Seeing the Wood from the Trees: Generalising OS MasterMap® Tree Coverage Polygons to Woodland at 1:50 000 Scale

Patrick Revell
patrick.revell@ordnancesurvey.co.uk
Ordnance Survey®, Research and Innovation, Romsey Road, Southampton, UK

Abstract

The OS MasterMap product from Ordnance Survey (OS) is sourced from a detailed geographic database with Great Britain coverage. OS are conducting research into deriving smaller scale cartographic products automatically from this database. This paper presents details of research into vector-based automated model and cartographic generalisation of the source topographic tree coverage polygons to woodland represented at 1:50 000 scale. The whole symbolisation process is addressed to ensure that the results conform as closely as possible with the current Ordnance Survey raster product specification.

Firstly, tree polygons are clustered into autonomous groups, then the polygons are amalgamated into woodland. For each wood, the bounding features are identified and then the wood boundary is simplified with respect to these features. The next step is to slice up the woods by the traversing network features, resolve the conflicts and place tree symbols in the generalised woods. The paper concludes by discussing how the results will be evaluated and what remains to be implemented.

1 Introduction

1.1 Generalising from the base scale

In 2001 Ordnance Survey launched the vector-based Topography Layer in the OS MasterMap suite of products. This product forms a complete coverage of Great Britain at a high level of detail; being surveyed at 1:1250 scale in urban areas, 1:2500 scale in rural areas and 1:10 000 scale in mountain and moorland areas. In 2003 OS MasterMap was enhanced with the addition of the Integrated Transport Network™ (ITN) Layer. ITN provides a complete vector road network of Great Britain and is consistent with the Topography Layer. OS MasterMap is supplied from a seamless national large scale database, which is constantly being updated with the latest changes (Ordnance Survey 2005a).

The creation of this national large scale geographic database introduced the possibility of automatically deriving smaller scale vector and raster products from a single source. This was a key step in realising the “capture once, use many times” philosophy of OS. The logical starting point for researching this possibility is to develop automated tools for deriving the traditional OS map scales from the base scale data. Maintaining links between the source and derived features will ensure precision and consistency with the base scale, coupled with vastly reduced update costs. Once the traditional products have been derived, the aim would then be to extend the tools for generalising to arbitrary scales and specifications.

1.2 Generalising to 1:50 000 scale

One of the key traditional products of OS is the 1:50 000 scale (1:50k) raster coverage of Great Britain. This retails as both the Landranger® paper map series and as 20 x 20km image tiles (Ordnance Survey 2005b). The raster data are currently revised and updated manually. A choice has been made to focus generalisation research on developing tools to achieve OS specification 1:50k data automatically from the base scale data. The large transition in scale should illustrate the common generalisation problems, which will be of benefit when the research moves on to investigate automated creation of 1:10 000 or 1:25 000 scale data.

Due to the popularity, longevity and iconic nature of the Landranger product, the 1:50k specification is to a certain extent “owned by the nation”, and thus is difficult to modify without good reason. Therefore the results of the automatic generalisation must conform as closely as possible to the current Ordnance Survey raster product specification. Extracts from the specification are provided in section 2.
The Clarity software from Laser-Scan has been selected as the platform for researching the generalisation tools. Clarity includes a re-implementation of the Agent core (Agent 2000; Regnauld 2002) and a Java API for customising the system (Laser-Scan 2003). The Agent core allows a hierarchical generalisation strategy to be developed, where the meso-agents in the upper levels are used to control groups of features which have a geographic or cartographic meaning. Meso-agents are able to create and control other meso-agents, while at the bottom of the hierarchy they control micro-agents. Micro-agents are responsible for the generalisation of individual cartographic features.

So far the Ordnance Survey research into 1:50k generalisation has produced promising results for urban buildings, rural buildings and roads (Revell 2004; Revell et al. 2005). The high proportion of map space occupied by woodland, and the close interaction with many other feature classes gives generalisation of the tree coverage polygons from the base scale a high priority as the next topic of research. Details of the source data model are provided in section 3.

1.3 The digital cartographic and landscape models

A digital cartographic model (DCM) can be defined as the digital equivalent of a specific map, amenable for visual communication (Brassel and Weibel 1988). As the current 1:50k product is raster, it was necessary to propose a new vector DCM for representing the woodland at this scale. The DCM presented in section 4.2 will support the whole symbolisation process, to ensure the result conforms as closely as possible with the current raster specification. This approach is consistent with recent moves towards database driven cartography (Hardy et al. 2004), as opposed to employing separate symbolisation software such as Mercator for map finishing.

