Douglas-Peucker algorithm drops all intermediate points if they fall within a tolerance band of the straight line, called the anchor-floater line, connecting the first and last points. This is reasonable (Visvalingam and Whyatt, 1991b). However, the assumption that the furthest point from this arbitrary anchor-floater line must be a critical point is questionable. The area-based method is based on the observation that it is easier to filter points on lines by a process of elimination rather than selection.

3.2 Effective Area

Size, as measured by area, sets a perceptual limit on the significance or otherwise of other perceptual indicators. It is also the most reliable metric for elimination since it simultaneously considers distance between points and angular measures. Area measures are widely used in traditional cartography (Robinson et al, 1984) and have been used for eliminating polygonal features by others working in Digital Cartography (Deveau, 1985). This concept may be extended to the simplification of linear features. The area-based algorithm associates with each non-terminal point its effective area. This is the areal displacement which would occur if that point alone was omitted from the current line (Figure 1a).

McMaster (1987) used least areal displacement from the original line for comparing the performance of line simplification algorithms. In the area-based algorithm the areal displacement is measured relative to the current, part simplified line, and not the original source line.

3.3 The Algorithm

The algorithm is as follows:

- * Compute the effective area of each point (see Section 3.2)
- * Delete all points with zero area and store them in a separate list with this area
- * REPEAT
 - * Find the point with the least effective area and call it the current point. If its calculated area is less than that of the last point to be eliminated, use the latter's area instead. (This ensures that the current point cannot be eliminated without eliminating previously eliminated points.) Delete the current point from the original list and add this to the new list together with its associated area so that the line may be filtered at run time.
 - * Recompute the effective area of the two adjoining points (see Figure 1b).

UNTIL

The original line consists of only 2 points, namely the start and end points.

The ability to rank points in some manner is useful in scale-free mapping since it expedites the filtering of points at run time. However, Visvalingam

