INSTITUTE OF TERRESTRIAL ECOLOGY
(NATURAL ENVIRONMENT RESEARCH COUNCIL)

Project. T02083j5

Countryside Survey 2000
Module 7

LAND COVER MAP 2000

THIRD INTERIM REPORT

incorporating the

Sixth Quarterly Progress Report

CSLCM/Int3/Prog6

R.M. Fuller, G.M. Smith, J.M. Sanderson,
R.A. Hill, A.G. Thomson & R.T. Clarke
Institute of Terrestrial Ecology
Monks Wood
Abbots Ripton
Huntingdon
Cambs PE17 2LS

Corresponding author:
R.M. Fuller
Section for Earth Observation
Institute of Terrestrial Ecology
Monks Wood
Abbots Ripton
Huntingdon
Cambs PE17 2LS

Telephone: 01487 773 381
Fax: 01487 773 467
Email: R.Fuller@ITE.AC.UK

31 August 1999

MAIN REPORT

SIXTH QUARTERLY PROGRESS REPORT
AND EXECUTIVE SUMMARY TO THE THIRD INTERIM REPORT

- This is the Sixth Quarterly Progress Report and forms the Executive Summary to the Third Interim Report on Land Cover Map 2000 (LCM2000), a part of Countryside Survey 2000 (CS2000). The Report covers work done to 31 August 1999.
- LCM2000 is making a census survey of the land cover / widespread Broad Habitats of the United Kingdom using satellite imagery and automated image processing techniques to achieve a classification accuracy of 90% for target classes.
- The early stages of LCM2000 involved procedural developments which are now in operational use.
- Pre-processing stages include atmospheric corrections (new ITE software) and terrain-induced illumination corrections (Cambridge University subcontract) and geo-registration.
- The project is using the Classification of Environment with Vector- and Raster-Mapping (CLEVER-Mapping) - an object-orientated, per-parcel approach to classification.
- Laser-Scan's Unix version of the prototype segmentation software, designed by Cambridge University colleagues, is in routine use.
- Segmentation consists of two separate stages: edge-detection to identify boundary features; and region growing from seed points.
- Experiments have examined: band selection for edge-detection and segmentation; choice of edge-detector; threshold values to identify edges; thresholds to control the degree of segmentation; post-segmentation boundary rejection and generalisation.
- Field reconnaissance data, collected in 1999, recording so far 83 thematic subclasses, bring overall coverage to *c.* 85% of the UK (Figure 1).
- The data are being used to train the classifier, with field-mapped land parcels being identified on segmented images as training polygons. The training is objectively based on the segments, identified individually by pointing to them, without the need to painstakingly draw the outline. The process of training is probably 20 times quicker than per-pixel training.
- Software developed by ITE is being used to review training areas, displaying image 'chips' to show the quality of the remotely sensed data in each training area. The operator compares the spectral signatures of training areas, defines spectral subclasses, rejects odd examples, selects and flags training polygons. The review facility saves much time in classification.
- The parcel structure of the database is being exploited to 'roll over' the classification results of one scene onto an overlapping unclassified neighbour, by identifying and labelling equivalent polygons on the new image. The training data review procedure ensures that the training set matches standards which an operator would apply.
- Classification is using the training polygons to interrogate the parent image to derive statistical measures for reflectances in each chosen band and for each spectral subclass. CLEVER-Mapping uses a shrinking procedure when extracting raster data for polygons, to avoid edge pixels and ensure the use of 'pure' core pixels of a cover type. The shrinkage is a dynamic process, minimising edge effects while ensuring an adequate sample in smaller polygons.
- The classification procedure applies a maximum likelihood algorithm using polygon mean statistics to select the most likely class in statistical terms. CLEVER-Mapping in IGIS also records the probabilities for the top five subclass options.

- A combination of external data, plus the class probabilities, is being used in knowledge-based correction, to identify land parcels classified with low confidence and / or those with classes out of their natural context. The additional use of class-probabilities makes it a particularly subtle yet powerful tool for post-classification correction.
- Per-pixel classifications record the natural heterogeneity associated with polygons.
- Comparative studies will use field survey data to validate and calibrate results. The former will assess how far extrapolation attaches the correct labels to areas of known cover. The latter will offer a method for 'translation' between detailed field classes and the more generalised target classes of LCM2000. Procedures have been proposed, but initial trials await delivery of digitised field data on Broad Habitat classes.
- Production has involved processing 11 scenes in parallel from initial atmospheric corrections through geo-registration, illumination correction, co-registration of summer-winter data, segmentation, classification and knowledge-based correction stages (Figure 1).
- Final classifications are available for 4 scenes with a 5th undergoing knowledge-based correction; preliminary classifications are available for a further 2 scenes, giving near-complete coverage of SE England, with much of the remainder of England and parts of Wales progressing through pre-processing stages (Figure 1).
- Two samples of a classified images, both covering an area about 20 km wide, appear in Figure 2. They clearly show the very clean and accurate classification which results from the per-parcel approach. The results are typical of those elsewhere.
- Initial classifications of widespread Broad Habitats have shown that, in principle, the Broad Habitat classification can be achieved very successfully, with further sub-division of Broad Habitat classes to meet wider user objectives. Checks show an accuracy in excess of 90%, assessed relative to training polygons.
- Overall success will only become clearly apparent once validation, based on field survey squares, starts to provide an objective and independent measure, starting in mid-autumn.
- The GANNT (Figure 3) records progress in scene-pair equivalents. The assessment has taken account of the development work on pre-processing and its contribution to overall progress.
- Methods developments went further than envisaged, are now complete, and all necessary tools and procedures have been fully tested in operation. Field reconnaissance is ahead of schedule. Image searches/acquisitions are half-completed and will quickly be completed once summer 1999 images reach Eurimage archive. Classifications are lagging behind the original schedule due to the large amount of development needed to bring them to operation. They are now progressing at a rate commensurate with completion in November 2000.

- Figure1. Map of coverage
- Fig 2a & b: classifications
- Fig 3. Gantt

Throw this page away. Figs become 7/8, 9/10 & 11/12 (one-sided) - numbered in roman numerals; and main report starts on 13 but starts ‘1’ in normal numbers.

MAIN REPORT

1. INTRODUCTION

1.1 Context

Countryside Survey 2000 (CS2000) is a major national audit of the land types, habitats, landscape features and plants of the British countryside at the end of the Millennium. Land Cover Map 2000 (LCM2000) is part of CS2000.

LCM2000 will provide a census of the countryside of the United Kingdom. LCM2000 outputs will be in the form of digital maps and databases, plus a range of derived products held in a geographical information system (GIS). The Consortium of Departments and Agencies who are funding LCM2000 is listed in Table 1.

Table 1. The Consortium of funding agencies for Land Cover Map 2000.

Countryside Council for Wales
Department of the Environment, Transport and the Regions
Environment Agency
Ministry of Agriculture, Fisheries and Foods
Natural Environment Research Council
Scottish Natural Heritage
Scottish Office
Welsh Office
Environment and Heritage Service (Northern Ireland)
Department of Agriculture for Northern Ireland
Ordnance Survey of Northern Ireland

1.2 Background

LCM2000 updates and upgrades the *Land Cover Map of Great Britain* (LCMGB), made in 1990-92 (Fuller *et al*., 1994). Refinements include:

- Improved accuracy of classification;
- Added thematic detail;
- Compatibility with other systems of environmental survey and evaluation;
- Closer integration between field and satellite data.

With these improvements in mind, ITE undertook a range of methodological developments, most important of which was CLEVER-Mapping - the Classification of Environment with Vector- and Raster-Mapping (Smith *et al*., 1998). This is a per-parcel procedure for classification, currently being used in the production of LCM2000.

1.3 Aims

The aims for LCM2000 are:

- To undertake a census survey of the land cover / widespread broad habitats of the UK at the turn of the Millennium;
- To apply the most appropriate satellite imagery and automated image processing techniques to achieve a classification accuracy of 90% for target classes;
- To produce and make available, under licence, a range of geographically referenced data outputs on land cover characteristics, tailored to the needs of Consortium members;
- To calibrate and validate satellite-derived classifications against ground reference data, publish results of the correspondence analyses and provide a guide to their interpretation.

This Report is the third six-monthly *Interim Report* on LCM2000 (see Fuller *et al.* 1998b, 1999a) and covers work done up to 23 August 1999. It is part of a series of reports on LCM2000, which are listed in the references.

2. REFERENCE DATA FOR MAP PRODUCTION

2.1 Images

Image acquisitions were described in earlier reports (Fuller *et al.* 1998c, 1999a & 1999a) which showed that well over half of the UK was covered by both summer and winter imagery by spring of 1999. No significant searches have been undertaken since then, as the Eurimage archive of quick-looks for summer 1999 will not be complete until mid to late September 1999; a search will be made then. Observations of weather on overpass dates suggest that most of the gaps in UK coverage will have been filled in 1999.