The DCM is populated at the end of the processing using a sequence of cartographic generalisation operations, which enhance the cartographic appearance of the data. Other operations are required before this stage is reached, which reduce the level of detail from the base scale data, whilst retaining positional accuracy. Such operations are termed model or statistical generalisation (Brassel and Weibel 1988). These operations are used to populate a digital landscape model (DLM), which for the purposes of this research is just a transitional step on the way to the DCM. Details of the DLM can be found in section 4.1.

It is useful to break the problem down into model generalisation for populating the DLM (section 5), followed by cartographic generalisation for populating the DCM (section 6), particularly since the transition is large from the base scale to 1:50k. At present it is unclear whether the DLM and DCM should persist after they have been initially populated. This decision is dependent upon wider Ordnance Survey operational decisions regarding derived data, such as the means by which updates are propagated. The paper concludes with a discussion of what remains to be implemented and how the results from the prototype will be evaluated.

2 The target specification

Information in this section is taken from the OS 1:50k raster specification (Ordnance Survey 2000). Numerical values for the threshold parameters exist, although they are not provided in this paper.

2.1 Woodland extents

All woodland is represented at 1:50k with a green filling as in Figure 2. Avenues of trees are not shown. Woods are always shown if the extents are greater than the minimum guaranteed retention size. Smaller woods and clumps of trees are included only if they form prominent landmarks and in which case must be enlarged to a minimum depiction size (smaller than the minimum guaranteed retention size).

The term “prominent landmark” is an example of the subjective elements in some of the current specifications. Detecting if a wood is a “prominent landmark” could potentially require an analysis of the terrain to determine the places from which the woods are visible. For the moment “prominent landmark” is assumed to mean that the trees are isolated from any other areas of woodland. This definition will probably require additional tuning during development of the algorithms.
There is a threshold for the minimum width of woods. Where necessary, woods or parts of a wood are enlarged to this minimum. This applies particularly to long narrow strips of woodland. If the wood is shown either side of a track where the wood fill would normally be shown over the track, the minimum width applies from the edge of the track to the perimeter of the wood. This is to prevent graphic conflict between the track edge and the perimeter of the wood.

Formal specification documents rarely include all the parameters required for automatic generalisation, since much of this knowledge is implicit in the skill of the cartographer and embedded in the current product. For example, there is no threshold value for minimum hole size in woodland, so observations had to be made by comparing the base scale data with existing 1:50k mapping. A suitable threshold was determined by calculating the sizes of the largest eliminated holes and the smallest retained holes for a sample region.

2.2 Woodland symbols

Coniferous, non-coniferous and mixed woodland types are represented using the symbols in Figure 1, although the accuracy is only that required to permit the correct positioning of tree symbols. Scrub is shown as non-coniferous when extensive and forming a dense cover. The terms “extensive” and “dense cover” are further examples of the lack of precision in the specifications, requiring specific work to tune the generalisation algorithms.

![Figure 1. Symbols used to represent types of tree coverage at 1:50k.](image)

All woods have the appropriate tree symbols, which are restricted to areas of green fill. The relative positions of the coniferous and non-coniferous symbols may vary in mixed woodland, although they tend to be plotted in pairs. Trees are positioned in a screen pattern right up to the wood boundary. Symbols may be broken by the perimeter detail. When adding trees the screen pattern is followed except in small woods where a complete tree is shown. In very small woods where there is no room to show a complete symbol the base of the symbol is positioned in the wood, showing as much of the symbol as possible.

2.3 Woodland boundaries

Where the woodland is unbounded by other map features, a thin pecked black line is used to represent the perimeter. Otherwise, where an obstructing feature (fence, hedge, wall or bank) encloses the woodland, the boundary of the green fill is represented as a continuous thin black line. This includes buildings, whose walls can form part of the perimeter of the wood. Where water forms the boundary, or a boundary fence is very close to water, the water is shown as forming the boundary. In this case, the perimeter of the wood is represented as a blue line in the appropriate gauge for the inland water.

