# Representing Forested Regions at Small Scales: Automatic Derivation from the Very Large Scale.

W. Mackaness, S. Perikleous, O. Chaudhry

The Institute of Geography, The University of Edinburgh,
Drummond St, Edinburgh EH8 9XP
Telephone: (44) 131 6508163
Fax: (44) 131 6502524
william.mackaness@ed.ac.uk

## 1. Context

Being able to view geographic regions at multiple levels of detail is essential to geographic inquiry (Sheppard and McMaster 2004). It affords meaning through the identification of pattern, and interpretation of a palimpsest of processes that operate at various scales (in both time and space). One class of feature commonly found on topographic maps is forest. At the finest scale it can involve the representation of individual trees. And at very coarse scale (say 1:250 000 scale) forests are shown in a way that enables broad classifications of land use, that 'map' to our conceptual understandings of what constitutes (prototypically speaking) 'forest' – such that we are able to conceptualise what is meant by the 'Amazonian Forest', or 'Sherwood Forest' (one of the largest forests in the UK). Rather than the redundancy of multiple databases (each recording forests at these different conceptual scales), surely it is more efficient to maintain a single, highly detailed database, that acts as single point of update? Then to apply generalisation algorithms that, metaphorically speaking, aggregate the detail of the tree, in order to see the forest? The creation of such a system can support integration/conflation of data at different scales (Weibel 1995), 'intelligent zoom' in interactive environments (seeing more detail as you zoom in to the map), scale dependent spatial analysis (analysis of data at a scale appropriate to the task) and exploratory data analysis. This paper presents a technique that automatically creates forested regions for visualisation at a scale of 1:250 000 from very detailed mapping – the Ordnance Survey's (OS) MasterMap (1:1 250/1:10 000) Forestry layer (a vector-based topography layer).

## 2. Methodology

The input data was OS MasterMap data of forest regions around Peebles, in the Scottish Borders (Figure 1). Some regions were used to parameterise the algorithm as part of the initial pilot study, whilst other regions were used to assess the success of that parameterisation process. In this research we used two types of forests: Coniferous and Non Coniferous trees.

### 2.1 Rich get Richer – Poor get Poorer

Initial work drew upon the work of Muller and Wang (1992) who developed an algorithm to generalise groups of lakes. The essence of their methodology was to rank the lakes in order of size, to define a midpoint in the ranking, and for those lakes greater than a certain area, they were further enlarged (by buffering), and those beneath the midpoint, for their areas to be reduced in size. Then any lakes falling beneath some prescribed visual tolerance (the size at which they were no longer discernable to the human eye for a given scale) were removed. Their 'area patch' methodology is analogous to the idea that 'the rich get richer, and the poor get poorer'. Their algorithm was re-implemented and applied to forest patches, but produced disappointing results (Figure 2). Though the algorithm presented here retains the idea that "the rich get richer and the poor get poorer" it was extended to additionally support aggregation of polygons and the elimination of unwanted detail.