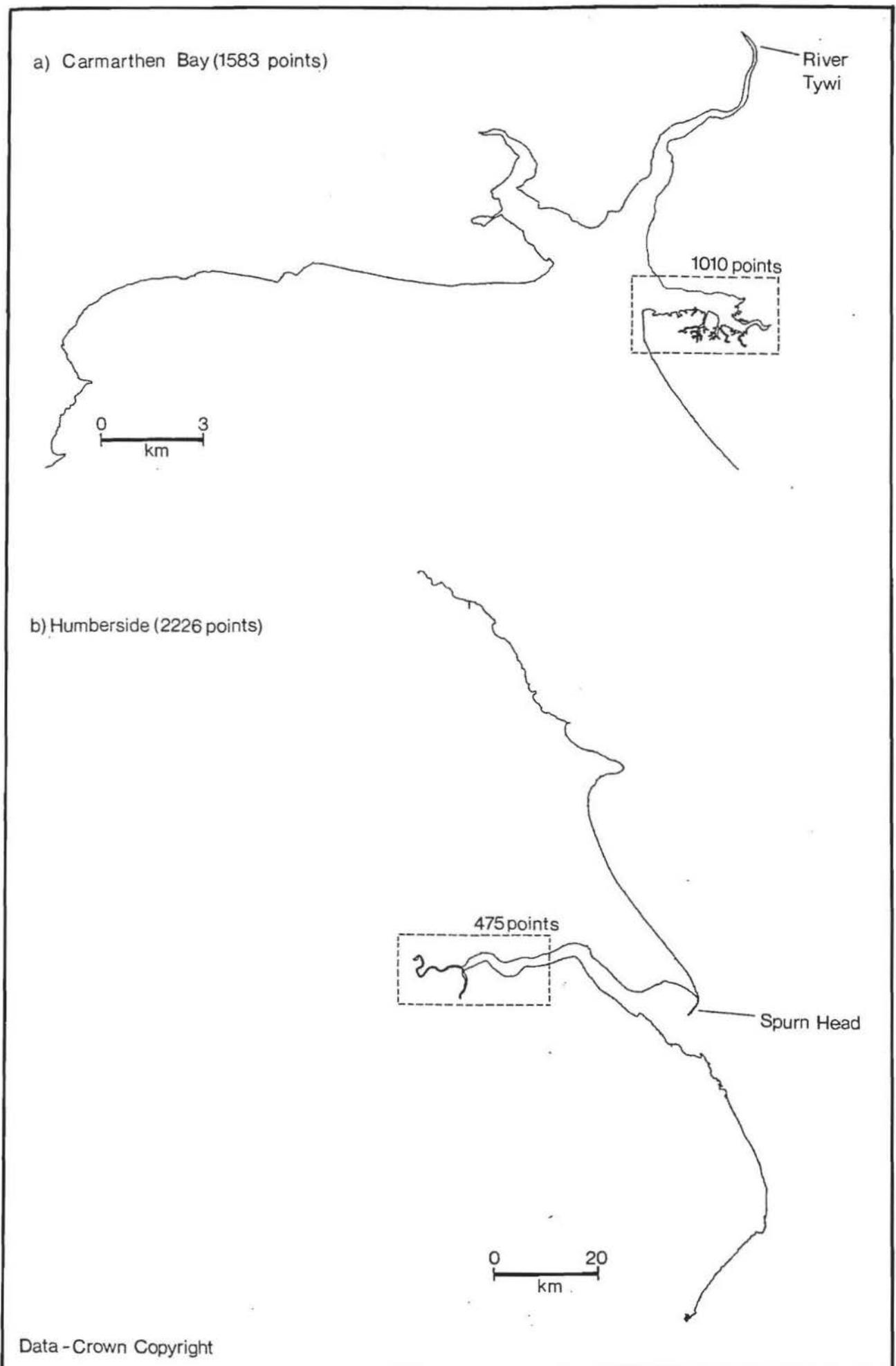


Figure 2 : Test Data

and Whyatt (1990) demonstrated why a fixed rank-order of critical points, no matter how they are devised, limit the scope for producing appropriate scale-related displays. However, at this early stage of experimentation with a relatively simplistic metric, we were more concerned with assessing whether the algorithm was capable of progressively eliminating features within features.

When expressing the above algorithm as a computer program, attention must be paid to the precision and range of co-ordinates to avoid problems of overflow and underflow. Visvalingam and Whyatt (1991a) pointed out that the articulation of an algorithm as a computer program requires a consideration of special geometric cases, numeric problems and digitising errors. For example, different implementations of the Douglas-Peucker algorithm produced different results.

3.4 Locally Derived Metric but Holistic View of Line

The processes of selection by the Douglas-Peucker method and elimination using effective area differ in two distinct ways. The Douglas-Peucker method requires the calculation of the distance of each point from an anchor-floater line. Therefore, all points are considered initially. Once the point of maximum offset is selected, the two parts of the line can be processed independently of each other. Thus, contrary to the claims made by others, the algorithm ceases to be global and holistic after the selection of the first point.

Even though the effective area is calculated using only three neighbouring points, in the area-based method, the process does require a holistic view of the line whilst progressively eliminating detail. It involves a comparison of all remaining points along the line when seeking to eliminate a point. (It is possible to optimise this computation).

4. DATA, RESULTS & DISCUSSION

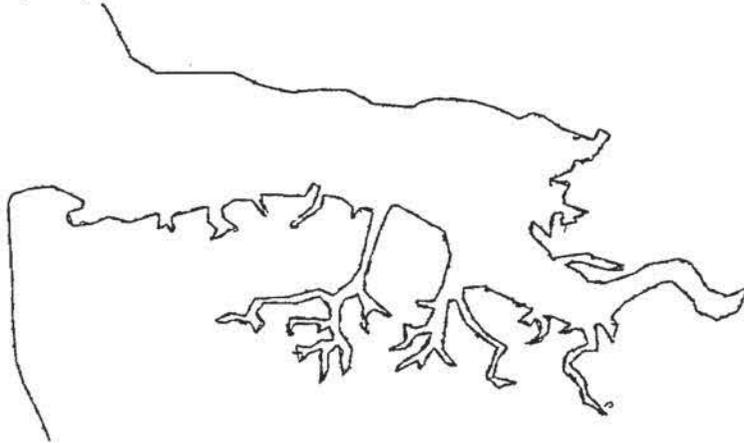
4.1 Data

Two complex stretches of coastline, around Carmarthen Bay and Humberside (Figure 2), were selected since they describe features within features. The data were extracted from the files containing the boundaries of British Administrative Areas (digitised from 1:50000 source maps by the Department of Environment and Scottish Development Department and held at the South West Universities Regional Computer Centre (Wise, 1988)). This data may be accessed without charge for purposes of academic research.

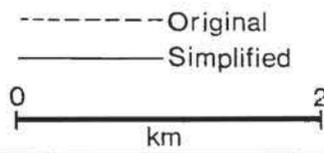
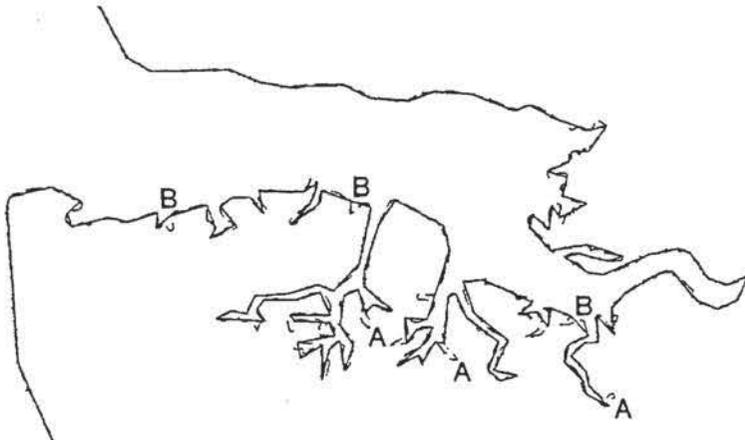
4.2 Results