![Figure 2. 1:50k extract showing fenced and unfenced tracks bordering woodland.](image)
Paths or dismantled railways are not used as boundaries for woods. When this occurs the wood boundary is shown by pecks, and the path or dismantled railway is shown either inside or outside the wood. Figure 2 shows that the green fill runs right up to the continuous perimeter lines, and otherwise runs up to the outermost pecked perimeter line.

### 2.4 Woodland traversing features

Rivers, canals and classified roads have two parallel casing lines and a coloured fill which breaks the green filling of woods as shown in Figure 3. Tracks and roads with no colour fill also have two parallel casing lines, but acquire a green filling inside when they cross the woodland. The casing lines for roads and tracks are solid when bounded by obstructing features and are pecked otherwise. Firebreaks are shown by two parallel pecked lines which contain no green fill to distinguish them from tracks. The gap between the firebreak pecked lines is shown to scale subject to a *minimum width*. Narrower firebreaks are enlarged to this minimum.

![Figure 3. Rivers, streams, tracks, paths and a road traversing woodland at 1:50k.](image)

Rivers are shown with a single thick black line which also visually breaks the green filling. Railways may be enclosed by cuttings or embankments, making for a wider break in the green fill. Rights of way (single pecked pink lines in Figure 2), paths (single pecked black lines in Figure 3) and streams (single blue lines in Figure 3) are plotted over the top of the green fill and do not break the visual appearance of continuous woodland.

### 3 The source data

#### 3.1 Woodland extents

In the source database a single vegetation coverage polygon is defined as an area whose boundary is completely surrounded by polygons of other classes, or by obstructing feature, path and water lines. This means that something which is perceived as a wood can be represented by many polygons in the source data, divided by network features such as roads, railways, tracks, paths, rivers and streams. The characteristics of the source data are illustrated in Figure 4.

#### 3.2 Woodland symbols

The vegetation polygons to be considered for generalisation to woodland at 1:50k are those with *descriptive term* attribute set to “Coniferous Trees” or “Nonconiferous Trees”. A single feature can have multiple values for the *descriptive term*, which is used to indicate when the coverage is mixed. The *descriptive term* can also be “Coniferous Trees (Scattered)” and “Nonconiferous Trees (Scattered)” and “Scrub”. Such polygons are only relevant if forming part of a denser surrounding tree coverage as per the specification, and will be ignored for the purposes of this research.
3.3 Woodland boundaries

Obstructing features are represented in the source database by linear features with physical presence attribute set to “Obstructing”. The edge of water is shown by using linear features with descriptive group attribute set to “Inland Water” and physical presence “Edge / Limit”. It is assumed that the buildings are generalised before the woodland, so that the sides of the generalised building polygons can be used to form part of the woodland perimeter. In addition woodland may be bounded by the woodland traversing features, discussed in the next section.

There may be narrow strips of other land types between the tree coverage and the bounding features, which will need to be eliminated by the generalisation. An example of such a scenario is shown on the right side of Figure 4.

![Figure 4. Disjoint tree coverage polygons (green) and obstructing features (black). Ordnance Survey © Crown Copyright. All rights reserved.](image)

3.4 Woodland traversing features

Linear features which traverse woodland are unfortunately not modelled as networks in the source database with the exception of the ITN road centrelines. Rail, inland water, paths and tracks are represented using disjoint polygons and lines. An algorithm has been developed for constructing a hydrology network from the inland water polygons and lines in the source database (Greenwood et al. 2003). It will be necessary to extend this algorithm to create the other networks which are required for generalising woodland, although this development is outside the remit of the current research.

Firebreaks are not modelled explicitly in the source database, so will require some work to distinguish them from other breaks in the woodland coverage such as glades and clearings. Firebreaks are very uncommon on the 1:50k product, so for the purposes of this research they are ignored.

4 The target data model

4.1 The DLM

This section contains details of DLM which is used as a staging post on the way to the 1:50k DCM.

4.1.1 Woodland extents

The extents of woodland should be modelled as polygon features, which are permitted to contain holes. The traversing networks, such as roads, should be ignored when determining the extents of woodland from the source data. From a spatial cognition point of view a woodland is usually perceived to be unbroken when a network feature passes through it (Bennett 2001), so this is an appropriate way of modelling the landscape. In addition, modelling the extents in this way will support creating of future cartographic products, allowing flexibility in which traversing networks are shown.
Each *woodland extents polygon* should have a reference to all the source tree coverage polygons which it contains. Maintaining references between the source and derived data is an important part of the proposed data model as shown by Figure 5. This approach will make later processing steps more efficient and will support incremental update of the derived data.

