CENTRE FOR ECOLOGY AND HYDROLOGY
(NATURAL ENVIRONMENT RESEARCH COUNCIL)

Project. T02083j5/C00878

Countryside Survey 2000 Module 7

LAND COVER MAP 2000

FINAL REPORT

CSLCM/Final

R.M. Fuller, G.M. Smith, J.M. Sanderson, R.A. Hill, A.G. Thomson,
R. Cox, N.J. Brown, R.T. Clarke*, P. Rothery & F.F. Gerard

Centre for Ecology and Hydrology, Monks Wood,
Abbots Ripton, Huntingdon, Cambridgeshire, PE28 2LS

* Centre for Ecology and Hydrology, Winfrith Technology Centre,
Winfrith Newburgh, Dorchester, Dorset.

Corresponding author:
G. M. Smith
Section for Earth Observation
Centre for Ecology and Hydrology
Monks Wood
Abbots Ripton
Huntingdon
Cambs PE28 2LS

Telephone:
(Direct dial) 01487 772479
Fax: 01487 773467
Email: gesm@ceh.ac.uk

1 March 2002

CONTENTS

EXECUTIVE SUMMARY

Introduction

- Land Cover Map 2000 (LCM2000)[1] provides a census of UK habitats and land cover as digital maps and databases plus a range of derived products held in a geographical information system (GIS). The map updates and upgrades the Land Cover Map of Great Britain, made in 1990-92.
- LCM2000 is a classification of spectral data recorded by satellites; external datasets add context to refine the classification. The process generated 72 class Variants, combined into 26 Subclasses, which in turn gave 16 Target classes and 10 Aggregate classes. Subclasses were combined appropriately to simulate 20 Broad Habitats (BHs) of the UK Biodiversity Action Plan. In the following text, LCM2000 classes are referred to in bold lettering and Broad Habitats in italics.

Methods

- LCM2000 used red, near infrared and middle infrared reflectance bands from summer and winter satellite images. The image data were calibrated to reduce atmospheric haze effects, masked to remove cloud, shadow and snow, corrected for differential illumination due to undulating terrain, and registered to the British and Irish National Grids.
- The pixels which make up the image were grouped into areas or segments broadly equivalent to land parcels (e.g. individual fields) using a process of image segmentation. The results were generalised to exclude small segments ≤0.5 ha.
- The classification of land cover types within segments avoided the use of edge-pixels with mixed spectral signatures by using the mean spectral response of pixels within the core of each segment.
- Ground reference data from field reconnaissance surveys were used to identify image segments of known land cover. These formed a sample of so-called ‘training areas’ used to calculate the spectral reflectance statistics for each land cover class.
- Classification used a maximum likelihood algorithm applied to each segment. The procedure compared the mean reflectances of unknown segments with the training set and recorded the most likely spectral subclass in statistical terms: in fact, it stored probabilities for the top five spectral subclasses, usually covering >90% of the probability distribution.
- For segments which were classified with low confidence or with classes out of their natural context, knowledge-based corrections (KBCs) were used to allocate an alternative class label, where more appropriate.
- The individual classified satellite scenes were ‘mosaicked’ together into 100 x 100 km squares. Residual cloud-holes were patched using the best available substitute images (e.g. single date classifications).
- Maps showing acid-sensitivity in the landscape were used to label semi-natural grasslands as ‘acid’, ‘neutral’ or ‘calcareous’. Geological maps showing peatland were used to distinguish between ‘heaths’ and ‘bogs’.

Output maps and data

- Map displays and printed maps use cartographic conventions which balance the reliability of mapping and the importance and extent of a class, whilst bringing out important patterns in the landscape. However, greater detail is available in digital products at Subclass and Variant levels and any user-defined colour scheme could be applied.
- At national scales, regional differences across the UK are evident. At regional scales, the inter-relations of habitats within the landscape are clearly apparent. At the level of the individual county, the continuity and fragmentation of habitats becomes visible. At local levels, the full

[1] A ‘Glossary of terms and acronyms’ appears in section 19, page 73

landscape structure becomes much clearer: in unenclosed landscapes, the mosaics of semi-natural habitats are recorded; cities and towns are subdivided into urban and suburban zones with open spaces; individual fields are evident; and larger linear features such as rivers and motorways can be seen.

- In a GIS, the full scope of the LCM2000 data becomes clear. Each segment is delineated by 'vector boundaries' recording the outlines in digital form. The database carries a range of attribute data for each segment. These describe its shape, size and location, source images and their dates. Thematic details and class probabilities are also recorded.
- The map structure, related directly to real features on the ground, can be used to help our understanding of the environment. It shows the inter-connectivity of landscape features, their immediate context and the wider neighbourhood in which environmental influences operate. The map helps us to see how ecological principles can explain patterns of biodiversity. It provides spatial data on land uses which influence hydrology. The data underpin climate models; they can also be used to predict the impacts of climate change on landscapes and ecosystems. The deposition of pollutants and their environmental consequences can be assessed in context. LCM2000 indicates land uses and values which underlie planning, environmental risk assessment and socio-economic modelling.
- LCM2000 offers a database structure upon which users can build. They may edit data to make corrections and refinements. They may add qualifying detail. They might update the information, recording changing land cover through time.
- A 25 m grid-based 'raster dataset' was derived from the Subclasses of the vector-based LCM2000. Generalised 1 km summary products are also available; they record class-dominance and summary percentage cover per 1 km^2 at the 26 Subclass level. Subclass and Aggregate class datasets are provided in the Countryside Information System which gives data-access to non-technical users.

Comparison with field survey

- CS2000 field survey data, covering a stratified random sample of 569 one-kilometre squares, provided information to assess the quality of LCM2000 in Great Britain (field data are not in appropriate format for similar analyses in Northern Ireland). The comparison was made for the BH classes derived by both surveys. The field data are not so-called 'ground truth'; therefore, the process of inter-comparison was one of 'calibration' rather than 'validation'.
- Comparisons of the field survey and LCM2000 generated 'correspondence matrices', one for each sample square. Results show similarities and highlight differences due to different spatial resolutions, time differences in surveys, class-definition differences, and errors in one or both surveys.
- Statistical procedures were developed to provide confidence limits for national and regional measures of correspondence between the two surveys. The estimated direct correspondence in Britain, at BH-level, is 54% (with the 95-percentile range estimated at 53-56%). It was known from the outset that there would be mismatches when the field survey and LCM2000 were compared at BH-level. The correspondence is higher for LCM2000 Target classes: 65% across Great Britain, 73% for England-Wales and 51% for Scotland for comparisons based on field survey land parcels. The largest differences were found in upland areas where field and satellite-based mapping were most problematic.
- Differences in resolution, the data-model and timing of surveys contribute to the differences between LCM2000 and field survey. The comparison as a whole suggests that LCM2000 may record Target classes with around 85% success. Individual classes fare differently.
- LCM2000 **Broadleaved and mixed woodland** is near identical in extent to the field survey coverage of the *Broadleaved, mixed and yew woodland* BH in the UK. However, direct agreement is lower because many woodlands and clearings are at or below the minimum mappable unit of LCM2000. *Coniferous woodland*, generally planted and in larger blocks, records similar coverages and a greater direct correspondence.

- **Arable and horticultural land** covers nearly a quarter of the UK according to LCM2000 and equivalent field estimates. Apparent confusions between the *Arable and horticultural land* and *Improved grassland* BHs in LCM2000 relate largely to rotation farming in squares where field and satellite survey-years differed.
- **Improved grassland**, covering more than a quarter of the UK according to LCM2000 and field survey, is the most extensive single cover class. Generally, improved swards were readily recognisable and well-classified on LCM2000. However, the distinction of ‘improved’ grassland from ‘semi-natural’ types could be both difficult and controversial.
- **Semi-natural grasslands** (and **Bracken**) present problems in their distinction. *Neutral*, *Calcareous* or *Acid grassland* BHs gave no consistent spectral characteristic by which to determine soil acidity; and external data were of limited value. *Bracken* was problematic because so much of the imagery used to make LCM2000 was recorded in May, when bracken on the ground was at a minimum; often such stands were recorded as the underlying *Acid grassland*.
- The coverage of the **Mountain, heath and bog** Aggregate class is the same by field and satellite surveys. However, at BH level, the components *Dwarf shrub heath* and *Bogs* are mapped differently. LCM2000’s distinction of **Bogs** using a peatland map gave a conservative estimate of the class. The *Montane* BH was defined by altitude criteria on LCM2000 which recorded more of this class than the field survey estimate. The LCM2000 and field surveys recorded the *Fen, marsh, and swamp* BH differently, mainly through differences in the treatment of rush-pastures.
- **Water (inland)** on LCM2000 is an aggregation of the *Standing open water and canals* and *Rivers and streams* BHs. There was no attempt to distinguish standing from flowing water but the overall coverage is similar to field estimates.
- The *Built up and gardens* BH is mapped in more detail by LCM2000. The field survey treated urban land as continuous without recording open spaces in the urban zone. LCM2000 recorded open spaces >0.5 ha and distinguished **Suburban and rural development** from **Continuous urban / industrial land**.
- **Coastal habitats** of *Supralittoral rock*, *Supralittoral sediment*, *Littoral rock* and *Littoral sediment* were recognised at BH level in LCM2000 Subclasses but shown on maps as the combined classes **Supralittoral rock and sediment** and **Littoral rock and sediment**. Differences between field and LCM2000 estimates relate the tidal state at the time of survey and, in part, to LCM2000’s greater geographical offshore coverage.
- The *Boundary and linear features* BH was not targeted by LCM2000. LCM2000 only includes linear habitats which have an area >0.5 ha: to have been resolved by the images they will also have been =2 pixels wide.

Calibration against field data

- Calibration of LCM2000 against field survey allows the generation of BH cover-statistics equivalent to those from sample-based field survey but benefiting from the comprehensive coverage of LCM2000. The resultant statistics are probably CS2000’s best current estimates for BH coverage at regional and national scales. The mean values compare very closely with field survey estimates, but with confidence limits which are much tighter (about half the field survey range).
- A weakness of the field survey has been its inability to make reliable estimates of cover at local scales. At Target class level, LCM2000 is likely, in general, to be 85% correct in its mapping, but with the potential for significant local errors. Local results are likely to be mostly reliable for Aggregate classes. Further work to provide and assess calibrated BH estimates at local levels is planned.

UK land cover

- According to LCM2000, more than half the UK is used for intensive agriculture or is developed. The remainder is largely semi-natural. Woodlands occupy a quarter of the semi-natural land, with **Broadleaved woodland** and **Coniferous woodland** about equal in extent. **Mountain, heath and bog** cover a third of the UK's low intensity land; semi-natural grass swards (including rougher examples of improved swards) form over a third of all semi-natural cover. **Coastal** habitats and **Open water**, while important, are small in extent.
- The four countries of the UK differ markedly. Intensive uses affect nearly three-quarters of England, about two-thirds of Northern Ireland and about half of Wales; in Scotland, less than a quarter is intensively farmed or developed.
- The semi-natural land of England is evenly split between woodlands and grasslands. In Wales, the balance is similar within a far greater extent. In Scotland, **Mountain, heath and bog** make up more than half of all the semi-natural land. Northern Ireland also has reasonably extensive **Mountain, heath and bog** and **Semi-natural grass** but, at the resolution of LCM2000, is notably short of woodland cover.

Land cover change

- The measurement of changes in land cover demands high levels of precision to map real differences and to distinguish them from survey errors and generalisations. In a comprehensive national survey, the necessary precision for change detection cannot be achieved consistently by satellite-based mapping alone. The LCM2000 classification sought to remove known deficiencies in the 1990 classification and to bring field and satellite surveys into closer match. The segment-based approach of LCM2000 generated different results from the 1990 raster product. These differences preclude direct comparisons with the 1990 product. It may be possible to select intelligently, from those differences mapped, the elements which are attributable to change and those attributable to error and / or differences in the data products. This approach will be the subject of research and development, beyond the scope of the production phase.

Conclusions

- In conclusion, LCM2000 has, for the first time, mapped the land cover of the whole of the UK from satellite images. The resultant vector data record the 'real' structure of the landscape and thus can satisfy wide ranging user-needs. LCM2000 offers so much more scope than the conventional per-pixel products of the earlier mapping. It has a detailed spatial resolution which is far better than other vector-based maps of full UK land cover. LCM2000's structural picture of the landscape shows the spatial inter-relations of parcels and habitats. It therefore lends itself much better to applied uses where patterns affect processes and determine their consequences.

1. INTRODUCTION

Land Cover Map 2000 (LCM2000)[2] is a part of Countryside Survey 2000 (CS2000 - Haines-Young *et al.* 2000). It provides a comprehensive survey of the countryside of the United Kingdom in the form of digital maps and databases plus a range of derived products held in a geographical information system (GIS). LCM2000 updates and upgrades the Land Cover Map of Great Britain (LCMGB), made in 1990-92 (Fuller *et al*., 1994a). Refinements include:

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

LCM2000 was funded by a consortium which also formed a steering group to the project:

- The Countryside Council for Wales;
- The Department for the Environment, Food and Rural Affairs;
- The Department of Agriculture and Rural Development (Northern Ireland);
- The Environment Agency;
- The Environment and Heritage Service, Department of the Environment (Northern Ireland);
- The National Assembly for Wales;
- The Natural Environment Research Council;
- Scottish Natural Heritage;
- The Scottish Executive.

The members of the Consortium have policy and / or operational remits which require, for their implementation, sound information on the status and trends in natural resources. LCM2000 provides comprehensive information on land cover and of widespread examples of terrestrial, freshwater and coastal Broad Habitats in the UK; it shows their patterns, inter-relations and environmental contexts. Such information helps users to take stock, develop their understanding of environmental processes, predict environmental impacts, model change, plan responses, devise management strategies and monitor their successes in operation.

2. AIMS

The LCM2000 project aimed to:

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

[2] A 'Glossary of terms and acronyms' appears in section 19, page 28

3. REFERENCE DATA FOR MAP PRODUCTION

3.1 Images

LCM2000 is made from combined summer and winter satellite image data recorded by, in order of preference, the Landsat Enhanced Thematic Mapper (ETM), the Thematic Mapper (TM), or the Indian Research Satellite LISS sensor. The target period for summer imagery was the main growing season for arable crops, from mid-May to late July, or rather later in Scotland excluding May but continuing into August. The target 'winter' period started at the time of the first frosts in Autumn 1997 (about October), and extended to late April 1998 in southern Britain and even into May in the Scottish Highlands (i.e. until deciduous trees were in full leaf). Compromises were allowed where these dates could not be met.

Under ideal conditions, it takes 46 satellite scenes to cover the UK. Just 49 were used for GB in LCMGB 1990 (Fuller, *et al.* 1994b). LCM2000 required 79 scenes (Figure 1 and Appendix 1) to complete coverage for the UK, due to the high incidence of localised cloud on all but one of the scenes used. The 79 images were combined as summer-winter composites, as far as possible, giving 78 'combinations' (compared with 32 combinations for Britain in LCMGB 1990 (Fuller *et al.* 1994b)). In all, 84% of the UK was mapped from summer-winter composite images (Figure 2), with 9% from summer-only data and 6% winter-only coverage (this compares with 87% summer-winter, 8% summer-only and 3% winter-only in Britain in LCMGB 1990).

Only 23% of the UK was mapped from images recorded in the target summer and winter periods of 1998 (compared with 58% of GB in 1990); no other 1998 summer-winter composite images were useable, even those outside the target seasons. However, for a further 22% of the UK, 1998 data were combined with images from other years. Hence, 45% the UK was mapped with a contribution from 1998 data. The remainder of the UK, 54% in all, used data entirely from other years. In 1990, 3% of GB remained unclassified due to cloud-cover. LCM2000 overcame similar problems using LCMGB data, upgraded to LCM2000 standards (i.e. built as segments >0.5 ha with LCM2000 cover types) to infill cloud remnants: 1% of UK coverage came from this source. The result gave complete UK cover.

3.2 External data

Ordnance Survey and Ordnance Survey Northern Ireland paper maps at 1:50 000 scale were used as a basis for geo-registration of images to the British National Grid. A digital terrain model (DTM) was used in pre-processing of satellite image data to model and compensate for the effects differential illumination due to undulating terrain. It also provided contextual data to aid classification.

An acid-sensitivity map (Hornung *et al.* 1995) was used to label semi-natural grasslands as 'acid', 'neutral' or 'calcareous'; this map defined acidity sensitivity classes as highly sensitive - pH <4.5 (i.e. truly acid), moderately sensitive - pH >4.5 and <5.5. (treated for these purposes as neutral but really slightly acid) and low sensitivity - pH >5.5 (really with neutral and calcareous components). A peatland mask, based on British Geological Survey drift mapping[3], was used to define 'heaths' and 'bogs'. CORINE data (based on LCMGB 1990 but developed with mid-1990s OS mapping) were used in some knowledge-based corrections of map outputs.

It should be realised that these external datasets were all collected with different resolutions, with different levels of generalisation, and their own potential inaccuracies, so the final accuracy of LCM2000 depends in part on their reliability (Smith & Fuller, in press).

---

[3] http://www.bgs.ac.uk/products/digitalmaps/digmapgb_drift.html

Figure 1. UK Map showing the mosaic patterns of images used in the construction of LCM2000.

Figure 2. UK Map showing the contributions of summer, summer-winter and winter images to the construction of LCM2000; also those areas not covered by 1998-2001 images and patched using enhanced Land Cover Map of Great Britain 1990 data.

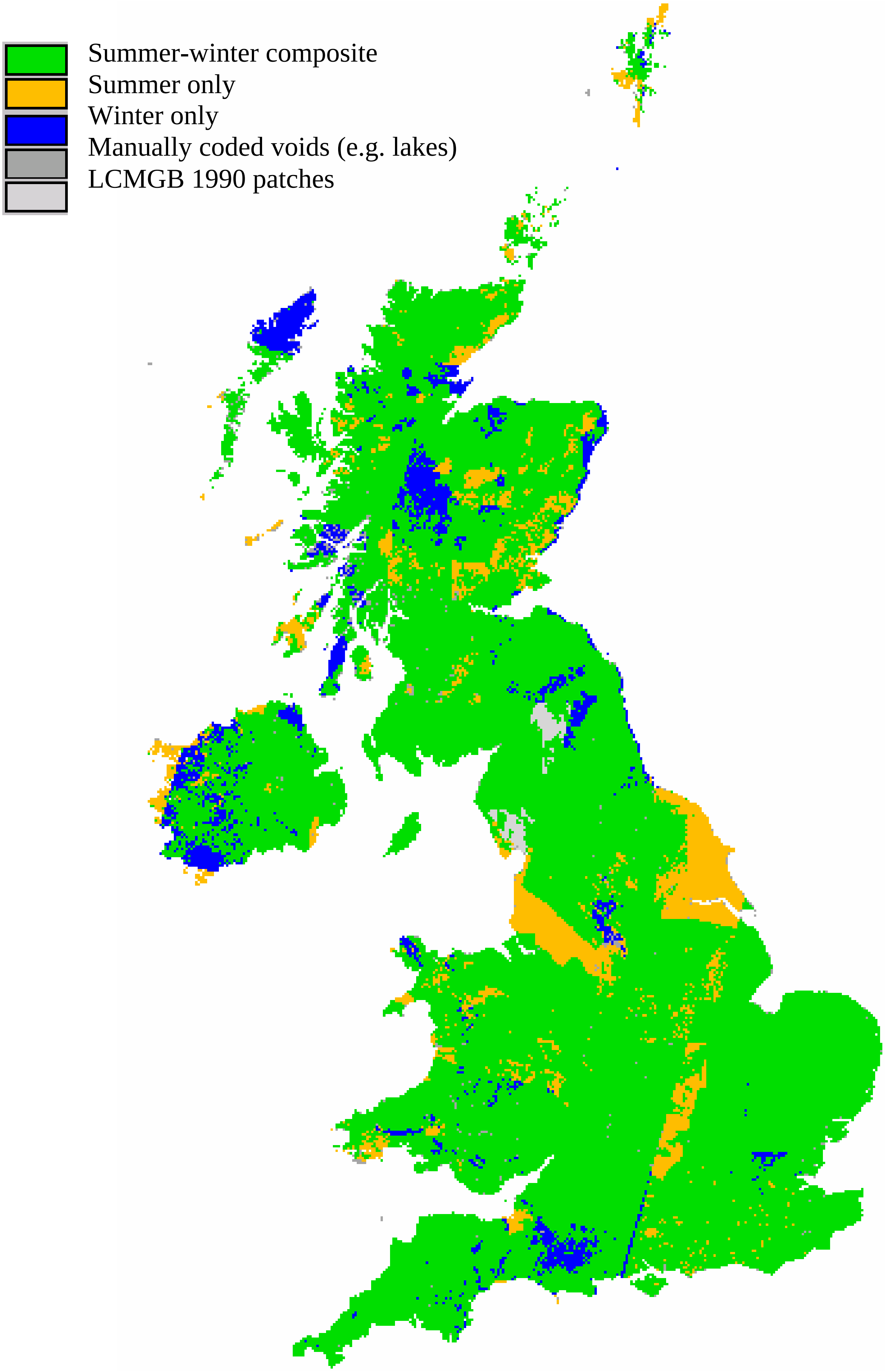


Table 1. Broad Habitats and their relation to LCM2000 Target classes, Subclasses and Variants (the Suclasses are shown in their map display colours).

BH	LCM Target class	LCM Subclasses	Variants	
22. Inshore sublittoral	Sea / Estuary	Sea / Estuary	sea	
13. Standing water/canals	Water (inland)	Water (inland)	water (inland)	
20. Littoral rock	Littoral rock and sediment	Littoral rock	rock, rock with algae	
21. Littoral sediment		Littoral sediment	mud, sand, sand/mud with algae	
		Saltmarsh	saltmarsh, saltmarsh (grazed)	
18. Supra-littoral rock	Supra-littoral rock and sediment	Supra-littoral rock	rock	
19. Supra-littoral sediment		Supra-littoral sediment	shingle, shingle (vegetated), dune, dune shrubs	
12. Bogs	Bogs (deep peat)	Bogs (deep peat)	bog: shrub, grass/shrub, undifferentiated (all on deep peat)	
10. Dwarf shrub heath	Dwarf shrub heath (wet / dry)	Dense dwarf shrub heath / Open dwarf shrub heath	dense ericaceous, gorse	open ericaceous
15. Montane habitats	Montane habitats	Montane habitats	montane	
1. Broad-leaved woodland	Broad-leaved wood	Broad-leaved / mixed woodland	deciduous, mixed, open birch, scrub	
2. Coniferous woodland	Coniferous woodland	Coniferous woodland	conifers, felled, new plantation	
4. Arable & horticultural	Arable and horticultural	Arable cereals	barley, maize, oats, wheat, cereal (spring), cereal (winter),	
		Arable horticulture	arable bare ground, carrots, field beans, horticulture, linseed, potatoes, peas, oilseed rape, sugar beet, mustard, non-cereal (spring), unknown	
		Non-rotational horticulture	orchard, arable grass (ley), setaside (bare), setaside (undifferentiated)	
5. Improved grassland	Improved grassland	Improved grassland	intensive, grass (hay/ silage cut), grazing marsh	
	Neutral / calcareous semi-natural / rough grasslands	Setaside grass	grass setaside	
6. Neutral		Neutral grass	rough grass (unmanaged), grass (neutral / unimproved)	
7. Calcareous		Calcareous grass	calcareous (managed), calcareous (rough)	
8. Acid	Acid grass and bracken	Acid grass	acid, acid (rough), acid with Juncus, acid with *Nardus/Festuca/Molinia*	
9. Bracken		Bracken	bracken	
11. Fen, marsh and swamp	Fen, marsh and swamp	Fen, marsh, swamp	swamp, fen/marsh, fen willow	
17. Built up areas, gardens	Suburban and urban	Suburban/rural developed	suburban/rural developed	
		Continuous Urban	urban residential/commercial, urban industrial	
16. Inland rock	Inland Bare Ground	Inland Bare Ground	despoiled, semi-natural	
20 relevant BHs	16 target classes	26 target/subclasses	72 target/subclasses/variants	

4. LAND COVER AND BROAD HABITATS

4.1 Background

As an aid to the implementation of, and reporting under, the UK Biodiversity Action Plan (BAP), the UK Biodiversity Group identified a framework of '**Broad Habitats**' to encompass the entire range of UK habitats (Table 1 column 1). The descriptions of Broad Habitats (see Appendix II) was developed by the Joint Nature Conservation Committee (JNCC: Jackson, 2000). LCM2000 aimed to contribute to the assessment of habitats by mapping, as far as possible, the widespread examples of terrestrial, freshwater and coastal Broad Habitats. While their mapping was always treated as a key objective, LCM2000 also aimed to record further details where possible, giving cover classes sought by other users.

> Hereafter, Broad Habitats are referred to simply as BHs. In order to distinguish BHs, italic text is used (e.g. the *Coniferous woodland* BH). LCM2000 classes are given in bold text (e.g. LCM2000 **Continuous urban land**). Where an LCM2000 class closely matches a BH class, the same nomenclature is used. (e.g. LCM2000 **Coniferous woodland**). Where the LCM2000 class, while broadly similar, differs in significant respects, the name is designed to reflect that difference (e.g. the LCM2000 class **Broad-leaved / mixed woodland** differs from the BH *Broad-leaved, mixed and yew woodland* in that yew woodland is not sufficiently extensive for consideration in LCM2000. In tabulations and figures, the BH nomenclature is sometimes abbreviated but any reference to a BH is a reference to the original BH class and name.

4.2 Broad Habitats and LCM2000 classes

LCM2000 is a thematic classification of spectral data recorded by satellite images; external datasets add context to help refine the spectral classification. The spectral classes defined in this process (Kershaw & Fuller 1992) can be combined into thematic components which can in turn be aggregated to build various classification schemes (Figure 3). LCM2000 aimed, where possible, to distinguish BHs. In practice, **Target classes** (Table 1, column 2) were considered the nearest match which could be achieved consistently and with a high level of accuracy. **Subclasses** (Table 1, column 3) were then defined to give, as far as possible, the full complement of BHs. Subclasses were mapped consistently throughout the UK, but sometimes with compromises on accuracy. Some BHs were subdivided where this was considered valuable for wider use of data. Thus class **Variants** (Table 1, column 4) are the thematic components of the BHs / Subclasses. They were recognised where possible but not necessarily consistently (e.g. individual crops could not be distinguished once harvested).

In practice, most BHs are readily identified by LCM2000. However, users should be aware of a few key differences between BH definitions and those of equivalent Target classes and Subclasses; differences in nomenclature aim to draw attention to those of definition. In Table 1, a hard line between Target classes or Subclasses shows a distinction which is generally reliable. However, a dotted line identifies situations where the distinction is more difficult. Because some BHs are distinguished using floristic characteristics, particularly the presence (not necessarily the dominance) of indicator species, LCM2000 distinctions may differ from those of field survey. Table 1 shows a mis-match in the 'read-across' between some BH and Target class distinctions.

The *Bogs* BH, for example, is characterised in the field by the presence of peatland indicator plant species; yet, it is often dominated by heathers. LCM2000 distinguishes heather-dominated **Bogs** using a peat map. Where this differs from floristic indications, LCM2000 may record the cover type as **Dwarf shrub heath -** hence the mis-alignment in Table 1 between columns 1 and 2. Note that the mis-match between a BH and a Target class applies in turn to its Subclasses (e.g. the *Bog* BH may have been confused with **Dense** and **Open** components of **Dwarf shrub heath**).

Figure 3. The hierarchical nature of the Land Cover Map 2000 classification system. LCM2000 is made up of 1000s of **Spectral classes**; these come together thematically as 72 class **Variants** of 26 **Subclasses**, the latter mapped consistently throughout the UK. These Subclasses are combined into LCM2000 **Target classes** which simulate the **Broad Habitat** classification, though with some differences. Target classes and Broad Habitats combine unambiguously into 10 **Aggregate classes**.

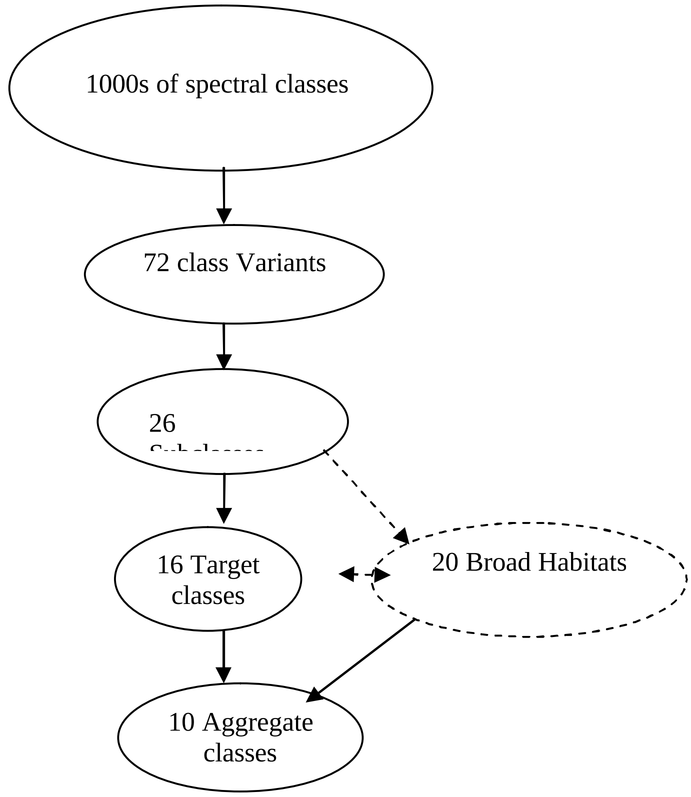


Table 2. Aggregate classes and their relation to Broad Habitats and LCM2000 Target classes.

BH (name abbreviated)	Aggregate classes	LCM2000 Target class
1. Broad-leaved woodland	Broad-leaved / mixed wood	Broad-leaved / mixed wood
2. Coniferous woodland	Coniferous woodland	Coniferous woodland
4. Arable & horticultural	Arable and horticultural	Arable and horticultural
5. Improved grassland	Improved grassland	Improved grassland
6. Neutral grass	Semi-natural / rough grass and bracken	Neutral / calcareous semi-natural / rough grasslands
7. Calcareous grass		
8. Acid grass		
9. Bracken		Acid grass and bracken
11. Fen, marsh and swamp		Fen, marsh and swamp
12. Bogs	Mountain, heath and bog	Bogs (deep peat)
10. Dwarf shrub heath		Dwarf shrub heath
15. Montane habitats		Montane habitats
16. Inland rock		Inland Bare Ground
17. Built up areas, gardens	Built-up & Gardens	Suburban and urban
13. Standing water / canals	Standing Open Water and Canals	Water (inland)
20. Littoral rock	Coastal	Littoral rock and sediment
21. Littoral sediment		
18. Supra-littoral rock		Supra-littoral rock and sediment
19. Supra-littoral sediment		
22. Inshore sublittoral	Sea	Sea / Estuary
20 relevant BHs	10 Aggregate classes	16 target classes

Several problems hinder the distinction of *Improved grassland* from semi-natural swards and the subdivision of the latter into *Acid*, *Neutral* and *Calcareous grassland* BHs. Rough (unmanaged) grasslands present particular problems: they include elements of 'improved' and 'semi-natural' swards which are spectrally indistinguishable. LCM2000 treats all rough grasslands as semi-natural and puts them in a Target class called **Neutral / calcareous semi-natural / rough grasslands**. There is then an attempt to distinguish improved and semi-natural components at the Subclass level, but neither the spectral nor available contextual data can fully match the floristic distinctions. Forced to allocate the rough grasslands to a single BH, LCM2000 first used the **Neutral grassland** category (generally appropriate according to JNCC definitions (Jackson 2000)). Then, because the BH classification distinguishes *Acid*, *Neutral* and *Calcareous grasslands* (again relying upon indicator species in field surveys) LCM2000 had instead to use contextual analyses, drawing upon an Acid sensitivity map (Hornung *et al.* 1995). Thus the original LCM2000 **Neutral grassland** may later have changed to a **Calcareous** or **Acid grassland**, in a calcareous or acid context. Furthermore, because the pH ranges of the Acid sensitivity map were not ideal, contextual analysis may have over-estimated calcareous and neutral components (see Appendix II). Another issue with semi-natural grasslands concerns LCM2000 **Acid grass** which may include some stands of the *Bracken* BH (too sparse or dissected to show clearly on early summer images).