2.2 Broad habitat specifications

The Specification for LCM2000 set out a plan to map, as far as possible, widespread examples of Broad Habitats, as defined under the Biodiversity Action Plan. A list of target cover classes was reported in detail (Fuller *et al*. 1998c). Amendments were agreed at subsequent Consortium Meetings. Field-meetings with Scottish Natural Heritage and the Countryside Council for Wales, plus reconnaissance surveys of summer 1999, have demanded further evolution of the classification to accommodate new cover types and to correct misunderstandings of Broad Habitat definitions. Finalised definitions are yet to be provided by the Joint Nature Conservancy Council. Therefore, the current list of LCM2000 Target Classes (Table 2) is provisional.

The following are key developments of the classification:

- A criticism of the LCMGB 1990, was its failure to distinguish ericaceous bogs from heaths. To improve classifications in LCM2000, members of the field reconnaissance team had a Technical Meeting with field surveyors from SNH, in mid-May, on Rannoch Moor. The Scottish meeting helped to clarify the very real problem of the LCMGB ‘bog’ and ‘heath’ classes. The LCMGB 1990 'Open dwarf shrub heath' was not entirely 'heath' but included both 'heath' and 'bog' types. To distinguish these in field reconnaissance, the Key to Broad Habitats (see Barr, 1998) is a good but necessarily simplistic guide, identifying species diagnostic of bog. The combination of additional field training at the Technical Meeting plus the Key allows the LCM2000 team to identify ericaceous and grassy bogs in reconnaissance. It was found that the distinctions could be made in passing without the need to closely examine the floristic composition. Checks against images, if universally repeatable, suggested that it would be possible to distinguish these classes spectrally.

- Unfortunately, a later field visit to Migneint with CCW showed that the situation was not quite as simple as that. The CCW team stressed how important the measured depth of peat was to their interpretation of the bog/heath categories. While key indicator species were often diagnostic and while overall appearance was strongly indicative, the use of an auger to test peat depth was still used as the ultimate arbiter by the CCW Phase I survey teams. The decision process beyond the 'quadrat' scale was made the more difficult because of the highly variable peat depths, precluding an easy categorisation of widespread cover types.
- The difference in approach reflects the different objectives of the SNH Peatland Survey team and that of the CCW team based upon Phase I phytosociological methods.
- The LCM2000 team were left with the conclusion that, while some bogs were unambiguously identifiable from a cursory glance, others required contextual information on peat depth, not obvious to the observer in passing and not likely to be detectable using present remote sensing capabilities. Therefore, the use of drift maps to add a 'knowledge-base' to the distinctions was deemed essential, though itself clearly subject to the uncertainties of drift mapping based upon point-sample surveying. It is expected that a combination of spectral reflectance, class-probability based on reflectance, and dominant drift per segment will give a relatively refined classification of bogs and heaths in LCM2000, improving substantially on 1990 records and, hopefully, offering the best national picture of these Broad Habitats, to date.
- Further changes were made in the tabulation of rough grass, marsh and coastal habitats based upon an improved understanding of definitions.
- Ultimately, the classification is based upon the chosen aggregation knowledge-based classes and subclasses, which can be thought of as the building blocks for this (and any other) classification. There are currently about 83 thematic subclasses (Table 3), many of which are subdivided spectrally within a scene. Once Broad Habitat definitions are fully spelt out, then minor modifications will allow simulation of the Habitats from the thematic and, if necessary, the spectral subclasses.

2.3 Ground reconnaissance data

Ground reference data collection in 1999 consisted of 5 reconnaissance surveys covering a total of 40% of the UK, bringing overall reconnaissance coverage up to *c.* 85% (Figure 1). Some 83 thematic subclasses have been recorded in field reconnaissance trips: each is identified in Table 3 and related to Broad Habitats in Table 2. The completion of these 5 surveys provides the ground reference data needed for classification of images until late spring 2000. Reconnaissance in northern Scotland and Northern Ireland will take place at that time, to ensure continuity.

Table2

Table 3

3. PRE-PROCESSING OF IMAGE DATA

Operational use of CLEVER-Mapping based upon image segmentation is (as has always been made clear) entirely new and unique to LCM2000. It has necessarily offered a steep learning curve and many procedural developments in the early production stages. Some such developments were in the form of software refinements; many of them required the development of macros to automate, as far as possible, the processes of image analysis; but others simply formalised and made objective the complete sequence of activities needed to conclude satisfactorily the making of LCM2000. It was recognised that the investment of time in such developments would later benefit productivity, though it did slow the early production phase.

The following sections outline the techniques and analytical 'tools' which now form the basis of the operational programme. Those issues discussed in earlier reports are only covered briefly. New developments receive more attention.

3.1 Haze, cloud and cloud shadow

Earlier reports outlined procedures implemented for the correction of atmospheric haze, cloud and shadow masking within scenes. These improvements have been incorporated routinely into the processing stream (Fuller *et al.* 1999a).

3.2 Geo-registration and resampling

Geo-registration registers the images to the British National Grid (BNG) with a 25 m output pixel size. This correction initially involves collecting a sample of ground control points on an image and the equivalent points on an OS 1:50 000 paper map. The image analysis system calculates a transformation between image and map, calculating the differences in scale, rotation and offset. The image is then 'resampled' to match the map projection.

When producing an output image to the reference projection, it is necessary to chose a resampling algorithm which can calculate realistic reflectances for each pixel, based on the reflectances of the surrounding pixels in the unregistered image. The conclusion, after trials of the resampling methods, was that cubic convolution produced better results for segmentation and this form of output has proved fully satisfactory for classification purposes.

3.3 Illumination correction

Landscape facets are illuminated differentially, according to whether they are horizontal, face the sun, or face away from the sun. This differential illumination can be modelled using a digital terrain model (DTM) and compensation duly made. Such correction is important if topographic facets of land surface are not to have a highly significant and perhaps dominating effect upon the results of image segmentation. ITE commissioned Cambridge University to operationalise software which was tested in the CLEVER-Mapping programme. A 50 m grid-based DTM has been purchased from EDX Engineering Inc., for use in terrain modelling. Terrain corrections are now routinely being used for all images: this not only compensates for the terrain but also corrects the illumination to give results equivalent to a nadir viewing angle with the sun directly overhead: the reflectances are all thus normalised and standardised, allowing inter-comparison within and between images. It follows that, from this corrected dataset, ITE is building an archive of spectral signatures for target classes, throughout the UK and in varying seasons.

3.4 Band combinations for analyses

Landsat Thematic Mapper (TM) offers 6 different spectral bands, from visible blue to mid-infrared, with a spatial resolution of 28.5 m and a thermal infrared band at a coarser 120 m spatial resolution. LISS offers 3 bands from green to near-infrared with a spatial resolution of 23.5 m, and middle infrared band with a spatial resolution of 70 m. A summer-winter composite of two TM scenes could give 12 usable bands; a TM-LISS composite gives 10 bands. Where, as in SE England, an extra TM scene is added, there are 16 available bands. Along margins, scenes overlap and several composites can contribute data, with perhaps tens of bands at some locations. All such data can be captured by IGIS to record spectral response per band per date. However, there may be disadvantages in using bands which potentially confuse rather than inform the process. Moreover, many of the processes are limited, in principle or by design, to operate on a smaller bandset (details are given under relevant sections of this report). The net result has been, as a rule, that LCM2000 has used the red, near-IR and middle-IR bands of the first choice summer and winter images for classification. Extra waveband information from additional dates (where available) has only been used where it is needed specifically to help distinguish spectrally similar cover types. Such use of extra data is made in direct response to problems which arise with particular scenes and particular date-combinations (Table 4).

Table 4. Scenes, locations, dates and supplementary images / bands used in production of Land Cover Map 2000.

Area ID	Location	Summer data[1]		Winter data[1]		Supplementary data used	Bands used
		Sensor	Date	Sensor	Date		
s0_23q3	NE Norfolk	TM	14/05/98	TM	14/02/98		
s0_24	Suffolk	TM	30/05/98	TM	14/02/98		
S5_30	Lincolnshire & Norfolk	LISS	16/05/98	TM	14/02/98		
S5_31 S5_31a	Cambs N London	LISS LISS	16/05/98 16/05/98	TM TM	14/02/98 14/02/98	09/08/98	NDVI as KBC[2]

[1] Bands used were:
segmentation: Summer R, MIR and Winter NIR;
classification: Summer and Winter R, NIR, MIR

[2] Normalised difference vegetation index used in knowledge-based correction:
NDVI is proportional to the quantity of vegetation cover and was used to distinguish arable areas, harvested and predominantly bare in August, from woodlands, in leaf in August, which were otherwise confused with arable crops on May images of Cambridgeshire.