### 4.1.2 Woodland boundaries

The boundaries of the woodland should be modelled as linear features. The inner and outer rings of each *woodland extents polygon* should be comprised of *woodland boundary line* features. Each *woodland boundary line* should be referenced to the obstructing feature, linear network feature or generalised building which it follows. The reference is not populated if there are no bounding feature in the source data. Each *woodland boundary line* should also have reference back to the *woodland extents polygon* to which it belongs and ideally it should be topologically structured against it.

![Figure 5. Simplified UML diagram linking the source classes (blue) to the derived classes.](image)

### 4.2 The DCM

This section details the 1:50k DCM which will be populated by transforming the DLM data. The classes devoted purely to the symbolisation are highlighted in yellow on Figure 5.

#### 4.2.1 Woodland traversing features

The traversing features of importance when generalising woodland are those which break the green fill and tree symbols on a 1:50k map. Paths, rights of way and streams are ignored since they have a thin representation, and attention is focussed on the thicker features which occupy a large amount of map space. Hence the rivers, canals, classified roads and railway networks should be converted to polygons to prevent the woodland fill, boundaries and tree symbols from occupying the same portion of the map. Each *woodland extents polygon* should have a reference to any such *traversing network polygons* which it contains.

Modelling the 1:50k cartographic representation of these networks explicitly as polygons will facilitate easy identification and resolution of problems such as narrow strips of woodland caused by the masking. This masking operation would normally be performed by separate symbolisation software, however it is believed that greater accuracy and control over the result can be achieved with an integrated solution.

#### 4.2.2 Woodland fill polygons

The remaining internal areas of the *woodland extents polygons* should be classified as *woodland fill polygons*. The *woodland fill polygons* are the areas where it is permitted to show the green fill and the tree symbols. The *woodland fill polygons* should have a reference back to the *woodland extents polygon* in which they are contained. Thus each *woodland extents polygon* will be completely covered by a tessellation comprised of *woodland fill polygons* and *traversing network polygons*, which can be easily identified by following references.
4.2.3 Woodland symbols

The coniferous and non-coniferous tree symbols should be point features, with a representation taken from a symbol library. The coniferous/non-coniferous tree pairs for mixed coverage should be modelled by a single point, with attributes describing the X-Y offset for the coniferous and non-coniferous symbols. The coniferous, non-coniferous and mixed tree point features are only permitted inside a woodland fill polygon and should have a reference to the appropriate woodland fill polygon.

5 Populating the DLM

5.1 Identifying clusters of tree polygons

The first step in generalising the tree coverage polygons from the source database is to break the problem down into manageable autonomous units. To achieve this, a generic clustering algorithm has been developed in Clarity. The basis for this clustering algorithm was an algorithm employed for detecting urban areas and rural building clusters (Revell 2004). The original algorithm operated by buffering the individual buildings, then dissolving the buffers together to form clusters. All the geometry processing was carried out in memory, so it was unsuitable for large datasets.

The generalisation research for woodland took the opportunity to make the existing algorithm scaleable and totally generic so it will cluster any group of point, line and area features. The algorithm still takes a buffer size parameter, although it now limits the in-memory buffering and dissolving operations to batches of 800 objects. After each 800 objects, the intermediate result is stored in the database. A spatial search is used to detect existing temporary database clusters with which they overlap and a combination is performed when the interim result is stored. Hence at any intermediate stage there are no overlaps between the cluster geometries. The process continues until all of the objects have been processed.

The algorithm takes optional parameters which allow holes to be removed from the clusters and the clusters to be simplified using a combination of Douglas-Peucker (1973) and morphological dilation/erosion. There is also an option to filter the result into a number of distinct feature classes according to the size of the cluster. These optional parameters are not used for the woodland generalisation, but are essential for the detection of urban areas and rural building clusters.

Figure 6. Results of clustering (blue) the base scale tree coverage polygons (green).

Ordnance Survey © Crown Copyright. All rights reserved.
Figure 6 shows the result of applying the clustering algorithm to an extract of tree coverage polygons from the source database. A reference is stored between the clusters and the objects they contain, to avoid expensive spatial searches in later processing steps.