Class Variants (Table 1 column 4, Appendix III) are shown according to their **best fit** with BHs. The 'read-across' in Table 1 shows the actual aggregations used to generate BHs for attribute coding in the GIS and for calibration. LCM2000 **Aggregate classes** (Table 2) combine Target classes and Subclasses to a simplified 10-class level where they compare closely with equivalent BH-aggregations: at this level, maps and statistics largely coincide. Thus Aggregate classes are used for reporting purposes. Further information of the classifications and correspondences is given later.

4.3 Ground reconnaissance data collection

Mapping was dependent upon the use of representative ground reference data as a sample of so-called 'training areas' from which to calculate image spectral reflectance statistics per class, per image waveband, per satellite scene, and per imaging date. These statistics were then used to allocate unknown areas to their most likely class. Ground reconnaissance surveys involved identifying the thematic class associated with each unique 'spectral class' on sample areas of image: for each combination of summer and winter data, the examples had to form a representative sample, offering an adequate number of pixels per segment, for a replication of sites, to allow accurate characterisation of a class's spectral response. Field reconnaissance took a generalised approach over a large area. Surveyors ensured, for each satellite scene, that they included, as far as possible, the full range of class Variants in all their spectral forms. These included sunlit and shaded examples, those based on variable species compositions, phenological variants, variable mosaics of vegetation types, and crops on varying soil backgrounds. In rarer and smaller examples, even a limited training sample sometimes proved impossible to locate. Under such circumstances, these rare class Variants may be absent from sections of map even though they are present somewhere in the locality.

Details were collected in a very few days for each scene using a broad stratification to select examples (as available) of coastal, lowland, marginal and upland terrain, covering arable, pastural and semi-natural types, and focusing attention on oddities within the images (assuming that the commonplace would be picked up routinely in passing). The field reconnaissance survey would ideally have matched the timing of the satellite summer-overpass. However, there were obvious difficulties in that images first had to be recorded, delivered, co-registered and printed before being used to direct the reconnaissance. The solution was either to survey in anticipation of probable imagery, or to survey after imagery, accepting the possibilities of change. LCM2000 adopted both principles. Arable areas were visited in 1998 in anticipation of 1998 imaging. Northern and western Britain, where land cover changes were likely to be fewer, were mostly covered in 1999. Scotland

and Northern Ireland were visited in 2000. Fieldwork required basemaps or images to annotate with cover attributes. Where new images were unavailable, 1990 images from LCMGB were used. They showed, with few exceptions, the field patterns of 1998 and the general zones of semi-natural cover and so were suitable for the recording purpose.

5. PRE-PROCESSING OF IMAGE DATA

Images from TM and ETM offer six different spectral bands, from visible blue to middle infrared (MIR) wavelengths, with a nominal spatial resolution of 28.5 m, and a coarser thermal infrared at 120 m. LISS offers three bands from green to near infrared (NIR) wavelengths with a spatial resolution of 23.5 m and MIR band at 70 m. All such data could have been used in the analyses. However, some bands may confuse rather than inform the process (e.g. those most affected by haze). Moreover, many of the production processes are limited in principle or by design to operate on a smaller bandset. LCM2000 has thus used the red, NIR and MIR bands of the first choice summer and winter images (Fuller & Parsell 1990, Fuller *et al*. 1994a) making 6 bands in total.

A range of pre-processing techniques (described in Appendix IV) was used for image analyses:

- The correction of atmospheric haze effects on images (Liang *et al*. 1997),
- Cloud and shadow masking within scenes,
- Snow masking in winter / spring scenes,
- Geo-registration and resampling to produce 25 m pixels aligned to the British and Irish National Grids,
- Correction of differential illumination due to undulating terrain,
- Resolution enhancement for LISS MIR data.

6. IMAGE SEGMENTATION

LCM2000's segment-based mapping was as extension of the procedures developed for 'CLEVER-Mapping' (the Classification of Environment with Vector and Raster mapping - Smith *et al.* 1997, Smith & Fuller 2001). The approach segmented an image into areas broadly equivalent to land parcels and vegetation patches. Analysis could then avoid mixed edge-pixels and use average spectral responses from within segments to improve thematic identification. The segmentation consisted of two separate stages: first, edge-detection to identify boundary features; second, region growing from seed points selected to avoid edges.

6.1 Band selection

It was only possible to use 3 of the 6 bands for edge-detection and segmentation. Mathematical combinations of bands within the same image were possible (e.g. band ratios and principle components analysis). However, trials showed the use of individual image-bands to give the most consistent results. The following rules applied:

- The summer image contributed two bands - red and infrared (with MIR for TM/ETM or NIR for LISS),
- The winter image contributed one band - NIR (which was generally the brightest),
- For single date cover, red, NIR and MIR data were used, whichever the sensor.

6.2 Edge-detection and segmentation

An edge detector was used to ensure that seedpoints were selected away from parcel-edges. The level of spatial subdivision was controlled by the operator. The aim was to ensure, as far as possible, that no complex (mixed) segments would result. An iterative process involved segmentation, inspection and, possibly, re-analysis with altered edge-image inputs and / or changed region growing / merging thresholds to derive acceptable segmentations. The segmenter proved robust in its tolerance of different edge thresholds, with substantial changes in the threshold needed to induce significant changes in output segments. The fine level of control which could thus be exerted was gratifying, suggesting that the segments represented real entities, not artificial features governed by highly sensitive input parameters.

The level of spectral distinctions between the various cover types varied according to their characteristics: e.g. the distinction between wheat and grass was often subtle, while a water body might have comprised a multitude of spectral variations based on depth, sediments and aquatic vegetation, which are of no consequence for the BH classification. It was necessary to choose a level of segmentation which separated all the Subclasses while avoiding the risk of grossly sub-segmenting entire features e.g. substantially subdividing fields.

6.3 Post-segmentation generalisation and boundary rejection

Segmentation results were simplified using spatial generalisation procedures, eliminating reject pixels and dissolving small polygons:

- Non-segment edge pixels were dissolved into adjoining parcels,
- Single-pixel islands were dissolved into their surroundings,
- Small segments <9 pixels (i.e. ≤0.5 ha) were attached to the neighbouring segment which was most similar spectrally,
- The procedure stores dissolved and aggregated polygons with potential for later analyses.

These steps incorporated miscellaneous pixels into segments, to reduce the final vector dataset to a manageable size. Because the mapping procedure took segment reflectances from core pixels only, it generally took sufficient account of added edge pixels to avoid difficulties. Once an acceptable segmentation of the images was achieved vector versions of the segments were created and the GIS database built. This was a simple procedure of raster-to-vector conversion, where the boundaries between segments were represented by vector lines. Excessive sub-segmentations, i.e. the sub-division of a single field, have not been simplified. This could be undertaken by 'intelligent' vector-generalisation methods, should users see fit. Occasional problems have been found with complex segments, clearly representing two cover types which were not separated by the segmentation routine. This problem could not be fixed for the production process; it probably affected <0.1% of polygons, so it was viewed as a negligible problem, lost alongside those of spectral classification.

An example of a segmented image appears in Figure 4. Further details on band-selection for segmentation, the segmentation process, generalisation and vector-conversion are given in Appendix V.

7. TRAINING THE CLASSIFIER

7.1 Interactive training

'Training' is the procedure by which a sample of known cover types is used to deduce the spectral characteristics of the cover types, for later extrapolation to classify examples of unknown land cover. Field reconnaissance data were used to direct such a process. Areas of known land cover, marked on field reconnaissance maps, were identified as training areas, objectively labelling image segments (instead of subjectively outlining the pixels of a cover type as is usual with pixel-based classifications). The process of training was quicker, so it was easier to build up a representative sample of training areas, each of which contained a large number of pixels. Additional 'check parcels' were defined at the same time for use in a preliminary validation of classification results.

A refinement built into LCM2000 was the opportunity to visualise and review training data as 'colour charts' representing the spectral characteristics of each training area (Figure 5). The operator compared and contrasted training areas of a class Variant, placed them into spectral subclasses, rejected odd examples and selected the finalised training set. The operator was able to review the training areas in any band-combination - summer, winter or as a composite - to ensure that the spectral subclasses were not mixed. These procedures, thought to be entirely new to operational image processing, contributed substantially to the quality improvements in LCM2000. The combination of automated training area delineation plus easy review, editing and sorting into spectral subclasses helped the team deal with the large number of image-composites used to build the UK coverage.

7.2 GIS 'self-training'

The class of a segment on one scene can be transferred to an equivalent segment on an overlapping, unclassified scene. Objective and automated comparisons of the datasets were used to locate near-identical segments and pick up an class label from the 'donor' scene. A segment-overlap >80% was needed to justify transfer. Labels attached with lower probabilities (P < 85%) were rejected. This 'self training' frequently helped to identify additional examples of rarer training types, improving the chances of defining a valid training set for extrapolation. 'Self training' data and original field training data were reviewed simultaneously to define spectral subclasses. This process helped achieve the best possible edge-matching across scene-boundaries.

8. CLASSIFICATION

8.1 Extraction of Subclass statistics

Training areas were used to derive statistical measures of reflectances, in each chosen band and for each spectral subclass. The segment-based approach used a shrinking procedure when extracting raster reflectance data, to avoid edge pixels; it ensured the use of 'pure' core pixels of a cover type. The shrinkage was made a dynamic process whereby the required amount of shrinking (25 m) was applied and, if insufficient raster data were collected, the shrinkage was reduced (by 2.5 m) and the raster extraction repeated. This process continued until enough data were extracted (minimum 4 pixels) or the shrinkage reached zero. The number of pixels extracted and the shrinkage achieved were stored as polygon-attributes in the GIS for future reference. Training areas were, as far as possible, those where 100% shrinkage was achieved.

8.2 Maximum likelihood classification

The classification procedure used a maximum likelihood algorithm (Mather 1997, Schowengerdt 1997) applied per-segment. When each segment was classified, its mean reflectances were calculated from shrunken segments; the shrinkage applied and the number of pixels extracted for classification were stored as attributes. The classification procedure compared the shrunken polygon's mean reflectances with the training set and recorded the most likely spectral subclass in statistical terms: in fact, it stored probabilities for the top five spectral subclasses, usually covering >90% of the probability distribution. Classification was an iterative procedure: each successive classification was visually inspected, the training set was edited as necessary and the classification re-run. Once a 'final' version was achieved, the classification of pre-labelled segments (i.e. training areas and check areas) was scored to check that they were being classified with 90% success. Only then was the per-segment classification passed on to later stages of knowledge-based correction. At this point, per-pixel classifications were also made, using the same training data, to record the natural heterogeneity associated with land parcels (but also the 'noise' known to be associated with per-pixel mapping).

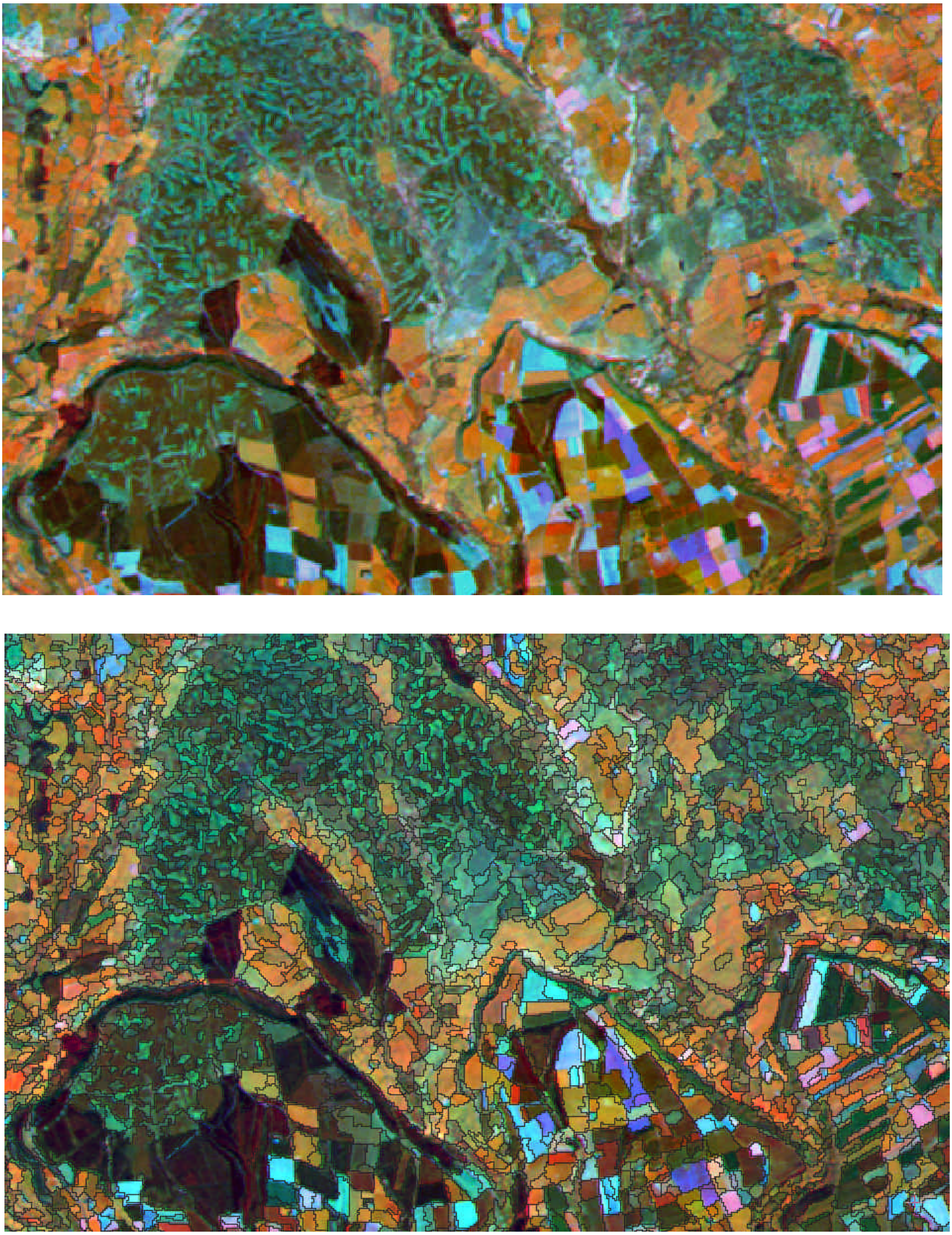


Figure 4. Top, a section (12km x 8 km) of summer-winter composite image covering the North York Moors; below, the same image segmented with vector-outlines to the image-segments overlaid.

Figure 5. An example of the colour charts used to sort training data at the Variant level into spectral subclasses (D - deciduous wood; L - littoral (mud); Ar - arable (rape); Ab - arable (barley); Gi - grass (improved); Ga - grass (acid); H - heather; Us -urban (suburban)). For each training area (e.g. D11), the display shows four colour blocks signifying: top, the mean spectral response in the summer image; second, the spectral response of the pixel most different from the mean in the summer image; third, the mean spectral response in the winter image; fourth, the spectral response of the pixel most different from the mean in the winter image; it also records the number of pixels used in calculating the mean spectral response. The training areas were selected according spectral homogeneity and sample size, and grouped according to spectral similarity. This approach also allowed the identification of problems within the training data. In the red box there is confusion between arable barley and improved grassland which may be due to a mis-identification during field reconnaissance. In the purple box is an area probably of acid grassland which has been labelled as heather, maybe a labelling error when transferring field annotations to the GIS.

8.3 Knowledge-based correction

Knowledge-based correction (KBC) procedures (Groom & Fuller 1996) were used to identify and re-label segments which were classified with low confidence and / or with classes out of their natural context. Some KBCs were applied universally. Others were general rules, used where needed but ignored where spectral classification had been successful. Some KBCs were one-off corrections: specific to a particular land cover type, on a particular image, and perhaps in just a small part of the landscape. Interactive KBC was sometimes used where generalised contextual rules could not be applied. Because of this variety of approaches, the following descriptions give examples rather than a complete picture. All KBC alterations are coded into the attribute data, attached to the relevant segments in the GIS database. The codes are cross-referenced to a tabular record giving the exact details of a correction. This information is available to users.

A rule applied consistently across the entire classification process concerned class-probabilities. Where any maximum likelihood class was allocated with a probability <50%, the other spectral subclass probabilities were summed to see if another BH was more appropriate (i.e. its spectral subclasses cumulatively took a dominant percentage of the overall probability). If so, the lead Subclass/Variant of the BH (i.e. usually the second choice class) took precedence. As an example, if a *Built up* subclass was allocated with 35% probability but *Arable* 'wheat', 'barley', 'oats' and 'linseed' had respectively 30%, 20%, 10% and 5% probabilities, the *Arable* BH was deemed more likely to apply, with 'wheat' as the most likely Variant.

Internal context was also used: arable land, surrounded by dominant urban cover, was corrected to the more likely **Suburban** Subclass; this rule was applied widely but had no relevance in semi-natural landscapes. Bare ground in the context of coniferous plantations, evidently felled, was recorded as such; this rule applied where felled plantations were commonplace. External context masks were also used to correct classes similarly 'out of context'. Coastal masking was used to exclude inland 'littoral' habitats and shoreline 'terrestrial' types (Groom & Fuller 1996); this rule was applied to any scenes with a coastline. The DTM helped to identify erroneous areas of lowland *Fen, marsh and swamp* above the floodplain or 'urban' areas, really bare ground, in high upland; such rules were only applied as necessary and then tailored to specific circumstances. An altitude of 600 m (see Ratcliffe & Thompson, 1988) was set for the *Montane habitats* class and applied to vegetation above this threshold; this rule was used consistently. The CORINE data were used to identify land classified as 'scrub' and 'grassland' which was in reality most likely to be orchard) i.e. the *Arable and horticulture* BH (with 'orchard' recorded as the class Variant); this rule only had real relevance in areas with a significant incidence of orchards. General manual recodes were also made where known inaccuracies could not be dealt with using external or internal context. The process was effective because the operator could define a region in which to apply the correction, use the attribute data to select the relevant polygons, then apply the correction. The procedure was a 'one-off', tailored to local circumstances: however, correction of arable fields mis-classified as urban / suburban land was a commonplace form of manual recoding.

KBC rules operated sequentially. As a consequence, a polygon might have changed class more than once: *Littoral sediment* in a terrestrial context might first have changed to *Built up*; then, because it was recorded at 400 m altitude, it was deemed more likely to be *Inland rock*. All contextual corrections were recorded in the GIS database, including the input and output classes of any changed polygons, and the rule(s) used in their alteration. Multiple corrections were recorded in sequence.

Table 3. GIS attributes available for classification and post-classification analyses

Unique identifier for every segment
Modal aspect
Mean elevation
Modal slope
Codes for ancillary information e.g. terrestrial or maritime
A priori land cover type (subclass level) used for training
Information on roll over and cropping of segment
Identifying code for segments used for training
Number of pixels in the segment
Number of pixels in the shrunken segment at classification
Percentage shrinkage achieved during classification
Top class returned by MLC
Probability value of top class returned by maximum likelihood classification
Top 5 classes and probabilities returned by maximum likelihood classification
Per-pixel maximum likelihood classification, top 5 classes and class-fractions in segment
Class returned by stage-1 KBC
Record of stage-1 KBC operations
Final land cover type prior to stage-2 (final) KBC
Record of stage-2 (final) KBC operations
Widespread Broad Habitat code (e.g. 2.3.4)
Hierarchical class code (WBHclass.subclass.variant e.g. Gss)
Generalised description of operations used to classify the segment
Top 5 classes and fractions of LCMGB 1990 within segment
Majority class from Land Cover Map 1990 within segment
Majority class from Corine Land Cover 1990 in LCM2000 segment

Table 4. Broad Habitats and their relation to LCM2000 display classes and approximate colours

BH	**LCM2000 display classes**
22. Inshore sublittoral	Sea / Estuary
13. Standing water/canals	Water (inland)
20. Littoral rock	Littoral rock and bare sediment
21. Littoral sediment	
	Saltmarsh
18. Supra-littoral rock	Supra-littoral rock and sediment
19. Supra-littoral sediment	
12. Bogs	Bogs
	Dense dwarf shrub heath Open dwarf shrub heath
10. Dwarf shrub heath	
15. Montane habitats	Montane habitats
1. Broad-leaved woodland	Broad-leaved / mixed woodland
2. Coniferous woodland	Coniferous woodland
4. Arable & horticultural	Arable and horticultural land
5. Improved grassland	Improved grassland
	Neutral grass (grazed / rough)
6. Neutral	
7. Calcareous	Calcareous grass (grazed / rough)
8. Acid	Acid grass (grazed / rough)
9. Bracken	Bracken
11. Fen, marsh and swamp	Fen, marsh and swamp
17. Built up areas, gardens	Suburban / rural developed
	Urban and industrial areas
16. Inland rock	Inland rock and bare ground
20 relevant BHs	21 display classes

03/01/2002

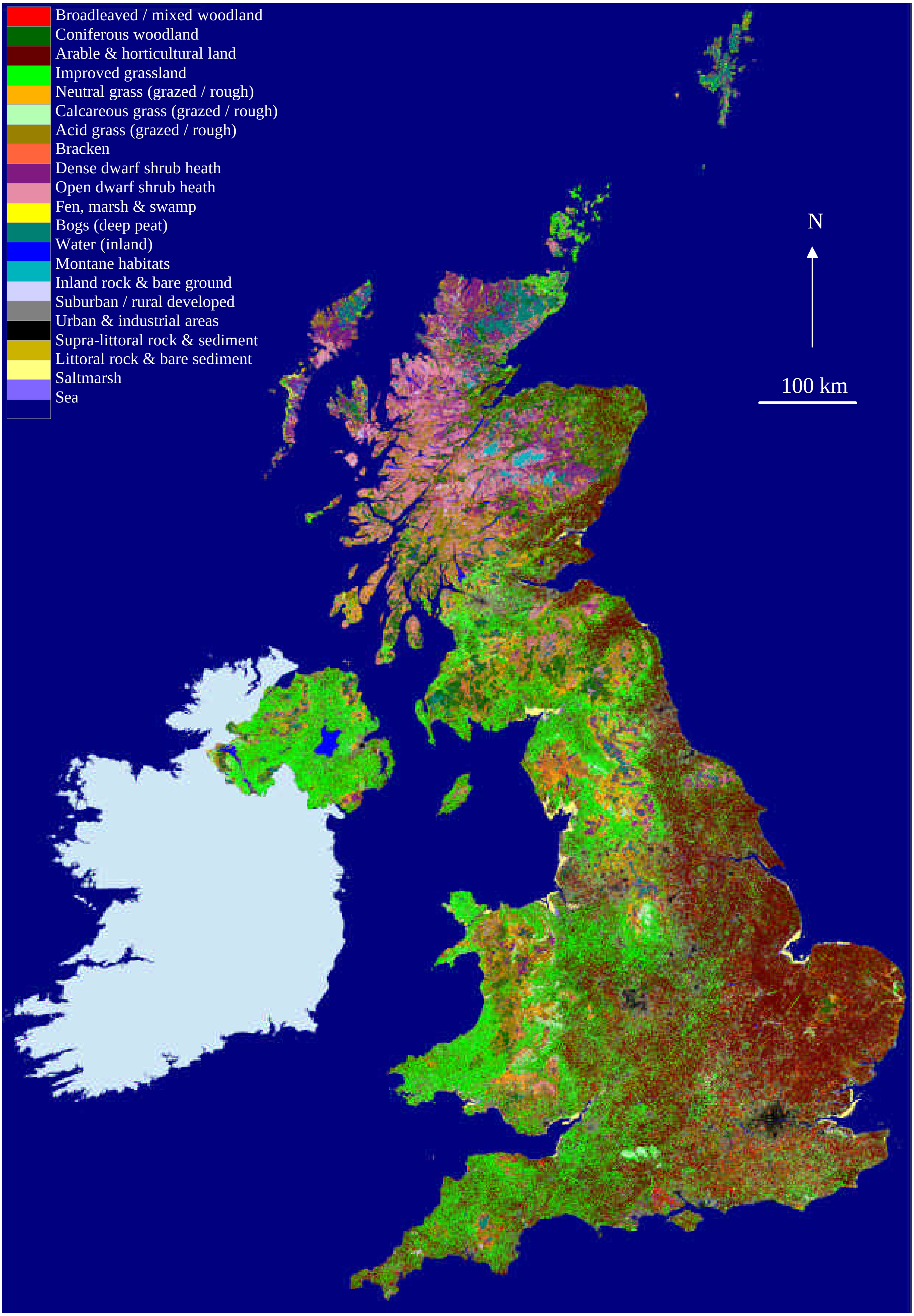


Figure 6. An overview of Land Cover Map 2000. This key also applies to Figures 7-10.

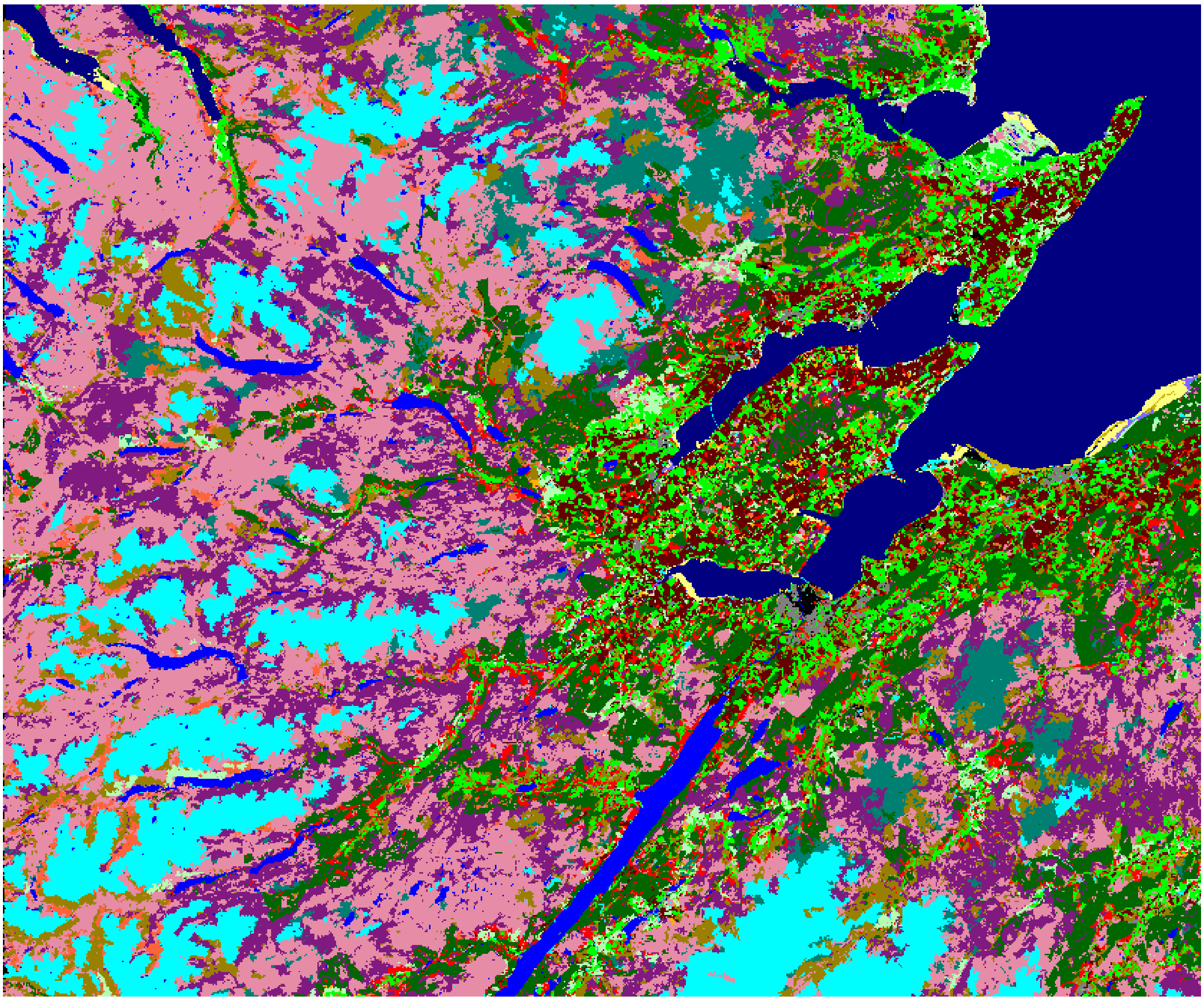


Figure 7. Above: a 100 km x 90 km area of Land Cover Map 2000 with Inverness and the Moray Firth to the east of centre; below left, a key to this and other map sections.

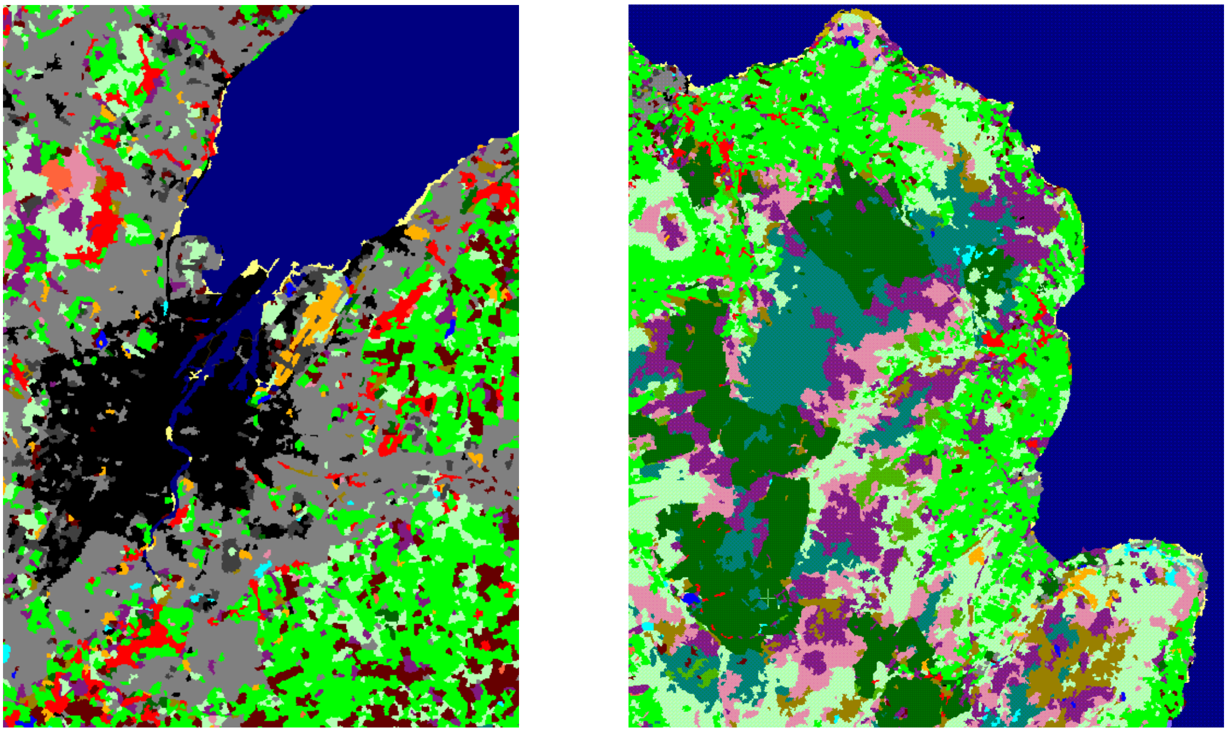


Figure 8. Land Cover Map 2000: a. 15km x 20km centred on Belfast and b. 20km x 25km of the Northern Antrim plateau.