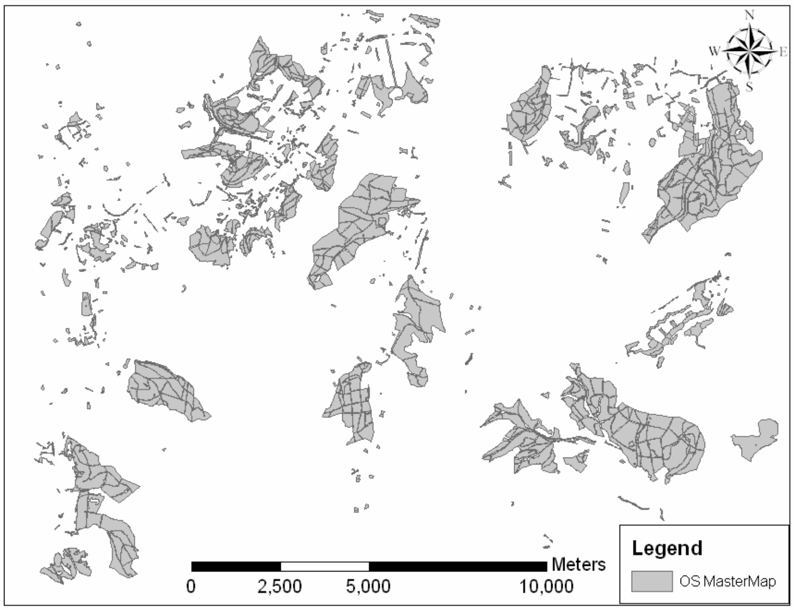


Figure 1: Example of Input Data (Woodland) from OS MasterMap (OS MasterMap© Crown Copyright Ordnance Survey. All rights reserved)

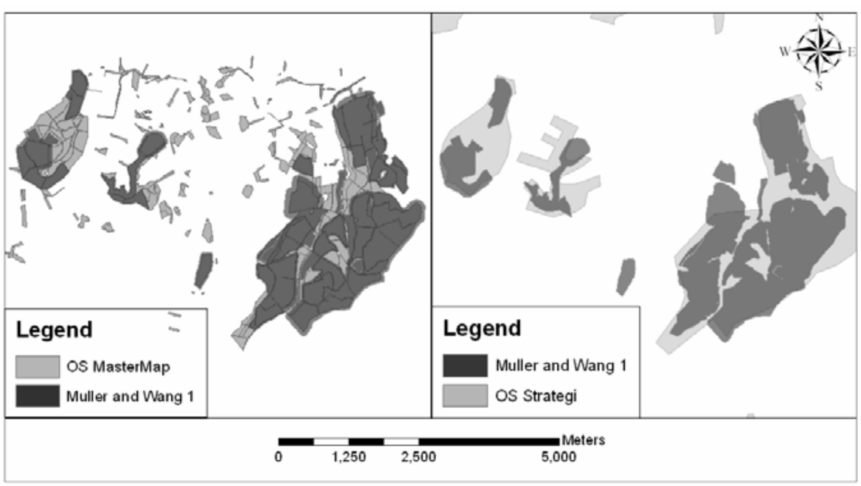


Figure 2: Unsatisfactory results from Muller and Wang's Methodology (OS MasterMap© Crown Copyright Ordnance Survey. All rights reserved)

The methodology presented here works by buffering (either enlarging or shrinking) the size of forest patches. When resulting patches overlap, they are joined (union'ed). Small holes within regions are 'patched' and reclassified to that of the surrounding patch. The algorithm takes into account groups of small patches. Where there is sufficient density of small patches, these are union'ed. As a final step, the areas are significantly shrunk, and then re-enlarged. The selection of various tolerances (how much to enlarge by, what size of hole is deemed to be too small, what is the tipping point at which some grow larger, and some grow smaller), was done by empirical analysis. In other words, various output was examined using different tolerances until the desired output was consistently produced. The key stages are visually summarized in Figure 3.


Figure 3: visual summary of the methodology (OS MasterMap© Crown Copyright Ordnance Survey. All rights reserved)

## 2.2 Pre selection and union

Original forest polygons provided by OS included very small patches which are not useful to a generalized solution but create high processing overheads. Therefore those patches had to be eliminated before starting any calculations, and patches within a certain distance threshold were aggregated. This reduced the number of patches by up to 90% depending on the amount of fragmentation and the density of patches.

The thresholds (used either for selecting or buffering) were empirically derived after numerous tests with small test data sets. For example inspection of the dataset revealed that there were many patches with an area less than 5000 m2(T 1) a size that is practically invisible to the human eye at 1: 250 000 scale. The buffering threshold (T 2) (used in step 1 and 2) was derived by inspection of dense collections of small areas and was set roughly to half the width of the roads. This threshold was trying to merge patches separated by narrow roads. The most important threshold, (T 4), was the size above which patches were expanded (Step 3). The threshold for elimination (Ws) was originally set by calculating the smallest visible patch found at 1:250 000 scale. The remaining threshold values where identified empirically.

