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 sedimen	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	dense ericaceous, gorse
		Open dwarf shrub heath	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**.

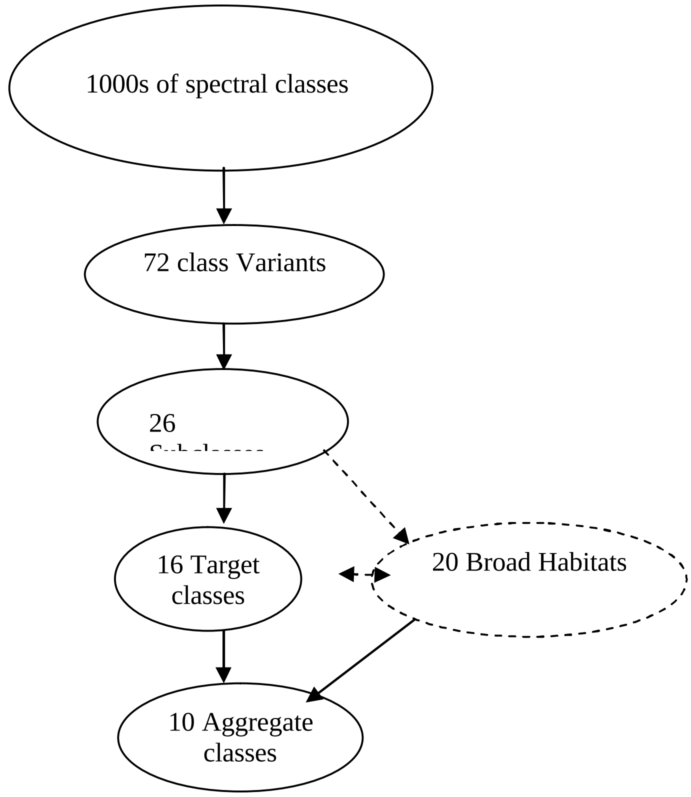

Table 2. Aggregate classes and their relation to Broad Habitats and LCM2000 Target classes.

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).