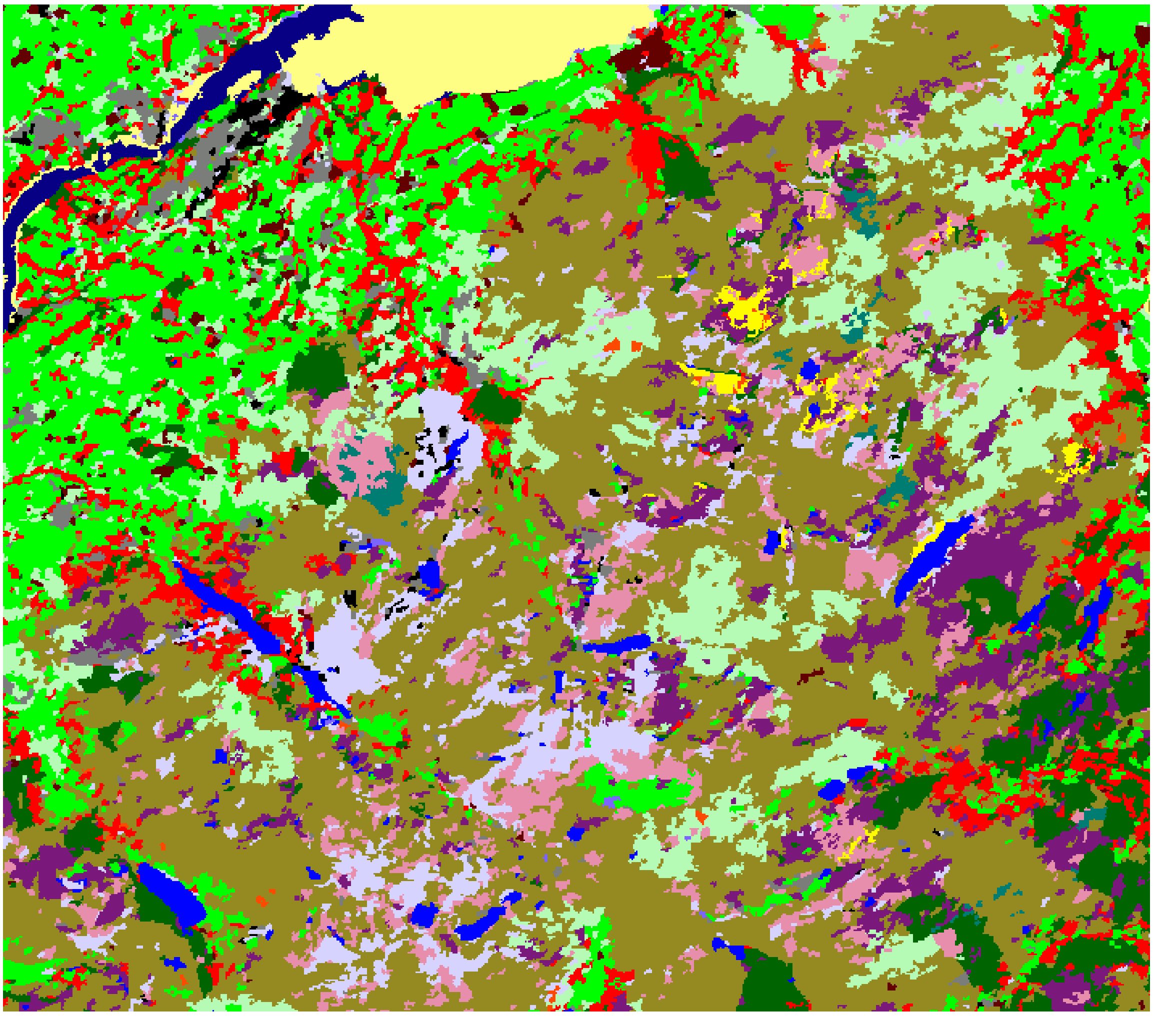


Figure 9. A 20km x 20km area covering Bangor and Snowdon.

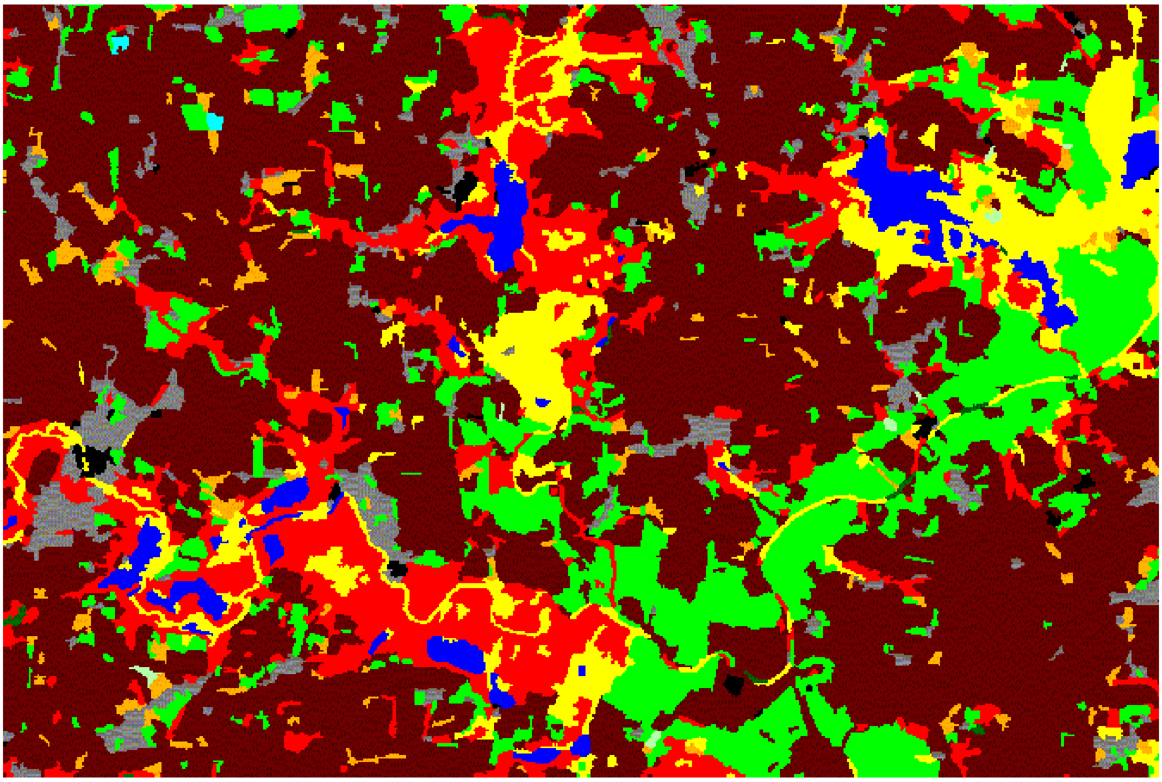


Figure 10. The Norfolk Broads from Wroxham to Hickling (16 km x 10 km).

9. MAP OUTPUTS

9.1 Building 100 km squares

Construction of the full UK database required the individual classified satellite scenes to be mosaicked together. Residual cloud-holes were patched using the best available substitute cover (e.g. single date classifications). Mosaics were built for 100 km sections. The joining of classified scenes was not a simple task. With vector data, the segments along the edge of one scene had to be made to butt-join neighbouring segments derived from another scene; yet their exact outlines may have differed, especially as seasons, sensors and years differed across the joins. When creating a 100 km square of LCM2000, all the available segments were loaded into a single dataset. A hierarchy was then created, based on classification quality and / or the best match to the target summer 1998 date. The dominant set of segments was then made to 'erode' the secondary set, deleting segments that were completely overlapped and cropping segments that were partially overlapped. The erosion procedure was applied through the hierarchy until a single set of segments remained, without overlaps, for the entire 100 km square. The cropped segments all carry a flag in the GIS database. Figure 1 reveals the complexity of the UK mosaic.

After completing each 100 km section, a final phase of contextual analysis was applied. The acid-sensitivity map was used to label semi-natural grasslands as probably 'acid', 'neutral' or 'calcareous'. The peatland mask was used to re-label grass 'moor' and 'heaths' on a peat substratum as 'bogs' and to 'correct' 'bogs' with no underlying peat to 'moors' and 'heaths', as appropriate. When 100 km squares were completed, the final attributes were assembled (Table 3). All the data used in the processing chain are retained for onward analyses.

9.2 Map display classes

Map displays use cartographic conventions which balance the reliability of mapping and the importance and extent of a class whilst bringing out important patterns in the landscape. Table 4 shows the colours used on hard-copy maps; it compares the map nomenclature with the BHs. Display formats were designed for national or regional plots and avoided the distinction of the rarest or most dissected classes which would be obscured at those resolutions. Thus, *Supra-littoral* BHs, mostly small in extent, were aggregated to **Supralittoral rock and sediment**. Because the mapping of BHs is not exact, some separate BHs are aggregated thematically in LCM2000 plots: the spectrally similar *Littoral rock* and *Littoral sediment* BHs were aggregated. Where there is not a direct match with the BH-classification, **Map display classes** may have been mixed at BH level; the **Neutral grass (grazed / rough)** category of LCM2000 includes setaside and other derelict grasslands, some of them 'improved'. Map display classes are essentially the Target classes, but some Subclasses are shown below the BH level where they are deemed widely useful and accurately distinguished. The **Saltmarsh** Subclass is shown specifically. The *Built up areas and gardens* BH is subdivided into important **Suburban / rural developed** and **Continuous urban** components. Dense and open Subclasses of **Dwarf shrub heath** are shown separately as they bring out very distinctive patterns (e.g. of muir burning) in what would otherwise be extensive tracts of seemingly uniform landscape. The mapping closely matches the widely familiar colour-scheme adopted for LCMGB 1990; the exceptions are the introduction of new colours for *Supra-littoral* classes and the distinction of semi-natural grasslands from *Improved grasslands.* While GIS-displays generally adopt the same colouring, detail is available at Subclass and Variant levels and any user-defined colour scheme can be applied.

Figure 6 shows the resulting land cover map as a generalised overview. Figures 7-10 show selected details. Further examples appear on the website (http://www.ceh.ac.uk/data/lcm/LCM2000.shtm). LCM2000 gives us a comprehensive picture of UK Broad Habitats, their extents and distributions.

At regional scales, the landscape patterns become apparent, in particular the inter-relations of BHs within the landscape. At the level of the individual county, not only are these patterns clear; so too is the continuity and fragmentation of the various BHs. As we focus from regional to local levels the landscape structure becomes much clearer. In unenclosed landscapes, the mosaics of semi-natural habitats are a key characteristic recorded by LCM2000. Cities and towns are subdivided into urban and suburban zones with obvious open spaces. At the local level, in enclosed landscapes, individual fields are evident. At full resolution, larger linear features such as rivers and motorways can be seen.

9.3 Data products

Once data are accessed in a GIS, the full scope of LCM2000 becomes clear. Each segment carries a range of attribute data that describe its shape, size and location; the source images and their dates are referenced. Thematic details include the BH, Target class, Subclasses and class-Variant (spanning 72 cover types); class probabilities are recorded. Other details include pixel-based scores of within-parcel heterogeneity and LCMGB and CORINE 1990 classes. A range of ancillary data can be included, for example, terrain, soils and geology. Analyses produce a detailed picture of the UK BHs, their patterns, inter-relations and environmental contexts, at a range of scales.

The 'Level-2 Dataset' comprises ARC/View 'Shape Files' which record a topologically structured vector dataset carrying the following attributes:

- A unique segment label;
- A total pixel count;
- A count of core pixels;
- A 'level-2' classification giving Target classes / Subclasses (26 types) coded to relate to the BH;
- A process history descriptor: scene-identifier, probability indices and KBC rules applied.

The 'Level-3 Dataset' for specialist use comprises data as above but with the following extra details:

- A 'level-3' classification giving Target classes / Subclasses / Variants (72 types overall) coded to relate to the BH;
- The top five per-pixel classes and their fractional cover within the segment.

A 'Raster Dataset' was derived from the Subclasses of the Level-2 Dataset to give a 25 m grid-based map incorporating the spatial refinements of LCM2000 (i.e. based on per-segment classification, and generalised accordingly).

The Raster Dataset was generalised to give 1 km 'Summary Products':

- Class dominance at the 26 Subclass level, stored as a single layer dataset,
- Summary % cover per 1 km^2, for the 26 Subclass types (i.e. a 26-layer dataset),
- The Class dominance dataset as a CIS 'census' file (http://www.cis-web.org.uk/)
- The Subclass dataset as a CIS 'census' file,
- An Aggregate class dataset as a CIS 'census' file.

9.4 Use of the data

LCM2000 is primarily a vector GIS database. It offers users a basic structure which they can use in conjunction with other datasets. The map structure, related directly to real features on the ground, can help our understanding of the environment. It shows the inter-connectivity of landscape features, their immediate context and the wider neighbourhood in which environmental influences operate. The map helps us to see how ecological principles can explain patterns of biodiversity. It may provide spatial data on land uses which influence hydrology. The data underpin climate models; they can also be used to predict the impacts of climate change on landscapes and ecology.

The deposition of pollutants and their environmental consequences can be assessed in context. LCM2000 indicates land uses and consequential values which underpin planning, risk assessment and socio-economic modelling.

LCM2000 also offers a structure upon which users can build. They may edit the data to make corrections, where local knowledge provides better information. They might update the information, recording changing cover through time. Appendix VI makes some recommendations for customising for use and development of the data. Further details on data dissemination are given in Appendix VII.

9.4 Cover statistics

LCM2000 can be used to generate cover statistics for any region, administrative or physiographic, within the UK. It can do so for any aggregation of classes. LCM2000 can generate cover data at Subclass level (Table 5) but these cannot be compared directly with FS-based BH data. At Aggregate class level (Table 5), the uncalibrated statistics from LCM2000 closely match FS data aggregated to the same level. If LCM2000 is used to generate basic (uncalibrated) BH data, then these results can be compared and contrasted with FS-estimates for BH cover. Table 6 gives approximate cover for BHs from uncalibrated LCM2000 data, compared with FS-estimates: statistics are for England, Wales and Scotland, for Britain as a whole and for Northern Ireland. The FS data are presented as the 95% confidence range (i.e. the mean ± 2 standard errors). Despite the often wide range for FS coverage at 95% confidence, only about 40% of LCM2000 results (those in bold) fall within the 95-percentile range. Some other values are close. However, some class estimates (*Neutral Grassland, Calcareous Grassland, Bracken*) disagree with FS results and others show gross differences in some parts of Britain (*Bogs* and *Dwarf shrub heath* in Scotland).

There is scope for substantial refinement through calibration of LCM2000 data. The FS uses the species and contextual characteristics of BHs to ensure generally that BHs are recognised to defined standards (Jackson 2000). However, the sample-based FS can only estimate cover statistics within fairly broad ranges (±30% as an average confidence limit for GB statistics but with a range up to ±89%); the FS records much more variable confidence limits (averaging ±40-54%) for individual countries (with many statistics showing ranges close to ±100%). The comprehensive coverage of LCM2000, if correctly calibrated, may generate the best estimates we can currently make for the cover of BHs in GB and its constituent countries.

10. CALIBRATION

10.1 Introduction

CS2000 FS data provided information to assess the quality of LCM2000. There were two basic objectives:

- To measure correspondences to get a broad picture of LCM2000 map-accuracy;
- To calibrate LCM2000 to the FS, to allow the generation of BH cover-statistics, equivalent to those of FS, from the comprehensive coverage of LCM2000 data.

The FS in Britain examined 569 one-kilometre squares, 519 of them in 1998, the others in 1999; it recorded much greater detail than LCM2000 (Barr, 1998). A separate survey was conducted in Northern Ireland but the data are not yet available in a digital format suitable for testing here. Surveyors recorded great detail: the data took the form of a ‘primary’ coding of cover types, together with a range of ‘secondary’ codes giving qualifying information. These data contributed, through objective evaluations of codes and attributes, to the generation of BH mapping from the basic field data.

Table 5. The coverage (km^2) of Subclasses from Land Cover Map 2000 (LCM2000) and of Aggregate classes (bold) from LCM2000 and field survey (FS) in the UK and constituent countries.

LCM2000 Subclasses	**England**	**Wales**	**England & Wales**		**Scotland**		**NI**		**UK**	
	LCM2000	LCM2000	LCM2000	FS	LCM2000	FS	LCM2000	FS	LCM2000	FS
Broad-leaved / mixed woodland	**10963**	**1609**	**12572**	**11710**	**2687**	**3000**	**341**	**510**	**15600**	**15220**
Coniferous woodland	**2990**	**1435**	**4425**	**3800**	**8454**	**9930**	**660**	**610**	**13540**	**14350**
Arable cereals	19873	181	20054		1799		0		21853	
Arable horticulture	27698	849	28547		5136		992		34675	
Non-rotational horticulture	695	0	695		408		0		1103	
Arable & Horticultural	**48266**	**1030**	**49296**	**46090**	**7342**	**6390**	**992**	**590**	**57630**	**53070**
Improved grassland	30183	7720	37903		10321		9049		57272	
Setaside grass	1777	18	1796		5		0		1801	
Improved grassland	**31960**	**7738**	**39699**	**44310**	**10326**	**10510**	**9049**	**5680**	**59073**	**60500**
Neutral grass	5008	1347	6355		4415		1742		12512	
Calcareous grass	7884	1470	9354		1293		460		11107	
Acid grass	2787	3188	5975		8508		1310		15793	
Bracken	706	294	999		893		17		1909	
Fen, marsh, swamp	180	16	196		1		80		278	
Semi-natural grass and bracken	**16564**	**6315**	**22879**	**15120**	**15110**	**14460**	**3610**	**3400**	**41600**	**33000**
Dense dwarf shrub heath	1331	580	1911		5163		453		7527	
Open dwarf shrub heath	1317	549	1867		16847		782		19496	
Bog (deep peat)	1056	58	1114		4020		523		5657	
Montane habitats	0	0	0		3971		0		3971	
Inland Bare Ground	1112	265	1377		785		78		2241	
Mountain, heath & bog	**4817**	**1453**	**6269**	**7722**	**30786**	**31260**	**1837**	**1670**	**38892**	**39230**
Water (inland)	**581**	**94**	**675**	**1060**	**1420**	**850**	**677**	**n/a**	**2771**	**n/a**
Suburban/rural developed	9527	689	10216		1169		439		11825	
Continuous Urban	4262	171	4432		314		67		4813	
Built up areas and gardens	**13788**	**860**	**14648**	**11800**	**1483**	**1510**	**506**	**n/a**	**16637**	**n/a**
Supra-littoral rock	0	0	0		19		1		20	
Supra-littoral sediment	85	33	118		48		12		178	
Littoral rock	5	5	10		46		13		68	
Littoral sediment	141	75	216		146		40		402	
Saltmarsh	201	43	244		31		1		276	
Coastal	**433**	**155**	**588**	**743**	**290**	**820**	**67**	**30**	**945**	**1593**
TOTAL	**130362**	**20689**	**151051**	**142355**	**77898**	**78730**	**17739**	**n/a**	**246688**	**n/a**

Footnotes:

LCM2000 statistics from a full count of cover based on a 25 m grid

FS statistics from Haines-Young *et al.* 2000.

n/a - not available

Differences between LCM2000 and FS statistics are explained in the text.

Coastal coverage is variable because of variable tidal states and differences in inclusion/exclusion of offshore inter-tidal areas

Total coverage differs according to the definition of the FS population of 1 km squares; in Northern Ireland some BHs are excluded from survey

Table 6. LCM2000 cover (km^2) for BHs (uncalibrated) for England, Wales, Scotland, Northern Ireland and GB, compared with field survey estimates at 95% confidence (i.e. +/- 2 standard errors).

	England			Wales			Scotland			GB			Northern Ireland		
	LCM 2000	**FS Low**	**FS High**	**LCM 2000**	**FS Low**	**FS High**	**LCM 2000**	**FS Low**	**FS High**	**LCM 2000**	**FS Low**	**FS High**	**LCM 2000**	**FS Low**	**FS High**
Broadleaved, mixed and yew woodland	**10963**	8460	11480	**1609**	1221	2255	**2687**	1980	4020	**15259**	12750	16670	341	418	608
Coniferous woodland	**2990**	1802	4158	**1435**	175	1461	**8454**	7232	12648	**12879**	10672	16808	**660**	461	763
Arable and horticulture	48266	39978	47802	1030	1706	2708	**7342**	4442	8358	**56638**	47944	57036	992	436	735
Improved grassland	31960	33314	39946	**7738**	6392	8960	**10326**	8422	12598	50024	50536	59104	9049	5329	6035
Neutral grassland	5008	2992	4828	1347	382	684	4415	1178	2182	10770	5046	7214	1742	2285	2792
Calcareous grassland	7884	20	680	1470	-3	51	1293	-198	738	10647	72	1228	460	5	14
Acid grassland	2787	2794	4846	3188	882	2420	**8508**	5542	9418	**14483**	10594	15306	1310	207	359
Bracken	706	1016	2304	294	512	1634	893	1092	2228	1892	3258	5522	17	28	61
Dwarf shrub heath	**2649**	2304	4936	**1130**	552	1898	22010	7794	12246	25788	12152	17588	1235	78	173
Fen, marsh and swamp	180	914	2046	16	307	929	1	2266	4474	197	4138	6802	80	445	615
Bog	**1056**	426	1534	58	153	1475	4020	17022	23758	5134	18696	25664	523	1249	1717
Standing open water and canals	**581**	14	1766	**94**	42	284	**1420**	218	1482	**2095**	790	3010	677		
Montane habitats	**0**	-16	16	**0**	0	0	3971	-138	1098	3971	-128	1108	0		
Inland rock	1112	46	194	265	13	85	785	160	600	2163	324	796	78		
Built up areas and gardens	13788	8434	12406	860	971	1789	**1483**	948	2072	16132	11028	15592	506	1114	NA
Supralittoral rock	0	62	218	0	22	88	19	326	814	19	502	1038	1		
Supralittoral sediment	**85**	20	460	**33**	9	131	**48**	20	440	166	180	880	**12**	3	29
Littoral rock	5	344	2016	5	40	318	**46**	-24	64	55	440	2320	13	0	11
Littoral sediment	**342**	252	468	117	35	105	**178**	124	296	**637**	490	790	41		

FS squares formed a stratified random sample, covering 40 strata or 'Land Classes' throughout Britain (Haines-Young *et al.* 2000). National and regional statistics were estimated by extrapolation of the sample observations, using the stratification. Thus, for any measure (e.g. the cover of a BH), a mean and standard error could be calculated for each stratum using the sample squares in that stratum. A national or regional estimate would then use the known extent of each stratum in the area of interest to weight its contribution to the calculation of an extrapolated statistic with confidence limits.

The FS data are not 'ground truth'. An independent quality assurance survey showed 88% repeatability for the primary codes from which, essentially, the BH labels were generated. In addition to the coding uncertainties, the mapping of boundaries in unenclosed uplands proved impossible to achieve with repeatable results; the relative proportions of BHs in squares were recorded similarly by different field surveyors but the exact distributions they recorded showed low correspondence. As a consequence, the FS adopted the 1990 outlines and codes, unless there had been evident changes in which case the maps were updated. Of course, the 1990 data had themselves been subject to the same difficulties as the CS2000 survey. Consequently, discrepancies between FS and LCM2000 in upland mapping cannot be attributed entirely or even predominantly to LCM2000 errors. The consequence of uncertainty in the field data is that the process of inter-comparison was one of 'calibration' rather than 'validation'. Indeed, inter-calibration is more appropriate when comparing two surveys with such different characteristics.

10.2 GIS Operation

ArcInfo coverage files, labelled with BHs, were generated for all 569 FS squares and equivalent LCM2000 map-sections. To accommodate residual errors in the geo-registration of satellite images, comparisons first assessed the need for a shift in x- and / or y-directions, by up to 3 pixels (75 m), when selecting LCM2000 extracts for 1 km squares. The 'true' position for the LCM2000 extract was taken as that with the maximum FS-to-LCM2000 correspondence. Figure 11 shows a plot of the shift-distances recorded in the 569 squares. The largest shifts were generally in uplands, where imaging view-angle and undulating terrain interacted to give displacements which could not be fully corrected when rectifying image-geometry. The overall mean shift-distance was 53 m, with 48% of squares shifted one pixel (25 m) or less in x- and or y-direction, and 62% shifted two pixels or less. There was no significant systematic error. Shifting increased correspondence by an overall average of just 2%. The comparisons operated using the shifted extracts (where appropriate) from LCM2000; Figure 12 shows example squares.

A comparison of FS and LCM2000 generated 569 correspondence matrices, one for each 1 km square. The evaluation included three types of comparison:

- **Per-pixel comparisons** between FS and LCM2000 maps; a direct overlay, with no regard for the structure of either dataset, where results show cumulative differences, i.e. due to: i. the FS's greater original spatial resolution, ii. time differences in surveys, iii. class-definition differences, iv. errors in one or both surveys.
- **Per-segment comparisons**, where labels in LCM2000 segments are compared with the segment's dominant class according to FS: a measure of how well the spectral-classification of a segmented image fared.
- **Per-parcel comparisons**, where FS land parcels and their classes were compared with a class-label for the parcel derived from LCM2000: a measure of how effectively the LCM2000 class labels could be transferred to conventional vector maps.

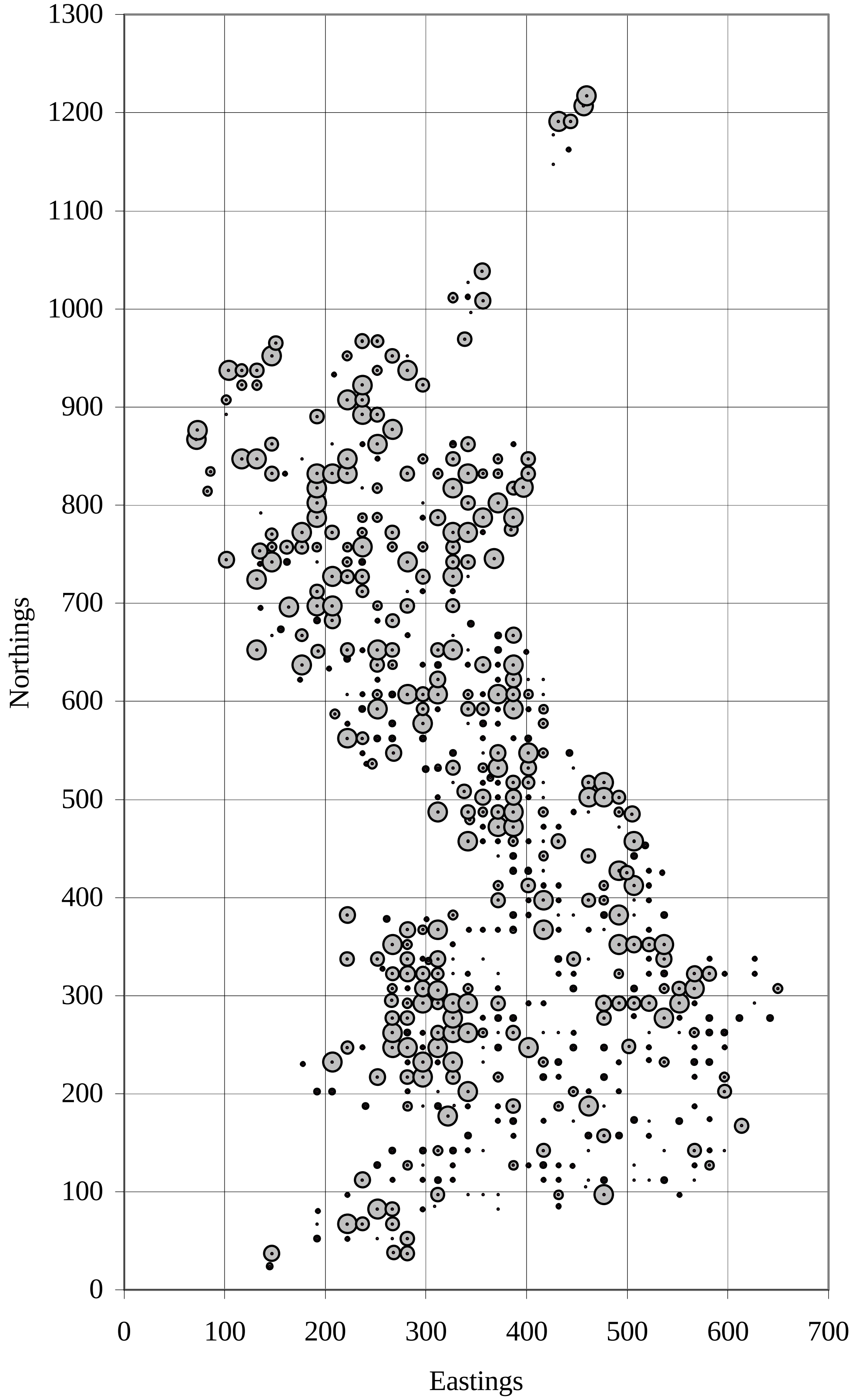


Figure 11. Shift distances, plotted in a Great Britain context, for the extracts from LCM2000 which were used in comparison with the CS2000 field survey squares. Point sizes represent shifts from the smallest with no shift to the largest representing a 3 pixel shift in x- and y-direction. Coordinates are British National Grid, expressed in kilometres.

The analyses, whether applied per-pixel, per-segment or per-parcel, used a raster-GIS approach. FS parcels and LCM2000 segments were sampled onto a grid with a 2.5 m cell-size. Per-pixel scores recorded the 160 000 sample-pixels in each 1 km square. The results for parcels and for segments were aggregated and weighted by area: i.e. an LCM2000 parcel-label was initially scored as 'right' or 'wrong' according to the FS; and the result was given a weight according to the number of 2.5 m pixels in that parcel; scores thus relate directly to the 1 km map area and continue to add in total to 160 000. Figure 13 shows the individual correspondence results for the 569 squares plotted spatially. It confirms the expectation that correspondences in uplands are generally less than those in lowlands.

Correspondences were calculated at various thematic levels:

- BH level but excluding *Boundary and linear features* and *Rivers and streams* (below LCM2000 resolution);
- BH level but generalising LCM2000 urban to match FS results (i.e. accepting that the FS records *Built up areas and gardens* without differentiating open spaces);
- Target class level, and allowing FS generalisation of urban;
- Aggregate class level, allowing FS generalisation of urban.