The results are encouraging and indicate that it is possible to model cartographic lines as consisting of features within features. The Carmarthen Bay test data shows a progressive elimination of size-related features (Figures 3 and 4). During the initial stages, the most noticeable changes occur in the boxed region in Figure 2 which contains many creeks. At all levels of generalisation indicated in Figure 3, the area-based method eliminates more points from the creek region than from the rest of the coastline; the number of points retained in the latter are listed below. Figure 3a shows that 77% of the points on the creek region may be eliminated without any significant departure from the original line. In Figure 3b, further elimination of points results in a shortening of the creeks (marked A) and a noticeable straightening of the coast (along B). After this, in Figure 3c, shorter creeks (marked C) and some headlands (marked D) are eliminated in their entirety. The result is a noticeable simplification of the coastline

a) 235 / 1010 (23%)



b) 162 / 1010 (16%)



Data: Crown Copyright



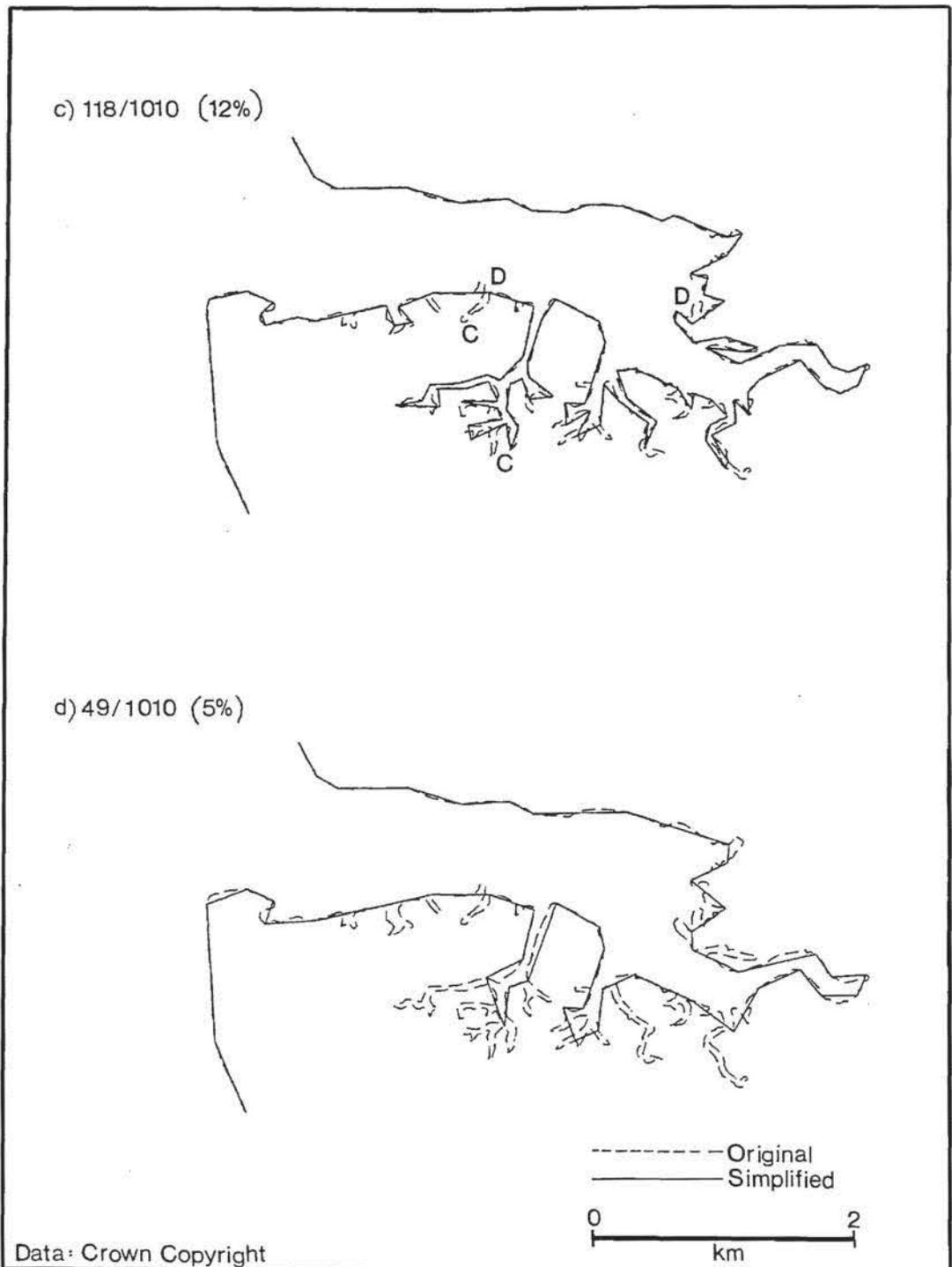


Figure 3 : Minimal Simplification and Typification
 Corresponding figures for all of Carmarthen Bay :
 a) 570/1583 (36%); b) 416/1583 (26%);
 c) 332/1583 (21%); d) 174/1583 (11%)