4. IMAGE SEGMENTATION

4.1 The segmentation software

Laser-Scan has implemented a fully operational version of the segmentation software, written originally in the Microsoft Windows environment by CUGD, but now in a Unix environment. It is in effect a ‘beta-test’ version, suitable for ITE operations, though not yet ready to market; it has, however, been fully tested and approved by ITE for LCM2000 operations.

The segmentation consists of two separate stages, i. edge-detection to identify boundary features and ii. region growing from seed points selected to avoid the edges identified in the first stage.

There were a number of methodological issues during the early production phase, including:

- Band selection for edge-detection and segmentation;
- Choice of edge-detection methodology;
- Thresholds to identify edges;
- Settings the degree of segmentation and the thresholds to achieve that;
- Post-segmentation boundary rejection and generalisation.

These issues have been discussed in previous reports but are briefly discussed in turn here.

4.2 Band selection

It is only possible to use three bands for edge-detection / segmentation. Therefore the segmentations are derived from a combination of one band from the winter image, and two bands from the summer image. If there are two scenes for either date, the first choice is to use the best summer scene with the best winter scene. If information from a third image is required the use of PC-1 will be investigated for each scene. In practice all scenes classified to date have used one band from a winter scene and two bands from a summer scene, with no need for alternatives.

4.3 Edge-detection

The segments are built around ‘seedpoints’ which have been selected as within a land parcel; a Sobell-based edge detector is used to ensure that the appropriate seedpoints are selected away from parcel-edges.

4.4 Segmentation and thresholds

The level of subdivision resulting from the segmentation procedure is in the control of the operator. By defining very tight thresholds, where small differences in reflectance values would define the spectral boundaries between segments, many very small segments can arise. In principle, this could reach the point where each pixel is itself a segment. Conversely, by allowing very loose thresholds, the parcels can be large, ultimately to the extreme of covering a whole image. The aim is to ensure a field-by-field segmentation, also separating zones within urban and suburban areas and subdividing heterogeneous semi-natural zones into meaningful segments.

Thresholds based upon reflectance differences between segments determine the amount of subdivision. Three different values are required, one for each band in the segmentation. An objective way of calculating the thresholds has been developed, using the means ± 2 standard deviations for each of the bands.

4.5 Post-segmentation generalisation and boundary rejection

It is possible, having derived an agreed segmentation, to simplify the results using spatial generalisation procedures - for example, eliminating reject pixels (not part of any segment); also dissolving very small polygons and growing adjoining regions into the dissolved areas, using intelligent spatial-contextual-analyses. Though these steps may incorporate odd pixels in a segment, they are necessary to reduce the final vector dataset to a manageable number of polygons.

The rules for post-segmentation generalisation and boundary rejection are as follows:

- Non-segment, edge pixels should be dissolved into adjoining parcels;
- Single-pixel islands should be dissolved into their surroundings;
- Small segments of less than nine pixels should be attached to their nearest neighbouring segments, using a spectral minimum distance rule;
- Segments < 16 pixels (<1 ha - the minimum mappable unit) might also be dissolved using a tighter spectral minimum distance rule.

4.6 Conversion to vector format

Once acceptable segmentations have been achieved, vector versions are created in a GIS database. This has been a simple procedure of raster-to-vector conversion where the boundaries between segments with different values in the segment images were represented by vector lines. Land parcel attributes, required in later analyses, have also been attached.

5. ORDNANCE SURVEY OF NORTHERN IRELAND VECTOR DATA

CLEVER-Mapping in Northern Ireland will aim to use vector-based CLEVER-Mapping built upon (Ordnance Survey Of Northern Ireland) OSNI vector map data (see Fuller *et al.* 1999c), similar to the procedures used in Jersey (Smith & Fuller, 1998 & in press). Trials have started to investigate the potential.

ITE was initially supplied with a 2.4 km x 3.2 km sample area of OSNI data, based upon the standard ‘tiles’ of vector data (08. km x 1.6 km). The NTF Level-2 format was used for export and import to Laser-Scan IGIS was achieved without obvious loss of linework but there were losses of detail. For example, it was observed that the original linework, which carried multiple attributes per line, only recorded single attributes upon import from the NTF Level-2 format. Features were interpreted as lines not areas; and a validation error was recorded because polygons did not close; the structured topology was lost.

It was also decided that the breakdown of all Northern Ireland into small tiles would prove problematic; building a continuous topology would be too great a task for inclusion in the LCM2000 work programme, and impossible to complete by the November 2000 launch date.

OSNI have suggested that they may be able to remove the tile frames and link polygon boundary features between tiles. The data would then be imported by OSNI as OSNI NTF data into ARC/Info, to build polygons and deliver shape files to ITE. These suggestions are provisional at present. OSNI will try a test area for ITE to examine. The way forward will be determined from there.

The Report of Fuller *et al.* (1999c) discussed the need for generalisation to protect OSNI interests and to produce simplified vector polygons commensurate with mapping from 20-30 m images. Such issues will be reconsidered once trials of frame-removal, boundary-linkage and polygon-building are complete and ITE has had the opportunity to evaluate the outputs for potential use in satellite-based CLEVER-Mapping.

6. TRAINING THE CLASSIFIER

Training for per-parcel analyses has operated in a similar way to that used in conventional maximum likelihood per-pixel classification. However the procedure has been objectively based on the land parcels. Land parcels can be identified individually by pointing to and highlighting them, without the need to painstakingly draw the outline. The process of training has therefore proved probably 20 times quicker than conventional manual definition, allowing a much larger sample.

Field reconnaissance data have been used to steer the training process, with land parcels which are recognisable on field-maps and images being identified as potential training polygons. Supplementary data were added by image interpretation, checked against external map data (e.g. in definition of urban areas, woodlands, water bodies and saltmarshes).

Training has ideally used a large number of extensive training areas. The team have attempted to identify at least 5 areas per subclass (usually many more), all examples broadly similar in size, such that none dominates the training set, and with, if possible, at least 30 pixels per training area. Sometimes, this has proved impossible. Some subclasses rarely offer extensive examples (e.g. scrub, bracken, evergreen broad-leaved woodland). Others, while nationally extensive, are locally rare (e.g. heaths). The team chose, where necessary and possible, to use a small training set. This sometimes proved successful, with extrapolations proving very realistic. Conversely, some extrapolations classified large areas of the scene based upon a very few pixels of training data. The only option here was to exclude the class altogether and accept that perhaps a small-scale of coverage of a class was missed rather than being exaggerated to a greater degree. Later knowledge-based corrections have sometimes allowed interactive procedures to make local corrections, where deemed necessary.

6.1 Reviewing training areas

A refinement built into IGIS operation has allowed the team to review training areas, after collecting the chosen samples. A procedure has been written by ITE in LULL (Laser-Scan’s macro language) which displays ‘image chips’ representing the remotely sensed data for each training area, side-by-side, on the screen. The operator has then compared and contrasted training areas and, where necessary, labelled the different thematic subclasses (Table 3) into a series of even more disaggregated spectral subclasses (see Kershaw & Fuller 1992). The operator has rejected odd examples and, as a result, has selected the resultant training sets. The operator has then reviewed the training areas both in summer and winter band combinations (something too time-consuming to consider in LCMGB 1990), to ensure that the spectral variant of thematic subclasses are not mixed in any image band-set. Additional information informed the operator of the size of each polygon, to give some idea of its contribution to the final classification. Training class statistics were displayed on the monitor, where needed to help evaluate the training set. These procedures, thought to be entirely new to operational image processing, have been developed specifically for LCM2000. They have also saved at least 2-3 days training for each scene pair. They have been enormously powerful in helping to refine the training sets. They are contributing substantially to the intended quality improvements in national land cover mapping.

6.2 IGIS 'self-training'

A consequence of using parcel-based procedures has been that classified land parcels from one image have helped train the classification of an overlapping, unclassified scene. Thus, areas of overlap have been used to locate near-identical polygons (>80% overlap) on the neighbouring scene and to pick up *a priori* class labels for use in classification. Probability rules have rejected the use of labels which were attached with low probability ($P < 85\%$). This 'roll over' of training data has frequently helped to identify additional examples of rarer training types, improving the chances of defining a valid training set for extrapolation. The rolled over training data and the training data attached from field reconnaissance are reviewed simultaneously to produce composite thematic / spectral subclasses.

6.3 Extraction of subclass statistics

Classification requires that the training polygons are used to derive statistical measures of reflectances, in each chosen band and for each spectral subclass. CLEVER-Mapping has used a shrinking procedure, when extracting raster data for the polygons, to avoid edge pixels and to ensure the use of 'pure' core pixels of a cover type. The shrinkage has been made a dynamic process whereby the required amount of shrinking was applied and, if insufficient raster data was collected, the shrinkage amount was reduced and the raster extraction repeated. This process is made to continue until enough raster data is extracted or the shrinkage amount reaches zero. The results of this procedure, in terms of pixels extracted and shrinkage achieved, have been reported and stored on the land parcel for future reference.