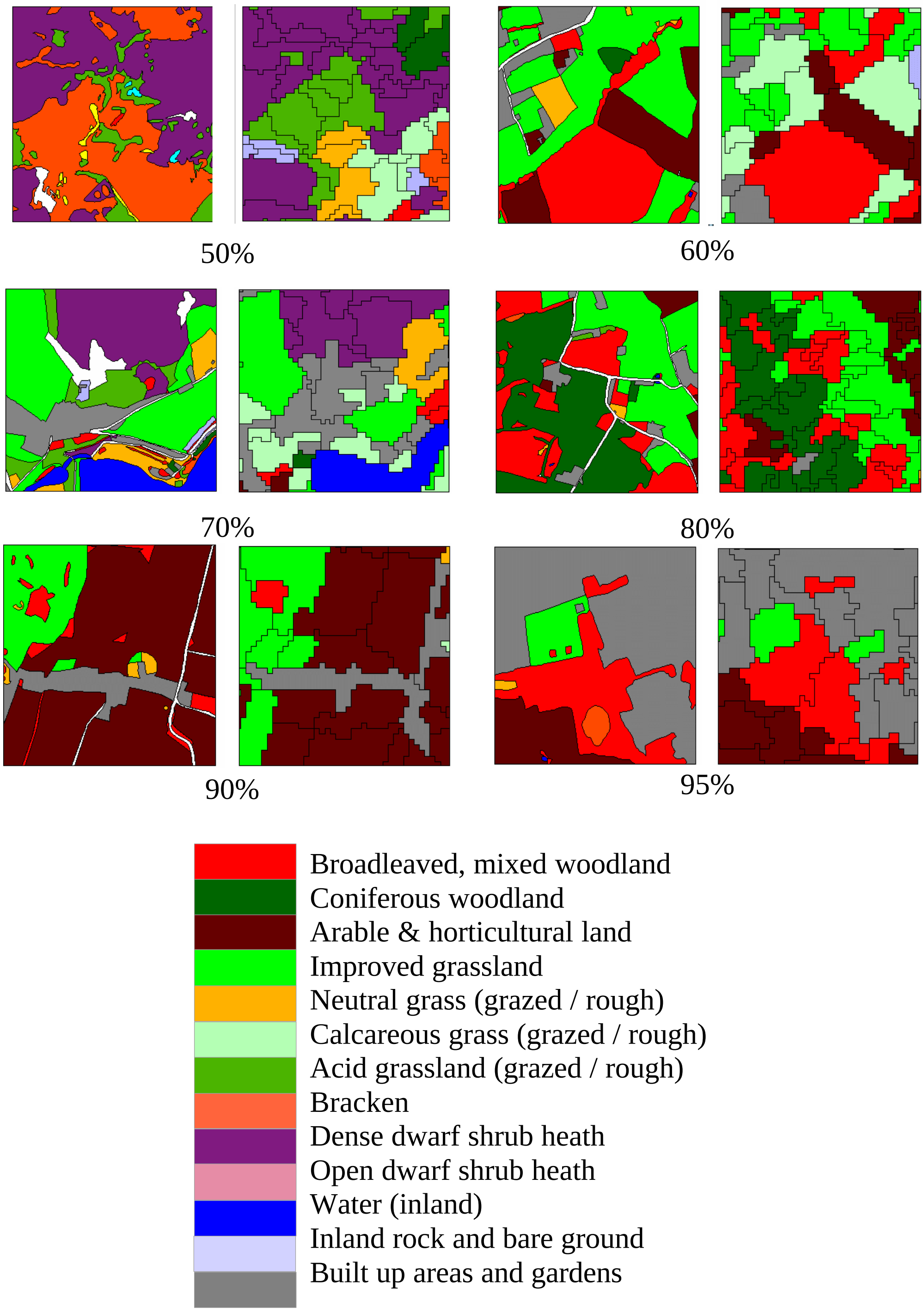
10.3 Confidence limits for measures of correspondence

A 'bootstrapping' procedure (Efron & Tibshirani 1998) was developed to provide confidence limits for measures of correspondence. Consider a stratum with *n* original sample squares: from these squares, a new random sample of *n* squares was taken with replacement (i.e. such that the same square might have been drawn more than once). The mean correspondence was recorded for the first new group of *n* squares. The process was then repeated with another sample of *n* squares giving another value for the mean correspondence in that stratum. This process was repeated 1000 times; and, from the 1000 estimates of correspondence, a mean value was calculated (tests with 2500, 5000, and 10000 runs demonstrated no noticeable improvement over the use of 1000-sample matrices). The 1000 different estimates of correspondence were then ranked from the lowest to the highest levels of correspondence and the 25th and 975th in rank were taken to represent the '95-percentile range' (i.e. that encompassing 95% of all estimates of correspondence). Bias-correction was applied to the confidence intervals to remove any bias which arises because the mean of the original sample squares differs from that of the bootstrap sample (Efron & Tibshirani 1998).

Correspondence estimates with confidence limits were generated for each of the 40 Land Classes in Great Britain. Correspondence-statistics could then be estimated for a part of Britain, using the 40 stratified estimates; the contribution of each stratum would be weighted according to its extent in the area in question. Correspondence assessments were made for: GB, England / Wales combined and Scotland. Table 7 gives overall correspondence by country, with per-pixel, per-segment and per-parcel comparisons. Per-pixel correspondence gave the lowest scores. It recorded every minor spatial difference between FS and LCM2000 products, even where these were inherent products of the mapping process. The per-pixel measure, for example, recorded differences in parcel outlines based on the 25 m image pixels as mismatches; also those due to differences in the minimum mappable unit (MMU) with FS maps recording parcels ≥0.04 ha against LCM2000 segments all >0.5 ha. Estimated per-pixel correspondence in Britain, at BH-level, is 54% (with the 95-percentile range estimated at 53-56%). In England and Wales the match is 60% (estimated range 58-62%). In Scotland it is lower at just 44% (range 40-47%).

LCM2000 segments labelled with FS classes are next in the level of correspondence. In Britain, the match at BH-level is 58% (range 57-60%). In England and Wales, the match is 64% (range 62-66%). In Scotland it is 47% (range 43-50%). This measure shows how the segmentation and

Figure 12. Pairs of field survey squares (left of a pair) and LCM2000 squares (right of a pair) illustrating correspondences from about 50% to 95%.



White areas on field-maps denote unclassified areas or complex mosaics

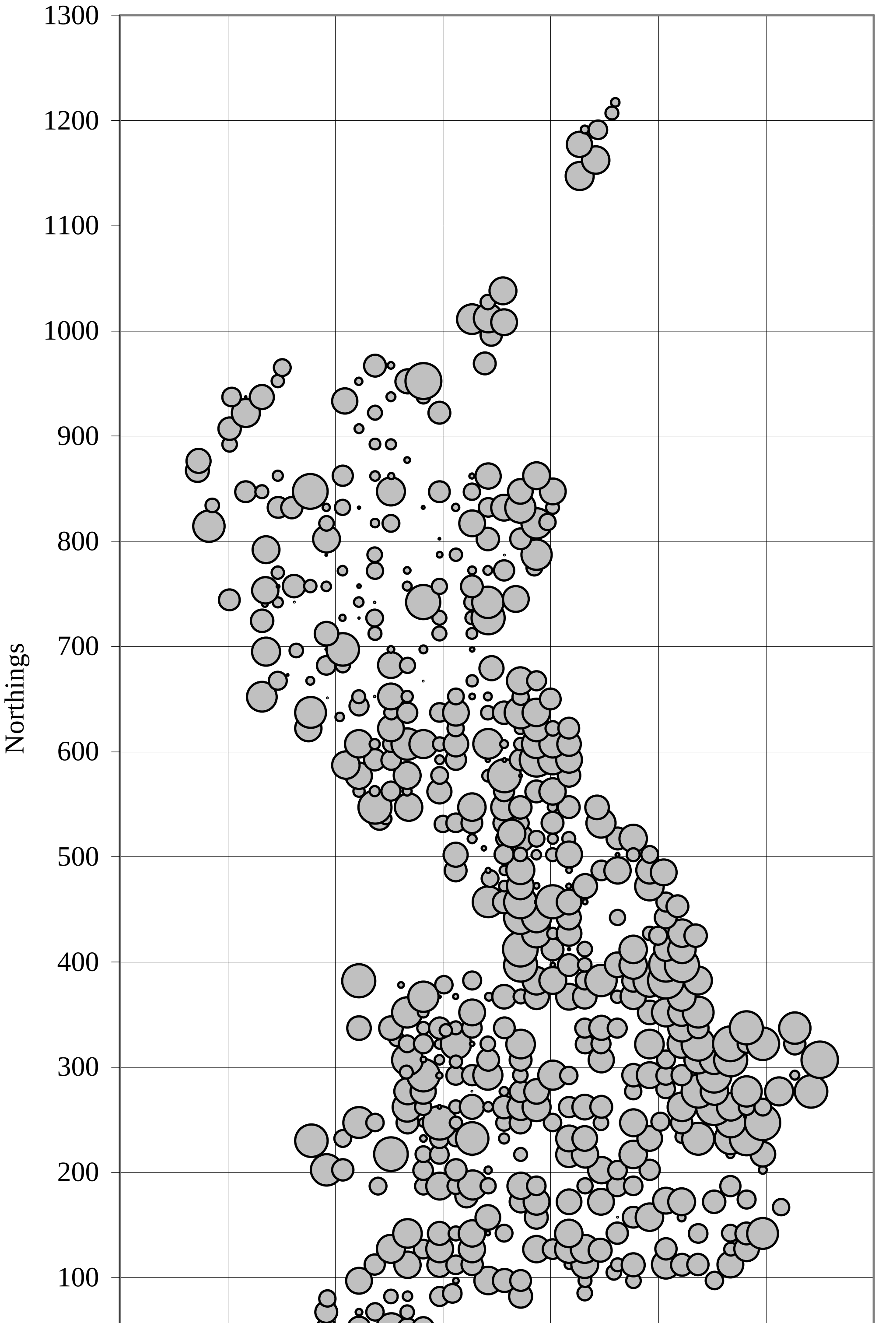


Figure 13. Correspondences recorded in LCM2000 comparisons with the Great Britain CS2000 field survey squares plotted in their GB context. Point sizes represent correspondences from the smallest on this plot at 13% to the largest representing 100% correspondence. Coordinates are British National Grid, expressed in kilometres.

spectral-classification of segments fared. It accommodated some differences in resolution, hence the improvements over per-pixel measures, but the labels drawn from FS mosaics of small parcels may have distorted the true picture of segment-dominance. Per-parcel correspondence gave the highest matches. Correspondence at BH-level in Britain is 62% (range 60-64%). In England and Wales the match is 69% (range 67-72%). In Scotland it is 48% (range 44-52%). FS parcels down to 0.04 ha MMU were labelled from generalised segments > 0.5 ha, thereby mis-labelling smaller landscape features such as ponds, shelter-belts and isolated farmsteads.

Table 7. Overall correspondence (%) from comparing LCM2000 with the CS2000 field survey squares in Great Britain: results are weighted estimates, based on 40 strata or 'Land Classes' and 'bootstrapped' to calculate confidence intervals (see text). Mean values, with 95-percent confidence limits, have been calculated for GB and constituent countries.

		Confidence interval (95%)	
	Sample	Bias corrected	
Analysis	mean	lower	upper
GB			
Per-pixel	54	53	56
Per-parcel	62	60	64
Per-segment	58	57	60
England & Wales			
Per-pixel	60	58	62
Per-parcel	69	67	72
Per-segment	64	62	66
Scotland			
Per-pixel	44	40	47
Per-parcel	48	44	52
Per-segment	47	43	50

It was known from the outset that there would be mismatches when FS and LCM2000 were compared at BH-level. At the Target class level, correspondence is higher than at the BH level: weighted correspondence is 65% across GB for parcel-based analysis, 73% for England-Wales combined and 51% for Scotland (due largely to bog-heath confusion and general problems in upland mapping). All three methods measure non-correspondences associated with i. the FS's greater original spatial resolution, ii. time differences in surveys, iii. class-definition differences and iv. errors in one or both surveys. To see how these differences contribute to the overall match or mismatch, it is necessary to examine the correspondence matrices for individual classes.

Table 8. Summary correspondence matrix (results expressed per 1000) made by from comparing LCM2000 with the CS2000 field survey squares in Great Britain: results are weighted estimates, using strata based on 40 Land Classes, bootstrapped' to calculate confidence intervals (see text).

LCM	CS2000	Broadleaved, mixed and yew woodland 1	Coniferous woodland 2	Boundary and linear features 3	Arable and horticulture 4	Improved grassland 5	Neutral grassland 6	Calcareous grassland 7	Acid grassland 8	Bracken 9	Dwarf shrub heath 10	Fen, marsh and swamp 11	Bog 12	Standing open water and canals 13	Rivers and streams 14	Montane habitats 15	Inland rock 16	Built up areas and gardens 17	Supralittoral rock 18	Supralittoral sediment 19	Littoral rock 20	Littoral sediment 21	Oceanic seas 22	Total
Broadleaved, mixed and yew woodland	**1**	33	4		3	4	2	1	1	1	0	0	0	0		0	0	3	0	0				**53**
Coniferous woodland	**2**	9	50		0	1	1	0	1	1	2	1	1	0		0	0	0	0	0		0	0	**66**
Boundary and linear features	**3**																							
Arable and horticulture	**4**	5	1		212	27	4	0	1	0	0	1	0	0		0	0	7	0	0		0		**259**
Improved grassland	**5**	7	0		11	181	13	1	5	1	0	5	1	0		0	0	6	0	0		0		**232**
Neutral grassland	**6**	3	0		3	10	2	0	13	3	3	4	5	0		0	0	1	0	0		0		**48**
Calcareous grassland	**7**	2	1		2	17	2	1	2	0	0	3	0	0		0	0	2	0	0		0		**34**
Acid grassland	**8**	2	1		0	5	1		17	5	9	3	17	0		0	0	1	0	0				**62**
Bracken	**9**	0	0			0	0		2	4	0	1	2			0	0	0						**9**
Dwarf shrub heath	**10**	4	3		0	0	0	0	5	3	35	4	53	0		0	0	0	0	0		0	0	**108**
Fen, marsh and swamp	**11**	0				0			0		0	0	0	0				0	0					**1**
Bog	**12**	1	0		1	0	1	0	2	0	6	2	16	0		0	0	0	0	0			0	**31**
Standing open water and canals	**13**	0	0			0	0		0	0	0	0	0	7		0	0	0	0				0	**7**
Rivers and streams	**14**																							
Montane habitats	**15**								2		5	0	3			2	0	0						**12**
Inland rock	**16**	0	0		2	0	0		0	0	0	0	0	0		0	0	0	0	0		0	0	**3**
Built up areas and gardens	**17**	1	0		4	2	1		0	0	0	0	0	0			0	38	0	0		0		**48**
Supralittoral rock	**18**	0								0			0			0			0	0		0	0	**0**
Supralittoral sediment	**19**	0				0	0			0								0	0	0				**0**
Littoral rock	**20**	0				0	0		0	0	0	0	0			0		0	1	0		0	0	**1**
Littoral sediment	**21**	0				0	0			0		0		0			0	1	1	0		4	1	**7**
Oceanic seas	**22**	0					0		0	0	0	0	0			0	0		1	1	0	2	12	**17**
Total		**68**	**61**		**238**	**249**	**28**	**3**	**52**	**18**	**61**	**23**	**98**	**9**		**2**	**2**	**61**	**4**	**2**	**0**	**6**	**14**	**1000**

BH level (excluding linear features)	**61.7**
Generalising urban	**63.6**
Target class level	**66.4**
Aggregate level	**74.1**

Summary correspondences (%) are given for Broad Habitats (yellow cells),and allowing for field survey generalisation of urban areas (+blue cells), at LCM2000 Target Class (+pink cells) and Aggregate class levels (+green cells)

Table 9. Summary correspondence matrix (results expressed per 1000) made by from comparing LCM2000 with the CS2000 field survey squares in England and Wales: results are weighted estimates, using strata based on 40 Land Classes, bootstrapped' to calculate confidence intervals (see text).

LCM	CS2000	Broadleaved, mixed and yew woodland 1	Coniferous woodland 2	Boundary and linear features 3	Arable and horticulture 4	Improved grassland 5	Neutral grassland 6	Calcareous grassland 7	Acid grassland 8	Bracken 9	Dwarf shrub heath 10	Fen, marsh and swamp 11	Bog 12	Standing open water and canals 13	Rivers and streams 14	Montane habitats 15	Inland rock 16	Built up areas and gardens 17	Supralittoral rock 18	Supralittoral sediment 19	Littoral rock 20	Littoral sediment 21	Oceanic seas 22	Total
Broadleaved, mixed and yew woodland	1	48	2		4	6	3	1	1	1	0	0	0	1		0	0	4	0	0				71
Coniferous woodland	2	8	20		0	0	0	0	0	0	0	0	1	0			0	0	0			0		30
Boundary and linear features	3																							
Arable and horticulture	4	7	1		289	34	5	0	1	0	0	1	0	0		0	0	9	0	0		0		347
Improved grassland	5	8	0		14	226	14	1	3	1	0	4	0	0		0	0	9	0	0				282
Neutral grassland	6	2	0		3	11	3	0	13	2	3	2	1	0		0	0	1	0	0				40
Calcareous grassland	7	3	1		3	25	3	0	2	0	0	1	0	0		0	0	3	0	0				41
Acid grassland	8	3	1		0	5	1		11	5	5	2	5	0		0	0	1						38
Bracken	9	0	0			0	0		2	5	0	0	0	0		0	0	0						7
Dwarf shrub heath	10	2	1		0	0	0		2	2	16	1	2	0		0	0	0	0	0				26
Fen, marsh and swamp	11	0				1			0		0	0	0	0				1						2
Bog	12	1			2	0	2	0	1	0	6	1	3	0		0		0						17
Standing open water and canals	13	0	0			0	0		0				0	6			0	0						6
Rivers and streams	14																							
Montane habitats	15																							
Inland rock	16	0	0		2	0	0		0	0	0	0		0		0	0	0	0	0		0		3
Built up areas and gardens	17	2	0		6	2	1		0	0	0	0	0	0			0	54	0	0		0		67
Supralittoral rock	18																							
Supralittoral sediment	19					0	0			0								0	0	0				0
Littoral rock	20	0				0	0											0	0	0			0	0
Littoral sediment	21	0				0	0			0		0		0			0	1	0	1		6	2	11
Oceanic seas	22	0					0										0		0	1		3	7	11
Total		**83**	**27**		**322**	**312**	**31**	**3**	**36**	**17**	**30**	**13**	**12**	**8**		**0**	**1**	**83**	**1**	**2**		**10**	**9**	**1000**

BH level (excluding linear features)	**69.3**
Generalising urban	**71.8**
Target class level	**74.5**
Aggregate level	**75.5**

Summary correspondences (%) are given for Broad Habitats (yellow cells),and allowing for field survey generalisation of urban areas (+blue cells), at LCM2000 Target Class (+pink cells) and Aggregate class levels (+green cells)

Table 10. Summary correspondence matrix (results expressed per 1000) made by from comparing LCM2000 with the CS2000 field survey squares in Scotland: results are weighted estimates, using strata based on 40 Land Classes, bootstrapped' to calculate confidence intervals (see text).

LCM	CS2000	Broadleaved, mixed and yew woodland 1	Coniferous woodland 2	Boundary and linear features 3	Arable and horticulture 4	Improved grassland 5	Neutral grassland 6	Calcareous grassland 7	Acid grassland 8	Bracken 9	Dwarf shrub heath 10	Fen, marsh and swamp 11	Bog 12	Standing open water and canals 13	Rivers and streams 14	Montane habitats 15	Inland rock 16	Built up areas and gardens 17	Supralittoral rock 18	Supralittoral sediment 19	Littoral rock 20	Littoral sediment 21	Oceanic seas 22	
Broadleaved, mixed and yew woodland	**1**	7	8		0	1	1	0	1	0	1	0	0	0		0		1						**21**
Coniferous woodland	**2**	12	104		1	1	2		2	1	5	2	2	0		0	0	1	0	0		0	0	**132**
Boundary and linear features	**3**																							
Arable and horticulture	**4**	2	2		70	15	4		0	0	0	0	0	0			0	3	0					**97**
Improved grassland	**5**	4	1		6	98	10	0	8	1	0	7	1	0		0	1	2	0	0		0		**140**
Neutral grassland	**6**	3	1		3	9	2	1	14	5	3	6	13	0		0	0	1	0	0		0		**61**
Calcareous grassland	**7**	1	2		1	2	1	2	4	0	0	6	1	0		0	0	1	0	0		0		**21**
Acid grassland	**8**	2	1		0	7	1		29	4	16	6	40	0		0	0	0	0	0				**106**
Bracken	**9**	1			0	0	0		1	3	0	1	5			0	0	0						**12**
Dwarf shrub heath	**10**	7	7			1	1	0	12	4	72	10	146	1		0	0	0	1	0		0	0	**261**
Fen, marsh and swamp	**11**								0		0	0	0						0					**0**
Bog	**12**	0	1			0	0		4	0	7	2	41	1		0	0	0	0	0			0	**58**
Standing open water and canals	**13**	0	0			0	0		0	0	0	0	0	9		0	0		0				0	**10**
Rivers and streams	**14**																							
Montane habitats	**15**								6		14	0	8			5	0	0						**35**
Inland rock	**16**		0		0	0	0		0	0	0	0	0	0			1	0	1	0			0	**2**
Built up areas and gardens	**17**	0	0		0	0	0		0	0	0	0	0				1	9	0			0		**12**
Supralittoral rock	**18**	0								0			0			0			0	0		0	0	**0**
Supralittoral sediment	**19**	0					0													1				**1**
Littoral rock	**20**					0	0		0	0	0	0	0			0		0	1	0		0	0	**2**
Littoral sediment	**21**	0														0		0	1	0		0	0	**1**
Oceanic seas	**22**	0					0		0	0	0	0	0			0	0		2	1	0	0	23	**26**
		39	**126**		**82**	**134**	**21**	**4**	**83**	**19**	**119**	**42**	**258**	**11**		**6**	**4**	**20**	**8**	**3**	**0**	**0**	**23**	**1000**

BH level (excluding linear features)	**47.6**
Generalising urban	**48.4**
Target class level	**51.5**
Aggregate level	**71.3**

Summary correspondences (%) are given for Broad Habitats (yellow cells),and allowing for field survey generalisation of urban areas (+blue cells), at LCM2000 Target Class (+pink cells) and Aggregate class levels (+green cells)

11. LCM2000 ASSESSMENTS AT COVER CLASS LEVEL

Correspondence matrices were combined via the stratification to give summary matrices for GB, England / Wales (combined) and for Scotland: a matrix was calculated for each stratum or Land Class based upon the sample squares in that Land Class; then the Land Class matrices were combined, each with a weighting according to its extent in GB, England / Wales, or Scotland. Tables 8, 9 and 10 give the summary matrices, based on per-parcel comparisons. The matrices are also summarised at the Aggregate class level, where LCM and FS match closely (Tables 11, 12 & 13). The following discussion examines similarities and some of the underlying cause of difference (NB no section numbers are given as, with sections based on classes, such numbers would conflict with Subclass numbers and thereby cause confusion).

Broadleaved and mixed woodland and **Coniferous woodland**: LCM2000 **Broadleaved and mixed woodland** and the FS *Broadleaved, mixed and yew woodland* show similar extents: UK coverage is 6.3% from LCM2000 and 6.2% from FS. However, direct agreement in the 569 squares is rather lower (44% of LCM2000 **Broadleaved /mixed woodland** is mapped similarly by FS). This is due partly to the fact that many woodlands are small, at or below the minimum mappable unit of LCM2000, and so they are excluded. Hence, a lot of FS *Broadleaved woodland* shows on the map as grassland or arable farmland, the typical situation for small copses and shelter belts. The converse also applies: that openings in a woodland, mapped by FS, are mostly too small to record on LCM2000, so woodlands may appear continuous where they are really open. Differences in the exact outlines of woodlands also contribute. *Coniferous woodland*, generally planted and in larger blocks, records similar coverages (UK 5.5% on LCM2000 and 5.8% through FS) and a far greater direct correspondence (70%).

Arable and horticultural land covers just over 23.4% of the UK according to LCM2000 and 21.5% by FS estimates. LCM2000's higher estimate relates in part to small features such as woodlands, prevalent in arable landscapes but generalised out by LCM2000. About 70% of LCM2000 **Arable and horticultural land** is coincident with FS *Arable*. There are apparent confusions between *Arable and horticultural land* and *Improved grassland* in LCM2000 mapping; these probably relate largely to rotation farming in squares where the survey-year differed in field and satellite surveys. However, there are also misclassifications of grass as arable and *vice versa*. Though this problem is relatively small, because grass and arable farming together make up so much of the UK, the misclassified elements contribute a significant proportion of the total map error. Confusion between arable and built up land is a small but nonetheless significant problem: it relates mostly to erroneous classification of satellite images, where part-grown or ripening crops have spectral signatures readily confused with those of partly vegetated suburban areas.

Improved grassland, 25.7% according to LCM2000 and 25.8% by FS, is the largest single Target cover / BH type in GB. Generally, it is readily recognisable and well-classified on LCM2000. However, the distinction of 'improved' grassland from semi-natural types can be both difficult and controversial. The 'improvement' of a grassland is a continuous process which may start with minor attempts at surface drainage and end with ploughing and reseeding. Other treatments include control of grazing, fertiliser application, liming and herbicide treatments. There is potential for reversion; and abandonment or extensification can give the impression (and the spectral character) of semi-natural swards. Field surveyors use species records and a range of contextual observations not open to the image analyst, though field surveyors also face difficulties in dividing the continuum. There is, inevitably, scope for differences in interpretation by FS and LCM2000; a significant proportion (near 20%) of FS 'improved grassland' is recorded by LCM2000 as semi-natural.

Tables 11-13. Summary correspondence matrices (results expressed per 1000) at Aggregate class level. The matrices were made by comparing LCM2000 with the CS2000 field survey squares. Results were calculated as weighted estimates, using strata based on 40 Land Classes, 'bootstrapped' to generate confidence intervals (see text).

Table 11. Great Britain

LCM2000 / **Field Survey**	Broadleaved / mixed wood	Coniferous woodland	Arable and horticulture	Improved grassland	Seminatural grass	Mountain, heath, bog	Standing open water	Built up areas and gardens	Coastal	Oceanic seas	**Total**
Broadleaved / mixed wood	33	4	3	4	5	1	0	3	0		**53**
Coniferous woodland	9	50	0	1	3	3	0	0	0	0	**66**
Arable and horticulture	5	1	212	27	6	0	0	7	0		**259**
Improved grassland	7	0	11	181	25	1	0	6	0		**232**
Seminatural grass	8	3	5	33	64	37	0	4	1		**154**
Mountain, heath, bog	5	3	3	1	20	121	1	1	1	0	**155**
Standing open water	0	0		0	0	0	7	0	0	0	**7**
Built up areas and gardens	1	0	4	2	1	0	0	38	0		**48**
Coastal	0			0	0	0	0	1	6	1	**9**
Oceanic seas	0				0	0			4	12	**17**
Total	**68**	**61**	**238**	**249**	**124**	**164**	**9**	**61**	**12**	**14**	**1000**
											74.1

Table 12. England and Wales

LCM2000	Broadleaved / mixed wood	Coniferous woodland	Arable and horticulture	Improved grassland	Seminatural grass	Mountain, heath, bog	Standing open water	Built up areas and gardens	Coastal	Oceanic seas	**Total**
Broadleaved / mixed wood	48	2	4	6	6	0	1	4	0		**71**
Coniferous woodland	8	20	0	0	1	1	0	0	0		**30**
Arable and horticulture	7	1	289	34	7	0	0	9	0		**347**
Improved grassland	8	0	14	226	25	0	0	9	0		**282**
Seminatural grass	8	2	5	42	52	14	0	5	0		**129**
Mountain, heath, bog	3	1	5	1	9	27	0	1	0		**46**
Standing open water	0	0		0	0	0	6	0			**6**
Built up areas and gardens	2	0	6	2	2	0	0	54	0		**67**
Coastal	0			0	0	0	0	1	8	2	**11**
Oceanic seas	0				0	0			4	7	**11**
Total	**83**	**27**	**322**	**312**	**100**	**43**	**8**	**83**	**13**	**9**	**1000**
											75.5

Table 13. Scotland

LCM2000	Broadleaved / mixed wood	Coniferous woodland	Arable and horticulture	Improved grassland	Seminatural grass	Mountain, heath, bog	Standing open water	Built up areas and gardens	Coastal	Oceanic seas	**Total**
Broadleaved / mixed wood	7	8	0	1	2	1	0	1			**21**
Coniferous woodland	12	104	1	1	7	7	0	1	0	0	**132**
Arable and horticulture	2	2	70	15	5	1	0	3	0		**97**
Improved grassland	4	1	6	98	26	3	0	2	0		**140**
Seminatural grass	7	4	4	17	87	79	0	2	1		**201**
Mountain, heath, bog	7	8	0	1	40	295	2	1	2	0	**356**
Standing open water	0	0		0	0	1	9		0	0	**10**
Built up areas and gardens	0	0	0	0	1	1		9	0		**12**
Coastal	0			0	0	0		0	4	0	**5**
Oceanic seas	0				0	0			3	23	**26**
Total	**39**	**126**	**82**	**134**	**169**	**387**	**11**	**20**	**11**	**23**	**1000**
											71.3

Summary correspondences (%) are given for Broad Habitats allowing for field survey generalisation of urban areas (i.e. total for shaded cells).

Semi-natural grasslands, bracken, fens and marshes present some problems in their distinction. Specific *Neutral, Calcareous* and *Acid grassland* BHs are not well separated by LCM2000. Differences relate first to problems with ‘rough grasslands’, some of which are semi-natural, others derelict and abandoned swards of unknown origin. The derelict swards may have been agricultural in origin, such as abandoned / neglected grazing or long-term setaside; they may be associated with tree-felling or new planting; or they may have arisen from development, for example, motorway verges, urban rough ground, derelict industrial sites, railway yards, vegetated dumps and quarries. Those associated with urban open-space go largely unmapped by FS. Most rough grasslands fit into the *Neutral grassland* BH, though some are ‘improved’. If the dividing line is hard to draw in the field, satellite imaging has even greater difficulties subdividing these habitats. The rough grasslands of LCM2000 were all treated, for BH mapping purposes, as *Neutral grassland*. In contrast, rough grasslands with species compositions indicating improvement were mapped by FS as *Improved grassland.* LCM2000 distinguished swards trained as ‘rough grassland’ at class Variant level; this distinction may give scope to refine the classification later, through integrated analysis of LCM2000 and FS data.

There is a general problem in separating *Neutral, Calcareous* or *Acid grassland* BHs. Quite simply, there is no consistent spectral characteristic which allows accurate distinction by image analysis. In order to comply as far as possible with the needs to map BHs, all LCM2000 semi-natural swards, whatever their origin, were subject to acid-sensitivity masking (Hornung *et al.* 1995) in order to allocate them to an acidity class. There were inevitable weaknesses in a mask based on a 1 km grid, which generalised soil heterogeneities at an even coarser scale and took no account of management practices such as liming. Moreover, the acid-sensitivity map defined acidity classes as ‘highly sensitive’ with a pH <4.5, ‘moderately sensitive’ with a pH between 4.5 and 5.5. and low sensitivity or pH >5.5. While the ‘acid’ label represents truly acid conditions, the moderately sensitive ‘neutral’ category is actually slightly acid and the low sensitivity ‘calcareous’ category includes neutral and calcareous elements. There was no affordable alternative available to the project. Not surprisingly, the results gave a poor match with FS data.