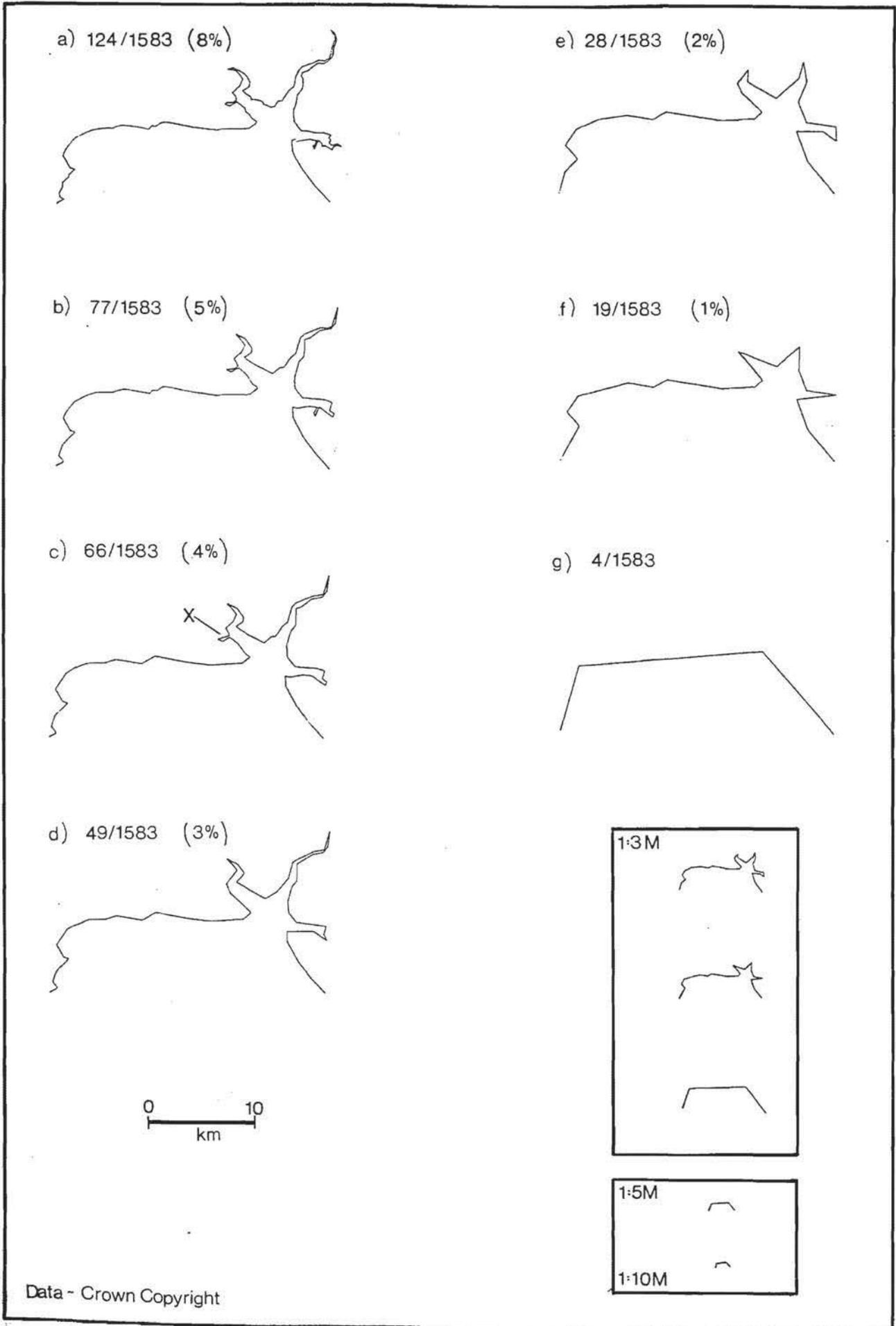
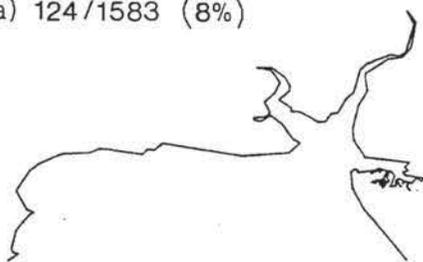
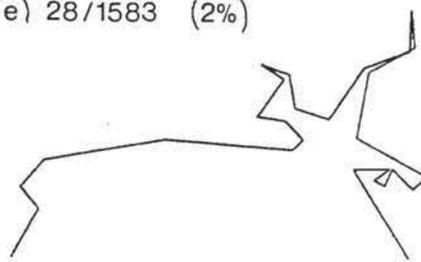


Figure 4 : Caricatural Generalisations

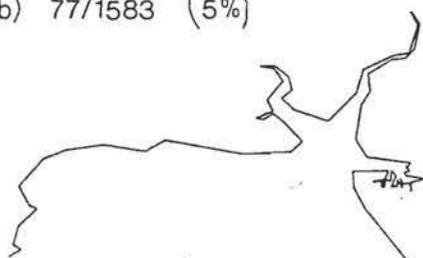
a) 124/1583 (8%)