7. CLASSIFICATION

7.1 Per-parcel maximum likelihood classification

A polygon can be used to interrogate any scenes which include that polygon: thus, in areas of overlapping scenes, multiple datasets might be drawn upon. However, in 1990, after wide-ranging experimentation (Fuller & Parsell, 1990) it was shown that red, near infrared and middle infrared bands were most useful in discriminating classes. These have been the automatic choice for current image analyses. Only where other bands were needed to separate target classes have they been called upon. The classification procedure uses the same maximum likelihood algorithm as does per-pixel classification. However, it applies the procedure to the polygon, using mean statistics for the raster data within the polygon to select the most likely class in statistical terms.

A substantial refinement has been that CLEVER-Mapping in IGIS records not only the most likely class and its probability but also records the probabilities for the top five subclass options, usually covering >90% of the probability distribution. This information has proved particularly powerful in later knowledge-based corrections.

Per-pixel classifications are also being made as a 'standalone' product and also to record the natural heterogeneity associated with CLEVER-Mapped polygons. Of necessity, refinement of this raster output has taken second place to that of the parcel-based classifications, with a lesser effort in knowledge-based corrections. The per-pixel classifications have not attracted the same level of attention as did the same output data of 1990, for the same reason that the parcel-based outputs of 2000 have been of much higher priority. The combination of vector parcels with their cover, plus the raster classification, will offer huge potential for refinement, if, when and where needed, once initial production is complete.

7.2 Knowledge-based correction

Knowledge-based correction (KBC) procedures are being used to identify and re-label land parcels which are classified with low confidence and / or with classes out of their natural context (e.g. urban areas in mountains). A basic procedure was developed for per-pixel use by Groom and Fuller (1996). Correction on a per-parcel basis allows more subtle and complex rules to be applied, providing a powerful tool for improving classification accuracy. The following KBC rules are being applied universally:

1. Where a subclass is allocated with a probability of <50%, the other subclass probabilities have been summed to see if any other Broad Habitat is more appropriate - for example, a polygon might be classified as 'urban' with 35% probability; wheat might be 30%, barley 20%, oats 10%, and oilseed rape 5%. Viewed simplistically (in statistical terms) but realistically in terms of likelihood, such a field is highly likely to be in the 'Arable ...' Broad Habitat. A procedure not unlike the 'single transferable vote' tots up the 2nd to 4th highest probability classes at the Broad Habitat level and, where another Broad Habitat 'wins' by a higher 'proportional representation' of the likelihoods, the procedure re-labels the polygon as the second ranked subclass in the probability listing.

2. Coastal masking has been used, just as in 1990, to preclude the recording of 'inland beaches' and 'urban areas' on the shoreline (Groom & Fuller 1996). Classes out of their natural context are reallocated to the most appropriate (i.e. spectrally closest) class. The coast mask is based upon the 1990 coastline (which was drawn interactively for all of Britain); it is updated and refined where necessary with the DTM 0 m contour as a guide and interactive delineation, applied per-polygon.

Most scenes present specific problems of their own, varying with the date of imagery and the imaging conditions. These problems need tailor-made solutions. The following rules have been applied selectively to correct scene-specific problems:

1. The DTM has been used with a threshold to identify erroneous areas of 'Fen, marsh and swamp' in the Cambridgeshire scene off the floodplain and out of their natural context, changing these to 'Neutral grassland'.
2. In the same scene, bare ground in the context of coniferous plantations, evidently felled, is recorded as the 'felled' subclass and so included in the 'Coniferous woodland' Broad

Habitat. Conversely, felled classifications outside of a coniferous context is more likely to rough grassland and converted to the appropriate class.

All contextual corrections have been recorded in the GIS database, to include the input and output classes of any changed polygons, and the rule(s) used in their alteration.

A number of procedures will be used universally rather than being applied per-scene. Drift maps will be used to qualify the categories of semi-natural grassland with ‘neutral’, ‘acid’ or ‘calcareous’ labels, as demanded by the Broad Habitat classification. Peatland drift will refine the classifications of bogs and heaths.

The KBC correction procedure might also be used, after calibration against CS2000 field survey data, to reduce biases, by seeking out polygons of over-estimated cover types and checking that they are classified with high levels of confidence. Similarly under-estimated types might be reviewed in polygons where they came second in likelihood, to see if there is good reason to correct the classification. The investigation of such a potential will not take place until late in the validation process.

8. MAP OUTPUTS

Samples of output, both covering an area about 20 km wide, appear in Figure 2. The samples cover a. the north Norfolk coast from Brancaster and Scolt Head in the west to Blakeney Point in the east; and b. the Thetford area of Norfolk, on the borders with Suffolk and Cambridgeshire. The samples were chosen because, between them, they illustrate a combination of 14 widespread Broad Habitats out the 18 possible types:

1. Broad-leaved woodland	11. Fen, marsh and swamp
2. Coniferous woodland	13. Standing water/canals
4. Arable & horticulture	14. Rivers and streams
5. Improved grassland	19. Littoral sediment
6. Neutral grassland	17. Supra-littoral sediment
8. Acid grassland	26. Inland rock
10. Dwarf shrub heath	27. Built up areas, gardens

It is worth drawing attention to a number of key features. The first and perhaps most obvious feature, to those familiar with the area, is the accuracy of the classification. There are small pockets of misclassification but the detail is predominantly correct.

In the north Norfolk map it is possible to see the sand and mud flats of the coast, protecting an outer fringe of dunes and shingle spits, which in turn offer shelter to extensive saltmarshes, only dissected by drainage creeks. The dunes at Holkham (centre) are in part pine-covered, with the marshes to landward of the dunes having been reclaimed such that most are now arable land but with blocks of semi-natural and improved grasslands also present. The shelter belt of deciduous woodland around Holkham Hall, just inland, is a characteristic feature, with the deer park to the north of the estate showing as grassland but with arable land dominating the rest of the grounds and the surrounding landscape. Built up areas include Wells (mid-east), just large enough to include an ‘urban’ core, with the smaller villages of Burnham Overy Staithe and Brancaster Staithe (upper mid-west) and the slightly larger village of Burnham

Market (mid-west), which serve to illustrate well the ability of the segmentation to build small polygons of these relatively small settlements.

The Thetford image shows the large blocks of coniferous plantation, with smaller parcels and fringing plantations of deciduous trees. The Little Ouse River runs through the forestry, approximately east-west, with patterns of marsh, rough grassland and lines of deciduous trees. The MOD Stamford Battle Area is a large area of acid grassland, distinctive in spectral signature and distinguished from the mainly rough grasslands and improved grasslands of the wider countryside. The River Wissey runs through the Stamford Battle Area, with a near-continuous margin of deciduous trees. Thetford town is obvious (S central) with the small urban core and the rather larger industrial areas to the south-west and the north. The other urban areas are RAF Lakenheath (SW), Brandon (mid-west) and Feltwell (west).

For those used to conventional per-pixel classification, a key feature of these sample map-areas is the lack of noise in the data, the very clean classification which results from CLEVER-Mapping. There is no tendency for artificial within-field heterogeneity; and little misclassification of mixed, between-field, pixels. Yet, the presence of fine detail is evident in the creeks of the north Norfolk saltmarshes, the rides in the Thetford forests and the runways of the Lakenheath airfield (SW corner of the Thetford map).

What is not apparent here is the land parcel structure underlying the generalised map depiction, where the arable land, for example, is subdivided into individual fields. What also is missing from this generalised display is the detail at subclass level, where, for example, the arable fields are subdivided into different crop types.

While these two examples were chosen for their distinctive mix of features, the detail and accuracy they portray are fully characteristic of the other areas which have been mapped. Other sites will be illustrated on the CS2000 web-site by 20 September 1999, prior to the Advisory Group meeting of 29 September 1999.

There are other Broad Habitats, not appearing in this illustration, nor yet on the web-site, which are upland types and which have not yet been covered by CLEVER-Mapping in LCM2000. The reason for this is the adoption of the ‘roll-over’ process. This process is valuable as it extends our knowledge of classes from one scene to another (and it was entirely logical therefore to start in Cambridgeshire where the team’s knowledge was greatest. Roll-over also ensures that ‘seamless joins’ are made as there is no tendency for differing results to abruptly meet as there would be with a piecemeal approach to the classification. The consequence of the approach is, however, that roll-over will be a progressive process, working westwards and northwards from our start in the south-east.