The *Bracken* BH was not a Target class of LCM2000. The Subclasses identified **Bracken** for BH-mapping purposes, but included only that which was growing in open conditions. Because so much of the imagery used to make LCM2000 was recorded in May, when the amount of bracken on the ground would be at the absolute minimum, there was often the tendency for stands to be recorded as the background *Acid grassland*.

The *Fen, marsh, and swamp* BH is defined as being ‘... characterised by a variety of vegetation types that are found on minerotrophic (groundwater-fed), permanently, seasonally or periodically waterlogged peat, peaty soils, or mineral soils. They include fens, flushes, marsh ... and swamps. This BH does not include neutral and improved grasslands on floodplains and grazing marshes, nor bogs, nor areas of carr woodland.’ The FS identifies much more *Fen, marsh, and swamp* (2.4%) than does LCM2000 (0.1%), largely because the surveys treat rush-pastures very differently, with LCM2000 including the these in the **Acid grassland** class while FS records them as *Fen, marsh, and swamp*. This distinction led to the FS recording a much greater extent of *Fen, marsh, and swamp* in 1998 than it had in 1990 and has raised questions over the classification. If these rush pastures are to be included in the *Fen, marsh, and swamp* BH, they are identified at LCM2000 Variant level: it will be helpful to check such examples against FS data during follow-up work on integration of the two datasets to see where patterns match.

Heath, bog and montane habitats presented problems in distinctions to BH definitions. LCM2000 targeted dwarf shrub and bog communities knowing that LCMGB of 1990 had already shown the problems in defining bogs to floristically-based standards. Dwarf shrubs grow on many bogs; but not all dwarf shrub coverage signifies bog; and not all bogs have dwarf shrub coverage. Underlying soil and moisture parameters, particularly peat-formation, dictate the classification. *Dwarf shrub heath* and *Bogs* BHs are mapped very differently by LCM2000 and FS. LCM2000 shows 11.1% cover of heath and 2.3% cover of bog. FS estimates 6.1% heath and 9.6% cover of bog. It is necessary to understand the BH definition in order to explain these differences.

JNCC (Jackson, 2000) state that the *Bogs* BH '... covers wetlands that support vegetation that is usually peat-forming ...' They note that the '... habitat type also includes modified bog vegetation that essentially resembles wet or dry dwarf shrub heath but occurs on deep acid peat which would have once supported peat-forming vegetation. Modified bog also includes impoverished vegetation dominated by purple moor-grass .... or hare's-tail cotton-grass ... Although there is no agreed minimum depth of peat that can support ombrotrophic vegetation, unmodified bog can be identified floristically by the presence of characteristic species such as cotton-grass ... and peat-forming sphagna ...' This use of indicator species is the main way that CS2000 field surveyors identified *Bogs*. The JNCC report goes on to say that 'Peat depth, although somewhat arbitrary, is used as the primary criterion to separate types of modified bog vegetation from the *'Dwarf shrub heath'* broad habitat type ... Therefore vegetation dominated by dwarf-shrubs, cotton-grass ... or purple moor-grass ... on peat greater than 0.5 m deep is classified as bog for the purposes of the Broad Habitat Classification.' Field meetings with conservation agency staff involved with Phase 1 survey gave support for such a definition.

With this in mind, peat depth was set as the main criterion for distinction of **Bogs** in LCM2000; a British Geological Survey map showing peat drift >0.5 m deep was used to determine the context of 'heath' and 'moor'. Any 'heath' or 'moor' on deep peat was recoded to bog and any 'bog' which did not coincide with peatland was recoded to heath or grass moor (depending on the key cover-component). In the event, the peat mask gave a very conservative picture of the true extent of peatlands: it generated a bog-coverage amounting to a quarter that of the FS estimate and much less than that suggested by Reid and Quarmby (1997). The issue is clearly controversial and needs careful examination. Bog surveys are currently being made by Countryside Council for Wales and Scottish Natural Heritage. Peatland is being mapped by Macaulay Land Use Research Institute and Soil Survey and Land Research Centre. It is proposed to re-examine LCM2000 bogs and heaths in a follow-up programme which will integrate LCM2000 with FS and external data. For the time being, the LCM2000 bog class is described as '**Bog (deep peat)**'.

Field reconnaissance for LCM2000 seldom visited *Montane habitats*, as they are not easily accessible. While the *Montane habitats* BH definition relies heavily on floristic composition (Jackson 2000), LCM2000 **Montane** cover was defined by altitude criteria, with all vegetated ground >600 m (see Ratcliffe & Thompson 1988) being treated as *Montane habitats*. LCM2000 records 1.6% coverage against the FS estimate of 0.2%. Clearly there is a difference and there must be questions over whether the altitude-based distinction is well made in this circumstance.

The *Inland rock* BH, while treated as part of the **Heath, bog and montane** Aggregate class, actually covers both natural and artificial exposed rock surfaces. Potentially, they include exposed mountain tops, screes and limestone pavements, as well as various forms of excavations such as quarries and waste tips and quarry waste. LCM2000 distinguished these components as **Inland bare ground**. However, it may erroneously have included temporary bare ground, particularly bare arable land, in this category, where contextual corrections have failed to identify the arable context. Unlike the FS, LCM2000 also mapped **Inland bare ground** in an urban context: this included gravel car parks, railway sidings and derelict industrial land. The consequence is that LCM2000 records four times as much **Inland bare ground** as the FS does *Inland rock*. The overall quantity is however small at 0.9% of LCM2000 or 0.2% of FS cover for the UK. Where, locally, coverage is extensive, it relates generally to the uplands and inclusion with the aggregate **Heath, bog and montane** habitats is appropriate.

It can be seen from the above that the **Dwarf shrub heath**, **Bogs, Montane** and **Inland bare** categories of LCM2000 do not directly match the BH definitions. It is for this reason that the classes are put into the **Heath, bog and montane** Aggregate class for some map-presentations and data tabulations.

Water (inland) on LCM2000 is an aggregation of the *Standing open water and canals* and *Rivers and streams* BHs. **Water (inland)** is mapped where >0.5 ha in extent and where its width substantially exceeds the 25 m input pixel size (only being recognised consistently where width is 2 pixels (50 m) and the area >0.5 ha). Smaller water bodies and narrower waterways are not included. There is no attempt to distinguish standing from flowing water. Despite these differences, the LCM2000 cover of **Water (inland)** is near identical to FS estimates for *Standing open water and canals* in GB. UK statistics are not available from FS as the NICS does not report on the class.

Built up and gardens are mapped in greater detail by LCM2000 than the FS *Built up areas and gardens*. LCM2000 distinguished open spaces >0.5 ha in the built landscape. FS treated urban land as continuous without recording open spaces in the urban zone. Consequently, the FS recorded more 'built up' land and less grassland, woodland and waterway. And as a result, FS *Built up* includes LCM2000 woodlands, grasslands and water. LCM2000 recognised urban areas comprising a mixture of built and vegetated surfaces as **Suburban and rural development** and those with little if any vegetation as **Continuous urban**.

Coastal habitats of *Supralittoral rock*, *Supralittoral sediment, Littoral rock* and *Littoral sediment* are, with the exception of some notable dune systems, shingle beaches and estuaries, relatively small habitats, often near to or below the resolution of LCM2000. They are treated as an **Coastal habitats** Aggregate class for reporting purposes; however, they are recognised at BH level in LCM2000 Subclasses and shown on maps as two classes: **Supralittoral rock and sediment** and **Littoral rock and sediment**. The distinction of these BHs (and aggregations thereof) relies upon definition of a high water mark, and use of OS 1:50 000 maps and local knowledge to indicate whether a surface is solid rock or sedimentary in recent origin. Neither type of information could be provided or applied with the precision and consistency to allow accurate separation at the resolution of LCM2000. As a result, the relative quantities recognised by LCM2000 and FS differed; they contribute a negligible amount to overall cover and overall non-correspondence. The greatest difference however relates to the tidal state at the time of imaging. As a result, some inter-tidal areas were under-represented; conversely, other offshore inter-tidal sediments, outside the FS population of terrestrial 1 km squares, were recorded by LCM2000 but not by FS. It must be recognised that neither survey provides nationally consistent and accurate estimates of coastal BHs.

The *Boundary and linear features* BH was not targeted by LCM2000. LCM2000 only includes linear habitats which have an area >0.5 ha: to have been resolved by the images they will also have been =2 pixels wide. Linear features were, however, mapped by the FS. As a consequence, the 2.5 m grid samples used in correspondence testing also recorded these items; they constituted about 3% of the landscape area. Because they were intentionally excluded in LCM2000, *Boundary and linear features* (and *Rivers and streams*) were excluded as distinct BHs in calibration.

12. LCM2000 ACCURACY?

The correspondence between LCM2000 and the FS is **not a measure of LCM2000's accuracy**. The FS does not provide 'ground truth': differences in resolution, the data-model and timing of surveys contribute to differences in correspondence. Nonetheless, it is possible to identify differences attributable to inherent characteristics of the surveys and others which relate to error

(Fuller *et al.* 1998). We might thus deduce a broad accuracy-value for LCM2000. Because LCM2000 did not directly map BHs, accuracy is best assessed at Target class level.

LCM2000 segments, compared with FS parcels, show a basic correspondence of 63.4% in per-parcel comparisons at Target class level (allowing for the FS generalisation of *Built up areas* and the LCM2000 omission of *Boundary and linear features* and *Rivers and streams*). As correspondence cannot realistically exceed the 88% repeatability of the FS, LCM2000 seems to be scoring at least 72% of its maximum potential. About 5% of the mis-match is explained by the 25 m grid underlying the image parcels, compared with the continuously variable structure of the field survey (if the field data are resampled onto the 25 m grid, the result shows 95% correspondence with the original input data). The MMU of LCM2000, which only records segments >0.5 ha, contrasts with the 0.04 ha MMU of the FS and explains many of the differences, especially for habitats which occur in less extensive stands (more than 4% of the area recorded by FS comprised parcels, not linear features, which were below the LCM2000 MMU). Time-differences explain other mis-matches: the FS was predominantly made in 1998; LCM2000 used images mainly from 1998-2001; (squares surveyed by field and satellite surveys in the same year are some 6% closer in correspondence than the national average). Evidently up to 15% of differences can be explained by the underlying structure of LCM2000 and, additionally, by its coarser MMU, and by date-differences. This suggest that LCM2000 may record Target classes with 87% success; to quote a figure of *c*. 85% accuracy at Target class level seems realistic.

13. CALIBRATION OF LCM2000 TO FIELD SURVEY

13.1 Operation

The calibration matrices, derived from inter-comparison of FS and LCM2000 data (see Section 11) were the basis of a calibration of LCM2000 cover-estimates to FS-equivalence. While a common classification was not a pre-requisite for inter-calibration, the BH classes were generated from both FS and LCM2000 data and were used for such calibration.

FS values for each land class in a given stratum or Land Class are estimated from LCM2000 values by multiplying the LCM2000 areas in each class by the calibration matrix, i.e.

$$FS = M \times LCM \qquad \text{Equation 1}$$

where *FS* and *LCM* are vectors of the proportions in each BH. M is an average calibration matrix, derived from a set of matrices M_1, M_2, . . . M_S, i.e. one for each of the *S* field squares belonging to the given stratum. Each element, M_{ij}, of the matrix *M* denotes the value for row *i* column *j* of the calibration matrix, i.e. the proportion of LCM type *i* classified as FS type *j*.

Consider the hypothetical calibration matrix below:

Calibration matrix	Field survey		
LCM2000	broadleaved	conifer	urban
broadleaved	0.75	0.10	0.15
conifer	0.10	0.85	0.05
urban	0.05	0.05	0.90

Results have been normalised so values sum to '1' across the rows. LCM2000 summary cover statistics (in italics below) are fed into the matrix:

LCM2000 Input cover values		Field survey broadleaved	Field survey conifer	Field survey urban
broadleaved	*1000*	750	100	150
conifer	*500*	50	425	25
urban	*200*	10	10	180
Calibrated output statistics		**810**	**535**	**355**

They are multiplied by the correspondence values to give output statistics which, if summed (bold), show how the same landscape might have been recorded by a comprehensive FS.

To calculate confidence limits for the calibrated estimates, bootstrap samples (Efron & Tibshirani 1998) of FS estimates are obtained by random simulated re-sampling of matrices M_1, M_2, . . . M_S with replacement, calculating the calibration matrix and applying the above calibration formula. The approach can be extended to allow for stratification, where a set of confusion matrices for each stratum is available. The FS squares were selected according to a stratified random sampling scheme, with the 40 Land Classes defining the strata. The national estimates of FS values are calculated by areally weighting the estimates derived separately for each stratum i.e.

$$FS = \frac{A_1 F_1 + A_2 F_2 + \ldots + A_n F_n}{A_1 + A_2 + \ldots + A_n} \qquad \text{Equation 2}$$

where A_i is the stratum area, and F_i is the vector of the proportions in each land cover type, of the i^{th} Land Class stratum. Each F_i is re-calculated for each bootstrap sample by randomly sampling from the set of confusion matrices from the given stratum (Equation 1). It was found that 1000 bootstrap samples were sufficient to calculate the 95% confidence intervals. Bias-corrected percentile limits were calculated to remove any bias that arises because the true parameter value is not the median of the distribution of estimates (Efron & Tibshirani 1998).

GB statistics on BHs were generated from LCM2000 calibrated cover, with confidence limits on each calibrated value. Weighted estimates with confidence limits were also calculated for England, Wales, England / Wales combined, and for Scotland (drawing upon Land Classes according to their presence and extent in each Country). A calibrated estimate of BH coverage was also made for Northern Ireland, assuming that the GB-calibration would apply; while there were weaknesses in this assumption, there is reason to believe that the same broad levels of over- and under-estimation applied when mapping BHs from satellite images. The calibration procedure generated calibrated statistics to match FS per-pixel cover, per-segment cover and per-parcel cover.

Table 14. Estimates of Broad Habitat cover for Great Britain from calibrated LCM2000 data: results give the mean of a bootstrapped estimate, with the bias-corrected 95-percentile range. Results are compared with estimates based on the sample-based field survey approach of Countryside Survey 2000: the 2 standard errors is used to estimate the 95-percentile range for the field survey.

	LCM	**LCM bias corrected confidence interval**		**FS**	
Broad Habitats	**mean**	**lower**	**upper**	**Total**	**2 x SE**
Broadleaved, mixed and yew woodland	16125	14679	17019	14710	1960
Coniferous woodland	12995	11677	14079	13740	3068
Arable and horticulture	54743	52433	57300	52490	4546
Improved grassland	59200	56811	61871	54820	4284
Neutral grassland	6674	5850	7525	6130	1084
Calcareous grassland	533	215	899	650	578
Acid grassland	12337	10699	14398	12950	2356
Bracken	3840	3140	4372	4390	1132
Dwarf shrub heath	14048	11909	16026	14870	2718
Fen, marsh and swamp	5273	4196	5846	5470	1332
Bog	22578	21037	26048	22180	3484
Standing open water and canals	1709	1433	1845	1900	1110
Montane habitats	606	117	1399	490	618
Inland rock	639	347	877	560	236
Built up areas and gardens	17652	16597	19812	13310	2282
Supralittoral rock	782	582	858	770	268
Supralittoral sediment	341	244	425	530	350
Littoral rock	0	0	1		0
Littoral sediment	977.2	568.8	1140.3	1380	940

Figure 14. A plot of the Broad Habitat cover estimates for Great Britain from calibrated LCM2000 and sample-based field survey as given in Table 12. The error bars show the 95-percentile range for each. The regression equation and R^2 values are given on the chart.

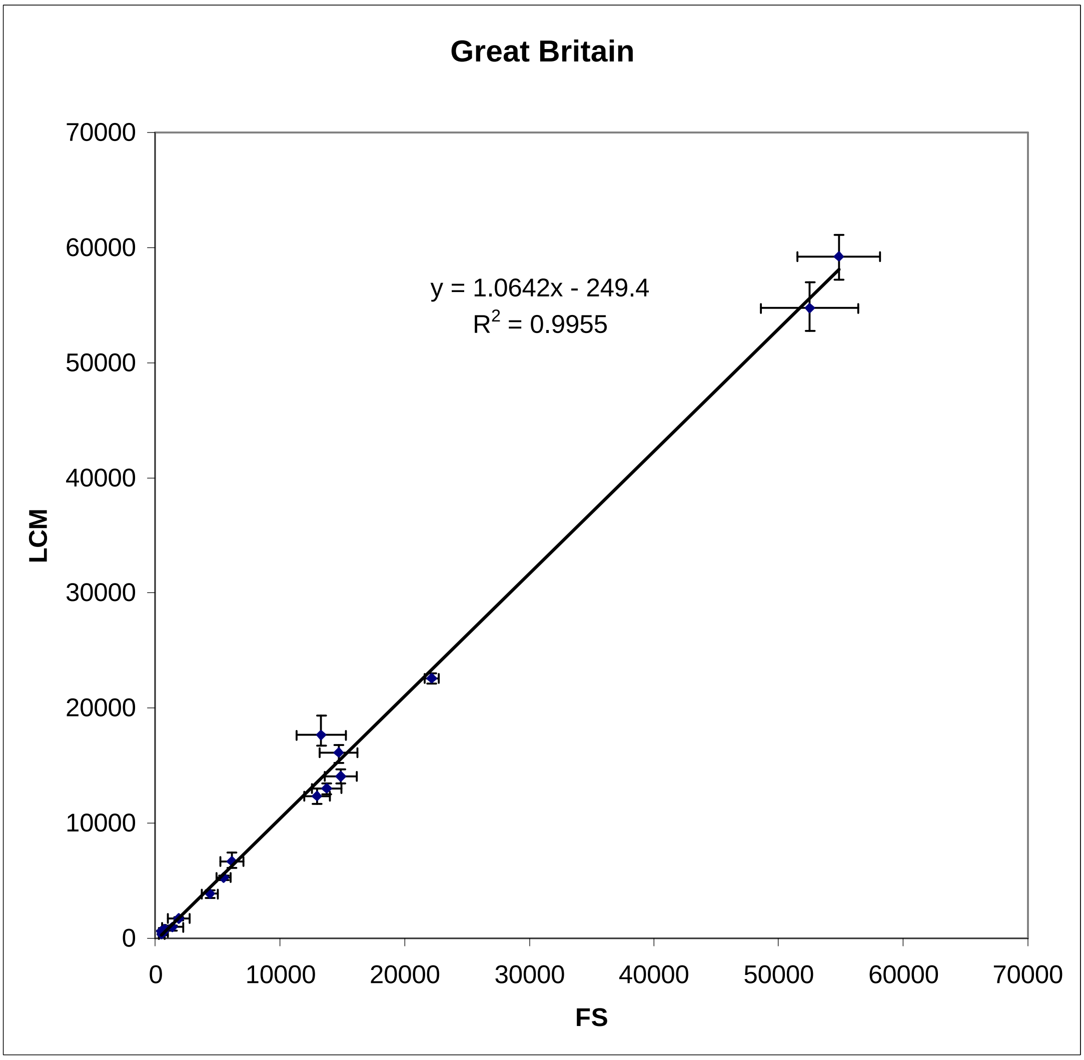


Table 15. Estimates of Broad Habitat cover for England from calibrated LCM2000 data: results give the mean of a bootstrapped estimate, with the bias-corrected 95-percentile range. Results are compared with estimates based on the sample-based field survey approach of Countryside Survey 2000: the 2 standard errors is used to estimate the 95-percentile range for the field survey.

	LCM	LCM bias corrected confidence interval		FS	
Broad Habitats	**mean**	**lower**	**upper**	**Total**	**2 x SE**
Broadleaved, mixed and yew woodland	10894	9974	11513	9970	1510
Coniferous woodland	2821	2317	3258	2980	1178
Arable and horticulture	47070	45062	49331	43890	3912
Improved grassland	40165	38151	42046	36630	3316
Neutral grassland	4260	3654	5029	3910	918
Calcareous grassland	377	102	726	350	330
Acid grassland	3969	3258	4637	3820	1026
Bracken	1501	1142	1795	1660	644
Dwarf shrub heath	2938	2300	3563	3620	1316
Fen, marsh and swamp	1331	1080	1483	1480	566
Bog	1060	603	1484	980	554
Standing open water and canals	512	357	601	890	876
Montane habitats	18	8	29	0	16
Inland rock	123	78	139	120	74
Built up areas and gardens	14435	13487	16080	10420	1986
Supralittoral rock	121	57	150	140	78
Supralittoral sediment	162	103	204	240	220
Littoral rock	0				0
Littoral sediment	812	462	955	1180	836

Figure 15. A plot of the Broad Habitat cover estimates for England from calibrated LCM2000 and sample-based field survey as given in Table 12. The error bars show the 95-percentile range for each. The regression equation and R2 values are given on the chart.

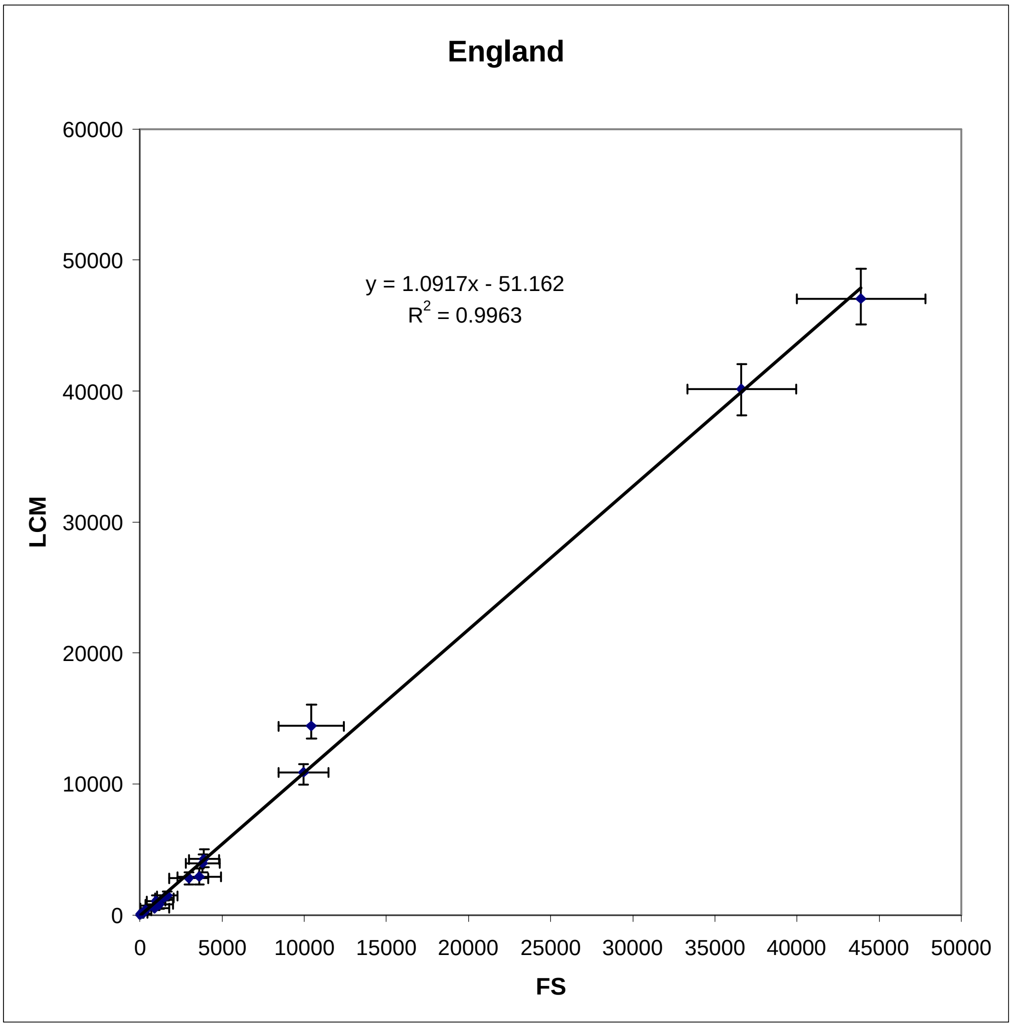


Table 16. Estimates of Broad Habitat cover for Wales from calibrated LCM2000 data: results give the mean of a bootstrapped estimate, with the bias-corrected 95-percentile range. Results are compared with estimates based on the sample-based field survey approach of Countryside Survey 2000: the 2 standard errors is used to estimate the 95-percentile range for the field survey.

	LCM	LCM bias corrected confidence interval		FS	
Broad Habitats	**mean**	**lower**	**upper**	**Total**	**2 x SE**
Broad Habitats	mean	lower	upper	Total	SE
Broadleaved, mixed and yew woodland	2016	1609	2376	1738	517
Coniferous woodland	820	502	1072	818	643
Arable and horticulture	1340	1151	1635	2207	501
Improved grassland	8145	7465	8633	7676	1284
Neutral grassland	590	487	653	533	151
Calcareous grassland	23	8	43	24	27
Acid grassland	1766	1258	2355	1651	769
Bracken	889	558	1237	1073	561
Dwarf shrub heath	963	649	1246	1225	673
Fen, marsh and swamp	578	397	761	618	311
Bog	785	462	1313	814	661
Standing open water and canals	122	101	143	163	121
Montane habitats	7	4	11	0	0
Inland rock	44	14	73	49	36
Built up areas and gardens	1228	1084	1513	1380	409
Supralittoral rock	46	22	63	55	33
Supralittoral sediment	43	21	65	70	61
Littoral rock	0.0				0

Figure 16. A plot of the Broad Habitat cover estimates for Wales from calibrated LCM2000 and sample-based field survey as given in Table 12. The error bars show the 95-percentile range for each. The regression equation and R^2 values are given on the chart.

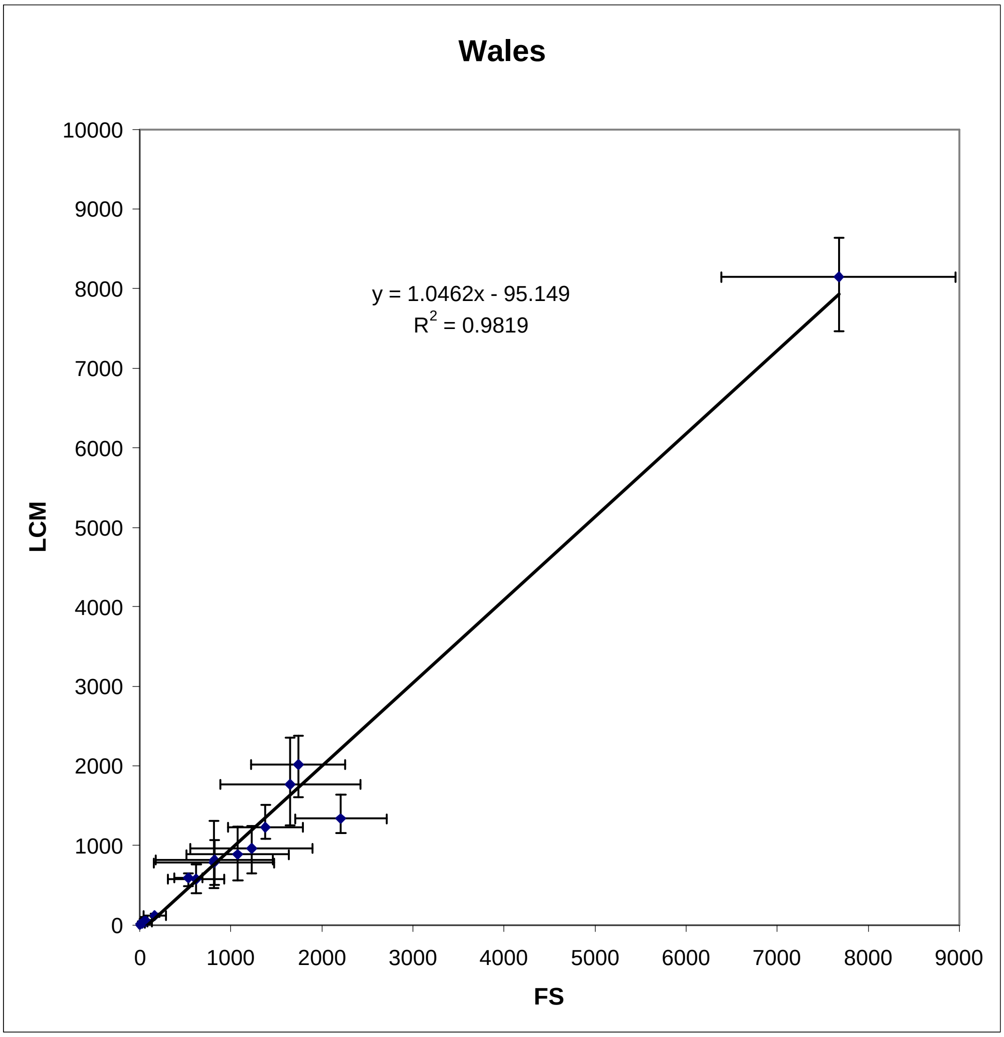
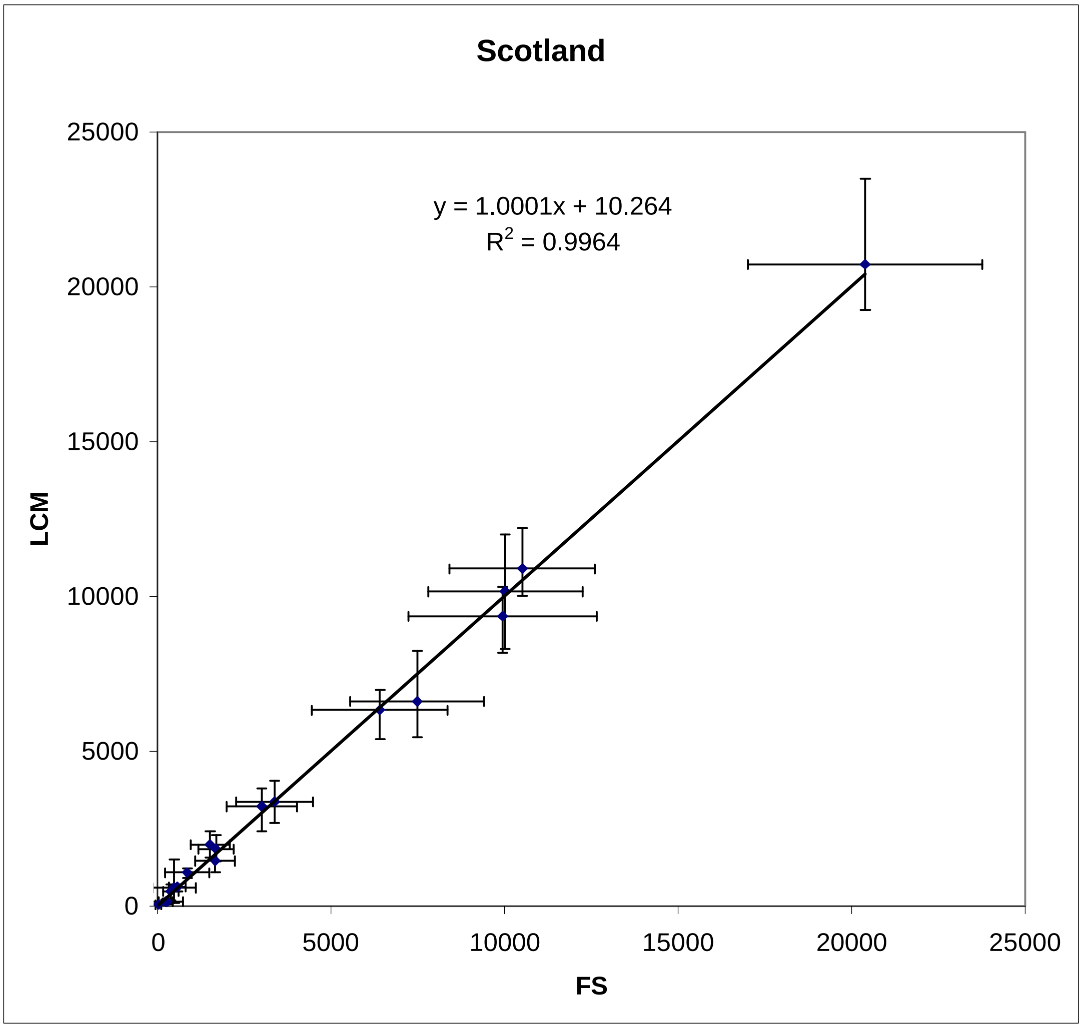


Table 17. Estimates of Broad Habitat cover for Scotland from calibrated LCM2000 data: results give the mean of a bootstrapped estimate, with the bias-corrected 95-percentile range. Results are compared with estimates based on the sample-based field survey approach of Countryside Survey 2000: the 2 standard errors is used to estimate the 95-percentile range for the field survey.