## 2.3 Promotion and Elimination

All the important data selection was done in steps 3-5 (Figure 3). Its effectiveness in choosing and creating the final woodland patches was crucial to the success of the overall process. Any patch larger than T4 was expanded by a blanket width defined by equation 1 (Muller and Wang ,1992).

$$\text{Blanket Width: } ti = (ci)(K)/\sqrt{|a_i - T4|} \qquad (1)$$

where K is the constant for scaling the blanket width

$K = t^*/\sqrt{(Max - T4)}$

t* = threshold provided by user( larger scale reduction = larger value of t*)

Max = Maximum patch area of the input data

T 4 = Threshold for expansion or contraction provided by user

and ci is the compactness index of patch ai

ci = ai/(pi/4π)

ai = Area of patch i

pi = Perimeter of area ai

π = Mathematical pi = (3.142)

Equation 1: Muller and Wang's equation for defining the amount of expansion.

The objective of the next step was to 'promote' small areas if they were part of a dense region. This was done by checking the remaining patches of the first filter to see if they were larger than 3Ws/4 (Threshold extra) (where Ws is the elimination threshold), and at the same time "close" to at least two other patches (Threshold close).Ordinarily, small isolated patches would be removed. But this intermediate threshold (T extra) meant that clusters of small patches were retained. The key idea was that if a patch contains two or more patches within a certain threshold distance (i.e., close) it is considered important and therefore is expanded to increase its chance of surviving the elimination filter that follows. The 'close' threshold was defined empirically and surviving patches were buffered (expanded) by a blanket width equal to ti (Blanket Width). All other patches not following this rule were contracted by a blanket width ti. Once this buffering process was completed, all those patches smaller than Ws were eliminated.

The result of these steps is a dataset with only large patches completely unified without holes, unless a hole is larger than threshold (T5). Patches tended to have sharp edges or unwanted cavities. These needed to be simplified. Simplification was achieved by making first a large positive buffering, a union operation, and then an equivalent negative buffering.

# 3. Model Implementation

The entire model was implemented using Eclipse Java Platform and Java Language programming. The Java Topology Suite's (JTS) libraries provided important methods such as the buffering and polygon union of overlapping and touching patches. JUMP (Java Unified Mapping Platform) was also used to view the input and output results. ESRI ArcGis and MapManager8 was used for initial data processing of OS supplied data, and resulting shapefiles. The algorithm presented here has recently been developed so that it is available as a JAVA generalisation algorithm for the WebGen (Generalisation Web Services) platform, enabling its deployment as a Web Service (http://www.ixserve.de/).

# 4. Case Studies and Evaluation

Figure 4 shows output from the algorithm for the input shown in Figure 1 (the two are overlaid in order to facilitate comparison). Evaluation was done firstly by a direct visual comparison between the output and the OS Strategi map (Figure 5 shows the results overlaid with OS Strategi data).

# 5. Evaluation

The algorithm was applied to a second geographic region (completely independent of the first) without any adjustment being made to the parameters. The results were equally encouraging. The results were examined by cartographers at the Ordnance Survey who commented that 'whilst the algorithm has performed well, there are slight question marks concerning the minimum size threshold' and that… 'although the granularity obtained was adequate, the outlines could be simplified slightly more'. It is always important to remember that these results are intended for display at 1: 250 000 (Figure 6). At this scale some of the discrepancies identified at the fine scale are no longer discernible. It is important that the algorithm is attempting to characterise a region.

