INSTITUTE OF TERRESTRIAL ECOLOGY

(NATURAL ENVIRONMENT RESEARCH COUNCIL)

Report to Directorate General XVI (European Commission)

DGXVI Contract

ITE Project T02086D1

GENERALISING THE LAND COVER MAP OF GREAT BRITAIN TO CORINE LAND COVER BY SEMI-AUTOMATED MEANS

Methods & Validation

D.S. Edmundson, F. F. Gerard, N. J. Brown, and R. M. Fuller

European Commission

Institute of Terrestrial Ecology
Monks Wood
Abbots Ripton
Huntingdon
Cambridgeshire PE17 2LS

4 March 1999

CONTENTS

EXECUTIVE SUMMARY

1. The European Commission (EC) programme ‘Co-ordination of Information on the Environment’ (CORINE) includes a major project to map the land cover and land use of member states.

2. The CORINE Land Cover map, generally produced in the late 1980’s - early 1990’s by visual interpretation and manual digitising, shows 44 cover types, in vector format (i.e. as digital map outlines) at 1:100 000 scale, with a minimum mappable unit of 25 ha.

3. The Land Cover map of Great Britain (LCMGB) of 1988-90 gives a raster (i.e. grid-based) map which records cover on 25 m cell size, identifying 25 cover-types, with a minimum mappable unit of 0.125 ha, showing landscape patterns at field-by field scale.

4. This projects aims to generalise the LCMGB to CORINE Land Cover format using a semi-automated generalisation procedure developed in 1994 by ITE.

5. The generalisation procedure has been made operational. It has been adapted to run on ARC/View and made more efficient.

6. Further refinement of the procedures have been implemented, including the updated generalisation criteria specified by the European Environment Agency for generalisation to CORINE Land Cover.

7. Outputs for Arable and Marginal landscapes have been evaluated by the production team, Cristina Seabra of the ETC/LC Technical Unit, and the results presented to the CLCGB Advisory Group.

8. Outputs for Pastoral and Upland landscapes will be evaluated by the production team and Cristina Seabra in the following month.

9. Operational production of CORINE Land Cover of Great Britain started in November. About half of GB, has been completed.

1 INTRODUCTION

1.1 The European Environment Agency (EEA) was launched by the European Union (EU) in 1993 with a mandate to co-ordinate and put to strategic use information of relevance to the protection and improvement of Europe's environment. The Agency carries out its tasks in co-operation with a European Information and Observation Network (EIONET). EIONET consists of national networks, organised by the Agency to help it retrieve information, and produce efficient and timely information on Europe's environment. To execute particular tasks, institutions or organisations have been contracted as European Topic Centres (ETC). There are today ETCs for Air Emissions, Air Quality, Catalogue of Data Sources, Inland Waters, Land Cover, Marine & Coastal Environment, Nature Conservation and Soil. The ETC on Land Cover (ETC/LC), led by the Environmental Satellite Data Centre (MDC) in Sweden, was established to provide accurate data on land cover in Europe, corresponding to needs across a wide range of applications.

1.2 A key activity of ETC/LC has been the completion of a European-wide inventory of land cover in 44 classes. This takes the form of a digital cartographic product, at a scale of 1:100 000. The inventory has been compiled, mostly in the late 1980's - early 1990's, using methods developed within the CORINE (Co-ordinating Information on the European Environment) experimental programme, undertaken by the Environment Directorate (DGXI) of the European Commission between 1980 and 1985. A major task of ETC/LC has been to develop and complete the Land Cover database begun within the CORINE programme. Today the CORINE land cover data base is operationally available for the greater part of the 3.5 million km² covered by the European Union and progress is being made, through the PHARE programme, in the production of maps to CORINE standards for the former Soviet Union states.

1.3 The land cover of Great Britain was mapped in detail by the Institute of Terrestrial Ecology using remotely-sensed data (Fuller *et al.* 1994). The British land cover map differs from CORINE in several respects, including its spatial resolution, the land cover classes mapped and the method of production.

1.4 The CORINE Land Cover map has generally been produced by visual interpretation of hard copy satellite images followed by manual digitising to give computer maps which show 44 cover types, as digital map outlines, in vector format at 1:100 000 scale, with minimum mappable units of 25 ha. The Land Cover Map of Great Britain (LCMGB) is a raster or grid-based product which records 25 cover-types, on 25 m grid, with minimum mappable units of 0.125 ha, showing landscape patterns at the field-by-field scale.

1.5 A pilot study successfully demonstrated semi-automated procedures to convert the LCMGB to CORINE specifications. These procedures involve generalisation from the 25 m resolution, reassignment of LCMGB classes to the CORINE categories, generation of CORINE mosaic classes from heterogeneous regions, and use of knowledge-based operations to add relevant land use information (Fuller & Brown 1994, 1996).

1.6 It was concluded that there would be significant financial benefits of automated conversion of the LCMGB to CORINE format, a process which was estimated to cost about 15% the price of new CORINE mapping of Britain. Conversion would ensure that CORINE land cover data for Britain were calibrated against the existing National map and against proven ground reference data available from Countryside Survey 1990. The approach would also ensure that information on CORINE land cover in Britain is entirely consistent with the national data-set and with the large number of uses to which these data have already been put.

1.7 This project therefore aims at converting the existing Land Cover Map of Great Britain to CORINE Land Cover format by semi-automated means. The work is jointly funded by the Directorate-General XVI for Regional Policy and Cohesion (DGXVI) of the European Commission (EC) and the UK Department of Environment and Transport (DETR), . The work is being carried out by the UK Institute of Terrestrial Ecology (ITE), .

1.8 The generalisation procedures and CORINE outputs are being evaluated by the Technical Unit of the ETC/LC represented by Cristina Seabra of CNIG, Portugal.

1.9 An Advisory Group with a flexible membership to meet specific advisory needs which prevail at any one time, oversees the production. The business of the Advisory Group is a two way process, mainly for ITE to assure deliverables are on time and of the quality required; and to advise or seek advice if and when problems arise. The projects advisory group members are:

Dr Andrew Stott	Department of the Environment, Transport and Regions
Dr Michael Alb<s	Commission of the European Communities, DGXVI
Dr Chris Steenmans	European Environment Agency
Dr Vanda Perdigvo	Commission of the European Communities, Joint Research Centre
Dr Stuart Gardiner	Scottish Office

2 METHODS

The generalisation methodology used to produce CORINE Land Cover was developed as part of a feasibility study in 1994. It is a semi-automatic procedure and comprises the following main steps:

- Removal of very small parcels;
- Extraction of 25 ha parcels with direct CORINE equivalence;
- Clustering of smaller parcels;
- Analysis and classification of mosaic parcels;
- Assignment of remaining small parcels to the most appropriate neighbouring class;
- Overlay onto the satellite images to check outputs;
- Use of ‘‘exogenous’’ data and expert interpretation to identify CORINE land uses;
- Smoothing of polygon boundaries;

The operation of these methods is documented in the ITE report to the Department of the Environment (Fuller & Brown, 1994) and further described in published papers by Fuller & Brown (1996) and Brown *et al.* (1996) with methodological refinements outlined by Gerard *et al.* (1996).