	LCM	**LCM bias corrected confidence interval**		**FS**	
Broad Habitats	**mean**	**lower**	**upper**	**Total**	**2 x SE**
Broadleaved, mixed and yew woodland	3215	2422	3804	3000	1020
Coniferous woodland	9354	8176	10296	9940	2708
Arable and horticulture	6329	5384	6984	6400	1958
Improved grassland	10892	10018	12191	10510	2088
Neutral grassland	1825	1438	2276	1680	502
Calcareous grassland	133	43	232	270	468
Acid grassland	6603	5448	8228	7480	1938
Bracken	1450	1081	1844	1660	568
Dwarf shrub heath	10147	8295	11996	10020	2226
Fen, marsh and swamp	3365	2672	4038	3370	1104
Bog	20733	19248	23495	20390	3368
Standing open water and canals	1075	906	1206	850	632
Montane habitats	581	90	1499	480	618
Inland rock	472	184	699	380	220
Built up areas and gardens	1989	1561	2401	1510	562
Supralittoral rock	615	456	679	570	244
Supralittoral sediment	135	83	189	230	210
Littoral rock	0	0	1		44
Littoral sediment	44	34	61	20	86

Figure 17. A plot of the Broad Habitat cover estimates for Scotland from calibrated LCM2000 and sample-based field survey as given in Table 12. The error bars show the 95-percentile range for each. The regression equation and R^2 values are given on the chart.



The calibrated estimates (Tables 14-17) are very close to direct FS estimates: when LCM2000 BH estimates are plotted (Figures 14-17) against FS estimates for GB and the 3 countries, slopes are near to unity and the intercepts close to the origin (see Table 18), illustrating that there is little bias in the estimation. R^2 values very close to '1' imply a close linear relationship.

Table 18. Slopes, intercepts and R^2 results obtained when calibrated LCM2000 Broad Habitat cover statistics are plotted against field survey estimates of the same cover statistics. Results are derived per-pixel, per-parcel and per-segment for Britain and constituent countries.

		Slope	Intercept	R^2
Great Britain	Per-pixel	0.936	289	0.996
	Per-parcel	0.965	-99	0.990
	Per-segment	0.884	947	0.994
England	Per-pixel	0.913	72	0.996
	Per-parcel	0.937	-112	0.991
	Per-segment	0.878	322	0.996
Wales	Per-pixel	0.939	107	0.982
	Per-parcel	0.978	65	0.967
	Per-segment	0.867	186	0.978
Scotland	Per-pixel	0.996	6	0.996
	Per-parcel	1.019	-98	0.990
	Per-segment	0.921	355	0.987

Because per-pixel estimates give direct equivalence to FS cover, the results which follow are based on per-pixel calibrations. Per-parcel estimates gave similar values, albeit with minor variations due to the bootstrap sampling. Segment-based bootstrapped estimates generate FS coverage but on the LCM2000 spatial framework (i.e. estimating what LCM2000 segments given FS labels would record as coverage). The values are so close that they support almost entirely the original field estimates; however, the LCM2000-based estimates produce tighter confidence limits. The near-universal correspondence is remarkable. While the derived datasets come together in the process of calibration, the measures of coverage are essentially independent. The extrapolation of sampled FS cover statistics via the Land Classes is an entirely different approach to the generation of the same statistics by use of a calibrated comprehensive survey from satellite imagery.

Calibration in Northern Ireland produced mixed results. For 11 BHs estimated by FS and LCM2000 alike, 9 values were closer to FS estimates after calibration, and 2 were apparently worse (Table 19). When these values were plotted (Figure 18) the original slope and intercept of 1.06 and -17 were adjusted to 0.84 and +214 respectively, further from the expected values of 1.00 and 0 respectively; and the R^2 value of 0.79 was marginally worse than the original 0.81. This impact on the regression lines is due substantially to the dominating effect of the points for the very extensive 'improved grasslands'. It is clear, however, that other factors are operating in Northern Ireland and that GB-based calibrations do not apply very effectively. For example, we know that, in GB, LCM2000 over-estimates semi-natural grasslands: calibration against FS serves to moderate the over-estimate and to generate more realistic statistics. In Northern Ireland, semi-natural grasslands are much more commonplace. The uncalibrated LCM2000 estimate for *Acid*, *Calcareous* and *Neutral grasslands* is 2783 km^2 in total; calibrated, the figure is 1459 km^2; the FS gives an estimate of 2830 km^2, much closer to the LCM2000 original. Clearly, calibrations made in GB do not apply in this example in Northern Ireland. It is highly likely that other calibration estimates can also be distorted rather than being improved.

LCM2000 can be used to generate calibrated data for any region, whether administrative or physiographic, within GB. Regional data are given (Table 20) for BH coverage in the six Environmental Zones (EZs) of GB, as used to report CS2000 FS results (Haines-Young *et al.* 2000). There must be serious questions as to whether the small size of the FS sample is adequate to allow estimation of cover statistics below this broad regional level.

13.2 Conclusions regarding calibration

The calibrated statistics, whether at national levels or broad regional scales, probably give CS2000's 'best' current estimates for the BHs at these levels. They certainly record mean values which compare favourably with FS estimates, but with confidence limits which are much tighter (about half the FS range). They take full account of the heterogeneity of the UK landscape and base the cover-estimates on comprehensive survey. While such a survey might itself show inaccuracies through satellite mapping, the calibration to the FS ensures that significant under- and over-estimates are taken into account. The close agreement of the statistics has much further ramifications: the weakness of the FS has been its inability to make reliable estimates of cover at local to regional levels. LCM2000 offers potential to do better, as it is providing site-specific survey. At Aggregate class level, LCM2000 is, even without calibration, likely to give results which would match a field survey recorded at this level of detail, albeit one with the lesser spatial resolution of LCM2000. At Target class level, LCM2000 is likely, in general, to be 85% correct, but with the potential, locally, for gross errors. LCM2000 calibrated against field data may give scope for local and regional census to BH standards. However, the calibration is only based on the 40 Land Class regions and it is likely that regionally-based calibrations cannot apply below, or perhaps even at, the regional level: similar problems to those seen with the Northern Ireland calibration are likely. Nonetheless, the local values are surveyed rather than estimated and hence are much more reliable than would be the equivalent field-based estimates. Thus local results are likely to very sound for Aggregate classes and reasonably so for Target classes. Whether there is scope to produce BH cover estimates at local level needs to be tested independently. It was not possible within the scope of LCM2000 production and reporting to validate this approach. However, such work is envisaged in follow-up research on integration.

Table 19. Coverage of Broad Habitats in Northern Ireland from raw (uncalibrated) LCM2000; calibrated using the calibration matrix derived for the UK; compared with Northern Ireland Countryside Survey cover estimates. The calibrated LCM2000 statistic is judged 'better' or 'worse' if it is closer to or further from the field survey estimate than the uncalibrated original.

Broad Habitats	**LCM raw total**	**LCM calibrated mean**	**NICS FS Total**	**Calibrated statistic**
Broadleaved, mixed and yew woodland	253	691	510	better
Coniferous woodland	520	569	610	better
Arable and horticulture	908	1654	590	worse
Improved grassland	7005	5772	5680	better
Neutral grassland	1073	526	2540	worse
Calcareous grassland	464	51	10	better
Acid grassland	1246	882	280	better
Bracken	17	253	40	better
Dwarf shrub heath	1042	713	130	better
Fen, marsh and swamp	0	416	530	better
Bog	440	1272	1480	better

Figure 18. A plot of the Broad Habitat cover estimates for Northern Ireland from calibrated LCM2000 and sample-based field survey as given in Table 17. The regression equation and R2 values are given on the chart.

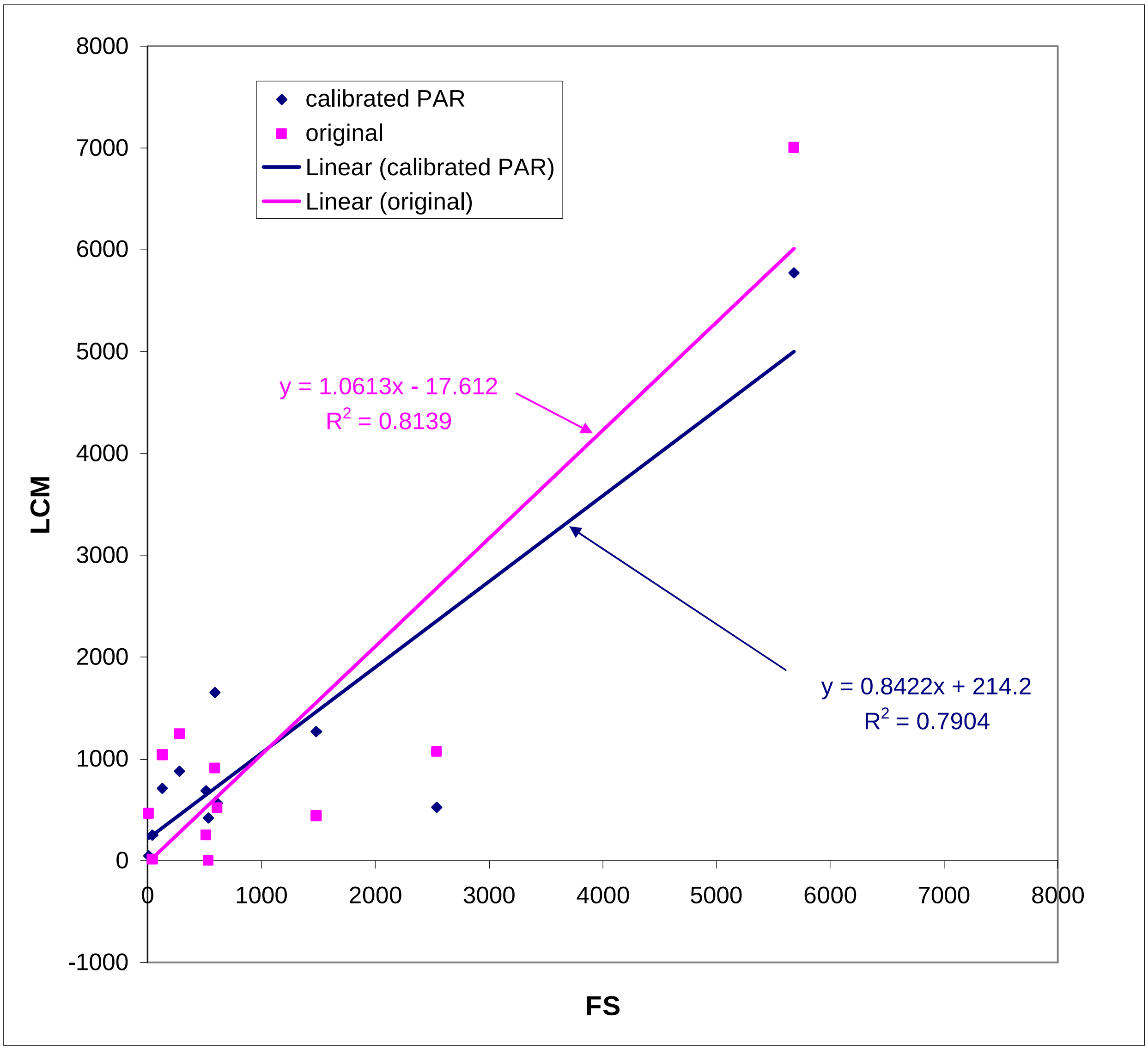


Table 20. Broad Habitat cover statistics (km^2) for the 6 Environmental Zones of CS2000, derived by calibration of LCM2000 cover to CS2000 field survey; also statistics estimated directly from the field survey with the ranges based upon the FS means plus or minus 2 standard errors

	EZ1					**EZ2**				
	LCM sample mean	LCM bias corrected 95% confidence lower	upper	FS lower	FS upper	**LCM** sample mean	LCM bias corrected 95% confidence lower	upper	FS lower	FS upper
Broadleaved, mixed and yew woodland	**6185**	5599	6911	4315	6937	**4737**	4216	5166	3642	5374
Coniferous woodland	**828**	550	995	341	1024	**1209**	866	1452	609	2639
Arable and horticulture	**33253**	31630	35034	29381	36186	**14505**	13124	16215	10507	15040
Improved grassland	**15252**	13611	16581	10815	15624	**25275**	23932	26888	21442	26142
Neutral grassland	**1955**	1540	2530	1147	2557	**1971**	1534	2350	1242	2316
Calcareous grassland	**296**	28	666	-51	584	**124**	57	253	1	221
Acid grassland	**317**	82	728	-56	538	**494**	300	668	219	972
Bracken	**146**	99	174	46	262	**613**	388	716	162	1593
Dwarf shrub heath	**72**	29	93	-42	286	**387**	152	660	42	1070
Fen, marsh and swamp	**256**	147	367	63	354	**740**	548	841	205	1447
Bog	**62**	28	100	-12	102	**54**	40	70	-78	496
Standing open water and canals	**259**	166	343	-213	1516	**191**	124	260	55	294
Montane habitats	**0**			0	0	**6**	2	10	0	0
Inland rock	**76**	58	81	7	142	**16**	6	31	4	28
Built up areas and gardens	**7093**	6203	8387	3518	6726	**8230**	7422	9424	4879	7915
Supralittoral rock	**0**			0	0	**217**	113	285	83	304
Supralittoral sediment	**1**	1	1	-1	3	**247**	156	329	24	583
Littoral rock	**0**			0	0	**0**			0	**0**
Littoral sediment	**94**	85	120	-296	862	**943**	517	1124	337	1817

	EZ3					**EZ4**				
	LCM sample mean	LCM bias corrected 95% confidence lower	upper	FS lower	FS upper	**LCM** sample mean	LCM bias corrected 95% confidence lower	upper	FS lower	FS upper
Broadleaved, mixed and yew woodland	**1949**	1454	2366	996	2147	**1394**	1166	1574	874	1486
Coniferous woodland	**1582**	1173	2128	527	2466	**2000**	1504	2460	768	2649
Arable and horticulture	**622**	364	957	185	889	**5447**	4669	6106	3649	7065
Improved grassland	**7575**	6539	8087	5655	8938	**6793**	6157	7632	5237	7957
Neutral grassland	**933**	639	1304	435	1185	**1263**	971	1626	818	1570
Calcareous grassland	**1**	0	2	0	2	**0**			0	0
Acid grassland	**4930**	4198	5908	3384	5890	**841**	515	1112	421	1257
Bracken	**1641**	1197	2204	1040	2358	**433**	235	640	188	691
Dwarf shrub heath	**3444**	2620	4097	2707	5633	**563**	286	796	104	1706
Fen, marsh and swamp	**939**	757	1196	683	1451	**723**	477	885	378	1048
Bog	**1730**	1172	2523	704	2385	**963**	623	1435	304	2040
Standing open water and canals	**183**	159	202	-41	500	**81**	35	117	-14	137
Montane habitats	**20**	10	32	-6	25	**6**	3	10	0	0
Inland rock	**78**	34	124	29	134	**279**	11	450	-10	372
Built up areas and gardens	**380**	272	464	157	407	**1566**	1130	1936	607	1659
Supralittoral rock	**2**	0	5	-2	5	**51**	38	66	2	107
Supralittoral sediment	**0**	0	1	0	1	**78**	21	128	-65	328
Littoral rock	**0**			0	0	**0**			**0**	0
Littoral sediment	**0**			0	0	**5**	2	9	-1	9

	EZ5					**EZ6**				
	LCM sample mean	LCM bias corrected 95% confidence lower	upper	FS lower	FS upper	**LCM** sample mean	LCM bias corrected 95% confidence lower	upper	FS lower	FS upper
Broadleaved, mixed and yew woodland	**814**	528	990	323	1094	**1055**	435	1684	221	2007
Coniferous woodland	**2947**	2288	3542	1892	5127	**4136**	3490	4867	2759	6674
Arable and horticulture	**722**	508	920	41	1951	**49**	9	104	-3	87
Improved grassland	**2908**	2364	3540	1613	4368	**1135**	606	1656	137	1705
Neutral grassland	**431**	236	612	61	700	**130**	57	197	22	198
Calcareous grassland	**155**	41	270	-201	737	**1**	0	2	-1	3
Acid grassland	**1191**	859	1475	921	2250	**4564**	3415	6225	3286	6829
Bracken	**594**	405	843	283	1062	**450**	253	635	219	877
Dwarf shrub heath	**2325**	1815	3067	1439	2969	**7388**	5421	8975	4983	8842
Fen, marsh and swamp	**1767**	1242	2391	786	2745	**828**	533	1098	510	1275
Bog	**8911**	8181	10074	6610	11068	**10794**	9407	12984	8003	12744
Standing open water and canals	**433**	307	556	125	591	**545**	447	596	-154	1010
Montane habitats	**34**	14	229	-6	12	**544**	62	1386	-139	1096
Inland rock	**25**	12	61	25	130	**166**	98	295	31	219
Built up areas and gardens	**310**	198	427	118	479	**101**	37	195	-7	158
Supralittoral rock	**599**	396	708	246	702	**102**	82	155	-25	117
Supralittoral sediment	**83**	54	111	20	164	**10**	10	10	-2	6
Littoral rock	**1**	0	2	0	**0**	**0**			0	0
Littoral sediment	**13**	0	36	-27	61	**0**			0	0

14. UK LAND COVER

Summary statistics for UK land cover are given in Figure 19a-e. It can be seen from these that more than 54% of the UK is used for intensive agriculture (**Arable and horticultural** land or **Improved grassland**) or is developed (**Built up land**). The remaining 46% is largely semi-natural. Woodlands occupy 25% of this semi-natural land with **Broadleaved woodland** and **Coniferous woodland** about equal in extent. **Mountain, heaths and bog** cover 34% of the low intensity land; semi-natural grass swards (including rougher examples of improved swards) form 36% of all semi-natural. **Coastal** habitats and **Open water**, while important, are small in extent.

The four countries of the UK differ markedly from each other. Intensive use for agriculture or development affects nearly three-quarters of England, about two-thirds of Northern Ireland and about half of Wales. In Scotland, less than a quarter is intensively farmed or developed.

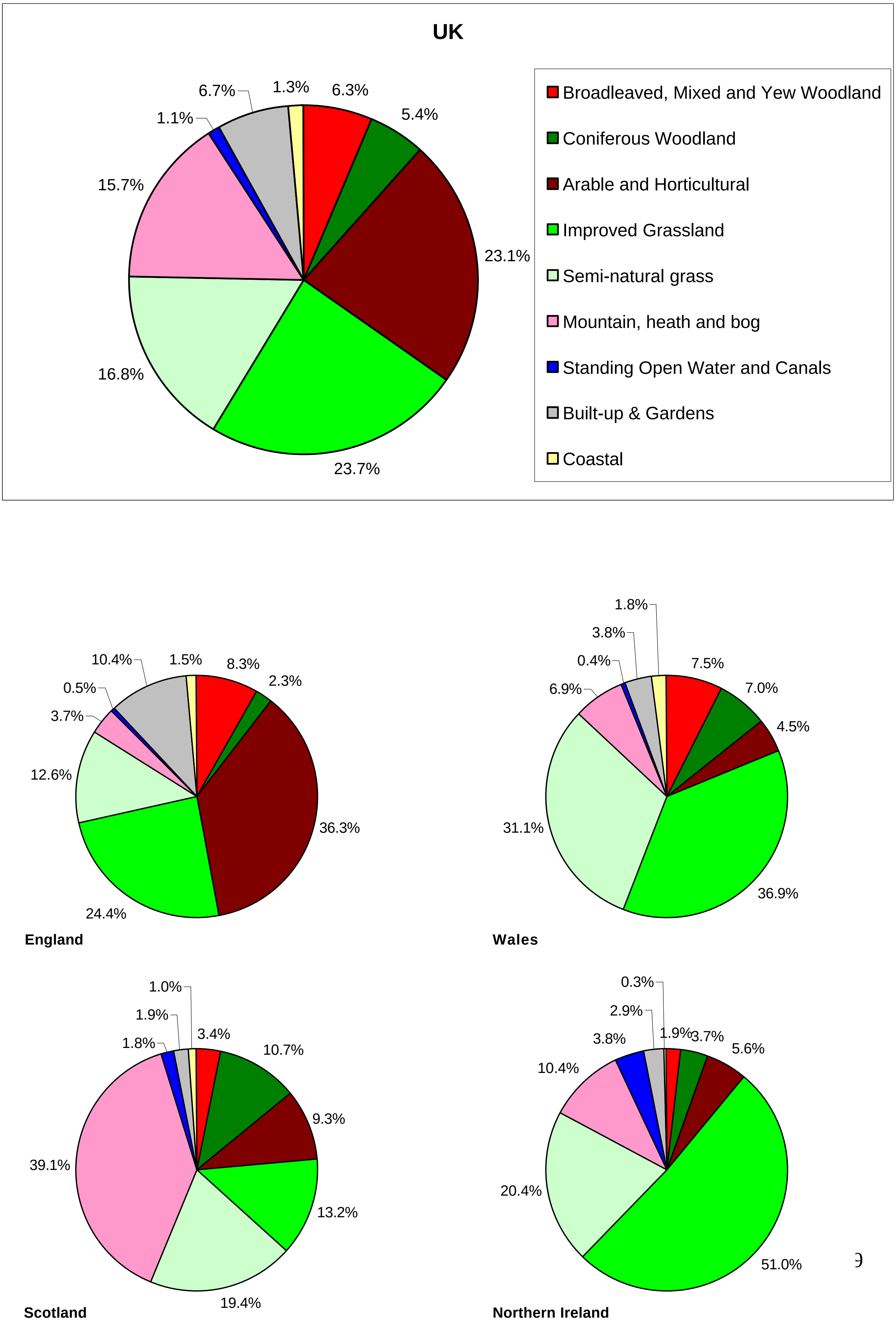
The semi-natural land of England is evenly split between woodlands and grasslands. In Wales, the balance is similar but with a far greater overall extent. Scotland is dominated by **Mountain, heath and bog** which makes up more than half of all its semi-natural land. Northern Ireland also has reasonably extensive **Mountain, heath and bog** and **Semi-natural grass** but, at the resolution of LCM2000 which excludes linear examples, is notably short of woodland cover.

15. CHANGE DETECTION

Landscape changes interest many users. The measurement of such changes demands high levels of precision to map real differences and to distinguish them from localised errors. Changes between LCMGB 1990 and LCM2000 were probably relatively small - FS suggests 17% at BH level - and detectable changes would generally have been exceeded by combined error rates. In a comprehensive National survey, the necessary precision for change detection cannot be achieved consistently by satellite-based mapping alone. The LCM2000 classification rightly sought to remove known deficiencies in the 1990 classification and to bring field and satellite surveys into closer match, even though the detection of change would be compromised. The segment-based approach of LCM2000 generated different results from the 1990 raster product. The classification based on BHs precluded direct comparison with 1990 classes.

Nevertheless, there will have been real changes in the period 1990-2000; and it may be possible to select intelligently, from those differences mapped, the elements which are attributable to change and those attributable to error and / or differences in the data products. The way to advance this work will be to use more intelligent approaches. The FS of 1990 and 1998 provided a measure of the expected directions and rates of change (Haines-Young *et al.* 2000). An intelligent approach might use these data. Calibration results identify LCM2000 under-estimates and over-estimates in 2000 which should be taken into account in analyses of change. The probabilities of classification recorded in LCM2000 point to possible errors in classification. All such clues could be used to select apparent changes which fit the known patterns of change. This approach will be the subject of research and development, beyond the scope of the production phase.

Figure 19. Pie charts showing the cover of Aggregate classes (uncalibrated statistics) for a. the UK and b-e. its constituent countries from LCM2000.



16. CONCLUSIONS

1. LCM2000 has, for the first time, mapped land cover for the whole of the UK from satellite images.
2. A period with a high incidence of cloudy weather in the target year of 1998 and subsequently doubled the number of images required to complete cloud-free coverage of the UK.
3. The procedures of LCM2000 proved robust and helped to overcome the difficulties of part-clouded data and the use of many different satellite images to build the complete UK mosaic.
4. A range of pre-processing procedures, new to LCM2000, gave improved data for classification. However, the use of haze correction was unable to counteract the worst effects in some images. The use of illumination correction for terrain-induced effects did not (and could not have) compensated for extremely low light levels in mid-winter coverage of north-facing slopes in more northerly latitudes.
5. The segmentation procedure proved robust and effective in segmenting the parcel-based structure of the UK landscape. Perhaps more surprisingly, segmentation was remarkably good at picking out urban, suburban and developed rural areas in the landscape. This ability to detect the major heterogeneity - the presence of the developed land - while aggregating the component pixels of the heterogeneous development parcel into a sensible structure, applied not only to the obvious examples of developed land but also applied when subdividing mosaics of cover in unenclosed semi-natural landscapes.
6. The training procedure based upon segments proved more efficient and more objective than normal manual methods for delineation of training areas. The review procedure based on ‘colour charts’, showing the spectral characteristics of training areas, allowed the construction of soundly-based spectral subclasses.
7. The spectral classification was used to build 72 thematic classes. These Variants were then aggregated at various levels: there are 26 Subclasses, mapped consistently across the UK. From these it is possible to simulate the 20 Broad Habitats, though with some departures from the standard definitions. The Subclasses also aggregated to 16 Target classes, mapped with consistency and a high level of accuracy. The BH classification and the Target classification give common classes at the Aggregate class level. Map display classes are those Target classes and Subclasses which are mapped reasonably accurately, are sufficiently widespread to show at the national scale and serve to bring out the patterns in intensively used and semi-natural landscapes.
8. Calibration suggests that LCM2000 maps Target classes with an accuracy level of >85%. This is probably less than the intended 90% accuracy, a consequence of the large quantity of imagery ‘off-target’ in terms of dates. Lesser accuracy is achieved for BHs. Indeed some BHs differ markedly from field survey ‘equivalent’ classes. The differences reflect the problems in defining BHs based upon physical properties of the substratum (soil type, pH and water content) and those best recognised by indicator species.
9. The process of calibration allows LCM2000 mapping to generate calibrated BH statistics with direct equivalence to field-surveyed BHs. It is believed that the comprehensive coverage of LCM2000 and the calibration based on field survey detail gives the ‘best’ estimates of BH cover and distribution. It is also believed that the combination of field detail and satellite coverage give scope for the generation of BH statistics down to the local level, something that neither survey could achieve on its own. This needs to be tested more fully.
10. LCM2000 is being made available to users at various levels of detail. The ‘Level-2 Dataset’ comprises ArcView ‘Shape Files’ as a topologically structured vector format carrying attributes which identify individual parcels, give a count of total pixels and core pixels, plus a process history descriptor (with a scene-identifier, probability indices and KBC rules applied). The Level-2 classification gives Target classes and Subclasses (26 types) coded to relate to BHs. The Level-3 Dataset gives details down to Variant level (72 types overall), also coded to relate

to BHs. The Level-3 dataset includes extra details on heterogeneity, recording the top five per-pixel classes and their fractional cover within the segment.

11. A ‘Raster Dataset’ derived from the Subclasses of the Level-2 Dataset gives a 25 m raster map based on per-parcel classification, and generalised accordingly.
12. The Raster Dataset is further generalised as 1 km products summarising: class dominance at the 26 Subclass level in a single layer dataset; also summary % cover per 1 km^2 for the 26 Subclass types (i.e. a 26-layer dataset) and Aggregate classes (a 10-layer dataset) in CIS format.
13. LCM2000 offers a data structure which can satisfy wide ranging user-needs. The vector-format of the data records the ‘real’ structure of the landscape, based upon land use and land cover parcels. It offers so much more scope than the conventional per-pixel products of the earlier LCMGB. With retention of all segments >0.5 ha, it offers far better resolution than other satellite-derived vector products such the manually mapped European CORINE Land Cover dataset with its 25 ha MMU.
14. LCM2000 records the structural patterns of the landscape and shows the spatial inter-relations of parcels and habitats. It therefore lends itself much better to applied uses where patterns affect processes, and an understanding the spatial inter-relations is a necessary precursor to understanding such processes and predicting their consequences.

17. GLOSSARY OF TERMS AND ACRONYMS

Acid-sensitivity map: (Hornung *et al.* 1995) used to label semi-natural grasslands as probably 'acid', 'neutral' or 'calcareous'; this map defined acidity sensitivity classes as highly sensitive - pH <4.5 (i.e. truly acid), moderately sensitive - pH >4.5 and <5.5. (treated for KBC purposes as neutral but really slightly acid) and low sensitivity - pH >5.5 (really including neutral and calcareous components).
Aggregate classes: 10 in number, combining Target classes and Subclasses to a simplified 10-class level where the resultant 'classes' compare closely with equivalent Broad Habitat-aggregations; at this level, maps and statistics from LCM2000 and field survey broadly coincide. Thus Aggregate classes are used for reporting purposes.
Attribute: refers here to a data item, held in the geographical information system, recording information about a GIS object; an attribute may be a numerical value (e.g. altitude), an alphabetical code (e.g. Aw = arable (wheat)) or a text string labelling or describing the parcel (e.g. the processing history descriptor).
BAP: UK Biodiversity Action Plan.
Bootstrapping: a statistical technique used to generate mean statistics and confidence intervals from sampled data. The bootstrap approach is based on the principle that, in the absence of any other knowledge about a population, the distribution of values found in a random sample is the best guide to the distribution in the population. The data values are sampled with replacement and the statistics computed through many iterations. These computed values, the bootstrap sample, are used to estimate properties of the statistic such as confidence intervals.
Broad Habitats: a classification by the UK Biodiversity Group to encompass the entire range of UK habitats as an aid to the implementation of, and reporting under, the UK Biodiversity Action Plan.
CIS: see Countryside Information System.
Class Variants: 72 in number, are a thematic aggregation of spectral subclasses and components of Target classes and Subclasses.
Core pixels: those pixels of a segment which extracted after shrinking the segment geometry to avoid edge-pixels; they were used in deriving training statistics and/or in deriving a segment's mean reflectance values for use in classification.
Countryside Information System: a Microsoft Windows-based program developed to give policy advisers, planners and researchers easy access to spatial information about the British countryside, especially Countryside Survey data.
CS1990: Countryside Survey 1990
CS2000: Countryside Survey 2000.
DTM: digital terrain model.
ETM: Enhanced Thematic Mapper, a sensor on the satellite, Landsat 7, recording visible and infrared reflectances.
EZ: Environmental Zone, a simplification of the 40 Land classes of Britain into 6 zones, with Northern Ireland forming a seventh.
FS: field survey (here referring specifically to the field element of Countryside Survey 2000).
GIS: geographical information system.
IRS: Indian Research Satellite
KBC: Knowledge-based correction.
Land classes: a stratification of the landscape of Britain to ensure that the sample field survey of CS2000 is representative of the range of different environments found in England, Wales and Scotland. There were 40 strata in CS2000, based upon the original 32 ITE Land Classes of CS1990; in 2000, the original 32 were subdivided to give classes which were specific to the 3 countries (i.e. individually, they did not cross borders): the resultant 40 classes (Haines-Young *et al.* 2000) are thus sometimes called National Land Classes.
LCM2000: Land Cover Map 2000.