5.2 **Amalgamating tree polygons into woodland**

Now the clustering is complete, each cluster is activated as a meso-agent, responsible for generalising all the tree polygons which it contains. Within the clusters it is necessary to amalgamate the tree polygons by proximity to form *woodland extents polygons*. The same proximity threshold is used for the amalgamation as for the clustering. However, the amalgamation may produce more than one *woodland extents polygon* per cluster depending on how the amalgamation algorithm operates. Various approaches were considered for the amalgamation including using a shrink-wrap hull or the dilation and erosion morphological operators.

Amalgamation using a constrained Delaunay triangulation was selected as the most flexible and accurate approach to amalgamation (Jones et al. 1995; Regnauld 2003). This is because it operates at the level of individual vertices; the triangles forming bridges between the features to be amalgamated. The algorithm can take full control over the criteria for which bridging triangles are selected, while maintaining the shape of the original features in the resulting amalgam. A new Java-based triangulation package has been developed for use with Clarity (Regnauld 2005), which provides the essential tools for implementing a triangulation-based amalgamation algorithm.

The first step is to dissolve polygons together which share a common boundary and store the result as temporary database objects. All of the tree polygons are now disjoint and the next stage is to build bridges between the polygons using a constrained triangulation.

Each triangle is processed and the triangle nodes are considered in three pairs. For each pair, a check is made to see if they belong to the same tree polygon. If two of the pairs connect distinct tree polygons and the remaining pair belongs to only one tree polygon, then the triangle is a candidate for a bridge. The shortest distance inside the triangle and between the tree polygons is then measured. If this distance is shorter than the maximum amalgamation distance, the triangle is retained. This approach works well, although some additional filtering is needed to remove triangles which conform to the criteria, but occur at the end of bridges as shown in Figure 7.

![Figure 7. Superfluos triangle (red) at the end of a bridge (yellow), requiring additional filtering.](Ordnance Survey © Crown Copyright. All rights reserved.)

When all the bridge triangles have been gathered together, the next step is to group them together into sets of adjacent triangles. This is easily achieved with an iterative method which makes use of a reference from each triangle to its adjacent triangles. The groups of adjacent triangles are dissolved together and the bridges are stored as temporary objects in the database. This prevents problems associated with holding too many geometries in memory.

The next stage is to dissolve the bridges together with the tree polygons and identify the new holes formed by the presence of the bridges. Such holes are shown in red in Figure 8. Holes in the original tree polygons are not included in this identification. The minimum woodland hole size is used as a parameter to instruct the algorithm to remove small holes. The filled holes are then stored as temporary bridge objects in the database, and dissolved together with any adjacent temporary bridge objects. Note that the holes which are junctions between tracks cannot be removed solely using the criteria of “having only bridges as neighbours”, since such a test would fail to identify T-junctions.
There may still be holes formed by the presence of the bridges which are above the minimum woodland hole size, but are very long and narrower than the maximum amalgamation distance. Such holes are bounded on one side by a temporary bridge, and on all other sides by a single tree polygon. The initial bridges were between distinct polygons, so did not detect these narrow holes, as shown by Figure 9. Such narrow holes are detected using another triangulation. Patches for these narrow holes are also stored in the database and dissolved together with any adjacent temporary bridge objects.

Now that all the holes have been filled, the temporary bridge objects in the database are checked to see if there are any one-triangle-bridges. A one-triangle-bridge is insufficient to connect two features, so a decision has to be made as to whether to retain or remove them on a case by case basis. If there are other more substantial bridges between two tree polygons, then any isolated triangles can be removed.

If there are no other more substantial bridges between the two features, then the two adjacent triangles must be inspected to see how much longer they are as compared to the maximum amalgamation distance. If they are not far off, then the one-triangle-bridge can be enlarged using the adjacent triangles to create a valid bridge. Otherwise the case is analogous to two buildings touching corner-to-corner, which is not perceived as a proximity conflict as in Figure 10. In such cases the one-triangle-bridge can be deleted.
The final step is to dissolve the bridges together with the temporary tree polygons to form a number of woodland extents polygons. At this stage a reference is created between the source tree coverage polygons and the woodland extents polygons in which they are contained. Results of the amalgamation are shown in Figure 11. Note that the inlet at the top of the source data has been closed off by the amalgamation to form a hole in the amalgam. It could be argued that the inlet should have been enlarged rather than eliminated. If the result is shown to be incorrect in the evaluation stage, the algorithm could be refined by detecting and removing short bridges connecting the current outer boundary to any inner holes.