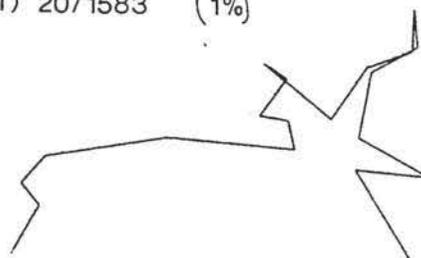
e) 28/1583 (2%)



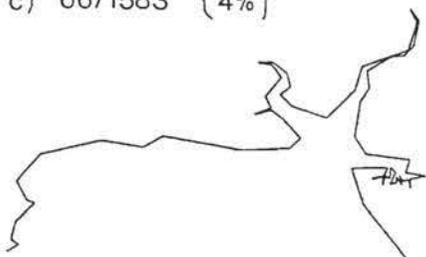
b) 77/1583 (5%)



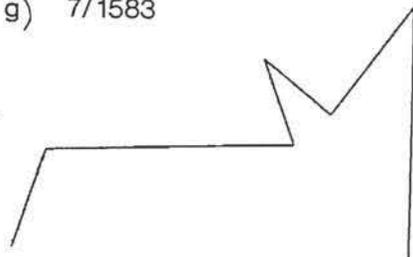
f) 20/1583 (1%)



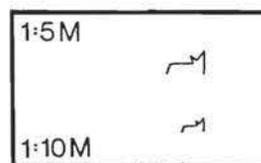
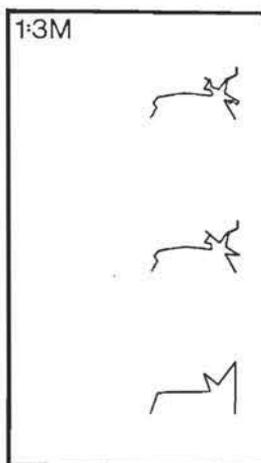
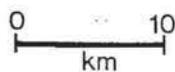
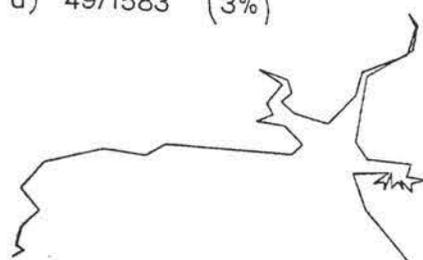
c) 66/1583 (4%)



g) 7/1583



d) 49/1583 (3%)



Data-Crown Copyright

Figure 5 : Output from Douglas-Peucker Algorithm

and creeks. On further generalisation, the presence of creeks is only intimated by the retention of a few of them; these have become noticeably widened and shortened (Figure 3d). The method first produces minimal simplification, in the realms of imperceptible generalisation as Jenks (1989) put it, then perceptible simplification followed by typification.

Figure 4 shows, at the 1:500 000 scale, the effect of further elimination. Figures 4a-c indicate that the generalisation and removal of the creeks precede the noticeable straightening of the coast at this scale and the elimination of the feature marked X on Figure 4c (which is actually an error introduced during digitising). There is then a progressive shortening of the three river estuaries (Figures 4d-f) into the distinctive three-pronged caricature seen on 1:1m and smaller scale atlas maps. The four-point representation of the bay would not be an inappropriate generalisation if the stretch of coastline occurred within a more extensive geographic area shown on very small scale, 1:10M, maps.

Figure 5 shows the results of the Douglas-Peucker algorithm using equivalent numbers of points. None of the maps are satisfactory and there is gross distortion of shape when less than 4% of points are used. Figure 6a shows the coastline depicted by the 200 points digitised by the Ordnance Survey from the 1:625 000 Route Planner map. Figure 6b shows that comparable results may be produced using just 77 points, i.e. 5% of points. In comparison, the Douglas Peucker method cannot achieve appropriate simplifications with 200, let alone 77, points (Figures 6c and d). When compared with simplifications produced by the Douglas-Peucker algorithm, using similar numbers of points, the area-based algorithm produces better balanced, aesthetically more pleasing and cartographically more appropriate simplifications. However, as with other point-based methods, it has its own merits and limitations.

4.3 Discussion

The method is attractive for the following reasons. It is very simple in concept and builds on existing ideas within cartography. It confirms the importance of size in geometric simplification. Minor features, such as spikes, are rapidly eliminated since they enclose very small areas. It achieves both minimal simplifications and caricatural generalisation. Entire sub-features are eliminated before features at the next higher level are noticeably modified. Since the test data demonstrates that the method performs adequately with more than three levels of features, it is reasonable to assume that the method will scale up to national and global levels. This needs to be tested. Attention needs to be paid to the implementation of the algorithm which must adjust the precision of co-ordinates during calculations to avoid overflow.

