This short article addresses some of the problems in cartographic generalisation, especially the mis-use of data point reduction algorithms for generalisation purpose. Special attention is paid to line generalisation. Through analysis, the author emphasizes that 'data reduction' is applied when the scale change is not of concern, while generalisation is applied due to the change in scales. Therefore, data point reduction algorithms should not be used for generalisation purposes. Also different benchmarks such as the manually generalised versions of existing maps should be used for the purpose of evaluating line generalisation algorithms instead of always using the original feature.

Some observations on the issue of line generalisation

Z. Li

Geodata Institute, University of Southampton, Southampton 509 5NH[1]

INTRODUCTION

Generalisation is a traditional topic in cartography. Numerous research papers on this topic have been published in cartographic and geographic journals since the 1950s. Also lengthy review papers have been produced by authors, such as McMaster (1987) and Brassel and Weibel (1988). Furthermore, this topic is becoming increasingly important with the advance of GIS (Geographical Information Systems) research and technology (Marble, 1984; Abler, 1987; Rhind, 1988). However, currently there are many confusions and misunderstandings, for example the mis-use of data point reduction algorithms for generalisation purposes which seriously impede the quality of existing research. The emphasis of this article which is aimed to clarify some of the confusion, is on the generalisation of line features.

A brief discussion of generalisation in general is given, followed by a discussion of the distinction between generalisation and point data reduction of line features. Further discussion of the comprehensiveness of the existing measures and other possible measures for the evaluation of line generalisation is given.

GENERALISATION IN GENERAL

Attempting to define what is meant by the term 'generalisation' proves to be difficult. In fact, "there is no unanimity on terms used to describe generalisation processes" (Keates, 1991). Each definition has its own emphasis or some personal preference.

Indeed, traditional cartographic generalisation is an inconsistent process of map data simplification. Therefore, it is not a purely scientific process but involves a highly artistic skill component, making cartographic generalisation rather subjective and very complex. Keates (1991) decided to describe it as a process of adjusting map information content instead of giving any more precise definition.

Due to its subjectivity there is no coherent theory to guide the cartographic generalisation process, although many researchers have made great efforts on this subject and some of them have been successful in identifying some sort of 'principle' or 'law' (Topfer and Pillewizer, 1964). Most have concluded that "The subject is a complex one and there will be a lengthy period of gestation before any coherent theory of generalisation emerges" (Steward, 1974).

Recently, Li and Openshaw (1993) have argued that generalisation is not only a cartographic process but it should also be regarded as a model creation process for the analysis of any spatial data. Therefore, it is a universal problem of spatial analysis. With such a view, they tried to find some sort of universal principle to make the generalisation process more scientific. Indeed, after analysing some natural phenomena, they identified a so-called *Natural Principle*. Based on this, a family of very promising algorithms for the generalisation of line features has been successfully developed (Li and Openshaw, 1992b) and techniques for 'area-patch' generalisation were implemented by Müller and Wang (1992). This 'natural principle' might well serve as a basis for the development of a comprehensive theory for generalisation.

THE GENERALISATION OF LINE FEATURES

The line is a basic type of cartographic feature, therefore the generalisation of line features has been and still is an important research topic. In fact, most of the literature in the area of generalisation is concentrated on line generalisation. A recent review paper by McMaster (1987) considered more than 40 such papers. In spite of all these efforts, there are many basic problems which the author feels need to be discussed urgently.

A major problem covers terminology and subsequent applications of algorithms. Line generalisation is equivalent to line simplification, but many have gone further to equate line generalisation = line simplification = data point reduction along digital lines. Thus data point reduction algorithms have been regarded in concept, and used in practice, as generalisation algorithms. For example, the Douglas data point reduction algorithm (Douglas and Peucker, 1973) has been considered by many as a standard routine for line generalisation and it has been implemented in many digital mapping and GIS packages under the command of "GENERALIZE".