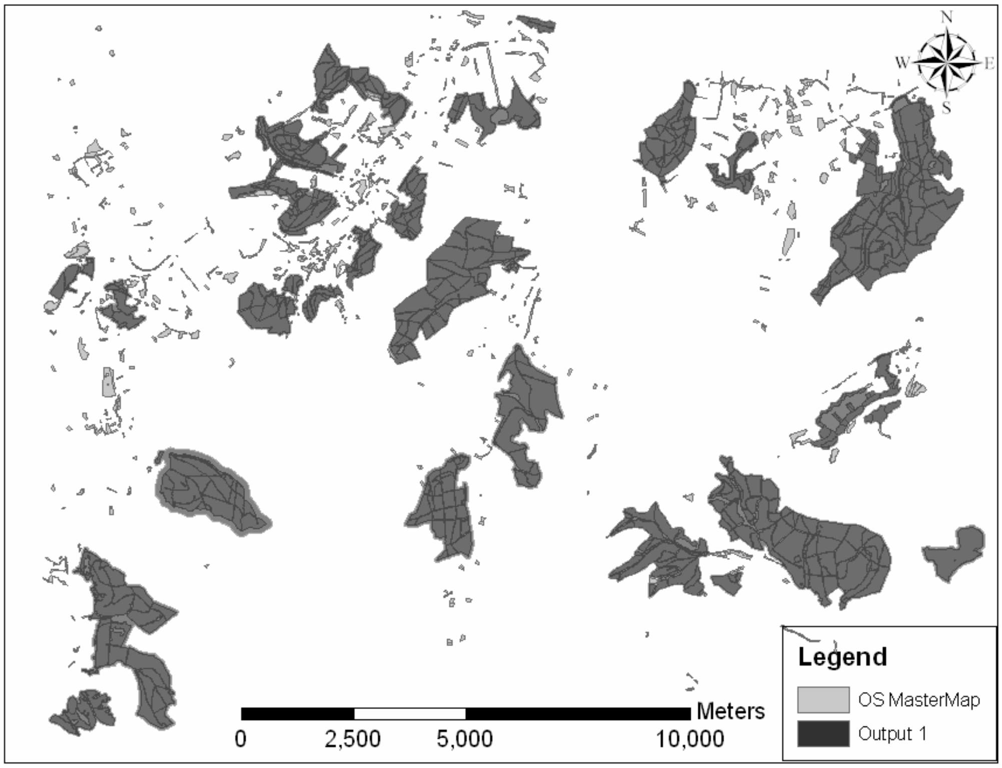


Figure 4: Output produced by the algorithm from input data in Figure 1. (OS MasterMap© Crown Copyright Ordnance Survey. All rights reserved)