![Figure 11. Results of combining disjoint tree coverage polygons into a single amalgam.](image)

5.3 Detecting woodland bounding features

Each newly created woodland extents polygon now becomes a meso-agent, responsible for managing all of the features to which it has a reference – its micro-agents. In the next stage each woodland extents polygon detects its bounding features. This is an important step since on the 1:50k maps the bounding features control the extents of the woodland in many cases. This and the remaining sections of the paper are theoretical discussions and the algorithms described have not yet been implemented.

Suitable bounding features must be allocated to the inner and outer rings of the boundary. Potential bounding features could be obstructing features, generalised buildings or any linear network features. Some boundary features will be entirely coincident with the perimeter of the woodland extents polygon, so these can be easily identified. In the remaining cases where the boundary features are not coincident, a buffer could be used for selecting candidates for the boundary in close proximity. These need to be mapped on to the perimeter of the woodland extents polygon, possibly by casting a normal outwards at regular intervals along the perimeter.

A check must be made to see if there is an intersection with the walls of a generalised building. The buildings are generalised prior to the woodland since they are small and define anthropogenic phenomena which the more continuous woodland coverage can surround with flood fill. The buildings generalisation does not take account of woodland, therefore some generalised buildings are likely to have eaten into the map space of the woodland extents polygons. For this reason the woodland extents will need to be modified, in addition to creating woodland boundary line objects consistent with the new perimeter. In some cases it might be more appropriate to adjust the position of the building and this issue will be investigated during the implementation.

The woodland extents polygon may need to be modified to run right up to the bounding features and appropriate woodland boundary line objects should be created. For example, if a track or dismantled railway runs along the edge of a woodland, and there is no obstructing feature in between, the woodland extents polygon should be expanded to include the track or dismantled railway.
Any remaining unallocated segments of the boundary should be assigned *woodland boundary line* objects with no reference to a bounding feature. The information gathered by this processing will be used to determine the colour, thickness and plot style (continuous/pecked) for each portion of the woodland boundary. All *woodland boundary line* objects should be topologically structured against the *woodland extents polygons*. This will allow consistency to be maintained for the shared geometry during generalisation.

6 Populating the DCM

6.1 Simplifying woodland boundaries

The portions of the woodland extents polygons shared with the linear network features and by generalised buildings are not available for simplification. The linear network features will be generalised by separate processes, so will have their own generalisation behaviour.

The parts which are available for simplification are those bounded only by obstructing features (represented as thin black continuous lines) and those which are unbounded (represented as thin black pecked lines). Unbounded sections shorter than the length of a peck should be reclassified as being bounded by an obstructing feature. Likewise bounded sections shorter than the length of a dash should be reclassified as being unbounded.

The level of detail for the woodland perimeter is greater than that required for 1:50k mapping, so the boundary geometry needs to be simplified. Each *woodland boundary line* is activated as a micro-agent responsible for simplifying its underlying topological links. Generalising the topological links will ensure that the *woodland extents polygon* is updated simultaneously. Without experimenting it is difficult to know which simplification algorithm will be suitable. Douglas-Peucker is a strong candidate for removing points, while a smoothing algorithm may be needed for making portions of the boundary less jagged and angular.

6.2 Dividing woodland by traversing networks

As stated previously, the rivers, canals, classified roads and railway networks should be converted to polygons to prevent the woodland boundaries, fill and tree symbols from occupying the same portion of the map. Each *woodland extents polygon* should have a reference to the *generalised water, road and railway polygons* which it contains.

The *woodland extents polygon* meso-agent can then perform a simple intersection operation to classify its remaining internal areas as *woodland fill polygons*. The *woodland fill polygons* are the areas where it is permitted to show the green fill and the tree symbols. A reference should be set between each *woodland extents polygon* and its *woodland fill polygon*. These new polygons should be topologically structured against the *woodland extents polygons* and *woodland boundary lines*, so that any modification maintains topological consistency between the features.

6.3 Generalising sections of woodland

The *woodland fill polygons* are then activated as micro-agents which check themselves for narrow corridors, juts and recesses and resolve any such problems. The triangulation is again of use here for measuring the proximity relationships between parts of the woodland boundary and detecting where there is conflict. Triangles can be used to form bridges to fill in narrow recesses and eliminate narrow juts using techniques similar to those of the amalgamation algorithm.