9. VALIDATION

9.1 Basic checks

Initial checks have used external datasets to check provisional classifications in an iterative process of classification, review and re-classification. The annotated images from the field reconnaissance have formed the first choice for checking. OS maps have also highlighted discrepancies in those classes which they depict (eg urban, water and woodland). Various external datasets will also be used. DETR have provided digitised urban boundaries for knowledge-based corrections and validation. SNH have passed on various maps of peatland. CCW have provided grid-referenced examples of Phase I cover types. Negotiations with relevant parties, over use of other external data, will continue.

9.2 Preliminary validation

The principles of the per-parcel validation have been tested in the Land Cover Map of Jersey (Smith & Fuller, 1998 & in prep.). An initial LCM2000 validation was initially applied to the areas under classification, using check-polygons (i.e. those with *a priori* labels but not used in training) to provide an independent, early assessment of quality. Even before knowledge-based correction, the correspondence was approaching 90 % at the widespread Broad Habitat level. The first classifications showed that, in principle, the Broad Habitat classification could be achieved.

Checking using known polygons not as training but as validation data was widely envisaged in the initial *Specification*. However, it has become clear that, in rarer and/or dissected cover types, the omission of some polygons from the training data can severely risk the use of an inadequate training set: it has often proved difficult to find 5 examples of a class containing at least 30 core pixels per polygon. It is for this reason that ITE has rejected the retention of independent examples of subclasses to save them purely for validation - their role in classification has been too important. The alternative, rejecting from the validation those subclasses / classes which are rarer, clearly had the potential to bias the validation towards the commonplace classes.

It was decided to use the full training set as a preliminary quantitative check on the classification of training polygons. It is recognised that this would also bias the result, in that data used to train the classifier are also used to check it. The accuracy measure can only be taken as an indication, therefore. Nonetheless, it is a valuable measure, prior to full and independent validation through CS2000 field data; it is also easily applied.

Table 5 gives results for the classification of 4 scenes. It records the results of correspondence at the subclass level, representing potentially about 40 cover types. At the simpler Broad Habitat level, with 18 widespread examples, correspondence is higher.

Table 5. An indication of accuracy, based on the classification of training polygons, for 4 scenes.

ID	Location	Per class correspondence (%)	Broad Habitat correspondence (%)
s0_23q3	NE Norfolk	85	93
s0_24	Suffolk	78	86
S5_30	Lincs & Norfolk	65	87
S5_31	Cambridgeshire	79	90

The results are simple scores of how *a priori* subclass labels for training areas are matched by the label finally allocated by maximum likelihood classification. It is, as noted, biased towards a higher correspondence value by the use of training polygons rather than an independent validation dataset. However, it also likely to underestimate the accuracy due to the lack of weighting for overall cover extent. So rare classes, usually less well mapped, contribute near-equally to the measure as do commonplace Broad Habitats. Later refinements will used a weighted measure as demonstrated in Jersey (Smith & Fuller, 1998 & in press). Accuracies are far greater than might be expected per-pixel; with 77% correspondence at the subclass level. It is clear that the value, averaging 89% is very close to the 90% target accuracy, measured per-polygon. It is considered that the use of a weighting for coverage and scene-size will, even prior to final ‘universal’ knowledge-based corrections, demonstrate that the required accuracy has been achieved. Results will be presented at the Advisory Group meeting of 29 September 1999. An objective measure of correspondence with independent check-data will only be made once digitised data for the field survey squares become available, later in the autumn.

9.3 Objective validation / calibration

Objective validation will use the independent sample of land cover data provided by the CS2000 field survey of 1 km squares. This procedure will be used to derive a method for ‘translation’ between detailed field classes and the more generalised target classes of LCM2000, providing assessments of bias in the LCM2000 classification, plus calibrated correspondence measures to allow inter-comparisons, integrated statistical uses and unified cover estimates at any scale of use; it will also provide improved extrapolation of field data in a spatial context and / or a better focus of LCM2000 data in a thematic context.

Comparisons have in the past used multiple regression methods to relate LCMGB cover to field observations (Fuller *et al*., 1998b). The use of such comparisons has been discussed with Peter Rothery (Statistician at ITE Monks Wood); and continued use of these is considered valid, useful and possible. However, a procedure has been suggested, based on the use of a confusion matrix of agreement (see example in Table 6) and using field-surveyed 1 km squares to estimate errors using bootstrap sampling (i.e. sampling with replacement) of the matrices for the individual 1 km squares in the stratum (i.e. the ITE Land Classes). The principles are outlined in the rest of this section.

Table 6. A hypothetical 'Confusion Matrix' of Agreement based on field-surveyed 1km squares.

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="6">LCM2000 widespread Broad Habitats (K) (1-18)</td></tr>
<tr><td>K=1</td><td>2</td><td>3</td><td>...</td><td></td><td></td></tr>
<tr><td rowspan="5">Equivalent
Field Survey
Broad Habitats
or e.g. Baseline
Classes (J)

(1-18)
or
(1-59)</td><td>J=1</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>2</td><td></td><td colspan="4" rowspan="3">P(J/K) =
proportion of pixels or area
classified as LCMGB cover type K
which are
Field Survey cover type J</td><td></td></tr>
<tr><td>3</td><td></td><td></td></tr>
<tr><td>...</td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

In Table 6:

(i) X_{TK} = Known total area in Geographic 'stratum' T which was classified as LCMGB type K

(ii) Estimate total area of FS type J in ‘stratum’ T by : $A_{TJ} = \sum_K (P(J/K) . X_{TK})$

(iii) Estimate errors from variation in Pr(J/K) between the 1km squares (which are the basic sampling units). Each 1km square gives rise to an estimate of the confusion matrix P(J/K) (albeit often ‘sparse’).

(iv) Sum estimates A_{TJ} over the ‘strata of interest’ to give $A_J = \sum_T A_{TJ}$

The errors (standard errors and confidence limits) can be estimated using bootstrap sampling (i.e. sampling with replacement) of the matrices P(J/K) for the individual 1 km FS squares in the stratum (e.g. ITE Land Classes (LC)). The B^{th} bootstrap estimate of the matrix P(J/K) for the stratum is the average of the matrices for the bootstrap sample of individual 1 km squares in the strata. Confidence limits for estimates A_J are obtained as the lower and upper 2.5% percentile limits of the frequency distribution of the M (200-500) bootstrap estimates.

Importantly, if P(J/K) is based on a GIS overlap in areas of the J and K classes within a square, or if a pixel grid is fine enough to include at least one pixel in each land cover polygon, then all FS cover types observed in the FS in that land class will be represented with some non-zero value in the matrix P(J/K).

(v) ‘Strata’ should, in theory, be ITE Land Classes from which FS squares can be treated as a ‘random sample’. However, to include enough averaging of individual 1km square matrices, it may be best to estimate bootstrap matrices P(J/K) for larger ‘strata’ such as CS90 ‘Landscape Classes’ or CS2000 ‘Environmental Zones’ which are amalgams of individual Land Classes and hence will involve averages over more squares. From the point of view of summing areas X_{TK}, this could still be done individual Land Classes, but all Land Classes in the same larger ‘stratum’ would use the same bootstrap estimate of the confusion matrix

This idea needs testing in operation, but it seems sensible, valid and appropriate in principle. There may be practical problems with having adequate validation data for rarer cover attributes, especially within each land class or aggregation of land classes. There is also the potential problem of variation in confusion matrices between satellite images; i.e. using bootstrapped confusion matrices built up from several images; and of partial extrapolation from some images and their accompanying field survey samples squares to estimate the confusion matrix and its bootstrap errors for another region.

This and other procedures will be investigated once sufficient sample data from field and LCM coincide. It is likely to be late October 1999 before such tests can commence. A Technical Advisory Group session will be devoted to validation procedures, once a representative selection of completed classifications and digitised squares become available and procedures have been tested.

10. ASSESSMENT OF WIDESPREAD BROAD HABITATS

The ‘assessment of Broad Habitats’ in the *Specification* envisaged an evaluation of the results of classification in regional and national contexts. This stage is clearly not yet realistically possible. However, the assessment needs also to look at the ability to map and measure these habitats as a pre-requisite to a geographical assessment.

The *First Interim Report* undertook a brief review of Broad Habitats with an assessment of their distinguishing features, difficulties in distinctions and their identification in relation to minimum mappable units, in both per-pixel and per-parcel measurement. This re-assessment simply serves to report experiences in production, difficulties in distinction, solutions where available and likely failures, where they exist, in full distinction of the widespread Broad Habitats.

1. Broad-leaved woodland generally presented no difficulties in distinction though, in the mid-Anglia scene of 16 May 1998 the images left confusion with arable crops, as both are actively growing in May. KBC has calculated a normalised difference vegetation index (NDVI - a measure proportional to the quantity of green vegetation) based on an August 1998 image to help distinguish crops and deciduous woodlands. In the other 5 scenes that have gone through classification stages the woodlands seem to very closely match field and OS records, without recourse to additional KBC. The expectation is that, by careful selection of dates and, if necessary, by use of additional data of other dates (planned within the budget of LCM2000), there will be no major problem distinction of this class.