As Li and Openshaw (1992a; 1992b; 1993) have always emphasized, it is quite misleading to use a data point reduction algorithm for generalisation purposes in theory, and very undesirable results will be inevitable in practice (see Li and Openshaw, 1992a; Wang and Müller, 1992). Conceptually, data point reduction is related to data compression and it is used to for the purpose of saving storage space and obtaining quick display, the purpose of which is to use a minimum number of points while the fidelity to the original line feature is kept as high as possible. Here, the number of points is of primary concern. Therefore, it is solely a digital processing problem: there is no such problem in conventional cartography. However, generalisation, the purpose of which is to modify the original feature to suit the presentation at smaller scale while the main characteristics of the original features are retained, always exists either in coventional cartography or in modern digital (or computer) cartography. Here, the main characteristics of the original feature are of primary importance but the number of points is not of concern at all. Thus, data point reduction is applied only when scale change is not of concern while generalisation is applied only due to the change in scale.

At this point, it may argued that the Douglas-Peucker algorithm has found success for generalisation purposes. However, the algorithm has its application to "the reduction of the number of points required to represent a digitized line or its caricature", as Douglas and Peucker (1973) originally and correctly pointed out. Visvalingam and Whyatt (1990) conclude that "the high performance of Douglas-Peucker algorithm on mathematical evaluations (as described by McMaster) may be interpreted as being indicative of its relative merits as weeding algorithm, but not necessarily as evidence of its superiority as a generalisation algorithm." Therefore, if it is misleadingly applied to generalisation, problems such as spatial conflict in the graphic results, is inevitable. Li and Openshaw (1992a) clearly illustrated that the algorithm designed for generalisation purposes will be capable of coping with even very complex line features while data point reduction algorithms will end up disastrously. *Figure 1* shows the differences in results with a medium complexity line feature. The result is by no means the fault of the data point reduction algorithm itself but the fault of the user who mis-used this algorithm.

MEASURES FOR EVALUATION PURPOSES

To judge or evaluate the performance of an algorithm for line generalisation, some measures or criteria are necessary. Visvalingam and Whyatt (1990) have concluded that existing measures "proposed by some other researchers are inappropriate, misleading and questionable". It is pertinent to briefly describe existing measures, to investigate whether Visvalingam and Whyatt are right or not, and then to discuss possible alternative measures.

It can be found that most investigators use the so-called 'areal displacement' and 'vector displacement' measures. They are the most popular two among the thirty measures which were designed by McMaster (1986). Actually, McMaster's thirty measures have taken into consideration the factors of line length, point coordinates, angularity, curvilinearity, vector difference, polygon difference and perimeter; indeed, almost every aspect of a line feature in the spatial domain. Through statistical analysis, the following six were considered as independent: percent change in the number of coordinates; percent change in the standard deviation of the number of coordinates per inch; percent change in angularity; total areal displacement per inch; percent change in the number of curvilinearity segments.

Visvalingam and Whyatt (1990) have systematically examined these measures and have produced many criticisms. For example, they state "since the angularity measure was strongly influenced by the presence of spikes, the preservation of angularity could not be regarded as a good indicator of the quality of simplification." They also point out that vector and areal displacements are inadequate since "two entirely different geometric shapes could result in the same amount of overall displacement". Visvalingam and Whyatt also criticise the dimensionality measure used by Müller (1987) and point out that "Müller's use of fractal dimension as an evaluative measure is equally questionable, given his own admission that traditional cartographers preserve neither statistical similarity nor fractal dimensionality when generalising lines."

In his original paper (1986), McMaster states that his measures are for the evaluation of line simplification. However, to many including the author, line simplification, as indicated previously, should be used to mean line generalisation; therefore, Visvalingam and Whyatt are correct in their criticisms. On the other hand, considering the fact that McMaster's definition of line simplification is "selecting a subset of the *original* coordinate pairs, while retaining those points considered to be most representative of the line" (Shea and McMaster, 1989), McMaster's measures are adequate for his purpose and his definition. Therefore, nothing is wrong with these measures themselves but, once again, a misunderstanding is due to the confusion in the concept. Similarly, the dimensionality might well serve as an evaluative measure for data reduction.