The procedure, initially designed to use ARC/Info and ARC/Grid functions, has been built into ARC/View processes as a sequence of scripts which produce consistent outputs, with the minimum of interactions (see further). However, results require inspection and fine-tuning of methods making some stages iterative. Other stages, especially the interpretation of CORINE’s land use classes (e.g. airports, recreation areas) need some interactive editing.

The above stages are applied to subsections based on 100 km square tiles of the LCMGB (with overlaps between subsections to ensure satisfactory subsequent edge matching).

3 METHODOLOGICAL IMPROVEMENTS AND ADDITIONS

The following paragraphs describe how the original generalisation procedures that were designed for use in ARC/Info have been adapted to run in ARC/View. This adaptation to ARC/View incorporated improvements in methodology as well as efficiency.

3.1 Grid or Vector

ARC/View handles Grid files in the same general manner as ARC/Grid. The original process involved generalising a given LCMGB tile using a combination of ARC/Info and ARC/Grid procedures on a Unix workstation. All processes were command-driven by typing in set functions at each stage of the conversion. The generalisation also involved five intermediate conversions from raster-to-vector and *vice versa* which were tedious, time consuming and had a weakness in the limited number of vectors that could be handled at any one time by ARC/Info. As the test area was relatively small in the first instance, it was not limited by numbers of polygons so the generalisation process was allowed to run smoothly. However, an increase in the test area to 50 km square of land, containing a greater variety of LCMGB landclasses, resulted in a total number of polygons exceeding the then 100 000 polygon limit of ARC/Info. This in turn led to a need for segmenting the land tile into a number of smaller sections, which introduced subsequent problems of edge matching.

ARC/View has much the same functionality as ARC/Grid but its object-oriented macro programming language ‘Avenue’ allows the user more control over functions (such as grid zonal functions, see below) and the linking together of various functions into sub-routines. The sub-routines can also be accessed via customised menu systems.

The initial changes made to the generalisation system were aimed at keeping the data in Grid format in order to avoid the imposed 100 000 polygon limit. This would also speed up the generalisation process by reducing the need for conversion between raster and vector formats and between systems. It would not compromise the quality of outputs or their correspondence with CLC; indeed, the refinements help to improve these.

3.2 Preliminary Thematic generalisation

The 25 LCMGB classes were simplified to use only the broad types needed to define CLC classes and mosaics.

3.3 Removal of small parcels and extraction of 25 ha parcels with direct CORINE equivalence

Instead of using ARC/Info ‘eliminate’ or ‘dissolve’ to remove polygons < 2 ha the ‘region group’ and ‘nibble’ commands were used in ARC/View to return a similar result in Grid format (Figure 1). Running this process in ARC/View rather than ARC/Info significantly reduces the time required of the user.

Growing polygons is a GIS procedure which has been developed for the CORINE ‘outliers’. By growing and shrinking regions, small clusters of regions can be merged and small isolations in grid zone boundaries can be ‘filled in’. This process, involving ‘cost distance mapping’ and ‘neighbourhood sum’ functions, is essentially unaltered except that the process is now fully automated. One click on the generalisation menu button executes the growing/shrinking procedure on all classes and creates a series of grid layers each of which contains the results of a particular class.

The merging of the grown class grids into a single dataset is time-consuming. The sequence of merging determines the priority of classes, with those merged later overwriting earlier class parcels where both have grown to bridge the same gap. The classes are thus merged in different orders of priority in the various UK regions, depending on the dominance of each class in the landcover map. The order of priority is defined by the dominance of the underlying LCMGB classes and on criteria included in the EEA generalisation manuscript. In the original process, this involved writing an ordered list of the class grids in an ARC/Grid command. In ARC/View a macro has been written that merges grids, depending on the order in which they appear in a view, thus speeding up the process and making it more flexible.

3.4 Clustering of smaller parcels and analysis and classification of mosaic parcels

The majority of changes made to the generalisation process have been in the final stage of the analysis and classification of ‘mosaic’ or composite polygons. In the original process, this was essentially an interactive process involving the user in visually determining the composition of mosaics – a laborious method suffering from the problems of human error and subjectivity.

Note:

The function **‘REGION GROUP’** groups all adjacent orthogonal pixels of the same class into regions (GRID B). Conditional rules are then applied on the regions based on their size (i.e. number of pixels). If regions are less than a given size (i.e. in this case 2 hectares) all pixels in the region are given a null value (corresponding to 0). If regions are greater than a given area they retain the class of the original grid (i.e. GRID A). The result is a mask grid (GRID C) which can be used in the nibble process.

The function **‘NIBBLE’** replaces the pixels of a Grid (GRID A) that were given null values in the Mask Grid (GRID C) with the values of their nearest neighbours.

These procedures are similar to the ones described in detail in Brown *et al.* (1996) for ARC/Grid.