LCMGB: Land Cover Map of Great Britain (the survey of 1988-92 which was part of Countryside Survey 1990).
Level-2 Dataset: ARC/View 'Shape Files' which record a topologically structured vector dataset carrying the following attributes: a unique segment label; a total pixel count; a count of core pixels; a classification giving Target class / Subclass (26 types) coded to relate to the widespread BH; a process history descriptor (with scene-identifier, probability indices and KBC rules applied).
Level-3 Dataset: ARC/View 'Shape Files' which record a topologically structured vector dataset carrying the following attributes: a unique segment label; a total pixel count; a count of core pixels; a classification giving Target class / Subclass / Variant (72 types overall) coded to relate to the widespread BH; a process history descriptor (with scene-identifier, probability indices and KBC rules applied).
LISS: the Linear Imaging Self-scanning Sensor of the Indian Research Satellite.
MIR: middle infrared.
MMU: Minimum mappable unit. LCM2000 retains all segments with 9 or more pixels (on the basis that a 3 x 3 pixel segment can contain a 'pure' core pixel); segments with ≤8 pixels (≤0.5 ha in area) are 'dissolved' into surrounding segments, with each pixel individually attached to the neighbouring segment which was most similar in spectral character.
NIR: near infrared.
Object: an item, in fact a polygon, in the LCM2000 GIS database.
Parcel: a parcel (sometimes called specifically a 'land parcel') is an area on the ground, often a field but perhaps with no boundary such as a woodland, a patch of relatively uniform vegetation (e.g. heath), a built up area, or a water body.
Peat drift: includes all superficial strata of peat, based on a British Geological Survey drift map (http://www.bgs.ac.uk/products/digitalmaps/digmapgb_drift.html). Peat drift is used to define bogs (i.e. with peat depth >0.5 m) as distinct from heaths and moors (with peat < 0.5 m).
Per-parcel comparisons: where field survey parcels and their classes were compared with a class-label for the same parcel derived from LCM2000.
Per-pixel comparisons: between field survey maps and LCM2000 maps - a direct overlay, with no regard to the structure of either dataset.
Per-segment comparisons: where labels in LCM2000 segments are compared with the segment's dominant class according to field survey.
Raster: a grid-based data structure used in GIS and image analysis systems.
Raster Dataset: specifically here the derived dataset from the Subclasses of the Level-2 Dataset giving a 25 m grid-based map incorporating the spatial refinements of LCM2000 (i.e. based on per-segment classification, and generalised accordingly).
Segment: the spectrally-defined aggregation of image pixels into a vector polygon, held in the GIS database; the term 'segment' is intended to distinguish the resultant feature from a 'land parcel' which refers to the actual feature on the ground. Often, segments record parcels, but with differences which relate to the underlying 25 m structure of the image.
Spectral classes: cover types with distinct spectral signatures: for example, 'shaded north-facing' and 'sunlit south-facing' are two distinct spectral classes of 'improved grass'; chalk, clay and peat soil background give different spectral classes of 'wheat'. Spectral classes may differ according to species content, crop variety, phenology, management practices, atmospheric haze, cloud shadow or any other factor with an impact on the recorded spectral reflectance.
Subclasses: 26 in number, an aggregation of the thematic class **Variants**; they were defined to give, as far as possible, the full complement of Broad Habitats; they also give details beyond the BH classification.
Summary Products: several products at 1 km: class dominance at the 26 Subclass level, stored as a single layer dataset; summary % cover per 1 km^2, for the 26 Subclass types (i.e. a 26-layer dataset); class dominance dataset in CIS format; the Subclass dataset in CIS format; an Aggregate class dataset in CIS format.

Summer target period: the main growing season for arable crops, from mid-May to late July in southern Britain, or later in Scotland, excluding May but continuing into August; 1998 was the first choice year, then 1999, 2000, and 2001.

Target classes: 16 in number, giving the nearest match to Broad Habitats which could be achieved consistently and with a high level of accuracy; where there are fundamental differences in the exact definitions of Broad Habitats and Target classes this is reflected by differences in the selected Subclass nomenclature.

Target period: see **‘**Summer target period’ and ‘Winter target period’.

Training: the procedure by which a sample of known cover types is defined in the image processing system to deduce the spectral characteristics of classes, to form the basis for automatic extrapolation by the system to classify examples of unknown land cover.

TM: Thematic Mapper, a sensor on the Landsat 5 satellite, recording visible and infrared reflectances.

Variants: see **class Variants**.

Vector: a digital line held as a series of x-, y-coordinates in a geographical information system.

Winter target period: from the time of the first frosts (about October) to late April in southern Britain and well into May in the Scottish Highlands (i.e. until deciduous trees were in full leaf); the target winter was from Autumn 1997 to Spring 1998, but winters of 1998-99, 1999-00 and 2000-01 were also considered.

18. REFERENCES

LCM2000 Reports (listed in chronological order)

1998

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

1999

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. Unpublished ITE report to the LCM2000 Consortium.

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.* Unpublished CEH report to the LCM2000 Consortium.

Fuller R.M., Smith, G.M., Sanderson, J.M., Hill, R.A., Thomson, A.G. & Clarke, R.T. 1999. *Countryside Survey 2000 - Part III. Module 7. Land Cover Map 2000. Third Interim Report incorporating the Sixth Quarterly Progress Report, CSLCM/Int3/Prog6.* Unpublished ITE report to the LCM2000 Consortium. 36pp.

Fuller R.M., Smith, G.M., Sanderson, J.M., Hill, R.A., Thomson, A.G. & Hall, M.W. 1999. *Countryside Survey 2000. Module 7. Land Cover Map 2000. Seventh Quarterly Progress Report, CSLCM/Prog7.* Unpublished CEH report to the LCM2000 Consortium.

Fuller, R.M., 1999. Module 7. Land Cover Map 2000. 3pp. In: *Countryside Survey 2000 Module Progress Reports, December 1999.* Unpublished ITE report to the DETR.

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

Fuller, R.M., 1999. Module 7. Land Cover Map 2000. 3pp. In: Hornung, M. (ed.) *Countryside Survey 2000 Third Integrated Progress Report*. Unpublished ITE report to the DETR.

Fuller, R.M., Smith, G.M. & Hill, R.A., 1999. *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).

2000

Fuller, R.M., Smith, G.M., Sanderson, J.M., Hill, R.A., Hall, M., Thomson A.G. & Brown N.J. 2000. *Countryside Survey 2000 Module 7. Land Cover Map 2000. Fourth Interim Report* *incorporating the Eighth Quarterly Progress Report. CSLCM/Int4/Prog8.* Unpublished ITE report to the DETR. 32pp

Fuller, R.M., Smith, G.M., Sanderson, J.M., Hill, R.A., Thomson, A.G., Gerard, F.F. & Hall, M.W. 2000. *Countryside Survey 2000 - Part III. Module 7. Land Cover Map 2000. Ninth Quarterly Progress Report, CSLCM/Prog9.* 5pp.

Fuller R.M., Smith, G.M., Sanderson, J.M., Hill, R.A., Thomson, A.G., F.F. Gerard & Clarke, R.T. 2000. *Countryside Survey 2000 - Part III. Module 7. Land Cover Map 2000. Fifth Interim Report. incorporating the Tenth Quarterly Progress Report, CSLCM/Int5/Prog10.* Unpublished CEH report to the LCM2000 Consortium. 30 pp.

Fuller, R.M., 2000. Module 7. Land Cover Map 2000. 3pp. In: *Countryside Survey 2000 Module Progress Reports, February 2000.* Unpublished ITE report to the DETR.

Fuller, R.M., 2000. Module 7. Land Cover Map 2000. 3pp. In: *Countryside Survey 2000 Module Progress Reports, April 2000.* Unpublished ITE report to the DETR.

Fuller, R.M., 2000. Module 7. Land Cover Map 2000. 3pp. In: *Countryside Survey 2000 Module Progress Reports, June 2000.* Unpublished ITE report to the DETR.

2001

Fuller R.M., Smith, G.M., Sanderson, J.M., Hill, R.A., Thomson, A.G., Gerard, F.F. & Clarke, R.T. 2001. *Countryside Survey 2000 - Part III. Module 7. Land Cover Map 2000. Sixth Interim Report. incorporating the Eleventh Quarterly Progress Report, CSLCM/Int6/Prog11.* Unpublished CEH report to the LCM2000 Consortium. 18 pp.

Fuller, R.M., 2001. Module 7. Land Cover Map 2000. 4pp. In: *Countryside Survey 2000 Module Progress Reports, January 2001.* Unpublished ITE report to the DETR.

Fuller, R.M., 2001. Module 7. Land Cover Map 2000. 4pp. In: *Countryside Survey 2000 Module Progress Reports, March/April 2001.* Unpublished ITE report to the DETR.

Fuller, R.M., 2001. Module 7. Land Cover Map 2000. 4pp. In: *Countryside Survey 2000 Module Progress Reports, May-August 2001.* Unpublished ITE report to the DETR.

LCM2000 and related publications (alphabetical order)

Carey, P.D., Brown, N.J., Singer, E.J., Hulmes, S., Scott, R.J., Smith, G. and Hill, M.O., 2000, *The practical constraints on securing locations for the farm-scale evaluations imposed by increased separation distances between non-GM and GMHT crops.* Report to DETR.

Fuller, R.M., Smith, G.M., Sanderson, J.M., Hill, R.A., & Thomson, A.G. In press. Land Cover Map 2000: a general description of the UK's new vector GIS based on classification of remotely sensed data. *Cartographic Journal.*

Haines-Young, R.H., Barr , C.J., Black, H.I.J., Briggs, D.J., Clarke, R.T., Cooper, A., Dawson, F.H., Firbank, L.G., Fuller, R.M., Furse, M.T., Gillespie, M.K., Hill, R., Hornung, M., Howard, D.C., McCann, T., Morecroft, M.D., Petit, S., Sier, A.R.J., Smart, S.M., Smith, G.M., Stott, A.P., Stuart, R.C., & Watkins, J.W. 2000. *Accounting for nature: assessing habitats in the UK Countryside.* DETR, London.

Smith, G.M. & Fuller R.M., 2000, Land Cover Map 2000: more than just an update. *SoC Bulletin,* **33** (2), 13-16.

Smith, G.M. & Fuller, R.M., In press., Land Cover Map 2000: A Data Resource For Scotland.

Smith, G.M. and Fuller, R.M., 2001, Land Cover Map 2000 and meta data at the land parcel level. *Proceedings of the RSPS meeting Uncertainty in Remote sensing and GIS, University of Southampton. 3-4 July, 2001.* (Book chapter)

Smith, G.M. and Fuller, R.M., In press. Land Cover Map 2000 and meta data at the land parcel level. *Proceedings of the RSPS meeting Uncertainty in Remote sensing and GIS, University of Southampton. 3-4 July, 2001.* (Book chapter)

Smith, G.M. In press. Land Cover Map 2000: a census of the UK countryside. IGGI News

Smith, G.M. In press. Land Cover Map 2000: a census of the UK from space. NERC News
Smith, G.M., Fuller, R.M. In press**.** Land Cover Map 2000: a data resource for Scotland. In: E. Mackey (ed) *The state of Scotland's environment and natural heritage.* Edinburgh: SNH.
Smith, G.M., Fuller, R.M., Hoffmann, A. & Wicks, T., 2000, Parcel-based approaches to the analysis of remotely sensed data. *Proceedings of the Remote Sensing Society Conference, Adding Value to Remotely Sensed Data,* Remote Sensing Society, Nottingham, CD-ROM.
Smith, G.M., Fuller, R.M., Sanderson, J.M., Hill, R.A. & Thomson, A.G., 2001, Land Cover Map 2000: a parcel-based map from satellite images. *Proceedings of the Remote Sensing and Photogrammetry Society Conference,* Remote Sensing Society, Nottingham.

Other references (alphabetical order)

Barr, C.J. 1998. *Countryside Survey 2000: Field Handbook. 3rd Draft.* Institute of Terrestrial Ecology, unpublished.
Ebron, B. & Tibshirani, R.J. 1998. *An introduction to the bootstrap.* Chapman & Hall: London.
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., Groom, G.B. & Jones, A.R. 1994a. 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., Groom, G.B. & Wallis, S.M. 1994b. The availability of Landsat TM images for Great Britain. *International Journal of Remote Sensing*. **15**, 1357-1362.
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.
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.
Hornung, M., Bull, K.R., Cresser, M., Ullyett, J., Hall, J.R., Langan, S., Loveland, P.J. & Wilson, M.J. 1995. The sensitivity of surface waters of Great Britain to acidification predicted from catchment characteristics. *Environmental Pollution*, **87**, 207-214.
Jackson, D.L. 2000. *JNCC Report No. 307 Guidance on the interpretation of the Biodiversity Broad Habitat Classification (terrestrial and freshwater types): definitions and the relationships with other habitat classifications.* Joint Nature Conservation Committee, Peterborough.
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.
Liang, S., Fallah-Adl, H., Kalluri, S., JaJa, J., Kaufman, Y.J. & Townshend, J.R.G. 1997. An operational atmospheric correction algorithm for Landsat Thematic Mapper imagery over the land. *Journal of Geophysical Research*, 102, (D14), 17173-17186.
Mather, P. 1987. Computer processing of remotely sensed images – and introduction. Chichester: Wiley.
Ratcliffe, D.A. & Thompson, D.B.A. 1988. The British uplands: their ecological character and international significance. In: *Ecological change in the uplands,* pp 9-36,edited by M.B. Usher and D.B.A. Thompson. Blackwell Scientific Publications, Oxford.
Schowengerdt, R.A. 1997. *Models and methods for image processing.* London, Academic Press.
Smith, G.M., & Fuller, R.M., 2001. Multi-sensor, high resolution, knowledge-based per-parcel classification of land cover: an example in the Island of Jersey. *International Journal of Remote Sensing.* **22**, 3123-3142.
Smith, G.M., Fuller, R.M., Amable, G., Costa, C. and Devereux, B.J., 1997, CLEVER-Mapping: An implementation of a per-parcel classification procedure within an integrated GIS environment. *Proceedings of the Remote Sensing Society conference, Observations and Interactions: RSS97,* Remote Sensing Society, University of Nottingham, 21 - 26.

19. ACKNOWLEDGEMENTS

Land Cover Map 2000 was funded by a consortium comprising:

- The Countryside Council for Wales;
- The Department for the Environment, Food and Rural Affairs;
- The Department of Agriculture and Rural Development, Northern Ireland;
- The Environment Agency;
- The Environment and Heritage Service of the Department of the Environment, Northern Ireland;
- The National Assembly for Wales;
- The Natural Environment Research Council;
- Scottish Natural Heritage;
- The Scottish Executive.

The British National Space Centre helped fund methodological research, undertaken jointly by CEH, Cambridge University Department of Geography (CUGD), Laser-Scan Limited and the Ordnance Survey within the CLEVER-Mapping programme; this became the basis of LCM2000. Laser-Scan Limited provided licences for use of its integrated geographical information system ‘IGIS’. Bernard Devereux and Carlos Costa of CUGD produced software for segmentation and for illumination correction. Ordnance Survey of Northern Ireland provided a digital elevation model used in image processing. The Department of Agriculture and Rural Development (Northern Ireland) provided soils data. Peter North (then of CEH, now of the University of Swansea) adapted atmospheric haze correction procedures to meet LCM2000 needs. The British Geological Survey provided geological ‘drift’ data used to check peatland classifications. The Department of Agriculture and Rural Development (Northern Ireland) provided soils data. The production team included Robin Fuller, Geoff Smith, Jane Sanderson, Ross Hill, Andy Thomson, Nigel Brown, Ruth Cox, France Gerard, Matthew Hall and Alastair Graham.

APPENDIX I. IMAGES USED IN CONSTRUCTION OF LAND COVER MAP 2000

Summer or winter	Sensor	Date	Full or part-scne	Primary data or infill	Satellite path	Row (standard scenes)	Latitude (special orders)	Longitude (special orders)
W	TM	21/11/1996	Full	Infill	203	-	51°57´N	2°19´W
W	TM	15/01/1997	Full	Infill	204	23		
W	TM	04/03/1997	Full	Primary	204	21		
S	TM	30/05/1997	Full	Infill	205	21		
S	TM	30/05/1997	Full	Primary	205	22		
S	TM	04/06/1997	Full	Primary	208	20		
S	TM	10/07/1997	Full	Primary	204	23		
W	TM	10/09/1997	Full	Infill	206	22		
W	TM	21/09/1997	Full	Infill	203	22		
W	TM	21/09/1997	Full	Infill	203	23		
W	TM	21/10/1997	Quarter	Primary	205	24 Q3		
W	TM	21/10/1997	Quarter	Primary	205	-	53°1´N	4°17´W
W	TM	28/10/1997	Full	Primary	203	-	51°24´N	2°39´W
W	TM	28/10/1997	Full	Primary	206	22		
W	TM	01/11/1997	Quarter	Primary	202	-	54°0´N	0°13´E
W	TM	01/11/1997	Full	Primary	202	-	52°46´N	0°30´E
W	TM	01/11/1997	Full	Primary	202	-	51°13´N	1°12´W
W	TM	01/12/1997	Full	Primary	204	22		
W	TM	12/12/1997	Quarter	Infill	201	-	51°7´N	0°22´E
W	TM	22/12/1997	Full	Primary	207	19		
W	TM	22/12/1997	Full	Primary	207	20		
W	TM	25/01/1998	Full	Primary	205	22		
W	TM	25/01/1998	Quarter	Primary	205	-	50°23´N	5°32´W
W	TM	03/02/1998	Full	Primary	204	23		
W	TM	03/02/1998	Full	Primary	204	24		
W	TM	03/02/1998	Full	Primary	204	25		
W	TM	14/02/1998	Full	Primary	201	-	52°48´N	1°1´E
W	TM	14/02/1998	Full	Primary	201	-	51°13´N	0°18´E
W	TM	28/02/1998	Full	Primary	203	22		
W	TM	28/02/1998	Full	Primary	203	23		
W	TM	28/02/1998	Quarter	Primary	203	-	55°37´N	1°24´W
W	TM	01/05/1998	Full	Primary	205	21		
S	LISS III	01/05/1998	Full	Primary	012	031		
S	LISS III	01/05/1998	Full	Primary	012	-	+ 20% S	
WS	LISS III	13/05/1998	Full	Primary	004	27		
S	TM	14/05/1998	Quarter	Primary	200	23 Q3	-	-
S	LISS III	16/05/1998	Full	Primary	015	030		
S	LISS III	16/05/1998	Full	Primary	015	031		
S	LISS III	16/05/1998	Full	Primary	015	032		
S	TM	19/05/1998	Quarter	Primary	203	-	55°15´N	1°36´W
S	TM	19/05/1998	Full	Primary	203	-	54°1´N	1°28´W
S	TM	19/05/1998	Full	Primary	203	-	52°26´N	2°14´W
S	TM	19/05/1998	Full	Primary	203	-	50°57´N	2°53´W
S	LISS III	20/05/1998	Full	Primary	011	031		
S	LISS III	20/05/1998	Full	Primary	011	-	+ 40% S	
S	TM	30/05/1998	Full	Primary	200	24	-	-
S	TM	30/05/1998	Mini	Primary	200	-	50°56´N	0°54´E

Summer or winter	Sensor	Date	Full or part-scne	Primary data or infill	Satellite path	Row (standard scenes)	Latitude (special orders)	Longitude (special orders)
S	TM	09/08/1998	Full	Primary	201	-	52°17´N	0°46´E
S	TM	09/08/1998	Quarter	Primary	201	-	50°57´N	0°0´E
W	TM	20/09/1998	Full	Primary	207	-	54 40' N	7 21' W
W	TM	16/11/1998	Full	Primary	206	21		
SW	LISS III	13/05/1999	Full	Primary	004	26		
S	TM	18/05/1999	Full	Primary	207	21		
S	TM	18/05/1999	Full	Primary	207	22		
S	TM	20/05/1999	Full	Primary	205	21		
S	L7ETM	29/07/1999	Full	Infill	207	19		
S	L7ETM	29/07/1999	Full	Primary	207	19		
S	L7ETM	29/07/1999	Full	Primary	207	20		
S	TM	30/07/1999	Full	Primary	206	20		
S	TM	30/07/1999	Full	Primary	206	21		
S	TM	30/07/1999	Full	Primary	206	22		
S	L5TM	01/08/1999	Full	Primary	204	020	56°50´N	1°41´W
S	TM	01/08/1999	Full	Primary	204	21		
S	TM	01/08/1999	Full	Primary	204	22		
S	L7ETM	10/09/1999	Full	Primary	204	23		
W	L7ETM	17/09/1999	Full	Primary	205	19		
W	L7ETM	17/04/2000	Full	Primary	202	023		
W	LISS III	04/05/2000	Full	Primary	010	26	+70%S	
W	L7ETM	05/05/2000	Full	Primary	206	20		
SW	LISS III	07/05/2000	Full	Primary	004	028	+80% S	
SW	LISS III	10/05/2000	Full	Primary	003	028	+90% S	
W	L7ETM	12/05/2000	Full	Infill	207	19		
W	L7ETM	12/05/2000	Full	Infill	207	20		
SW	L5TM	13/05/2000	Full	Primary	206	019		
S	L7ETM	29/07/2000	Full	Primary	207	022		
S	LISS III	08/08/2000	Full	Primary	010	024		
W	L5TM	26/09/2000	Full	Primary	206	018		
S	L7ETM	12/05/2001	Full	Infill	202	022		
S	L7ETM	13/05/2001	Full	Infill	202	023		

SW/WS codes indicate borderline decisions with actual use first
TM is Landsat 5 Thematic Mapper
L7ETM is Landsat 7 Enhanced Thematic Mapper
LISS III in the Indian Research Satellite ….
Special orders are those shifted from standard path-row format

APPENDIX II. A BRIEF REVIEW OF BROAD HABITATS WITH AN ASSESSMENT OF THEIR DISTINGUISHING FEATURES IN RELATION TO LCM2000 MAPPING.

1. Broad-leaved, mixed and yew woodland	Broad-leaved woodlands are characterised by stands >5 m high with tree cover >20%; scrub (<5 m) requires cover >30% for inclusion in this BH. Such fine distinctions cannot be made through remote sensing. It is a particular problem, albeit relatively rare, that open-canopy woodland (stands with trees <<50%) are in the BH; they may not be mapped consistently, due to the dominance of the non-woodland plants. Stands with near-closed canopies can be interpreted straightforwardly in the field and pure examples can normally be found for training the classifier. Broad-leaved evergreen trees (a part of this BH) rarely occur in stands >1ha, suitable for training and thereby appropriate for classification. Mixed woodland (with >20% broadleaved trees) was trained separately though, where individual stands of broad-leaved or evergreen trees exceeded the minimum mappable unit, they were treated as separate blocks within the woodland: in many parts of the UK, truly 'mixed woodlands' as opposed to those with mosaic-blocks of broadleaved and coniferous trees, are unusual.
2. Coniferous woodland	Coniferous woodland includes semi-natural stands and plantations. Cover should be >20%. The recognition of coniferous woodland is generally straightforward. Rare examples of open canopy semi-natural pinewoods may have been classified according to the dominant understorey class. The BH includes new plantation and recently felled areas (this is a class where the BH definition is based on land use, i.e. forestry, rather than cover). New plantations, predominantly heather and/or grass, for example, are recorded as such by the spectral classification of image data. New plantations are only be recorded as conifers when tree cover is sufficient to strongly influence the reflectance. LCM2000 includes newly felled areas. Once they are fully recolonised by rough grass, heath or scrub, they are recorded according to that cover. Deciduous larch is discernible from other deciduous trees and generally, correctly, included with other conifers.
3. Boundaries and linear features	Only the largest of linear features (e.g. shelter belts, motorways) might be mapped by the classification of satellite images. The field survey provides by far the best information on these BHs.
4. Arable and horticulture	This Broad Habitat includes annual crops, perennial crops such as berries and orchards, plus freshly ploughed land, annual leys and rotational setaside. Distinction of rotational setaside relies heavily upon the summer-winter composite images to demonstrate the seasonal characteristic and thereby help spectral distinction. Orchards with a ground flora are hard to distinguish and the class relies upon knowledge-based corrections using interpretations made for CORINE Land Cover mapping (Brown & Fuller, 1996). Setaside vegetated with ruderal weeds and rough grassland are included with the improved grass BH, but distinguished by LCM2000 at the subclass level.

6. Neutral grassland	The three semi-natural swards are the converse of the above and rely upon the assumptions as above. Where pH is known, separate field-identification, training and classification is used. However, a soil 'acid sensitivity' map is the deciding factor when distinguishing neutral from calcareous and acid grasslands. Under the final-stage 'knowledge-based correction' pH >4.5 and < 5.5 denotes 'neutral' soils. This range is not ideal tending towards the acid side of 'neutral'; however, the acid sensitivity map only offers 3 classes and this range, described as 'moderately sensitive' to acidification, is the nearest category to neutral. Using this pH level will tend to over-estimate 'neutral soils' and identify them in borderline acid situations.
7. Calcareous grassland	The same details apply as did to neutral grass but with the pH > 5.5. Again, the range is far from ideal; the acid sensitivity map describes this pH range as having 'low sensitivity' to acidification. This will tend to over-estimate 'calcareous soils', including neutral examples; however, there was no better alternative available.
8. Acid grassland	As above, but pH <4.5 denotes 'acid' soils. This range is appropriate.
5. Improved grassland	Improved grasslands will be distinguished from semi-natural grass. The criteria used by field surveyors (dominance of palatable grasses) also gives the grasslands a distinct spectral signature. It is recognised that management practices (heavy grazing) can obscure this dominance and might cause mis-classifications with semi-natural swards. However, the field training course and trial reconnaissance surveys suggest separation is feasible. If accuracies are lower than the intended 90% per-parcel, then the target classification will be that of the Specification (without distinction between semi-natural and improved swards), but the distinction will be retained at the subclass level.
5. Improved grassland continued	Integration of the broad assessment with specific field estimates might prove especially powerful as a guide to the spatial distributions and quantities of the various agricultural grasslands. Setaside grass, though to be included in this category, may be confused with rough neutral grass once well-established.
6. Neutral grassland	The three semi-natural swards are the converse of the above and rely upon the same assumptions as above. Where pH is known, separate field-identification, training and classification is used. A soil 'acid sensitivity' map is the main way of distinguishing neutral from calcareous and acid grasses: under the final-stage 'knowledge-based correction' pH >4.5 and < 5.5 denotes 'neutral' soils.
7. Calcareous grassland	The same details apply as did to neutral grass but with the pH > 5.5.
8. Acid grassland	As above, but pH <4.5 denotes 'acid' soils.
9. Bracken	There were problems in the accurate mapping of bracken in 1990 so it was not written into the Specification as a 'target class'. However, dense bracken is distinguished (excepting woodland stands) at the subclass level; it should be recognised that bracken often fails to offer stands sufficiently extensive for classification and training.
10. Dwarf shrub heath	This Widespread Habitat is essentially an aggregation of LCM1990's *Open* and *Dense Shrub Heaths*. This means that the Habitat could generally be identified on LCM2000 with no particular difficulties. However, the Broad habitat classification treats ericaceous vegetation on peat > 0.5 m depth as 'bog'. A drift map showing peat-soils is used to distinguish heaths from ericaceous bogs.

11. Fen, marsh and swamp	This Habitat includes fen, fen meadows, rush pasture, swamp, flushes and springs. Apart from rush pasture, examples of the Habitat are relatively rare, and seldom extensive enough to map as pixels, let alone polygons and records for Britain are likely to be localised. Though there are indications that dominant rush cover influences the spectral characteristics of a parcel enough to make the distinction, the final accuracy with which rush pastures is distinguished will only be apparent after validation.
12. Bog	The bog category includes ericaceous, herbaceous and mossy swards in areas with a peat depth > 0.5 m. The peat drift maps are the final control over the bog category. Areas classified as 'bog' but with <0.5 m are corrected to grass moor or heath, according to dominant cover type.
13. Standing open water and canals	Water bodies > 0.5 ha are readily mapped. There will be few if any canals which can be mapped at satellite image scales - they effectively form linear features.
14. Rivers and streams	Only the widest of rivers (>50 m) are shown accurately, though such information might be drawn from other maps. They will not be distinguished from class 13. Standing water, except perhaps contextually (e.g. through use of digital maps of rivers).
15. Montane habitats	This class should be clearly identifiable by context and the presence of vegetation cover at a sparse level should distinguish Montane habitats from 26. Inland Rock.
16. Inland rock	This Habitat includes natural and man-made bare ground.
17. Built up areas and gardens	This Habitat is a combination of Suburban / rural development and Continuous urban categories of 1990. LCM2000 identifies these as subclasses. It records the heterogeneity of urban land, e.g. the vegetation cover in parks and larger gardens, bare urban ground and the tillage of allotments, in more detail than is required by the Broad Habitat classification.
18. Supra-littoral rock	Distinction between rock from sediment is done contextually, by defining a vector region encompassing rocky coastlines. Distinction of supra-littoral needs us to define a high water mark: this is only straightforward for major features. However, rarely are there extensive areas of supra-littoral rock.
19. Supra-littoral sediment	Sedimentary coasts are also defined interactively. Large areas of supra-littoral sediment occur as beaches, mudflats, dunes and shingle beaches. Distinction of the supra-littoral component uses the terrestrial mask, derived from LCMGB 1990, updated with changes, where appropriate.
20. Littoral rock	These classes are those in the maritime mask zone on a rocky coastline. They are generally more extensive than supra-littoral rock and thus more readily mappable from satellite images.
21. Littoral sediment	Littoral sediments are those in the maritime zone, on sedimentary coasts; they may be very extensive. Saltmarsh is included with this Broad Habitat but mapped as a separate subclass by LCM2000.
22. Inshore sublittoral sediment	All areas of sea and estuary class are assumed to be inshore and sublittoral sediment, without distinction of rocky substrata.
23. Inshore sublittoral rock, 24. Offshore shelf sediment, 25. Offshore shelf rock, 26. Continental shelf slope and 27. Oceanic seas are irrelevant in the context of a land cover map.	