Thus, existing measures are for the evaluation of data point reduction algorithms and new measures or criteria need to be designed for generalisation purposes. The starting point could be to consider how the generalisation should be evaluated. Li and Openshaw (1992b, 1993) decided to use the manually generalised versions of the original line features at corresponding scales as benchmarks, instead of always comparing the results with the original feature, which is a common practice in the case of evaluating data point reduction algorithms. For example, if a line feature is generalised to suit a scale of 1:50,000 then the same feature at 1:50,000 scale map which was manually generalised should be used for comparison. In such a way, McMaster's thirty measures make sense here. Müller's dimensionality will also be appropriate. Further, Fourier spectra of both manually generalised and automated generalised line features could be produced for a more comprehensive comparison (see Li and Openshaw, 1992b). Indeed, an overlay of the Fourier spectra of the

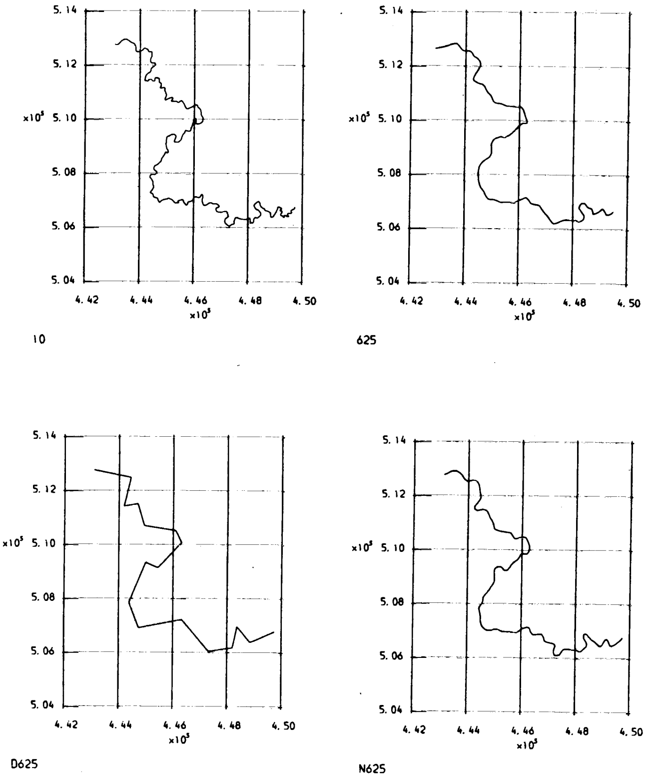

Figure 1. The generalisation of a river feature with medium complexity, by both manual and two automated algorithms, from 1:10,000 scale to 1:625,000 scale. (a) A river feature digitised from 1:10,000 scale OS topographic map; (b) The same river feature digitised from 1:625,000 scale map, which was manually generalised; (c) The river features generalised by a data reduction algorithm (Douglas-Peucker), from 1:10,000 (i.e. the data of *Figure 1a*) to 1:625,000 scale; (d) The river feature generalised by a generalisation algorithm (Li-Openshaw), also from 1:10,000 scale (i.e. the data of *Figure 1a*) to 1:625,000 scale.

results generalised by various algorithms could clearly show the differences in their nature. *Figure 2* is just such an example, showing the spectra of the three generalised results shown in *Figure 1*. Of course, other measures are also possible but it is not the main purpose to discuss this matter in more detail.

CONCLUDING REMARKS

In this short article, the author tries to distinguish line generalisation and line data point reduction, thus clarifying the confusions created in literature and in the practice of system implementation. The main points can be summarised as follows:

(1) Generalisation and data point reduction have different theoretical bases and should serve for different purposes. The current practice of using data point reduction algorithms for generalisation purposes is misleading although some degree of generalisation effect may be created through such a mis-use.

(2) The primary concerns are different for generalisation and point data reduction. For the latter, the number of points is of most concern while for the former it is not of importance at all. Instead, the main characteristics of the original features are important.

(3) Different benchmarks should be used for generalisation and data point reduction. For the former, the manually generalised versions of the features on existing maps could be the appropriate candidate, while for the latter, only the original feature is the appropriate one.

(4) Existing measures will still make sense if the proper benchmark as suggested above is used, but Fourier spectra could even more clearly show the difference in the nature of the generalised results.