2. Coniferous woodland: no problems observed except in occasional (and negligible) records of conifers on strongly shaded slopes with other vegetation; this is probably due to the relatively less effective operation of the terrain correction facility in predominantly flat lowlands, due to the shortage of differently angled facets for calibration of the correction algorithm. Felled conifer forests (treated as the Conifer Broad Habitat) demand contextual analysis of bare ground to identify their forest-context: this works well. There are no significant problems in mapping this Broad Habitat which have yet come to light.

4. Arable & horticulture: this category is generally identifiable and sub-divisible into specific crop types. No major difficulties have been encountered in classifications of extensive arable land.

5. Improved grassland: except insofar as the definition of ‘improved grassland’ is in itself problematic (and as yet the exact Broad Habitat definitions have not been supplied), there is no evidence of systematic difficulties in the spectral separation of the improved grassland Broad Habitat.

6. Neutral grassland, 7. Calcareous grassland and **8. Acid grassland** have, to date, mostly been mapped as unimproved and rough grasslands, without necessarily distinguishing their status in pH terms. The definition of unimproved and rough grasslands is open to some debate and applying such definitions rigorously through remote sensing may prove even more problematic. For LCM2000 purposes, these categories include grasslands with limited management by grazing or hay-cutting and situations with no apparent management such as natural (e.g. dunes and grass heaths) or abandoned swards (e.g. *Dactylis - Arrhentherum* swards). Drift maps will be used to qualify all of these. There will be residual problems due to the relatively poor resolution of such maps. But it has always been made clear that estimation of the stock of neutral, acid and calcareous grasslands will need supplementary data from the CS2000 field survey.

9. Bracken was not listed as a target class in the Specification. It is present in stands which are insufficiently extensive for classification and training, certainly in the 6 scenes of SE England which have been analysed in the early stages. Extensive stands in northern and western parts of the UK will allow its inclusion (though not woodland stands). For consistent portrayal at the UK level, it may be necessary to aggregate bracken with acid grasslands and treat it as a subclass of that target class. Again, the field survey will be needed to estimate the stock of bracken.

10. Dwarf shrub heath has presented no difficulties in training and classification in the SE England scenes. Difficulties in distinction from the Bog Broad Habitat (section 2.2) may cause future problems, though at this stage these cannot be reported.

11. Fen, marsh and swamp have, so far, been recorded extensively in Broadland and parts of the Fens. The category shows spectral overlap with rough grasslands, and accurate classification has benefited from the contextual use of the DTM to distinguish floodplain fens and marshes from the more widespread rough grasslands of higher and drier land.

12. Bog: no bogs have been recorded in the scenes mapped to date. Attention is drawn to notes in section 2.2 on bog classifications.

13. Standing water/canals: areas of standing water which are large enough to map have caused no problem in identification; the coastal mask removes any potential confusion with oceanic seas and estuarine waters.

14. Rivers and streams are rarely wide enough to map; as a result, the very few wider examples are not distinguished from standing waters. Other contextual data could later be used to make this Broad Habitat distinction, if needed.

15. Montane habitats - not yet mapped.

16. Supra-littoral rock - current coverage only includes the soft coasts of eastern Britain.

17. Supra-littoral sediment such as shingle beaches and sand dunes have been mapped separately, using a combination of spectral distinction of the sparsely vegetated surfaces, with maritime context to identify that they are not ‘inland bare ground’.

18. Littoral rock - current coverage only includes the soft coasts of eastern Britain.

19. Littoral sediment: due to wetness, intertidal sediments have spectral signatures which separate them from similar supralittoral examples; saltmarshes also have characteristic spectral signatures and have been mapped as a separate entity. The coastal mask is preventing potential confusion with inland surfaces showing similar spectra. As a result, no difficulties have been encountered in distinguishing this Broad Habitat.

25. Oceanic seas are readily distinguishable through the characteristic spectral signature of water in a maritime context. It is assumed that this class includes estuarine water (though this could be separated contextually if demanded by the Broad Habitat classification.

26. Inland rock: bare ground usually carries a characteristic signature, and coastal masking precludes accidental inclusion of supralittoral examples. It is not clear from the Broad Habitat nomenclature whether it is intended to include bare sand, mud, mineral workings etc. These are currently being treated as inland bare ground for expected inclusion in this Broad Habitat. However, bare arable land can be confused with these and there may need to be distinction, if possible, through contextual means.

27. Built up areas, gardens have proved to be very successfully distinguished during segmentation, and classified through characteristic spectral signatures. They are routinely being sub-divided into urban, industrial and suburban types, with good success.

Details of how these Broad Habitats are subdivided into subclasses / variants are given in Table 2.

In conclusion, the target classification is similar to that used in 1990, but with the changes required by the need to map Broad Habitats. Initial classifications have shown that, in principle, the Broad Habitat classification can be achieved, with the few *provisos* outlined in the original *Specification*. Overall success can only become clearly apparent once validation, based on field data, provides an objective measure of success.

11. THE PRODUCTION SCHEDULE

Production has involved processing scenes in parallel from initial atmospheric corrections through geo-registration, illumination correction, co-registration of summer-winter data and segmentation stages, rather than the procedures implied in the initial GANNT prepared for the *Specification*. The map (Figure 1) shows the state of processing. The revised GANNT records progress in scene-pair equivalents (Figure 3), using the scene-pair processing timetable to rate progress for each scene pair in turn and adding the elements together to establish progress overall. The progress has also taken into account the development work on pre-processing and its expected contribution to overall progress.

Equipment installation was completed in the first quarter. The **field course** was undertaken in May 1998. **Class finalisation** was all but completed in July 1998, but evolution / clarification of Broad Habitat definitions has kept this an open task; field visits with SNH and CCW also refined the team's understanding of bog / heath Broad Habitats; definitions are still to be published by JNCC, leaving this issue still open, despite operation of the Broad Habitat classification. **Methods refinements** involved procedural developments, software refinements, the development of macros to automate the processes of image analysis and other improvements simply to formalise and make objective the complete sequence of activities. **Field reconnaissance** has been completed for about 85% of the UK, with visits to NW Scotland and Northern Ireland needed to complete the exercise. The **Image search** and **image purchases** are half-completed and will quickly be finished once summer 1999 images reach the Eurimage archive. **Image analysis** has lagged behind the planned schedule with the equivalent of 8 scene-pairs processed, compared with an expected 11 pairs; however, the current rate of production now is commensurate with the target completion date. **Validation methods development / testing** awaits the delivery of sample field data, but plans are well advanced for the analyses; **validation / calibration** is similarly dependent. **Change detection trials** will take place later in the programme. Creation of **summary 1 km data & export to CIS** will take place once the LCM2000 is complete; it a process undertaken on various occasions and can be considered routine and very quick. The **assessment of widespread Broad Habitats** envisages two types of assessment: the first, assessment of the effectiveness of mapping the selected Broad Habitats is complete in subjective terms and demonstrably effective, but objective assessment will require the results of validation. The second phase assessment will be an evaluation of the distribution and pattern throughout the UK, including regional and national statistics. The production of **Quarterly statements** has been as planned. **Technical meetings** have taken place to schedule, with two field meetings in the last period. **Reports** are being produced as planned, with extra reports as contributions to the wider Countryside Survey. **Six-monthly Consortium meetings** have been rescheduled, with Consortium approval, to take account of duplication through Advisory Group meetings and Technical Meetings. The **submission of publications** has started with a submission on CLEVER-Mapping to the *International Journal of Remote Sensing* (Smith & Fuller, in prep.). **Preparation of general display and launch materials** was envisaged for later in the proceedings; currently material is displayed on the CS2000 web-site, this is being updated with sample outputs, and various display outputs are available in hard and soft copy formats; presentations on LCM2000 are becoming regular events.

12. CONCLUSIONS

- Summer-winter TM images cover most of England and Wales. It is believed that reasonably good weather in the northern and western UK in summer 1999 will, with substitute scenes as necessary, complete coverage for full UK mapping.
- A digital terrain model (DTM) for Britain is routinely contributing to pre-processing and to post-classification, knowledge-based corrections.
- Software refinements, macros which automate the processes of image analysis and other procedural developments have been formalised and are now successfully built into routine operations.
- The pre-processing and segmentation improvements, which delayed the start of production mapping, are now demonstrably operational and their refinement is producing time savings at post-classification stages.
- Field reconnaissance data have been collected for *c.* 85% of Britain.
- The field data are being used to train the classifier, with field-mapped land parcels objectively transferred onto the image segments. The training process has proved to be probably 20 times quicker than per-pixel training.
- The review of training data, using display image ‘chips’, has allowed the operators to refine high quality training sets very quickly, substantially reducing the number of iterations in the training-classification-review procedure, saving time in production.
- The interrogation of the image, to derive statistical measures for reflectances in each chosen band and for each spectral subclass using a dynamic shrinking procedure to ensure the use of ‘pure’ core pixels of each cover type, is fully operational.
- The maximum likelihood classification uniquely, in IGIS CLEVER-Mapping, is recording the probabilities for other major subclass options.