At this stage the *woodland fill polygons* below the *minimum guaranteed retention size* need to be evaluated. If they do not share their *woodland extents polygon* with any other *woodland fill polygons*, they should be enlarged to the *minimum depiction size*, since they can be considered to be “prominent landmarks”. Otherwise if they are small areas forming part of a larger wood, they should be deleted and the dependant *woodland extents polygon* and *woodland boundary lines* updated to reflect the change.

Seeing the Wood from the Trees: Generalising OS MasterMap®

Tree Coverage Polygons to Woodland at 1:50 000 Scale

© Crown Copyright 2005
Remaining narrow juts and corridors should be enlarged, but if possible the shape of the boundary should be preserved. For this conflict resolution it may be possible to use an energy minimisation technique such as beams (Bader 2001) or snakes (Galanda and Weibel 2003), applying forces where the gaps are too narrow. The special case where a track runs through a narrow corridor can also be treated in this way. The algorithm should be applied to the underlying topological links so that all of the dependant features are generalised simultaneously.

6.4 Positioning of tree symbols

The final task for the woodland fill polygon micro-agents is to create suitably placed tree point objects for the symbols inside their extents. The classification (coniferous, non-coniferous, mixed) can be determined by following the database references back to the source tree coverage polygons. Point objects can be created following the original classification and using a repeating screen pattern.

It is unlikely that the initial positions chosen for the tree symbols will be optimal. Tree symbols should be suitably spaced from each other and the symbols should be unclipped by the boundary if possible. Where the symbols are clipped, as much of the symbol should be shown as possible, with the base of the symbol taking priority. This is an optimisation problem, so would be best suited to an approach based on cartographic constraints and solved using a method such as least squares adjustment (Harrie 1999; Sester 2000) or simulated annealing (Ware et al. 2003). Simulated annealing is a strong candidate since it is straightforward to implement and has a strong track record for resolving placement problems, such as the Maplex text placement software from ESRI (2005).

![Figure 12. 1:50k extract illustrating the required transformation from the result in Figure 11.](image)

The appearance of the intended result is illustrated in Figure 12 to allow comparison with the current results Figure 11. The purple vignette indicates National Trust land, which is not currently available in the source database, so hasn’t been investigated during the research.

7 Conclusion

An approach has been detailed for all of the steps required to generalise the tree coverage polygons from the source database to OS specification 1:50k. So far generic clustering and amalgamation algorithms have been developed, which will be useful for generalising other feature classes. Population of the DLM will be completed when the woodland boundary features are identified. The next stage of the research will involve transforming the DLM to the 1:50k DCM.

When this development is completed, the results will need to be evaluated. One approach would be to compare the automated result against current 1:50k raster mapping. The drawback of this is that the automated system will consistently replicate the same solution, whereas for a manually created product, the chosen solution is dependent on the art of each individual cartographer. A better approach would be to get OS cartographers to comment on whether the result is cartographically acceptable and whether it is consistent with the current 1:50k style.
The work so far has illustrated some interesting memory management issues which occur when processing large groups of geometries. The research has been very instructive on the considerations for interaction between feature classes during generalisation – still a largely unexplored topic in generalisation research. When this work is complete it will stand alongside the building and roads results to further demonstrate the potential for generalising OS base scale data to 1:50k.

8 References


DOUGLAS, D. and PEUCKER, T. (1973), Algorithms for the reduction of the number of points required to represent a digitised line or its caricature, The Canadian Cartographer, Vol. 10, No. 2, pp. 112-122.


REGNAULD, N. (2005), Spatial structures to support automatic generalisation, to be presented at the ICC Conference, La Coruña, Spain.


REVELL, P., REGNAULD, N., THOM, S. (2005), Generalising OS MasterMap topographic Buildings and ITN road centrelines to 1:50 000 scale using a spatial hierarchy of agents, triangulation and topology, to be presented at the ICC Conference, La Coruña, Spain.


Ordnance Survey, the OS Symbol, OS MasterMap and Landranger are registered trademarks of Ordnance Survey, the national mapping agency of Great Britain. This article has been prepared for information purposes only. It is not designed to constitute definitive advice on the topics covered and any reliance placed on the contents of this article is at the sole risk of the reader.

The author would like to thank the two anonymous ICA reviewers for their detailed, thorough and helpful comments on the initial version of this paper.