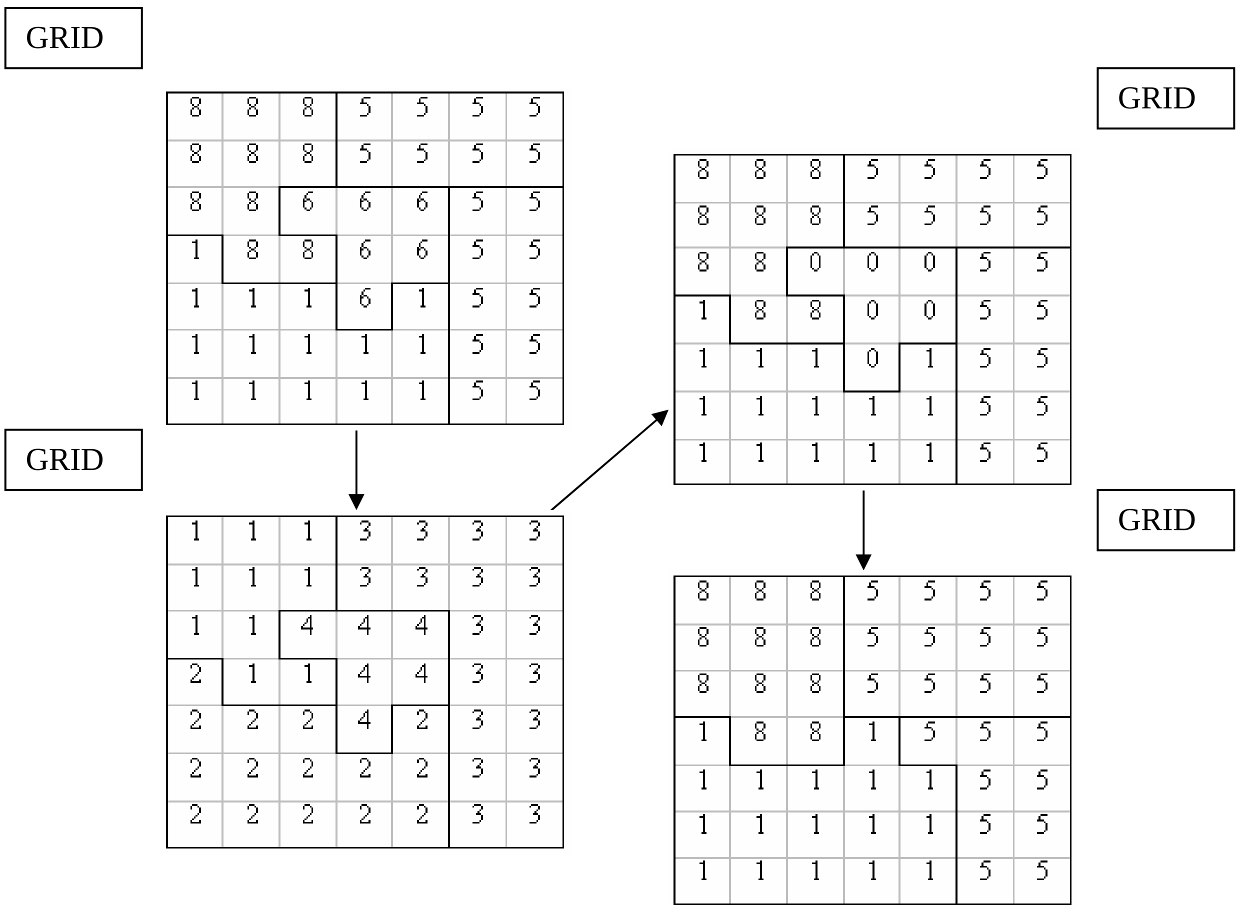

Figure 1. The region group and nibble process.

Using zonal geometry and zonal statistics functions in ARC/View it is possible to calculate the exact composition of mosaics in terms of % class composition. The results are stored in a Table and queried using the given CORINE rules to determine the correct landcode and thus Corine code. The whole process of mosaics has also been automated. When heterogeneous parcels are being classified into CORINE mosaic classes, criteria are incorporated which include information on parcel sizes and fragmentation, using the rules defined by the EEA manuscript. A schematic overview of the procedure is shown in Figure 2.

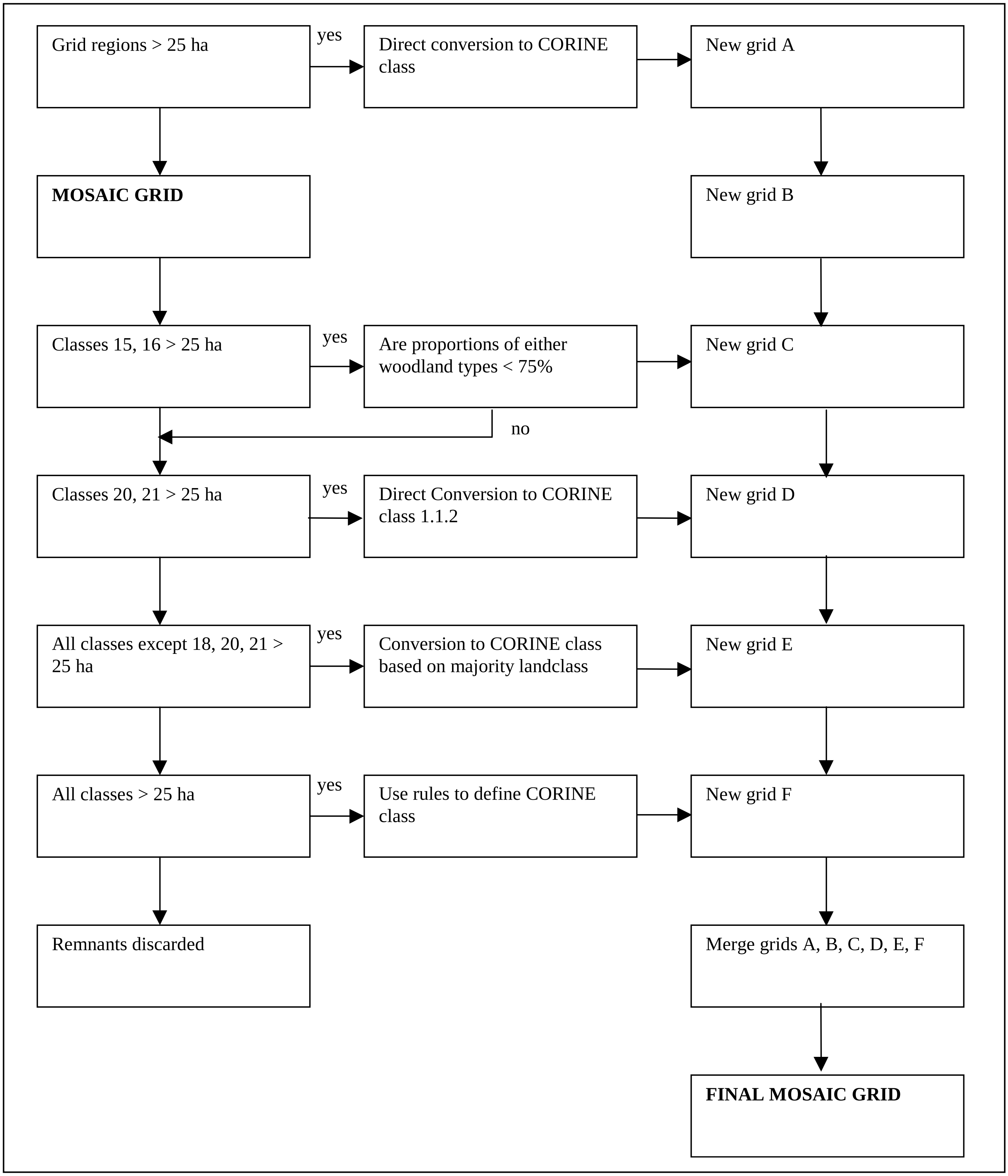

Figure 2. Schematic overview of the mosaic process.