- A combination of internal context, external spatial data and class probabilities is being used in knowledge-based correction (KBC), to identify land parcels classified with low confidence and / or those with classes out of their natural context; and KBC is used to correct any errors which are apparent.
- Validation has been designed as a two-stage process: first, using training-polygons to broadly check the results per-polygon; second, using field survey data to fully and objectively calibrate results, offering a 'translation' between detailed field classes and the generalised target classes of LCM2000.
- Overall production has involved processing 14 scene-pairs in parallel, from initial atmospheric corrections through geo-registration, illumination correction, co-registration of summer-winter data and segmentation stages. Final classifications are available for 4 scenes with a fifth at the KBC stage; preliminary classifications are available for a further 2 scenes, with the remainder progressing through various stages of pre-processing.
- Initial classifications of widespread Broad Habitats have shown that, in principle, the Broad Habitat classification can be achieved very successfully, with further sub-division of Broad Habitat classes to meet wider user objectives. Checks show an accuracy of 90 % assessed relative to training polygons, prior to 'universal' KBC procedures.
- Overall success will only become clearly apparent once validation, based on field survey squares, starts to provide an objective and independent measure of success, starting in mid-autumn.

The GANNT (Figure 3) records progress in scene-pair equivalents. The assessment has taken account of the development work on pre-processing and its contribution to overall progress. Some tasks are a little ahead of schedule. However, because developments have been extended to refine methods, automate procedures, and reduce post-classification analyses, on balance, the production has got behind schedule. However, the current rate of throughput, resulting from refined pre-processing procedures, is making good that delay.

13. REFERENCES

LCM2000 Reports

Fuller R.M., Gerard, F.F., Hill, R.A., Smith, G.M., Thomson, A.G., 1998b. *Countryside Survey 2000 - Part III. Module 7. Land Cover Map 2000. First Progress Report.* Unpublished ITE report to the LCM2000 Consortium.

Fuller R.M., Gerard, F.F., Hill, R.A., Smith, G.M., Thomson, A.G., 1998c. *Countryside Survey 2000 - Part III. Module 7. Land Cover Map 2000. First Interim Report incorporating the Second Quarterly Progress Report. CSLCM/Interim1*. Unpublished ITE report to the LCM2000 Consortium.

Fuller R.M., Smith, G.M. & Hill, R.A. 1998d. Land Cover Map 2000. In: *Countryside Survey 2000 Report CSJMT 7/4. Part III. Module 7.* Unpublished report to the Joint Management team of Countryside Survey 2000. 1p.

Fuller R.M., Smith, G.M. Hill, R.A. & Sanderson, J.M. 1999b. Land Cover Map 2000. In: Hornung, M. (ed.) *Countryside Survey 2000: Second Integrated Progress Report.* Unpublished ITE report to the DETR.

Fuller R.M., Smith, G.M., Sanderson, J.M., Gerard, F.F., Hill, R.A., Thomson, A.G., 1999. *Countryside Survey 2000 - Part III. Module 7. Land Cover Map 2000. Second Interim Report incorporating the Fourth Quarterly Progress Report. CSLCM/Interim2*. Unpublished ITE report to the LCM2000 Consortium.

Fuller, R.M., Smith, G.M. & Hill, R.A., 1998e. *Countryside Survey 2000 Report CSJMT 8/6. Land Cover Map 2000.* Unpublished report to the Joint Management team of Countryside Survey 2000. 2p.

Fuller, R.M., Smith, G.M. & Hill, R.A., 1999a. *Countryside Survey 2000. Land Cover Map 2000. The Land Cover Map Of Northern Ireland. An Extension To Land Cover Map 2000. Preliminary Report*. Unpublished ITE report to the Northern Ireland Members of LCM2000 Consortium (and provided to the full LCM2000 UK Consortium).

Fuller, R.M., Smith, G.M., Hill, R.A., Thomson, A.G. & Gerard F.F. 1998f. Module 7. Land Cover Map 2000. In: Hornung, M. (ed.) *Countryside Survey 2000 First Integrated Progress Report*. Unpublished ITE report to the DETR.

Smith, G.M., Fuller R.M., Sanderson, J.M., Hill, R.A. & Thomson, A.G., 1999. Land Cover Map 2000: Fifth Quarterly Progress Report. CSLCM/Prog5.

Other references

Barr, C.J. 1998. *Countryside Survey 2000: Field Handbook. 3rd Draft.* Institute of Terrestrial Ecology, unpublished.

Fuller, R.M. & Parsell, R.J. 1990. Classification of TM imagery in the study of land use in lowland Britain: practical considerations for operational use. *International Journal of Remote Sensing*, 11, 1901-1917.

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

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.

Groom, GB. & Fuller, R.M. 1996. Contextual correction: techniques for improving land cover mapping from remotely sensed images. *International Journal of Remote Sensing*. **17**, 69-89.

Kershaw, C.D. & Fuller, R.M. 1992. Statistical problems in the discrimination of land cover from satellite images: a case study in lowland Britain. *International Journal of Remote Sensing*, 13, 3085-3104.

Smith, G.M., & Fuller, R.M., 1998, *CLEVER-Mapping of land cover in Jersey. Final Report*. Institute of Terrestrial Ecology Report to the States of Jersey.

Smith, G.M., & Fuller, R.M., In prep., Multi-sensor, high resolution, knowledge-based per-parcel classification of land cover: an example in the Island of Jersey. Submitted to the *International Journal of Remote Sensing.*