Thus, cartographic lines may be modelled as consisting of overlapping hierarchies of nested features. The explicit representation of features will enable the user to associate weights to these to modify the normal geometric processing. The algorithm may be used for achieving both scale-dependent and scale-independent generalisations (Robinson et al, 1984). Except when eliminating the larger features, the most noticeable changes occur when size-related features are dropped in their entirety. Unlike the Whirlpool program, based on Perkal's algorithm, the area-based method does not leave lakes at the head of the estuary (Beard, 1991). In between the removal of features at different levels of the hierarchy, the appearance of the map is not substantially affected by the elimination of further points. Thus, features at each level of the hierarchy can be represented by a minimal or a larger set of points. A minimal set appears to be adequate for reduced smaller-scale displays while a larger set may be more appropriate for scale independent generalisation. The algorithm is also capable of suggesting typifications.

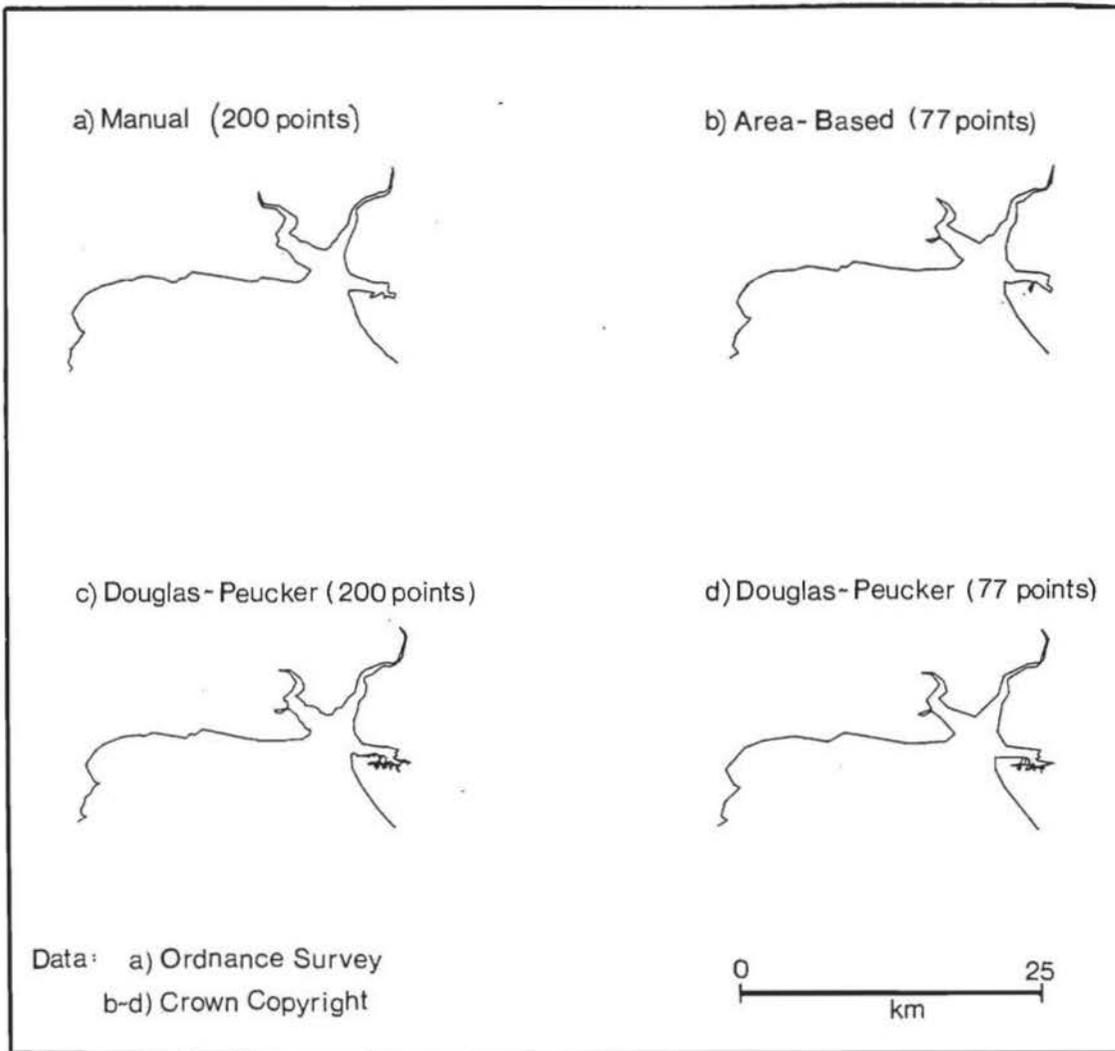
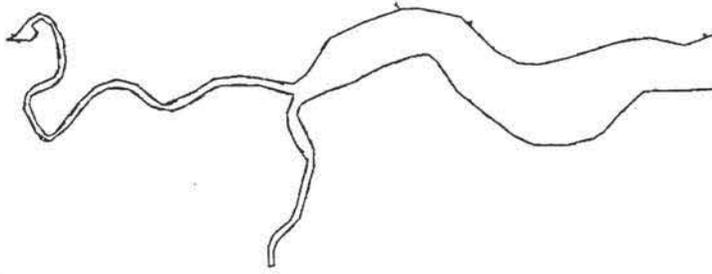
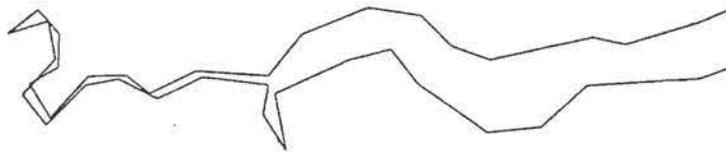