3.5 Use of ‘exogenous’ data and expert interpretation to identify CORINE land uses

Certain CORINE classes cannot be defined directly from computer algorithms. Features such as Port Areas, Airports, Sport and Leisure facilities, Dump Sites and Mineral Extraction Sites require exogenous data. The key sources which are used to help digitise the land use classes are the Batholemews point data set, maps from Ordenance Survey, a map showing mineral extraction sites from the British Geographical Society and a map showing registered dumpsites from the Environment Agency. These maps help locate the relevant land use classes on the Satellite imagery and LCMGB which are then manually digitised on top of the LCMGB. The digitised polygons are introduced at the beginning of the generalisation process (i.e. prior to removing polygons < 2 ha) and are given priority throughout the generalisation process. The CORINE classes which are being included through manual digitising are listed in Table 1.

Table 1. CORINE level 3 classes which will be included through manual digitising

manual interpretation/digitising prior to automated generalisation	manual revision/digitising after automated generalisation
<u>Land use classes (1.*.*)</u> 1.2.1 Industrial or commercial units 1.2.2 Road and rail network and associated land 1.2.3 Port areas 1.2.4 Airports 1.3.1 Mineral extraction sites 1.3.2 Dump sites 1.3.3. Construction sites (where identify-able) 1.4.1 Green urban areas 1.4.2 Sport and leisure facilities <u>Agricultural areas (2.*.*)</u> 2.2.2 Fruit trees and berry plantations (probably only large orchards) <u>Wetlands (4.*.*)</u> 4.1.1 Inland marshes	<u>Woodlands (3.1.*)</u> <u>Wetlands (4.*.*)</u> 4.2.3. Intertidal flats <u>Water bodies (5.*.*)</u> 5.1.1 Water courses 5.1.2 Water bodies

3.6 Assignment of remaining small parcels to the most appropriate neighbouring class

The small parcels, that remain after the classification of mosaic parcels, were originally dissolved, ‘nibbled’ into the adjacent parcels by using a combination of ‘nearest-neighbour’ and ‘cost-distance’ rules. The ‘nearest-neighbour’ rule tends to split the parcel up into sections which are dissolved into the adjacent parcels, creating new boundaries. In cases where a small parcel has more thematic affinity with one of its neighbouring parcels, this approach is not appropriate. The procedure was altered to include a form of thematic generalisation. The removal of small parcels is first based on their thematic content. A small parcel is joined with the neighbouring parcel that has the highest thematic similarity or affinity at CORINE level one (see Figure 3). If the small parcel has no thematic affinity with any of its neighbours or is an island surrounded by a single cover type, 'nibble' is applied.

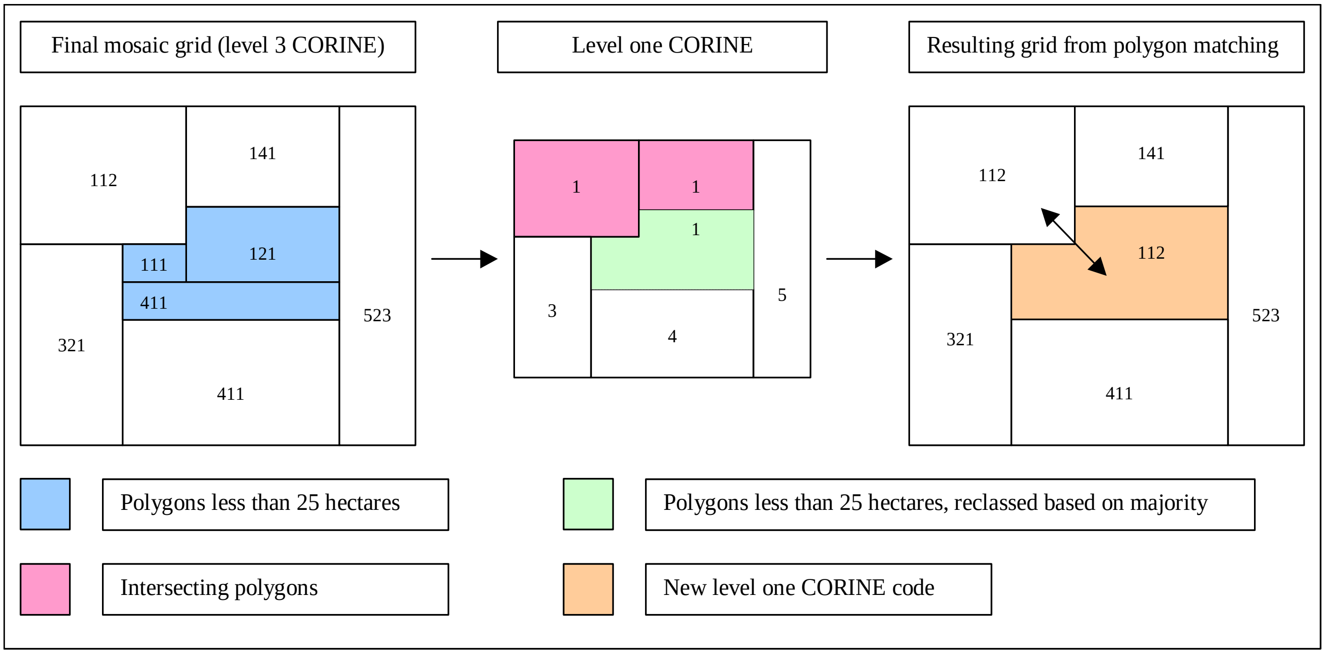

Figure 3. Schematic view of polygon matching algorithm

4 EVALUATION AND CALIBRATION

The procedure developed to validate the CORINE Land Cover map is described in the CORINE technical guide. The method uses extensive field surveying and aerial photography to checking the accuracy of the CORINE Land Cover map product. The Land Cover Map of Great Britain is based on satellite imagery dated between 1989 and 1990. There is no substantial set of aerial photography available for that period and the size (1 km^2) of the 508 Countryside Survey field samples collected in 1990 is inadequate for a validation exercise as described in the technical guide. Since the accuracy of the Land Cover Map of Great Britain is known to be between 80% to 85% (Fuller *et al.* 1998) the way forward is to first minimise the difference between semi-automatic and manual generalisation outputs and then to evaluate the impact of the generalisation process itself.

To assess the differences between automated and manual outputs, areas were plotted at 1:1000 000 scale onto A3 size paper giving study areas of 43 km x 29 km, covering the four main landscapes in GB (i.e. Arable, Pastural, Upland and Marginal landscapes) were selected. Together, the validation areas contain a variety of CORINE land-use and land-cover classes. Table 2 lists the percentage of land covered by each of the four landscape types in GB.