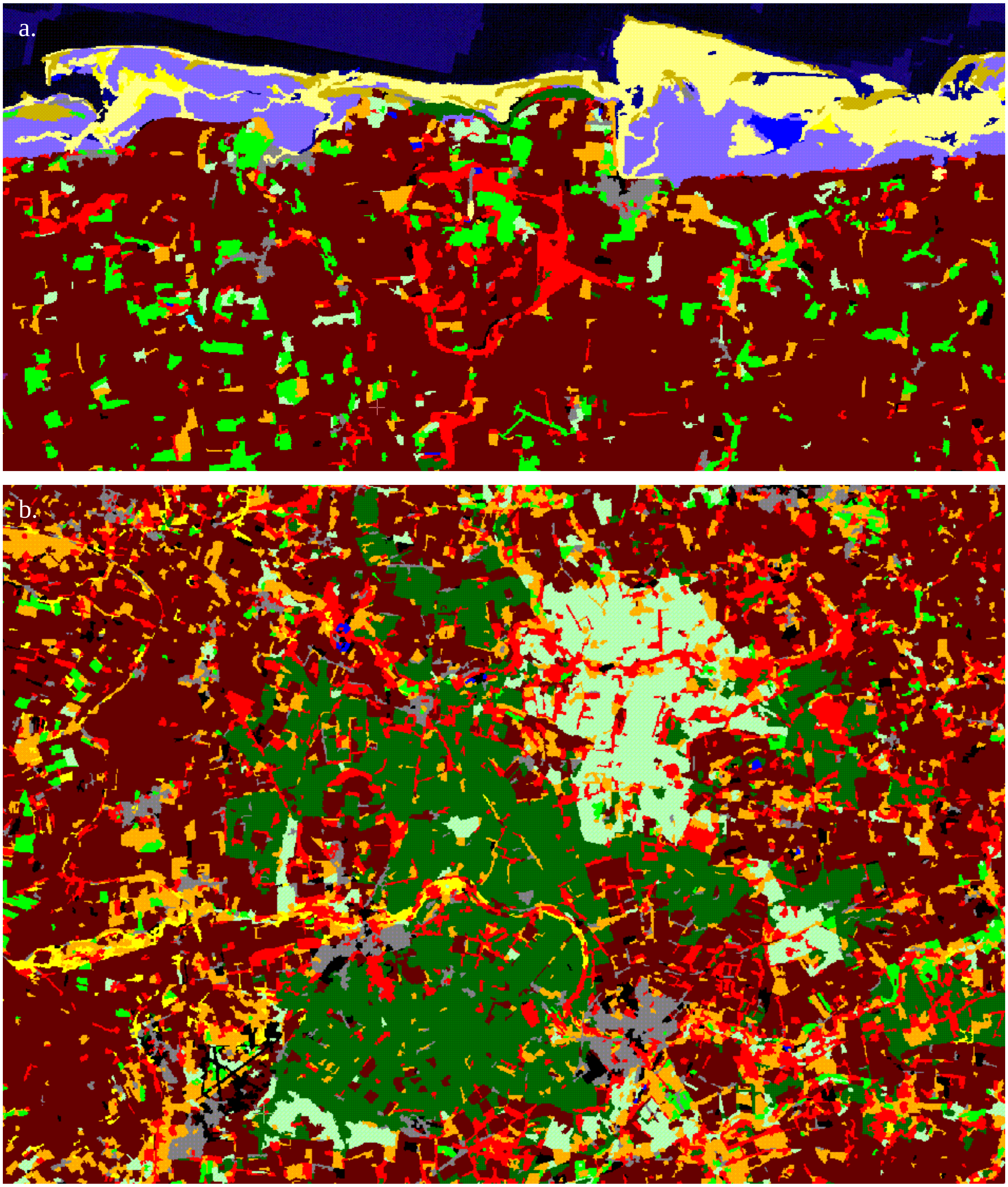

Figure 2. Sample outputs from Land Cover Map 2000, showing: a. the North Norfolk coast with Scolt Head Island (west), Holkham beach (centre) and Wells, Stiffkey and Morston marshes through to Blakeney Point (east); and b. the Breckland north of Thetford. Key Broad Habitats shown are: standing water (blue); littoral sediment (cream); supralittoral sediment - e.g. saltmarsh (pale blue) and fringing dunes (khaki); fen/marsh/swamp (yellow); neutral grassland (light tan); acid grassland (pale green); improved grassland (light green); arable and horticulture (brown); broad-leaved woodland (red); coniferous woodland (dark green); built up areas - suburban (grey), urban and industrial (black).

Table 3. Thematic subclasses and their codes, as recorded on field reconnaissance and used as a primary input to training and classification.

Arable:

Aw	Wheat
Ab	Barley
Ar	Oil seed rape
Ap	Potatoes
As	Sugar beet
Af	Field beans
Al	Linseed
Ao	Arable oats
Ah	Horticulture
Ac	Carrots
Aq	Peas
Am	Maize
Ax	Mustard
Aba	Arable bare
Ast	Cereal stubble
Se	Set-aside
Ss	Set-aside (sprayed)
Sb	Set-aside (bare)
Sv	Set-aside (vegetated)

Grass:

Gl	Ley
Gn	Neutral
Gi	Improved
Gu	Unimproved
Ga	Acid
Gc	Calcareous
Gr	Rough / unmanaged
Gj	With dominant Juncus
Gm	Moor (Nardus/Molinia)
Gmm	Grass moor molinia
Gmn	Grass moor nardus
Gh	Hay

Wood:

C	Conifer
Cl	Larch
Cn	Recent (<10yrs)
M	Mixed
Mn	Recent (<10yrs)
D	Deciduous
Dp	Poplar
Dn	Recent (<10yrs)
Dr	Rhododendron
E	Evergreen
Sc	Scrub

Wood (continued):

O	Orchard
On	Orchard (new)
Ov	Vineyard
Oh	Hop
Fd	Felled

Heath/Marsh:

H	Heather & dwarf shrub
Hd	Dry heath
Hw	Wet heath
Hg	Gorse
Ha	Arctic heath
Hb	Burnt heather
Hbg	Burnt heather now grass
Hga	Heather grass
Br	Bracken
F	Fen / swamp
Fm(g)	Fen marsh (grass)
Fw	Fen & willow
Bo	Bog
Bh	Bog (Heather dom.)
Bg	Bog (Grass dom.)
Bb	Blanket bog
Z	Montane habitats

Coastal:

Ls	Littoral sand
Lm	Littoral mud
Lr	Littoral rock
Sm	Saltmarsh
Smg	Saltmarsh grazing
Sr	Sub littoral rocks
Sd	Sand dune
Sds	Sand dune with shrubs
Sh	Shingle
Shv	Shingle vegetated
Ws	Sea
We	Water estuary

Urban & other:

U	Urban
Us	Suburban
Ui	Industrial urban
Ud	Despoiled land
Ba	Bare
W	Water
Wf	Water flooded

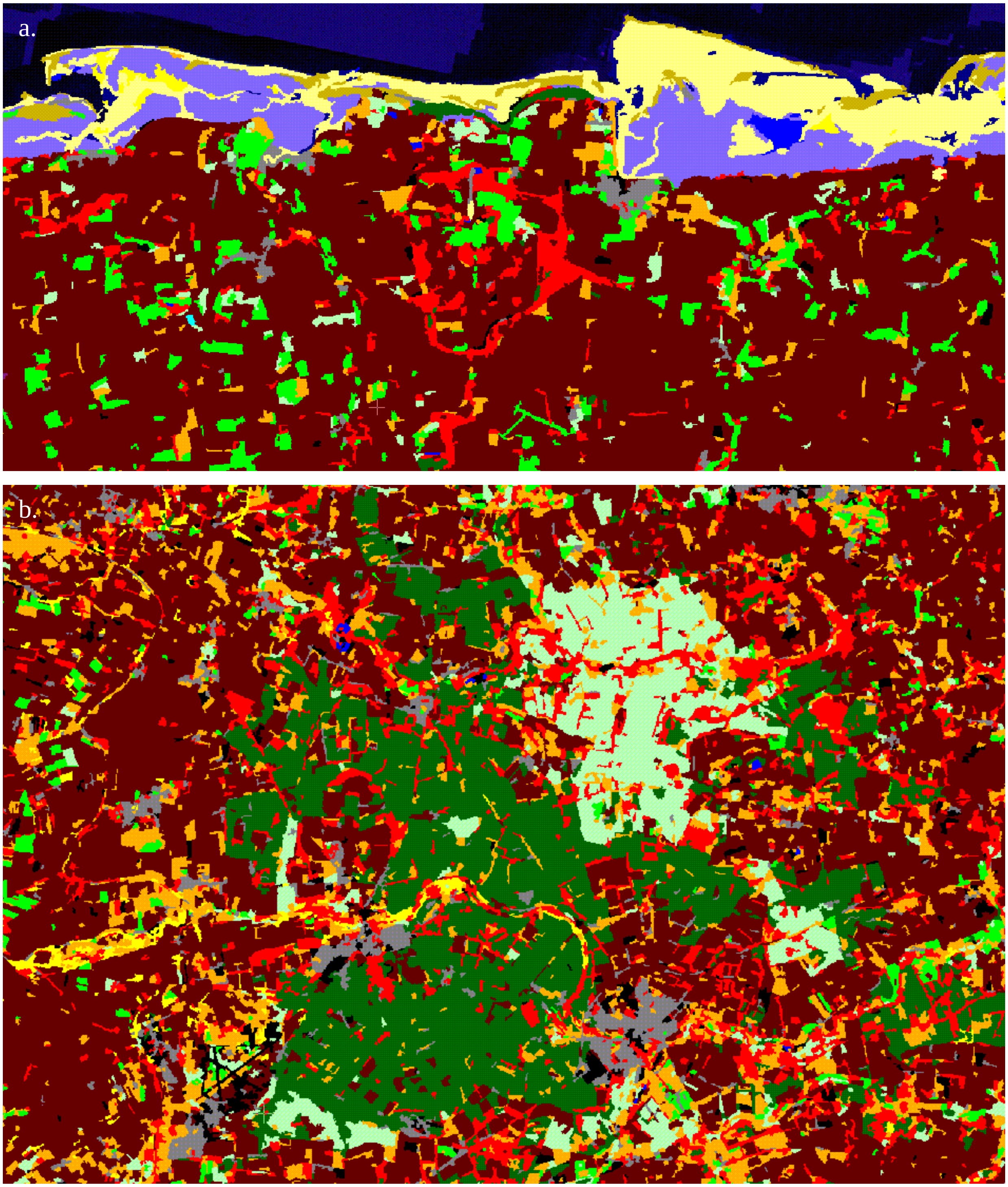

Figure 2. Sample outputs from Land Cover Map 2000, showing: a. the North Norfolk coast with Scolt Head Island (west), Holkham beach (centre) and Wells, Stiffkey and Morston marshes through to Blakeney Point (east); and b. the Breckland north of Thetford. Key Broad Habitats shown are: standing water (blue); littoral sediment (cream); supralittoral sediment - e.g. saltmarsh (pale blue) and fringing dunes (khaki); fen/marsh/swamp (yellow); neutral grassland (light tan); acid grassland (pale green); improved grassland (light green); arable and horticulture (brown); broad-leaved woodland (red); coniferous woodland (dark green); built up areas - suburban (grey), urban and industrial (black).

Figure 3. GANNT - The timetable for Land Cover Map 2000 (UK) and estimated progress in production (to the end of August 1999)