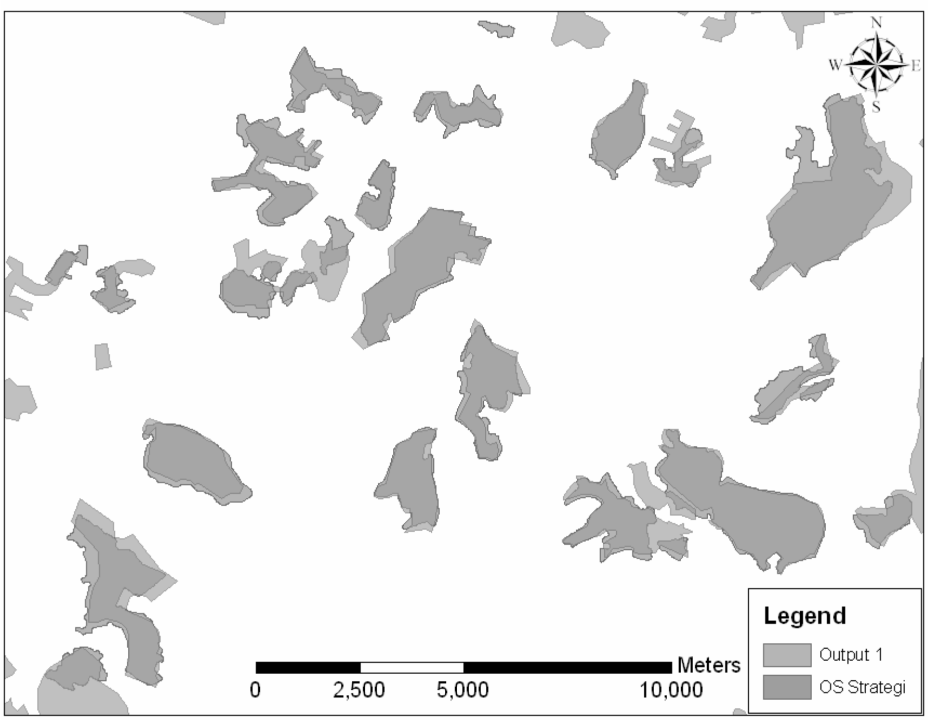


Figure 5: Output from the algorithm overlaid with OS Strategi data. (Mapping is Ordnance Survey ©Crown Copyright. All rights reserved).

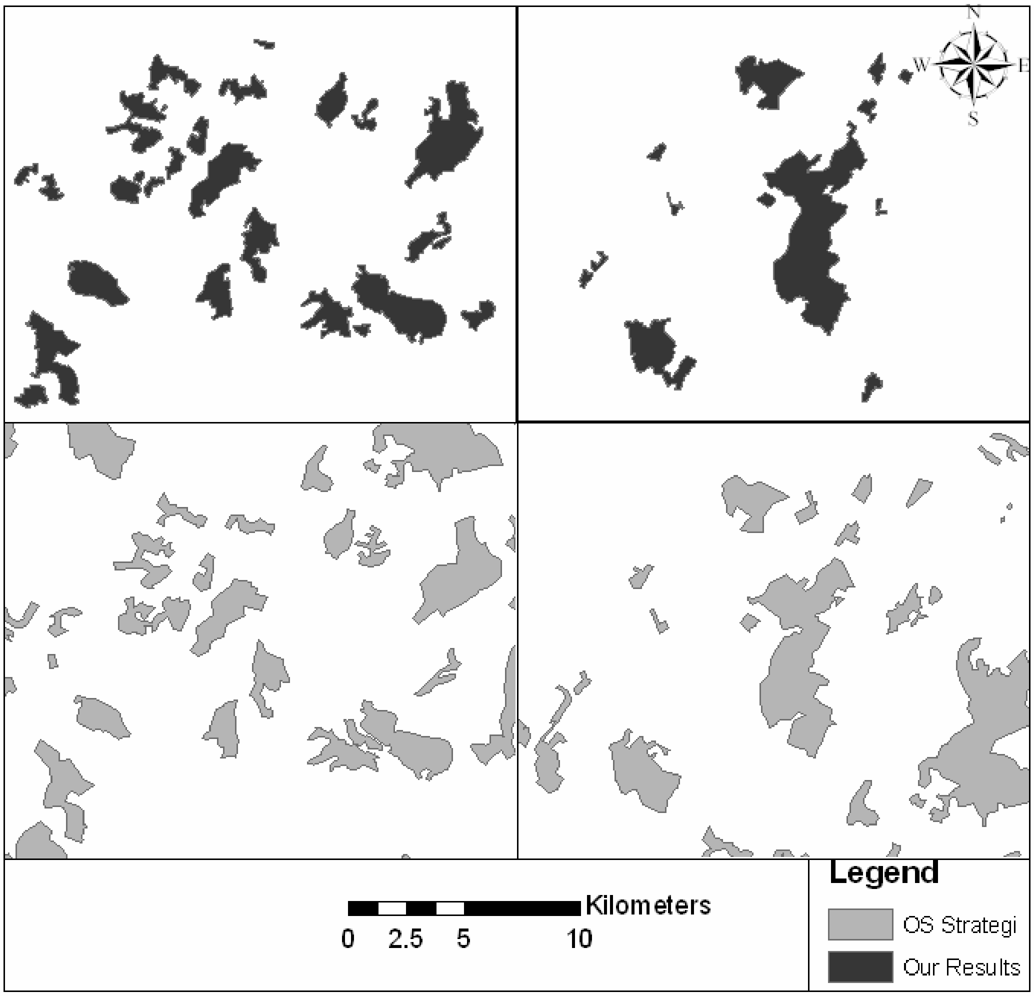


Figure 6: Output and OS Strategi at 1:250 000 scale – spot the difference? (Mapping is Ordnance Survey ©Crown Copyright. All rights reserved).