Table 2. Percentage of land covered by each of the four landscape types in GB

Landscape type	% coverage
Arable land	34
Pastoral land	29
Upland	21
Marginal land	16

Once complete, the CORINE map of GB will be evaluated by:

(i) comparing the class statistics of the LCMGB with these of the CORINE Land Cover map,

(ii) evaluating the changes in class statistics caused from one generalisation step to the other,

(iii) following the generalisation procedure in a set of Countryside Survey 1 km squares and assessing the impact of the overall process and the individual generalisation steps.

4.1 Assessing the differences between automated and manual output

Interpretation of CORINE classes was based on the same Landsat TM images which were used in the original per-pixel classifications. The CORINE manual interpretation was compared with the CORINE output based on the improved semi-automated procedure (see above). Comparisons were carried out, both visually (i.e. qualitatively) and through correspondence tables (i.e. quantitatively).

To date, the validation has been completed for two areas: an area of Arable landscape in Cambridgeshire and of Marginal/Upland landscape in West-Scotland; also for half of a third area of predominantly Marginal landscape in Wales. Two additional areas have been selected, one in Devon and one in North-east Scotland to represent predominantly Pastural and Upland landscapes respectively. The validation for the Marginal land in Wales, Pastural land in Devon, and the Upland in Scotland will be completed in March.

The effect of smoothing, a process which is carried out during the final stage of conversion, has also been assessed by comparing the CORINE versions of the section representing Arable land before and after smoothing.

4.1.1 Arable landscape (Cambridgeshire)

The visual quality of the automated output (Figure 4) is excellent, closely matching the manual version (Figure 5). It maintains the more complex outline of the original input, but this detail is not excessive and does not merit further generalisation. Indeed, it could be argued that the manual output oversimplifies outlines. There is no need to remove detail, just for the sake of it.

The correspondence matrix appears in Table 3. It shows that overall the correspondence is 875 per thousand pixels, a good result unlikely to be exceeded. This is because, in recording this correspondence, it should be recognised that deficiencies were noted in both the manual and automated product. The manual output includes a small number of polygons that fall below the 25 ha minimum. They represent 0.8% of the total area mapped. Also, some polygons marginally larger than 25 ha were accidentally omitted. The manual recording completely overlooked the presence of *1.4.1 Green urban areas* despite their being commonplace in Cambridge and Huntingdon (1.9% of the map area according to automated results); manual recording also failed to identify the mosaic habitats: *2.4.2 Complex cultivation patterns*, *2.4.3 Agriculture with .... natural areas* and *3.1.3 Mixed forest* which, with the benefit of hindsight, should have been considered (and which totalled 11.6 λ according to automated methods). The exact position of boundaries is inevitably questionable and the generalisation was of necessity subjective in operation. The manual procedure also involves several stages: hard copy output with potential distortion of paper prints, drawing with likely discrepancies in outlines, and digitising, all of which added to those discrepancies. Registration and, rubber-sheeting, carried out to achieve full registration to UTM and accommodate geometric discrepancies in traced film outlines did eliminate some misplacements but definitely not all non-systematic errors introduced during the manual procedure (Figures 6 and 7). This is not to argue that manual procedures were less accurate than automated one: simply to point out that this manual exercise was subject to all the same potential errors which are recorded in other countries' manual CORINE mapping. If manual/visual interpretation achieved 95% accuracy, then a correspondence of 87.5% would imply a possible accuracy of 92% on the part of the automated procedure.

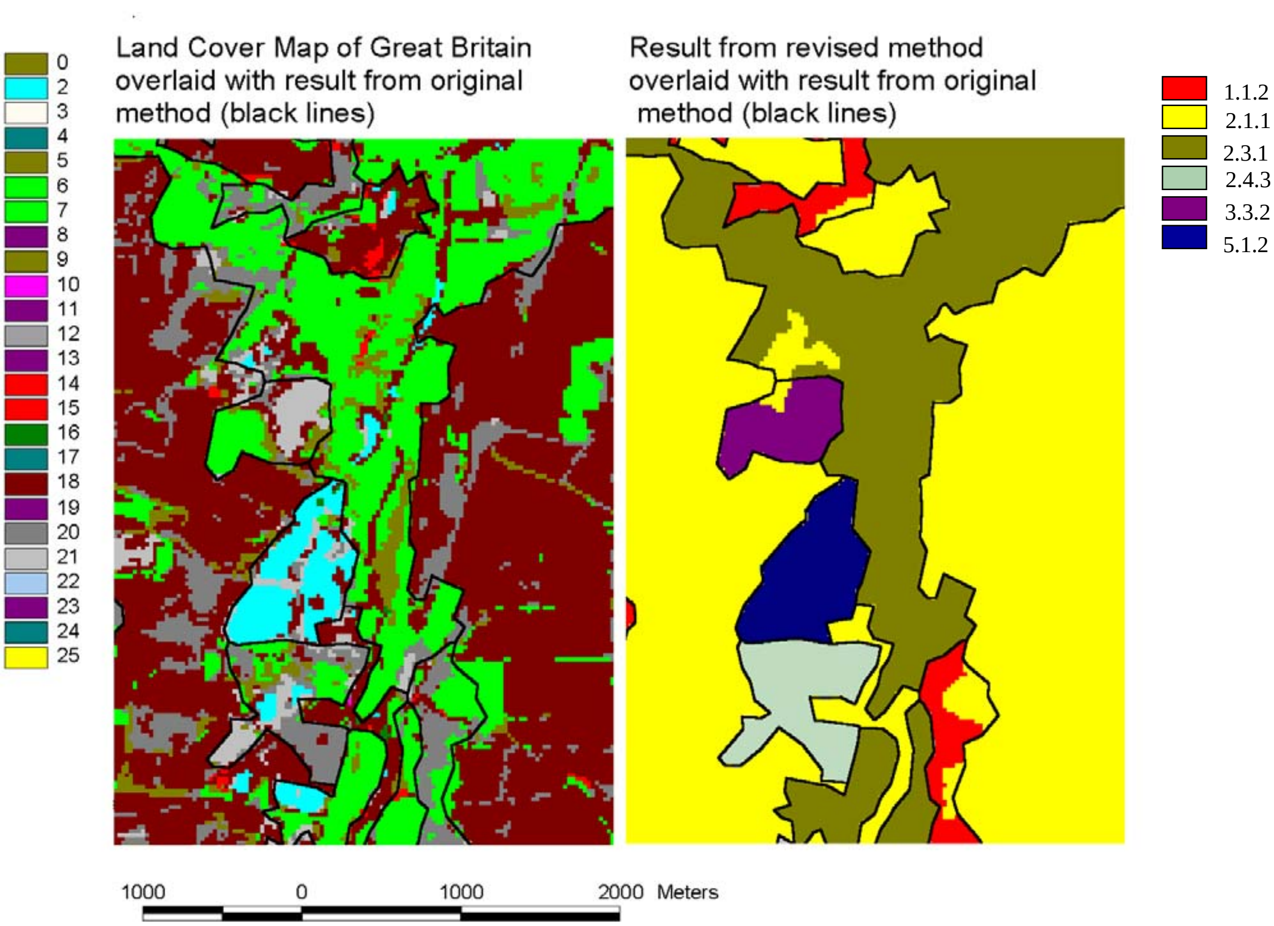

Figure 8. Comparing the LCMGB and the results from a. the original semi-automated generalisation method (black vector lines) and b. the improved method with thematic as well as spatial generalisation (colour plot).