APPENDIX III. LCM2000 CLASS VARIANTS MAPPED ONTO BROAD HABITATS, WITH CODES, CODE NUMBER, APPROXIMATE COLOUR AND THE ACTUAL RED-GREEN-BLUE COLOUR MIX (0-255) USED ON MAPS

BH (name abbreviated)	Variants	Alpha-code	Number	R	G	B
22. Inshore sublittoral	sea	We	22.1.1	0	0	128
13. Standing water/canals	water (inland)	W	13.1.1	0	0	255
20. Littoral rock	rock	Lr	20.1.1	255	255	128
	rock with algae	Lra	20.1.2			
21. Littoral sediment	mud	Lm	21.1.1			
	sand	Ls	21.1.2			
	sand with algae	Lsa	21.1.3			
	saltmarsh	Lsm	21.2.1	128	102	255
	saltmarsh (grazed)	Lsg	21.2.2			
18. Supra-littoral rock	rock	Sr	18.1.1	204	179	0
19. Supra-littoral sediment	shingle (vegetated)	Sh	19.1.1			
	shingle	Shv	19.1.2			
	dune	Sd	19.1.3			
	dune shrubs	Sds	19.1.4			
12. Bog	bog (shrub)	Bh	12.1.1	0	128	115
	bog (grass/shrub)	Bhg	12.1.2			
	bog (grass/herb)	Bg	12.1.3			
	bog (undifferentiated)	Bo	12.1.4			
10. Dwarf shrub heath	dense (ericaceous)	H	10.1.1	128	26	128
	gorse	Hg	10.1.2			
	open	Hga	10.2.1	230	140	166
15. Montane habitats	montane	Z	15.1.1	0	180	190
1. Broad-leaved woodland	deciduous	D	1.1.1	255	0	0
	mixed	Dm	1.1.2			
	open birch	Db	1.1.3			
	scrub	Ds	1.1.4			
2. Coniferous woodland	conifers	C	2.1.1	0	102	0
	felled	Cf	2.1.2			
	new plantation	Cn	2.1.3			
4. Arable & horticultural	barley	Ab	4.1.1	102	0	0
	maize	Am	4.1.2			
	oats	Ao	4.1.3			
	wheat	Aw	4.1.4			
	cereal (spring)	Acs	4.1.5			
	cereal (winter)	Acw	4.1.6			
	arable bare ground	Aba	4.2.1			
	carrots	Ac	4.2.2			
	field beans	Af	4.2.3			
	horticulture	Ah	4.2.4			
	linseed	Al	4.2.5			
	potatoes	Ap	4.2.6			
	peas	Aq	4.2.7			
	oilseed rape	Ar	4.2.8			
	sugar beet	As	4.2.9			
	unknown	Au	4.2.10			
	mustard	Ax	4.2.11			
	non-cereal (spring)	Ans	4.2.12			
	orchard	Ado	4.3.1			
	arable grass (ley)	Agl	4.3.2			
	setaside (bare)	Asb	4.3.3			
	setaside (undifferentiated)	Ase	4.3.4			
5. Improved grassland	intensive	Gi	5.1.1	0	255	0
	grass (hay/ silage cut)	Gih	5.1.2			
	grazing marsh	Gim	5.1.3			
	grass setaside	Gis	5.2.1	255	177	0
6. Neutral	rough grass (unmanaged)	Grn	6.1.1			
	grass (neutral / unimproved)	Gn	6.1.2			
7. Calcareous	calcareous (managed)	Gc	7.1.1	180	255	180
	calcareous (rough)	Grc	7.1.2			
8. Acid	acid	Ga	8.1.1	153	128	0
	acid (rough)	Gra	8.1.2			
	acid with *Juncus*	Gaj	8.1.3			
	acid *Nardus/Festuca/Molinia*	Gam	8.1.4			
9. Bracken	bracken	Gbr	9.1.1	255	100	60
11. Fen, marsh and swamp	swamp	Fs	11.1.1	255	255	0
	fen/marsh	Fm	11.1.2			
	fen willow	Fw	11.1.3			
17. Built up areas, gardens	suburban/rural developed	Us	17.1.1	128	128	128
	urban residential/commercial	U	17.2.1	0	0	0
	urban industrial	Ui	17.2.2			
16. Inland rock	despoiled	Id	16.1.2			
	semi-natural	Ib	16.1.1	210	210	255
20 BHs	72 target/subclasses/variants					

02/02/2002

APPENDIX IV. PRE-PROCESSING OF IMAGE DATA

Pre-processing of the image data can correct geometric distortions, remove systematic noise, eliminate unwanted changes in response due to differential illumination and normalise data from different images to physical units of reflectance rather than the arbitrary engineering units of the raw data. Image data in units of reflectance directly compare to other images in the same units or other spectrally calibrated spectral data sets. Pre-processing helped to improve the classification within images and the consistency of this classification between images. Pre-processing stages were:

- The correction of atmospheric haze,
- Cloud and shadow masking within scenes,
- Geo-registration and resampling,
- Correction of differential illumination effects on undulating terrain,

Atmospheric correction

Atmospheric haze both attenuates the amount of light reaching the surface and also scatters light into the sensor which has not interacted with the surface. These effects are most pronounced in shorter wavelengths and distort the information recorded by the sensor. Various algorithms were available to model the effects of the atmosphere on the light passing through it. The algorithms generally identify areas in the image for which the true reflectance can be estimated and in this way assess the distorting characteristics of the atmosphere at the time of imaging. These characteristics can then be used to model the atmosphere and remove its effects from the image.

Liang *et al.* (1997) developed such a method for TM data which automatically seeks out examples of cover types whose reflectances can be estimated and interpolates the 3-dimensional atmospheric characteristics between the examples to correct the image. The software was made available for CEH use. It was modified to accommodate differences in the European TM supply format. It was made operational and tested with full TM scenes. The correction was highly effective, removing not only haze but even penetrating areas of thin cloud on the uncorrected image.

The correction of IRS LISS data could not directly use this software which relies upon the presence of the greater number of bands of TM data. However, the same principles could apply: other bands, sensitive to atmospheric haze, could be used to estimate atmospheric characteristics and thereby compensate for its effects. These developments of the Liang software were made by CEH and, for the first time, made the procedure operable on LISS data.

The procedures for atmospherically correcting both TM and IRS LISS data were built into the processing line. Correction of atmospheric haze generally worked routinely and effectively. However, areas with worse than average haze were usually treated as cloud-covered and masked out, for later infill with substitute data. There was a problem in winter images of upland areas where the moorland vegetation is deciduous grass or bog. The haze-correction algorithm relied on finding local examples of green vegetation with predictable spectral characteristics, to help calibrate the correction. If, as in some uplands, green vegetation was absent, the algorithm did not work correctly and corrupted the data. Under such circumstances, there was no alternative other than to abandon atmospheric haze corrections. This did not preclude classification, which can compensate for the offset in values caused by haze. It did, however, mean that data values were not calibrated ‘reflectances’ and could not be directly inter-compared from one scene to another, to examine phenological effects on spectral signatures.

Illumination correction
Undulating terrain is illuminated differentially according to whether facets of terrain are horizontal, face the sun, face away from the sun and, if the latter, do so sufficiently to be shaded from direct solar illumination. Differential illumination and its consequent effects on radiation recorded by the sensor can be modelled using digital elevation models (DEM) and compensated for, offering corrected data based upon a theoretical horizontal surface illuminated from directly overhead. Such correction is important if facets of land surface are not to have a highly significant and perhaps dominating effect upon the results of segmentation.

CEH contributed financially to the operationalisation of software, developed by Cambridge University in the CLEVER-Mapping programme. The software offered the option of full National terrain correction prior to segmentation and classification. The software operates within an Erdas framework using the Spatial Modeler [*sic*] of Erdas.

The processing used a 50 m, grid-based, DTM, with integer heights in metres. The DTM was contracted by EDX Engineering Inc. from OS 1:63 360 maps. In comparisons with the OS equivalent, the data are shown to have a root mean square error averaging just 2.5 m. This DTM costs about half the price of the OS product, without restrictions on use, nor royalties to pay.

Terrain corrections were routinely used for all images: this not only compensated for the terrain but also corrected the illumination to give results equivalent to a nadir viewing angle with the sun directly overhead: the reflectances were all normalised and standardised, allowing inter-comparison within and between images.

The correction of differential illumination in undulating terrain worked well for summer data, with terrain effects on summer scenes barely perceptible after correction. Winter data suffered much poorer illumination, especially in infrared bands where the dynamic range of reflectances was very low and compensations could not accurately mimic non-shaded ground. The consequence was that shaded terrain facets, which gave rise to spectrally determined segments, were classified using spectral variants of the relevant class and these were aggregated thematically, post-classification (Kershaw & Fuller, 1992).

Cloud and cloud-shadow removal
Cloud and shadow masking used reflectance recorded in TM band-3 (red) and radiance in band-6 (thermal) and ratios of these. Red response was always high; because clouds are cold, thermal response was low; and the ratio gave a range which was nearly unique to cloud. By inspecting the images, thresholds could be set. A mask was created where the image fell within the pre-specified threshold in all three cases (red, thermal and ratio values). The mask was then grown, by 8 pixels to eliminate thin undetected cloud around edges. An offset for the cloud shadow was measured by inspection. The two masks, cloud and shadow, were added together: the result showed clear ground, cloud and / or shadow. The mask was used to include / exclude sections of image in subsequent processing. All these stages were automated in an analytical model, with only brief inspections needed to pick thresholds prior to the analysis. Cloud-masking worked well. On scenes where there were several layers of cloud at varying heights, the projection and masking of the position of the cloud-shadow proved particularly difficult. However, where necessary, manual intervention ensured that reject areas are correctly defined.

Geo-registration
Geo-registration allowed the images, usually corrected by the suppliers to compensate for the Earth’s rotation and curvature, to be registered to the British and Irish National Grids (BNG & ING) with a suitable output pixel size. This correction involved collecting a sample of ground control points (GCPs) on an image and, using a digitising tablet, the equivalent points on an OS / OSNI

1:50 000 paper map; once an image was registered to a map, then images which overlapped were better registered to the geo-corrected image.

The image analysis system was made to calculate a transformation between image and map, calculating the differences in scale, rotation and offset. The normal form of transformation was a polynomial model, relating x-image to X-map and y-image to Y-map, with a cross-over term; this meant that x on the image can change in scale as y on the image changes and *vice versa*. The more GCPs that were chosen, then potentially the more complex the model which related x-, y-image to X-, Y-map. A first order model assumed a simple linear scale difference plus rotation and offset; second order model allowed the scale to vary, gradually, increasing or decreasing across the image; more complex algorithms could accommodate highly variable changes in scale, due to sensor movement (unlikely) or terrain (more usual). These issues were all addressed in LCMGB 1990 (Fuller *et al.* 1994) and the problems might be thought of as having been solved; but LCMGB always used directly overlapping TM scenes, viewed from the exact same position in space. Due to the difficulties of obtaining images in the poor summers of 1998 and 1999, LCM2000 used some adjoining TM scenes, viewed from paths about 90 km away; it also combined TM and LISS images with different viewing geometries and pixel sizes. A complex series of experiments have sought the best possible combination of GCP collection strategies (image-to-map and image-to-image, varying the distribution and number of GCPs) and transformation order (first, second etc. to about seventh order). In brief, the results showed that:

- Primary corrections should be image-to-map,
- In any areas where images overlapped (including same season and contrasting seasons), further GCPs should be collected image-to-corrected-image,
- Overly complex polynomial models merely served to accommodate minor misplacements in GCPs, worsening geo-registration,
- There should always be an ample excess of GCPs for the chosen model (more GCPs were needed for higher order models) to ensure that minor errors in GCP location could be detected as residual errors in the fit of the model,
- This excess of GCPs better ensured that errors are not accommodated by the transform, and that such distortions did not get incorporated into an erroneous model and a poor rectification.

Resampling
When producing an output image with pixels located relative to the reference projection, it was necessary to chose a resampling algorithm which could calculate realistic reflectances for each pixel, based on the reflectances of the surrounding pixels in the unregistered image (the true grid will be offset from the arbitrary pixel grid). The nearest neighbour resampling algorithm takes, as the name implies, the nearest pixel from the unregistered image and assigns its reflectance values, unaltered, to the pixel in the output cell. Other methods interpolate between the surrounding pixels in the unregistered image: bilinear interpolation simply mixes surrounding reflectance values in proportion to their probable contribution to the output pixel; cubic convolution uses a mathematical weighting to improve the interpolation. These methods are written up in all standard remote sensing text books (e.g. Lillesand & Keiffer 1994, Schowengerdt 1997); they are mentioned here because the methods of GCP collection, the polynomial modelling and the resampling method could very substantially influence the geometric correspondence between winter and summer images. This in turn substantially affected the image segmentation, the residual proportion of edge pixels and the final classification outputs. The resampling algorithm could exaggerate problems or, better, could resolve them: so, for example, nearest neighbour resampling could enhance within-field heterogeneity and cause artificial over-segmentation; but, on single-date images, it may have 'smudged' mixed boundaries less than cubic convolution. Conversely, co-registration of summer and winter scenes, using nearest neighbour resampling, could have result in apparent mis-

registrations, because the nearest neighbour method, when extracting reflectance values from the unregistered image, was forced to round up or down its x- and y-location to take the nearest whole pixel; thus, co-registered scenes may have exhibited the cumulative effects of having half-pixel displacements in opposite x- and y-directions, giving potentially a v2 pixel difference or 1.4 pixels displacement, one image relative to the other. The conclusion, after trials of the resampling methods, was that cubic convolution produced better results cosmetically, which are also better for segmentation purposes. The geo-registration process worked satisfactorily and co-registration was achieved with generally less than a pixel of visible misregistration between summer-winter paired scenes.

APPENDIX V. SEGMENTATION

Segmentation procedures, written by CUGD, were only in prototype form when the project started. The software had been demonstrated, but was only capable of segmenting scenes of about 1000 pixels square. The process ran in the Microsoft Windows environment, preventing easy use with the other processing modules, such as Erdas and Laser-Scan IGIS, which run on Unix workstations. Laser-Scan implemented an operational version of the segmentation software for the Unix environment: it was in effect a 'beta-test' version, suitable for CEH operations, though not yet ready to market; it was, however, fully tested and approved by ITE for LCM2000 operations.

The segmentation consists of two separate stages, first, edge-detection to identify boundary features and, second, 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.

LCM2000 was based on CLEVER-Mapping (the Classification of Environment with Vector and Raster mapping - Smith, 1997). The procedure segmented the image into areas broadly equivalent to land parcels and vegetation patches. The segmentation consisted of two separate stages: first, edge-detection to identify boundary features; second, region growing from seed points selected to avoid edges.

Band selection

It was only possible to use 3 of the 6 bands for edge-detection and segmentation. As the classification was derived from a combination of summer and winter images, it was decided that the winter image would contribute just one band, and summer image the other two. In addition to the use of pure bands, mathematical combinations of bands within the same image could have offered substitutes. Vegetation indices (e.g. Lillesand & Keiffer 1994) would produce a one-band image with resulting pixel values closely related to vegetation cover: in tests made under LCM2000 developments, it was found that the index might fail, for example, to separate a dense grassland from an adjoining woodland, so this option was discounted. A principle component (PC) analysis (e.g. Lillesand & Keiffer 1994) can summarise the key characteristics of a multi-spectral image in fewer bands than the original image: thus, the first PC band might encapsulate as much as 80% of the variation across the image. It was found that the first PC might show more detail than does any single band, though the distinctions between red and infrared reflectance could be more useful than use of PC-1 and PC-2.

The following rules applied:

- The summer image contributed two bands - red and infrared (MIR for TM/ETM, NIR for LISS),
- The winter image contributed one band - NIR (which is generally the brightest),
- For single date cover, red, NIR and MIR data were used, whichever the sensor.

Edge-detection

The segmentation procedure built polygons around 'seedpoints' which were selected as being within a segment or a land parcel; an edge detector was used to ensure that the appropriate seedpoints were selected away from parcel-edges. The CUGD software used a Sobel edge detection algorithm (e.g. Mather 1987) to define the mixed boundary pixels and to avoid heterogeneous within-parcel pixels. The Sobel method gave outputs which weighted the strength of apparent edges and mosaics, giving a grey-tone image, from white where edges are strongest to black where no edges are detected. The algorithm, as implemented by LSL, sometimes formed double-lines of pixels along each side of certain linear features, leaving a void of 'non-edge' pixels which might be selected erroneously to form seed points within the linear feature. Experiments with other edge detectors showed that a filter which measured local variance, thereby measuring local heterogeneity, was an alternative to the Sobel method.

For segmentation, the resultant edge images (Sobel and variance-based) were 'density-sliced' to 'threshold' out the key edges, leaving the remainder as 'within field' pixels. Thresholding to extract 'edge' and 'non-edge' pixels from continuously variable variance images was also the subject of experimentation: the best results were those which left 'voids' in urban areas, sufficient to form seed points, while identifying most key linear features in the wider countryside (e.g. field boundaries, roads and narrow rivers) as such. The software offered advice on thresholds, based upon examination of the histogram of output values, defining a threshold which was shown to achieve the stated objective.

Segmentation and thresholds

The level of subdivision resulting from the segmentation procedure was 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 could arise. In principle, this could reach the point where each pixel was itself a segment. Conversely, by allowing very loose thresholds, the parcels could be large, ultimately to the extreme of covering a whole image. The aim was to ensure a field-by-field segmentation, also separating zones within-urban and suburban areas and subdividing heterogeneous semi-natural zones into meaningful segments.

The edge-image could affect the outputs, for example:

- edges and linear features themselves formed polygons;
- linear features also linked areal features one to another (e.g. two suburban areas linked into one polygon by a connecting line of mixed boundary pixels).

A wide-ranging series of trials, tested Sobel and variance edge-detectors with a range of threshold values to reject or include pixels as 'real' edges or not. The Sobel algorithm, using the 'recommended' threshold gave generally the best results, with no consistent improvement, often the contrary, when over-riding the use of the recommended value. The variance image was density sliced at varying levels to provide a range of edge-inputs for segmentation. The Sobel detector, while visually no better, was nonetheless far superior in its definition of edges/seedpoints for the subsequent segmentation; this is not surprising in that it had been 'tuned' for that purpose by the software developer.

The segmenter appeared to be fairly robust in its tolerance of different edge thresholds, with substantial changes in threshold needed to induce commensurately large changes in output segments: the fine level of control which could thus be exerted was gratifying, suggesting that the segments were real entities, not ephemeral features governed by highly sensitive input parameters.

Unfortunately, the level of spectral distinctions between the various cover types could vary according to the type: the distinction between wheat and grass might have been subtle, while a single suburban zone might comprise a multitude of spectral variations of no consequence for the target classification. It was necessary to chose a level of segmentation which was adequate to separate all the Target classes and Subclasses while avoiding, as far as possible, the risk of sub-segmenting entire features (e.g. substantially subdividing fields).

During the early phase of LCM2000 it was necessary to learn how best to cope with these problems and to choose objective criteria to select the thresholds which i. achieve a field-by-field separation, ii. grow urban and suburban areas into segments, rather that recording complex mixtures of pixels, and iii. which adequately separate what should be individual segments but which were, potentially, erroneously linked by ‘linear features’ (e.g. suburban areas linked by the road between them). Indeed, it was necessary to avoid the artificial growth of lines of highly mixed pixels which could form complex linear polygons, comprising various types of vegetation, often with roads and water features, which tended to spread across large tracts of image, sometimes picking up adjacent segments of similar spectral response.

Thresholds based upon reflectance differences between polygons determined how fine or coarse the segmentation was. The selection of the optimal threshold was, as above observations imply, a process fraught with some difficulty. Three different values are required, one for each band in the segmentation. A series of experiments was undertaken to test varying thresholds and to examine the results. Variable thresholds were tested in sequence: it was decided that these should be different per band, reflecting the data range to be found in the band. Determination of the range included the visual examination of histograms showing reflectances per band: there was a need to take account of long tails in a histogram distribution (where a very few pixels distorted the measured range) by clipping the bottom and top 5% of data values to assess the range occupied by 90% of the reflectance data. An objective way of achieving this, used means ± 2 standard deviations to gather similar information.

Different proportions of the potential range of thresholds were tested for their effect on output segmentations. While wholly objective criteria for segment-assessments could have been used, these would have relied upon the use of scarce validation data (e.g. field survey squares), too few to ensure widespread testing and, anyway, needed as an independent source of validation. In this instance, it was necessary to use semi-objective procedures, where rules have been defined (as above) but the success in their application was tested by inspection. Essentially, the aim was to ensure that, as far as possible, no complex (mixed) polygons would result. To achieve the desired objectives, required an iterative process of trial, inspection, re-analysis with altered edge-image inputs and/or segregation thresholds, re-inspection etc. until an acceptable segmentation results: the process was no different in principle to the iterative and subjective processes of training and classification used in the conventional per-pixel processes (Kershaw & Fuller, 1992), in LCMGB 1990 (Fuller, 1994) and in many other mapping exercises.

By trial and observation, it was decided that the edge-threshold determined by the software could not discernibly be improved upon; and a threshold equal approximately to one-fifth the reflectance range was appropriate for segmenting each band. These levels allocated about 95% of all pixels to segments; they gave field-by-field segmentations though often sub-segmenting fields; this over-segmentation was necessary to prevent growth of mixed chains of edge pixels; urban areas were adequately subdivided to at least match target classes. It was decided that any simplification of unecessary sub-segmentations would be better undertaken by more ‘intelligent’ generalisation methods or in vector analyses, post-classification.

Post-segmentation generalisation and boundary rejection

It was necessary, having derived an agreed segmentation, to simplify the results using spatial generalisation procedures - for example, eliminating reject pixels (not part of any segments); also dissolving small polygons and growing adjoining regions into the dissolved areas, using intelligent spatial-contextual-analyses. Though these steps incorporated odd pixels into polygons, they were necessary to reduce the final vector dataset to a manageable number of polygons. The CLEVER-Mapping procedure, which captured polygon reflectances from core pixels only, took sufficient account of edge pixels in a parcel to avoid difficulties. However, because removal of edges potentially lost useful data about landscape patterns, the information on dissolved polygons was retained; and per-pixel analyses were used to capture information on heterogeneity.

After a series of experiments, it was decided that:

- non-segment edge pixels would be dissolved into adjoining parcels,
- single-pixel islands would be dissolved into their surroundings
- segments < 9 pixels (≤0.5 ha - the minimum mappable unit) were also dissolved, being attached to their nearest neighbouring segments using a spectral minimum distance rule,
- because the segmenter worked on absolute difference, brighter fields were sub-segmented most despite a proportional difference in mean reflectances per band being <5%; these sub-segments were aggregated prior to classification using proportional difference to identify polygons to be joined,
- the generalisation procedure kept track of dissolved and aggregated polygons, should later analyses require inspection and classification at the detailed level.

Conversion to vector format

Once an acceptable segmentation of the images was achieved then it was necessary to create vector versions of the segments and build the GIS database. This was a simple procedure of raster-to-vector conversion where the boundaries between segments with different values in the raster images were represented by vector lines. The procedure attempted to create vector polygons giving the 'land parcels' required for classification. Therefore the vector lines representing the boundary of a segment were built into polygons. Attributes to hold information required by the land parcel could then be attached.

APPENDIX VI: ADDITIONAL NOTES FOR LCM2000 DATA USE

These notes summarise and emphasise some aspects of the LCM2000 specification, provide guidance when interpreting the data and outline good practice recommended by the LCM2000 production team.

LCM2000 Specification

LCM2000 maps land cover. This may be synonymous with land use (e.g. arable crop cover denotes arable land use) but often land use cannot be inferred (e.g. grass used for recreation is much like that which is grazed).

> ***LCM2000 sets a minimum mappable area as > 0.5 ha.*** Parcels and linear features of 0.5 ha and less were dissolved into the surrounding landscape during the production process using spectral or thematic proximity rules. In very limited areas some parcels of less than 0.5 ha may remain due to processing difficulties.

The LCM2000 classification is a hierarchy of Target classes, Subclasses and class Variants. The Target classes and Subclasses are configured, first, to generate widespread examples of Broad Habitats as defined by the Biodiversity Action Plan and, second, to extend the classification for wider use of the land cover data. Class Variants, if provided, give additional information, but these are not necessarily recognised with the accuracy nor consistency of Target and Subclasses.

Guidance for interpretation

The ability to distinguish land cover at the class Variant level will be dependent upon the dominant land cover at the time of imaging. When, for instance, crops have been harvested, distinctions are neither demanded nor attempted during LCM2000 production. Where possible, contextual information may be have been used to allow class separation, but these records may not be up to date.

Accuracy or correspondence. Users should take care not to refer to inaccuracy if they mean differences due to data model, scale, resolution, interpretation, class-definition, target classes etc. LCM2000 incorporates inevitable inaccuracies, but they may not be the major cause where it fails to match user needs.

Good practice

LCM2000 may be customised. It is more than just a map. It is a data storage and analysis framework which not only provides inputs into environmental applications, but can also be the starting point or foundation on which further research and applications work is based. The user is encouraged to alter and expand their version of the database as they see fit.

LCM2000 has unique labels. As part of the production process each parcel was given a unique label which was stored in the *Segid* attribute. All users of the LCM2000 vector products will receive data containing this attribute. It is recommended that the *Segid* attribute is retained within the LCM2000 dataset and any developments of it. This will allow unambiguous communications back to the LCM2000 data management team and between LCM2000 users.

LCM2000 production philosophy. When developing the methods for LCM2000 production it was decided to retain as much information as possible. This has resulted in LCM2000 being a dataset deep and rich with information. It is recommended, when altering LCM2000 data, that a similar philosophy is followed. Use parcel level meta-data attributes to record its lineage or process history. When changing attributes originally supplied with the LCM2000 data, keep a copy of them in a new attribute or put changes into a new attribute.

Document changes. Fully documenting changes to a database is essential for efficient and effective data management and security. This information will help when proving feedback to the LCM2000 data management team.

Data Assessment Report (DAR). To effectively manage feedback from users of LCM2000 data a DAR form is provided with the data or via the LCM2000 web site. The DAR is not only a means for communicating problems which may be identified in LCM2000, but also more general comments on the dataset, helpful hints on analysis and applications, suggestions for future releases and queries about good practice. The DARs will be reviewed by the LCM2000 data management team. The results of the review will feed into the development of future releases of LCM2000 and future UK land cover mapping activities. Some information may be distributed to current and future LCM2000 users to improve understanding and use of the data.

APPENDIX VII: DISTRIBUTION OF LCM2000 DATA

Strategy for distribution

Vector datasets

In its distributable form the LCM2000 vector data set consists of 6.6 million land parcels each of which will have either 5 or 7 attributes depending on the level of detail. Level-3 datasets (as supplied to the consortium) when held as ArcView Shapefiles require approximately 4.5 Gb of storage space (equivalent to ~ 10 CDs). Few if any organisations will have the capability to hold and process the whole LCM2000 as a single dataset.

LCM2000 will be delivered to consortium members as 100 km tiles (Appendix VII Figure 1). Some merging of 100 km tiles with small areas of cover was undertaken to simplify construction. Any land parcels that cross a tile boundary are present in both the squares each side of a boundary. The LCM2000 therefore consists of 45 Shapefiles, with the largest single Shapefile (total storage space for .dbf, .shp and .shx files) is 202 Mb and this file contains 364,845 land parcels. The LCM2000 vector data will be provided as compressed Shapefiles on CD-ROM representing a storage space of approximately 1.1 Gb (2 CDs).

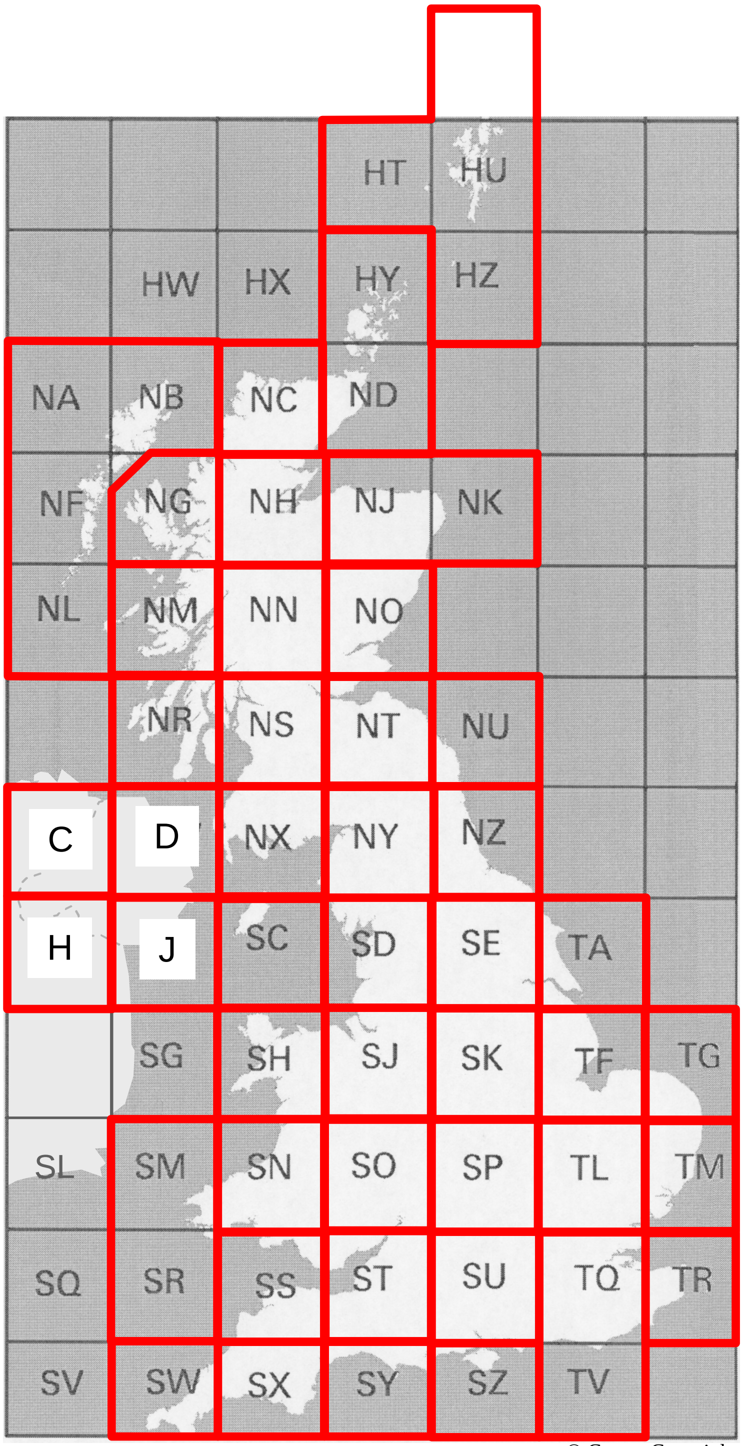
Raster datasets

The LCM2000 raster data was created by converting the land parcels within the vector data set into a 25 m grid. The raster results for each of the 100 km tiles were merged together into separate datasets for Great Britain and Northern Ireland. Each of the grid cells within the raster data sets records the dominant land cover at that location in terms of LCM2000 Subclasses. The raster data is provided in two formats; a flat binary file with one byte per grid cell and an ERDAS Imagine file. The storage space required for the GB and Northern Ireland files is 1.2 Gb and 55 Mb respectively. The LCM2000 raster data will be provided as compressed files on CD-ROM representing a storage space of approximately 120 Mb (1 CD).

Documentation

A full set of documentation will be provided on each CD. This will include:

Data formats (generic documentation for raster and vector products)
File formats
Attribute specification
Class list and definitions
Keys for the process history descriptor attribute
Health warnings / good practice / helpful hints
Data Assessment Report form



Appendix VII, Figure 1. A map of the 100 km tiles of the UK showing the areas available in LCM2000. The tiles grouped within heavy outlines were merged to simplify construction and data supply.