# 6. Conclusion

The two case studies confirm the general applicability of the algorithm and point to future work (such as improved smoothing and scalability of the algorithm). By making this algorithm available via the webGen JUMP service, it is hoped that other researchers can utilize and make further improvements to the algorithm.

# 7. Acknowledgments

Our thanks to the Ordnance Survey in support of this project.

# 8. References

BUTTENFIELD, B.P and MCMASTER, R.B, 1991. Making rules for Knowledge Representation, Longman:London

DOUGLAS, D.H. and PEUCKER, T.K., 1973. Algorithms for the reduction of the number of points required to represent a digitized ligne or its caricature. The Canadian Cartographer, pp 112-123.

GOLD, C.M., 1998. The Quad-Arc data structure. 8th International Symposium on Spatial Data Handling; (Vancouver)1998. pp. 713-724.

GOLD, C.M., NANTEL, J. and YANG, W., 1996. Outside-in: an alternative approach to forest map digitizing. International Journal of Geographical Information Systems, pp. 291-310.

MACKANESS, W. and MACKECHNIE , G.A., 1999.Automating the Detection and Simplification of Junctions in Road Networks, Geoinformatica, 3:2, pp 185-200

EDWARDES A, MACKANESS WA. and URWIN T (1998) Self Evaluating Generalisation Algorithms to Automatically Derive Multi Scale Boundary Sets. In: Poiker T, Chrisman N (eds) 8th International Conference on Spatial Data Handling, Vancouver, pp 361-372

MULLER, J.C and WANG, Z.,1992. Area Patch Generalization: A competitive approach, The Cartographic Journal vol 29 pp 137-144

PERIKLEOUS, S. (2006) Automatic Derivation of Forest Patches for Representation at 1:250 000 Directly from OS Master Map. Unpublished MSc Thesis, University of Edinburgh, Edinburgh.

REGNAULT, N., 1996. Recognition of building clusters for generalisation. Proceedings of 7th International Symposium on spatial data handling, pp 1-13

REVELL, P., 2005.Seeing the Wood from the Trees: Generalising OS MasterMap® Tree Coverage Polygons to Woodland at 1:50 000 Scale , 8th ICA Workshop on Generalisation and Multiple Representation, July 7-8, 2005, A Coruña, Spain

SHEPPARD E, MCMASTER RB (2004) Scale and Geographic Inquiry: Nature Society and Method. Blackwell Publishing.

THOM, S. 2005. ‘A strategy for collapsing OS integrated transport network dua carriageways’ In 8th ICA Workshop on Generalisation and Multiple Representations. July 7-8, 2005, A Coruna, Spain.

VISVALINGAM, M., and WHYATT, J. D., 1993. Line Generalisation by repeated elimination of points, Cartographic Journal, 30 (1) 46 - 5.

WARE, J.M., JONES, C.B. and THOMAS N., 2003 Automated Map Generalisation with Multiple Operators: A Simulated Annealing Approach. International Journal of Geographical Information Science, pp 743-769.

WEIBEL, R., 1995. Three essential building blocks for automated generalization. In: GIS and Generalisation - Methodology and Practice. Edited by Muller, J.C., Lagrange J.-P. and Weibel, R. pp. 56-69.

WEIBEL R. and JONES C.B., 1998 Computational perspective on map generalization, GeoInformatica, 2, pp. 307–314.

## Biography

Stathis Perikleous graduated from the MSc in GIS at The University of Edinburgh in September 2006. William Mackaness is a lecturer and Omair Chaudhry is a PhD student in the Institute of Geography at The University of Edinburgh.