The indicated level of accuracy exceeds normal expectations and gives considerable faith in the automated procedures. A closer look at class-level correspondence highlights the specific differences. However, it must be recognised that rarer classes (at least those which are rare in this study area) cannot be adequately assessed, as minor boundary differences for few smaller parcels may cause major quantitative differences; also the chance inclusion/exclusion of just a single polygon can cause a large percentage difference. In the following listings, the percentage values are, first, the proportion of manually interpreted cover which is recorded, second, the proportion of automated cover which matches the manual result:

1.1.1 *Continuous Urban* (71% / 61%) and 1.1.2 *Discontinuous Urban* (54% / 71%)

- Manual interpretation was particularly difficult and is unlikely to have achieved more than 80-85% success.
- With the dissected patterns of small villages in the study area, minor geometric displacements between manual and automated outputs could contribute substantial differences.
- The results suggest that overall correspondence is compromised by these factors: overlap of manual and automated classifications, both with 85% accuracy, might achieve only 72% correspondence, close to some of the values obtained.
- The manual method has been more generous in its inclusion of pasture and amenity grassland around villages as part of the *Discontinuous Urban*.
- More pixels have been assigned by the automated method to *Continuous Urban* and less to *Discontinuous Urban*.
- The difference in overall cover is small with manual recording giving 5.3% D*iscontinuous Urban* as opposed to the automated method's 4.1%; Figures for C*ontinuous Urban* are 0.39% and 0.45% respectively.

1.2.1 *Industrial* (100% / 99%), 1.2.4 *Airports* (89% / 86%), 1.3.1 *Mineral* 93% / 94%) & 1.3.3 *Construction* (87% / 85%), 1.4.1 *Green urban areas* (0% / 0%) and 1.4.2 *Sports and leisure* (74% / 58%)

- These land use classes were all captured manually, generally from the screen image but, where necessary, with revision from the hard copy.
- The two methods give very similar results except that the manual procedure overlooked the need to record 1.4.1. and scored a lower land use for *Sports and leisure* (4.2% cover v. 5.5% by automated interpretation).

2.1.1 *Arable* (92% / 96%)

- The revised semi-automated method has reduced confusion between 2.1.1 and 1.1.1. Overall correspondence was very high.

2.3.1 *Pasture* (79% / 63%)

- The manual procedure generalised more of the pasture out of the classification, increasing the arable coverage. It is not clear that the manual method was necessarily the most correct.

3.1.1 *Broad-leaved* (51% / 80%), 3.1.2 *Coniferous* (41% / 32%) and 3.1.3 *Mixed forest* (0% / 0%)

- The dissected patterns of *Broad-leaved* and *Coniferous* forests have led to substantial differences.

- In the revised automated method, the areas of *Mixed forests* are being identified (0.06 %). No *Mixed forest* areas were identified by the manual interpretation: an oversight.

3.3.2 *Moor/heath*

- No *Moor/heath* was recorded by manual interpretation. A very small coverage 0.08% was derived from automated conversion, which is in fact correct.

3.2.1 *Natural grasslands* (71% / 34%)

- The manual interpretation missed a significant area of long-term set-aside causing confusion in 2.1.1.

5.1.2 *Water bodies*(70% / 88%)

- The manual method recorded more Water bodies by aggregating smaller lakes <25 ha into larger ones >25 ha. This only raises total cover from 0.9% by automated conversion to 1.1% by manual interpretation.

The effect of smoothing was found to be minimal. Comparison between the CORINE output before smoothing and after smoothing gave an overall correspondence of 99%.

4.1.2 Marginal land (West-Scotland)

Unlike the Arable landscape in Cambridgeshire, the Marginal landscape of western Scotland contained few CORINE 1.*.* classes that required manual digitising. For correct identification of the CORINE classes 3.1.1 (*Coniferous woodland*) and 3.1.2 (*Broadleaved woodland*), manual digitising was necessary to ensure the inclusion of 'recently logged areas', which on the LCMGB are identified as bare. The Scottish test area contains a much larger proportion of semi-natural cover classes, such as *natural grasslands* (3.2.1), *moors and heaths* (3.2.2), *transitional wood* (3.2.4) and *peat bogs* (4.1.2). Another difference between the two areas is the large number of small and large lakes present in the Scottish test site.

When assessed visually, the CORINE maps from semi-automated and manual product show a good match across the area (Figure 9). The percentage correspondence achieved for the Marginal land area in western Scotland is 836 per thousand (Table 4). From Table 4, it is clear that the main differences between the manual and semi-automated methods are caused by the mismatches between boundaries of *natural grasslands* and *moors and heath*. The mismatch caused a loss of circa 10% correspondence. For interpreters, the identification of boundaries between natural vegetation types is a major challenge as the transition between natural vegetation types is usually gradual, resulting in fuzzy boundaries. The boundary identified by the semi-automated method is based on a set of rules which interrogate the underlying land cover class of the LCMGB and the cover composition of each polygon > 25 ha. Although, in this case, neither of the two methods consistently produces the 'correct' answer, the semi-automated approach has the advantage of being objective in the way it identifies boundaries between natural vegetation types. The many lochs in the landscape were successfully generalised by the semi-automated procedure.

4.1.3 Marginal Landscape (Wales)

As the validation for the Marginal landscape area in Wales is only partially completed, it is impossible to draw final conclusions from the results. However, the area for which the validation is complete can give a preliminary indication of the results. The selected landscape is very different from the test areas in Cambridgeshire and western Scotland. The Welsh area has steep and narrow built-up valleys with mineral extraction sites (coal mines) and slag heaps, some of which are still in operation, others of which are abandoned. The hill tops and slopes show an intricate pattern of heather moor, semi-natural grasslands and improved pastures with pockets of woodland and smaller copses (very small woodlands).