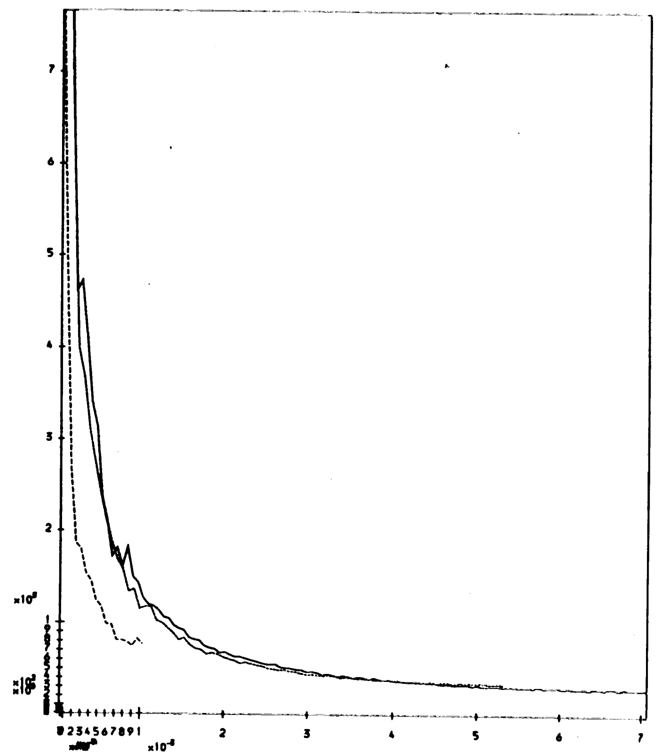

Figure 2. The Fourier Spectra of the three generalised results shown in *Figure 1*. Continuous Line: For the manually generalised result (i.e. *Figure 1b*); Dotted Line: For the results obtained by a generalisation algorithm (i.e. *Figure 1d*); Pecked Line: For the results obtained by a data reduction algorithm (i.e. *Figure 1c*).

REFERENCES

Abler, R., 1987. The National Science Foundation National Center for Geographic Information and Analysis. *International Journal of Geographical Information Systems*, 1 (4): 303–326.

Brassel, K. & Weibel, R., 1988. A review and conceptual framework of automated map generalisation. *International Journal of Geographical Information Systems*, 2 (3): 229–244.

Douglas, D. & Peucker, T., 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.

Keates, J., 1991. *Cartographic Design and Production*. 2nd Edition. Longman Scientific & Technical. 261 pp.

Li, Z., & Openshaw, S., 1992a. Algorithms for automated line generalisation based on a natural principle of objective generalisation. *International Journal of Geographical Information Systems*, 6 (5): 373–389.

Li, Z., & Openshaw, S., 1992b. A comparative study of the performance of manual generalisation and automated generalisation of line features. *AGI Year Book 1992–1993*.

Li, Z., & Openshaw, S., 1993. A natural principle for the objective generalisation of digital maps. *Cartography and Geographic Information Systems*, 20 (1): 19–29.

Marble, D., 1984. Geographic information systems: an overview. *Proceedings of Pecora 9 Conference*, Sioux Falls, S.D: 18–24. Also reprinted in *Introductory Readings in Geographic Information Systems*, edited by D. Peuquet & D. Marble, 1990. Taylor & Francis.

McMaster, R., 1986. A statistical analysis of mathematical measures for line simplification. *The American Cartographer*, 13 (2): 103–116.

McMaster, R., 1987. Automated line generalisation. *Cartographica*, 24 (2): 74–111.

Shea, K. & McMaster, R., 1989. Cartographic generalisation in digital environment: When and how to generalise. *Auto Carto 9*: 56–67.
Müller, J-C., 1987. Fractal and automated line generalisation. *The Cartographic Journal*, 24 (1): 27–34.
Müller, J. & Wang, Z., 1992. Area-patch generalisation: A competitive approach. *The Cartographic Journal*, 29 (3): 137–144.
Rhind, D., 1988. A GIS research agenda. *International Journal of Geographical Information Systems*, 2 (1): 23–28.
Steward, H., 1974. Cartographic generalisation: Some concepts and explanation. *Canadian Cartographer*, 11 (1): 1–77.
Topfer, F. & Pilewizer, W., 1964. The principle of selection. *The Cartographic Journal*, 3 (1): 10–16.
Visvalingam, M. & Whyatt, J., 1990. The Douglas-Peucker algorithm for line simplification: Reevaluation through visualisation. *Computer Graphics Forum*, 9 (3): 213–228.
Wang, Z. & Müller, J., 1992. Complex coast line generalisation. *Unpublished Manuscript, ITC*. Submitted to *Cartography and Geographic Information Systems*.

NOTES

[1]Present address: Department of Photogrammetry and Cartography, Technical University of Berlin, Sekr EB9, Strasse des 17 Juni 135, Berlin 12, Germany.