Figure 6 : Comparison of (a) manual simplification with output of (b) area-based algorithm and (c & d) Douglas-Peucker algorithm

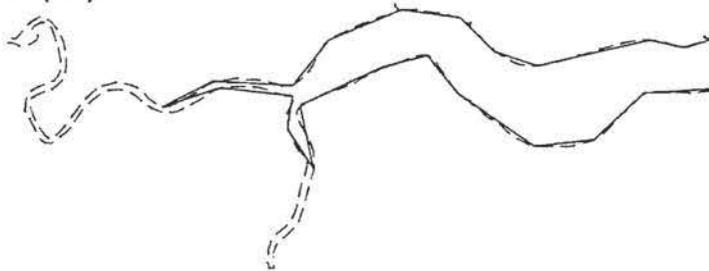
a) 106 / 475 (22%)



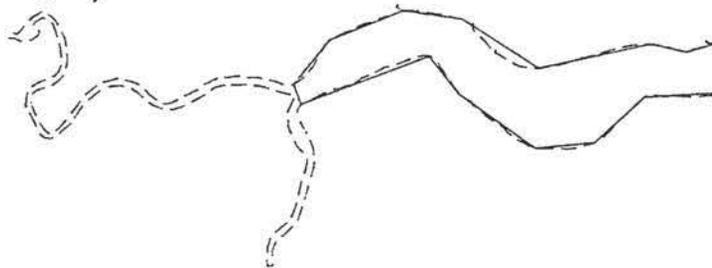
b) 40 / 475 (8%)



c) 24 / 475 (5%)



d) 16 / 475 (3%)



Data: Crown Copyright

----- Original
——— Simplified

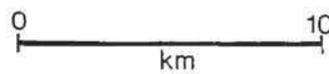


Figure 7 : Progressive Simplification of Rivers at the Head of the Humber Estuary

For example, in Figure 3d, the presence of an irregular coastline of creeks is only suggested.

Like all simple point-based algorithms, the area-based method has limitations. These outstanding problems raise some interesting issues. At present the algorithm considers only the size or area of triangles. Manual generalisation considers the size, shape and the geometric and geographic context of features consisting of several points. We were surprised therefore that this germ of an algorithm performed as well as it did. However, the cut-off values, yielding the characteristic caricatural shapes, were found by trial-and-error using interactive software. Topfer's law (Topfer and Pillewizer, 1966) is inappropriate at this level since the number of required points is determined by the original shape of the geomorphic features. Since caricatures consist of a minimal set of points, the inclusion/omission of even one point can alter the shape of the feature. Thus, although the algorithm for ranking points is objective, even if somewhat simplistic and arbitrary, the selection of cut-off values remains subjective; it is directed by purpose and influenced by the cartographer's wider knowledge of geography. However, even without further refinement, the algorithm can be used to derive an approximation of the required shape. The cartographer can then adjust the shape if necessary. None of the maps included in this paper have been modified.

As with many other point-based methods this algorithm can also cause complex lines to cross each other. Figure 7a shows that the banks of the Yorkshire River Ouse can stand some minimal simplification before they start to cross (Figure 7b). Similar crossings were observed on other narrow features, such as Spurn Head at the mouth of the Humber estuary (Figure 2b). In the past this has been regarded as a test of an algorithm's performance, mainly because there was limited scope for resolving the problem. However, in the context of the area-based method, crossing lines flag the need for more intelligent generalisation. Figures 7a-d point to different considerations requiring different approaches to generalisation. Figure 7a is sufficient if all that is required is minimal simplification. At grosser levels of simplification, it may be sufficient to show only the head of the Humber estuary (Figure 7d). At intermediate levels different strategies may be adopted, depending on the purpose of the map. If there is no reason for showing the rivers, only their mouths need to be indicated as shown in Figure 7c. If it is necessary to show the rivers, they can either be exaggerated by displacing the banks or be shown by an abstracted line. A similar approach can be adopted with other special features like Spurn Head. Line crossings indicate a need for drastic modification but this has not been routinely undertaken in line simplification in the past since it is computationally very demanding. Using the area-based method it is possible to isolate features, such as rivers and spits; even after minimal simplification, a number of points which share the same rank (seventeen on the Ouse) are eliminated together. There is scope for more efficient and intelligent detection of line crossings. Remedial action can then be based on traditional cartographic principles. Thus, in the context of the area-based method, line crossings are not defects but rather clues for prompting intelligent generalisation.