The general patterns of CORINE classes have been picked up by both methods (Figure 10). Visual examination of the resulting maps reveals one main difference however: the manually digitised CORINE classes, especially the *mineral extraction sites* (1.3.1), vary from the outlines of the semi-automated method. The manual digitising of these land use classes was carried out independently (i.e. by different people) for each of the methods. The difference results from the fact that neither interpreter is familiar with the area, that the spectral signature alone is inadequate for interpretation, that there is no external data which consistently identify the location and extent of these land use features (which anyway are dynamic, soon rendering OS map data out of date). The pattern of *mineral extraction sites* represents two interpreters best efforts at delineation and differences are probably attributable to equal levels of mis-interpretation on the manual stage of semi-automated mapping and on the independent manual output used to compare with that.

The correspondence calculated for the area of Wales for which the validation is completed is 650 per thousand. The complicated nature of the landscape with the narrow valleys and the intricate mixture of natural vegetation types are the main factors for the low correspondence value. In this area, many polygons are thin slivers so that a minor spatial shift between the two maps (manual and semi-automated) would result in low correspondence for these polygons. Moreover, many polygons were near the 25 ha size limit and were wrongly excluded or included by the manual interpreter. Also, in natural areas, like for the West-Scotland area, the identification of boundaries between natural classes was very subjective.

4.2 Conclusion

The LCMGB is estimated to be 80%-85% accurate. Supposing (for the moment) that CORINE from manual interpretation were to have had an accuracy of say 90%-100%, and that the semi-automated generalisation procedure perfectly mimics the manual generalisation process, we would expect an overall percentage correspondence to be 72%-85%. The achievement of 84% correspondence for Marginal land of West-Scotland is as expected and the 87.5% correspondence for Arable land is better than expected. The latter results suggests that generalisation to CORINE format is removing erroneous 'noise' in the per-pixel classification. In the Welsh site, manual interpretation was especially difficult with huge uncertainties as to the interpretation of features, the level of generalisation required, the best aggregations of smaller features and the exact CORINE class to use. It is probable that correspondence of 650 pixels per thousand results from two independent surveys each being around 80-85% correct. It is clear that scarce cover types, especially those which are also dissected into many small units at, near, or below the 25 ha minimum mappable unit are highly variable in their interpretation and automated conversion. A more complete analysis will ensue, once generalisation is completed for the whole study area.

It is not possible to say at this stage of the validation exercise what level of consistency will be achieved throughout Britain. However, it seems likely that the dominant cover types will be automatically mapped with similarly high levels of accuracy but that scarce habitats within a region will continue to cause more difficulties. The average of the correspondence results achieved for the Arable and Marginal landscape types, weighted by the % land in GB that is covered by the types is:

$$\frac{0.34 * 87.5 + 0.16 * 74.5}{0.5} = 83.34$$

where:

74.5% is the average of the % correspondence achieved for two Marginal land test sites 65% (Wales site) and 84% (West-Scotland site).

87.5% is the % correspondence achieved for the Arable land test site.

0.34 and 0.16 is the proportion of land in GB covered by Arable landscapes and Marginal landscapes respectively.

The overall level of correspondence at 83 % plus a very similar appearance of both manual & automated outputs, give faith in the procedure. Later tests in other terrain, pastoral and upland, will hopefully confirm that this level of accuracy is maintained and probably enhanced throughout Britain.

5 TIMETABLE

The project is on schedule (Figure 11). The generalisation method has been adapted for use in ARC/View while at the same time made operational and efficient.

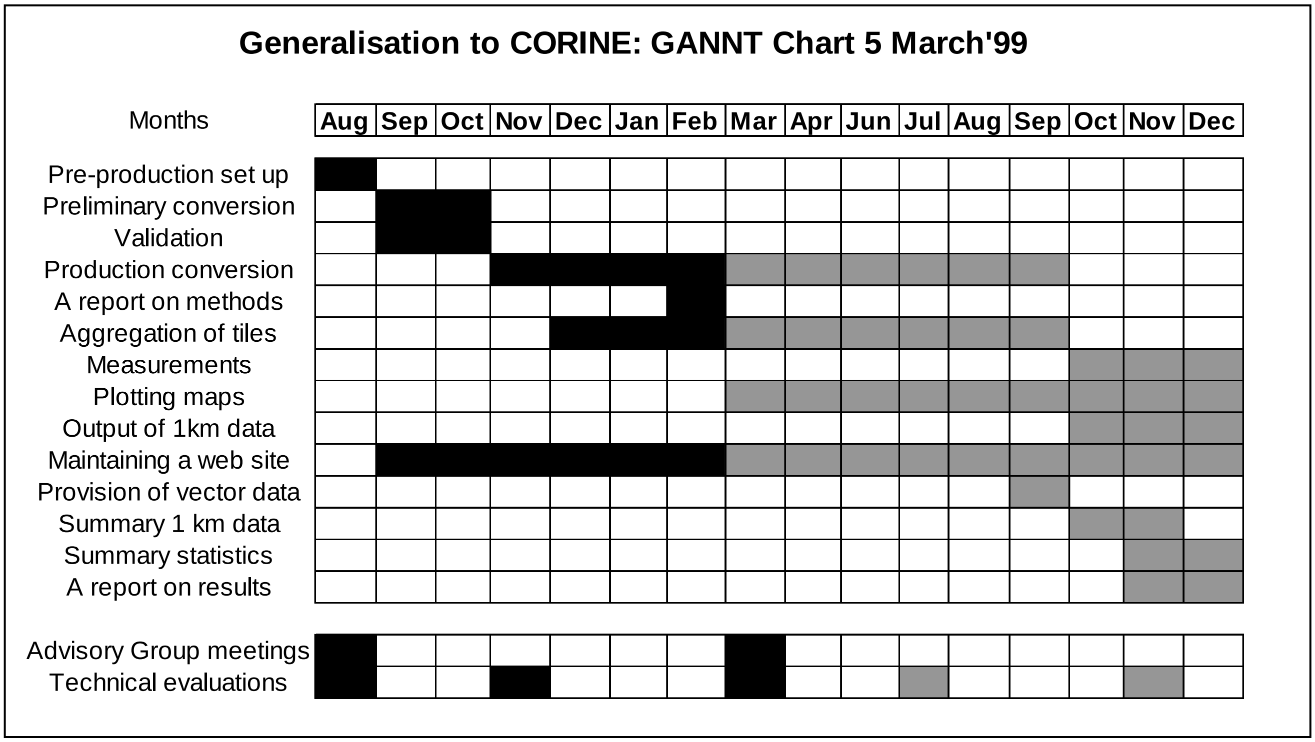

Figure 11. Project GANNT chart showing progress (in black) marked against the time schedules of the original proposal

6 REFERENCES

Fuller R M & Brown. N J. 1994. *A CORINE map of Great Britain by automated means: a feasibility study*. Unpublished Institute of Terrestrial Ecology report to the Department of the Environment.

Fuller, R.M., Groom, G.B. & Jones, A.R. 1994. The Land Cover Map of Great Britain: an automated classification of Landsat Thematic Mapper data. *Photogrammetric Engineering & Remote Sensing.* **60**, 553-562.

Fuller, R.M., Wyatt, B.K. & Barr, C.J. 1998. Countryside Survey from ground and space: different perspectives, complementary results. *Journal of Environmental Management,* **54**, 101-126.

Fuller, R. M. & Brown, N. 1996. A CORINE map of Great Britain by automated means. Techniques for automatic generalization of the Land cover map of Great Britain. *International Journal of Geographical Information Systems,* **8**, 937-953.

Gerard, F., Fuller, R.M., Brown, N.J., Jaakola, O., Ostling, M., Rosengren, M., Arozarena, A. & Bayarri, E. 1996. *Work Programme 6. Research and Development. Task 6.3. Spatial Generalisation*. Final Report to the Joint Research Centre of the Commission of European Communities. Institute of Terrestrial Ecology.

Table 3. Correspondence per 1000 pixels between manual and semi-automated method for the Arable land test site in Cambridgeshire

REVISED AUTOMATED / Corine Value	MANUAL: 111 - cont urb	112 - discont	121 - indust.	124 - airport	131 - mineral	133 - constr	141 - green urb	142 - recreat.	211 - arable	231 - pasture	242 - cultivate	243 - agric/nat	311 - broadlf	312 - conifer	313 - Mixed forest	321 - nat grass	324 - scrub	332 - moor/hth	512 - wtrbody	cover / 1000	% auto correct
111	**2.8**	1.4							0.0	0.3										4.5	**61**
112	0.9	**28.9**		0.2	0.0	0.0		0.0	9.2	1.6			0.2						0.0	41.0	**71**
121		0.0	**1.5**						0.0	0.0										1.5	**99**
124		0.1		**9.0**					1.2	0.2										10.5	**86**
131	0.0	0.0			**4.9**				0.2	0.1									0.1	5.2	**94**
133		0.0				**1.3**			0.2	0.0										1.5	**85**
141		0.0							0.3	1.5										1.9	
142		0.7						**3.2**	1.0	0.7										5.5	**58**
211		9.2		0.3	0.0	0.0		0.3	**717.0**	17.3			3.3	0.2		0.3			1.6	749.5	**96**
231	0.2	11.5		0.0	0.2	0.2		0.8	39.2	**91.1**			0.6			0.0			1.1	144.8	**63**
242	0.1	1.0		0.2	0.1				7.2	1.2			0.1	0.0		0.2			0.4	10.5	
243									0.3	0.1			0.0						0.2	0.5	
311		0.1							1.1	0.0			**5.5**	0.2						6.9	**80**
312									0.3	0.1			0.4	**0.4**						1.2	**32**
313									0.1				0.3	0.2						0.6	
321		0.2							2.4	0.1						**1.4**			0.0	4.1	**34**
324									0.1	0.0			0.4							0.5	
332				0.3					0.5											0.8	
512		0.1			0.0				0.6	0.4			0.0						**7.8**	8.9	**88**
Total	3.9	53.3	1.5	10.1	5.2	1.5		4.3	780.9	114.6			10.7	1.0		2.0			11.1	1000	
% manual correct	**71**	**54**	**100**	**89**	**93**	**87**		**74**	**92**	**79**			**51**	**41**		**71**			**70**		

Total pixels interpreted = 1224000000.0 **Total matching = 875 pixels / 1000**

Table 4. Correspondence per 1000 pixels between manual and semi-automated method for the Marginal land test site in West-Scotland.

MANUAL

AUTOMATED

Corine Value	112 - cont urb	142 - recreat.	231 - pasture	311 - broadlf	312 - conifer	321 - nat grass	322 - moors / heath	324 - trans wood	331 - beaches	332 - bare rock	412 - peat bogs	512 - wtrbody	523 - sea	cover / 1000	% auto correct
112 - cont urb	**0.9**	0.0	0.3	0.0	0.0	0.4	0.0	0.0	0.1	0.0	0.0	0.0	0.0	1.8	**52**
142 - recreat.	0.0	**0.0**	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
231 - pasture	0.1	0.2	**33.6**	0.0	0.0	7.6	3.7	0.0	0.0	0.0	0.1	0.1	0.6	46.0	**73**
311 - broadlf	0.0	0.0	0.1	**0.0**	0.0	0.1	0.4	0.0	0.0	0.0	0.3	0.0	0.0	0.9	**0**
312 - conifer	0.0	0.1	1.7	0.0	**223.7**	5.2	4.4	0.0	0.0	0.0	0.0	0.5	0.3	236.0	**95**
321 - nat grass	0.1	0.0	6.0	0.0	0.0	**168.9**	44.5	0.0	0.1	0.0	0.0	0.3	0.7	220.7	**77**
322 - moors / heath	0.1	0.0	9.3	0.0	0.0	57.2	**248.7**	0.0	1.2	0.0	1.5	0.8	0.9	319.8	**78**
324 - trans wood	0.0	0.0	0.1	0.0	0.0	0.2	2.5	**29.1**	0.0	0.0	0.0	0.0	0.0	31.9	**91**
331 - beaches	0.5	0.0	0.0	0.0	0.0	0.1	0.0	0.0	**1.3**	0.0	0.0	0.0	0.3	2.2	**58**
332 - bare rock	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	**0.0**	0.0	0.0	0.0	0.3	**0**
412 - peat bogs	0.0	0.0	0.0	0.0	0.0	1.2	1.3	0.0	0.0	0.0	**0.0**	0.0	0.0	2.5	**1**
512 - wtrbody	0.0	0.0	0.8	0.0	0.0	1.3	1.0	0.0	0.0	0.0	0.0	**29.3**	0.0	32.4	**91**
523 - sea	0.1	0.0	0.4	0.0	0.0	3.0	1.1	0.0	0.5	0.0	0.0	0.5	**99.9**	105.5	**95**
cover / 1000	1.9	0.2	52.3	0.0	223.7	245.6	307.6	29.1	3.3	0.0	2.0	31.6	102.7	1000.0	
% manual correct	**50**	**0**	**64**		**100**	**69**	**81**	**100**	**39**		**1**		**97**		

Total pixels interpreted = 1200884375

Total matching = 836 pixels / 1000