The area-based method also provides scope for achieving slightly different shapes using the knowledge that combinations of positive and negative areas indicate different types of features. Thus, navigational charts may choose to weight spits and headlands as more important than same-sized landward intrusions.

5. CONCLUSION

People have little difficulty in abstracting the features described by the geometry of lines. At present, computer simplification of lines is driven by

geometric rules of thumb. Much effort is being expended on codification into computer-readable form of existing knowledge (Buttenfield and McMaster, 1991). It is not enough to code already existing specifications on the 'what' and 'when' of generalisation. Intelligent generalisation requires a deeper understanding of 'why' and 'how to'. The automatic generalisation of lines awaits algorithms for structuring the point samples into overlapping hierarchies of features. The features may then be used to link in geographic labels and functional information to ignore or guide the underlying geometric reasoning. The novel area-based method for line generalisation, presented in this paper, has revealed the potential for modelling cartographic lines as consisting of a geometric hierarchy of features within features. Being a point-based algorithm, it has some inherent limitations. However, even in its present primitive form, it is already capable of producing acceptable simplifications and caricatural generalisations. It can be used to achieve both scale-dependent and scale-independent generalisation of features at a given level of the hierarchy by using either their minimal or a fuller description. It is also capable of producing typifications. Furthermore, this approach provides sufficient clues for isolating distinctive features for omission, exaggeration or skeletal representation. It opens up opportunities for further research in Digital Cartography.

ACKNOWLEDGEMENTS

We are grateful to the UK Science and Engineering Research Council for the award of a studentship to Duncan Whyatt, to the South West Universities Regional Computer Centre (SWURCC) for providing access to the DoE/SDD boundary files and to the Ordnance Survey for permission to use the 1:625 000 digital data. These data remain the property and copyright of the Crown.

REFERENCES

- Attneave, F (1954) "Some Informational Aspects of Visual Perception",
Psychological Review 61 (3) 183 - 193
- Beard, K M (1991) "The Theory of the Cartographic Line Revisited",
Cartographica 28 (4) 32 - 58
- Butenfield, B P and McMaster, R B (1991, eds.) Map Generalization :
making rules for knowledge representation (Longman, UK)
- Deveau, T J (1985) "Reducing the Number of Points in a Plane Curve
Representation", Proc. Auto Carto 7, Washington D.C, 152 - 160
- Douglas, D H and Peucker, T K (1973) "Algorithms for the Reduction of the
Number of Points Required to Represent a Digitised Line or its
Caricature", The Canadian Cartographer 10 (2), 112 - 122
- Jenks, G F (1981) "Lines, Computers and Human Frailties", Annals of
the Assoc. of American Cartographers 71 (1) 1 - 10
- Jenks, G F (1989) "Geographic Logic in Line Generalisation",
Cartographica 26 (1) 27 - 42
- McMaster, R B (1987) "Automated Line Generalisation", Cartographica 24 (2)
74 - 111
- Ramer, U (1972) "An Iterative Procedure for the Polygonal Approximation of
Plane Curves", Computer Graphics and Image Processing 1, 244 - 256
- Robinson, A H, Sale, R D, Morrison, J L and Muehrcke, P C (1984)
Elements of Cartography (Wiley & Sons, New York) 5th Edition
- Topfer, R and Pillewizer, W (1966) "The Principles of Selection", The
Cartographic Journal 3 (1) 10 - 16
- Visvalingam, M and Whyatt, J D (1990) "The Douglas-Peucker Algorithm for Line
Simplification: Re-evaluation through Visualisation", Computer Graphics
Forum 9 (3) 213 - 228
- Visvalingam, M and Whyatt, J D (1991a) "Cartographic Algorithms: Problems of
Implementation and Evaluation and the Impact of Digitising Errors",
Computer Graphics Forum 10 (3) 225 - 235
- Visvalingam, M and Whyatt, J D (1991b) "The Importance of Detailed
Specification, Consistent Implementation and Rigorous Testing of
Cartographic Software", in K. Rybaczuk and M. Blakemore (eds) Mapping
The Nations vol 2, 15th Conference of Int. Cartographic Assoc.
(September, Bournemouth) 856 - 859
- Waugh, T C and McCalden, J (1983) GIMMS Reference Manual (4.5), Gimms Ltd,
30 Kier Street, Edinburgh EH3 9E4
- White, E R (1983) "Perceptual Evaluation of Line Generalisation Algorithms",
Unpublished Masters Thesis, University of Oklahoma

Whyatt, J D. and Wade, P R (1988) "The Douglas-Peucker Line Simplification Algorithm", Bulletin of The Society of University Cartographers 22 (1) 17 - 25

Whyatt, J D (1991) "Visualisation and Re-evaluation of Line Simplification Algorithms", Unpublished Ph D Thesis, University of Hull, 259 pp

Wise, S (1988) "Using the DoE and SDD Digitised Boundary Data at SWURCC", User Guide, Bath University Computing Services, Bath, Avon