Land cover map 2007: using OBIA for LCM2007

Clare Rowland[1], Dan Morton[1], Chris Marston[2], Claire Wood[1]

[1]Centre for Ecology and Hydrology, Lancaster Environment Centre, Lancaster, LA1 4AP, UK
Email: clro@ceh.ac.uk
[2]APEM Ltd., Riverview, A17 Embankment Business Park, Heaton Mersey, Stockport, SK4 3GN, UK

Summary

Land cover map 2007 (LCM2007) is an object-based land cover map for the UK containing around 10 million objects. The LCM2007 spatial framework is based on the generalisation of national cartography products (OS MasterMap for Great Britain and Ordnance Survey Northern Ireland for NI). 34 composite images (based on summer and winter data) were classified using a maximum likelihood classifier. Areas where composite data were not available were filled with classifications from single-date data. A set of knowledge-based enhancements (KBEs) were then applied to refine the classification using ancillary data sets, including soil and altitude data. The final product showed a correspondence of 83%, when compared to 9127 ground reference polygons. A range of LCM2007 data products are available ranging from the full vector data set, with 10 attributes per polygon, to a 25m raster data set and a series of 1km raster products.

1   Introduction

Land cover mapping is one of the most developed fields in terrestrial remote sensing. Global land cover maps have been available for over 20 years and together with regional and local land cover maps are an important starting point in many environmental projects and analyses; they form a basis against which other data sets may be integrated and understood. Over the past 20 years there have been many developments in classification methods, including the development of Object-based Image Analysis (OBIA). These developments are reflected in the UK Land Cover Map (LCM) series. The first map, LCM1990 (formerly LCMGB), was a per-pixel product for Great Britain. It was followed by LCM2000 an object-based classification for the UK, where the objects were derived from image segments. LCM2007, like LCM2000, is an object-based (or parcel-based) classification for the UK, but improved so that generalised national cartography forms the basis of the objects. Objects for OBIA are typically based on image-segmentation, but may be derived from digital map data. LCM2007 represents a hybrid approach, with objects being based on generalised cartography supplemented with image segmentation, exploiting the benefits of both approaches.

This paper describes the creation of LCM2007, with reference to the data sets required and the production of the spatial framework. The classification procedure is then described, along with the use of a series of knowledge-based enhancements to refine the precision and accuracy of the classification. Finally, the validation of LCM2007 is discussed and conclusions are drawn. The paper ends with a brief summary of how LCM2007 data sets may be accessed.

2   Data Sets

A range of different types of data set were required to create LCM2007, including satellite imagery, national cartographic products, agricultural census data, ground reference data,

digital elevation model (DEM) data, soil data sets and urban extent (Morton *et al.*, 2011). The data sets were used to aid the pre-processing, the choice of classification training areas, the knowledge-based enhancements and the validation.

2.1 Data Sets

The satellite images used to produce LCM2007 were acquired from the Landsat-TM5, IRS-LISS3 and SPOT-4 and SPOT-5 sensors, these sensors all have pixel sizes of 20-30m. AWIFS data, with a pixel size of 60m, was used when other imagery was unavailable. LCM2007 was created by classifying multi-temporal composite images. These images are based on the red, NIR and MIR bands of one summer image and one winter image creating 6-band 2-date images. To complete LCM2007, 34 composite images were required, which were supplemented with some single-date images to gain full UK coverage (Fig. 1a). 91% of the UK was mapped with composite images for LCM2007 (Fig. 1b), compared to 84% and 87% for LCM2000 and LCM1990 respectively (Fuller *et al*, 2002).
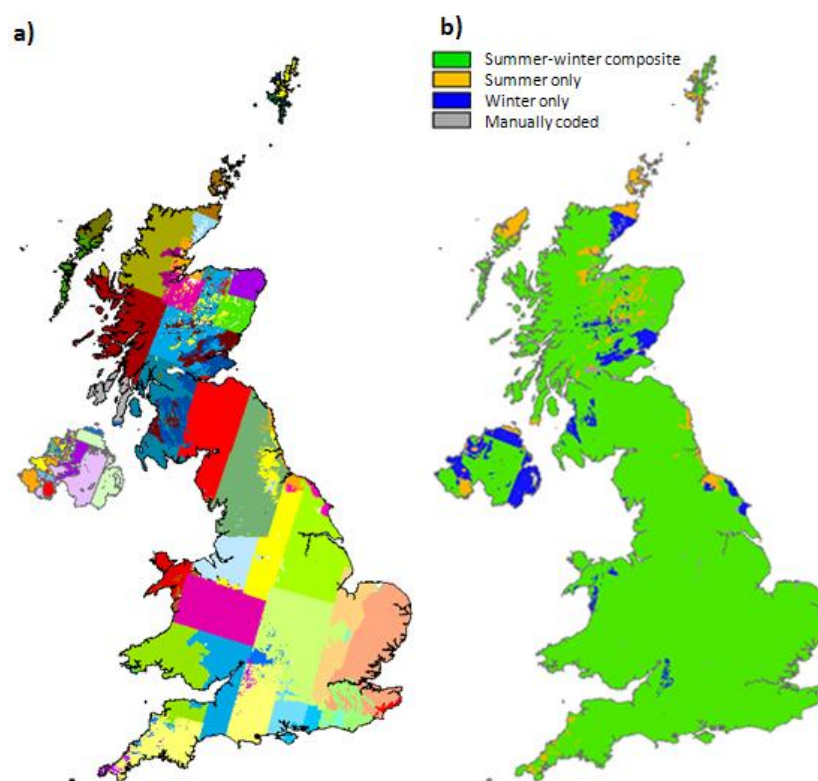

Figure 1. Satellite images used to create LCM2007, a) the mosaic of classified images used to produce full UK coverage for LCM2007 (different colours show different images) and b) area covered by composite images, single-date images and no data areas.

2.2 Ground reference points

The ground reference points were collected, during a series of field trips in 2007 and 2008, using a tablet PC-based Digital Data collection system developed for CS in 2007. The tablet was connected to a GPS and had a bespoke GIS package which was automatically updated with the tablet/car location, enabling users to easily record land cover at points along the field trip routes. The ground reference points were distributed widely across the UK and included rarer land cover types.

2.3 National cartography

Mapping in the UK is divided between the Ordnance Survey, who cover Great Britain, and Land & Property Services (LPS) for Northern Ireland. The starting point for the GB spatial framework was the Ordnance Survey MasterMap topography layer (OSMM) which is a very detailed product containing over 400 million individual features. The starting point for the NI spatial framework was the LPS's Large-scale Vector digital database, which includes both topographic and cartographic map features comprising lines, points and polygons.

3 Methods

Figure 2 shows the key processing stages in the production of LCM2007. The methods are described briefly in this paper, but for full details see Morton *et al,.* (2011).
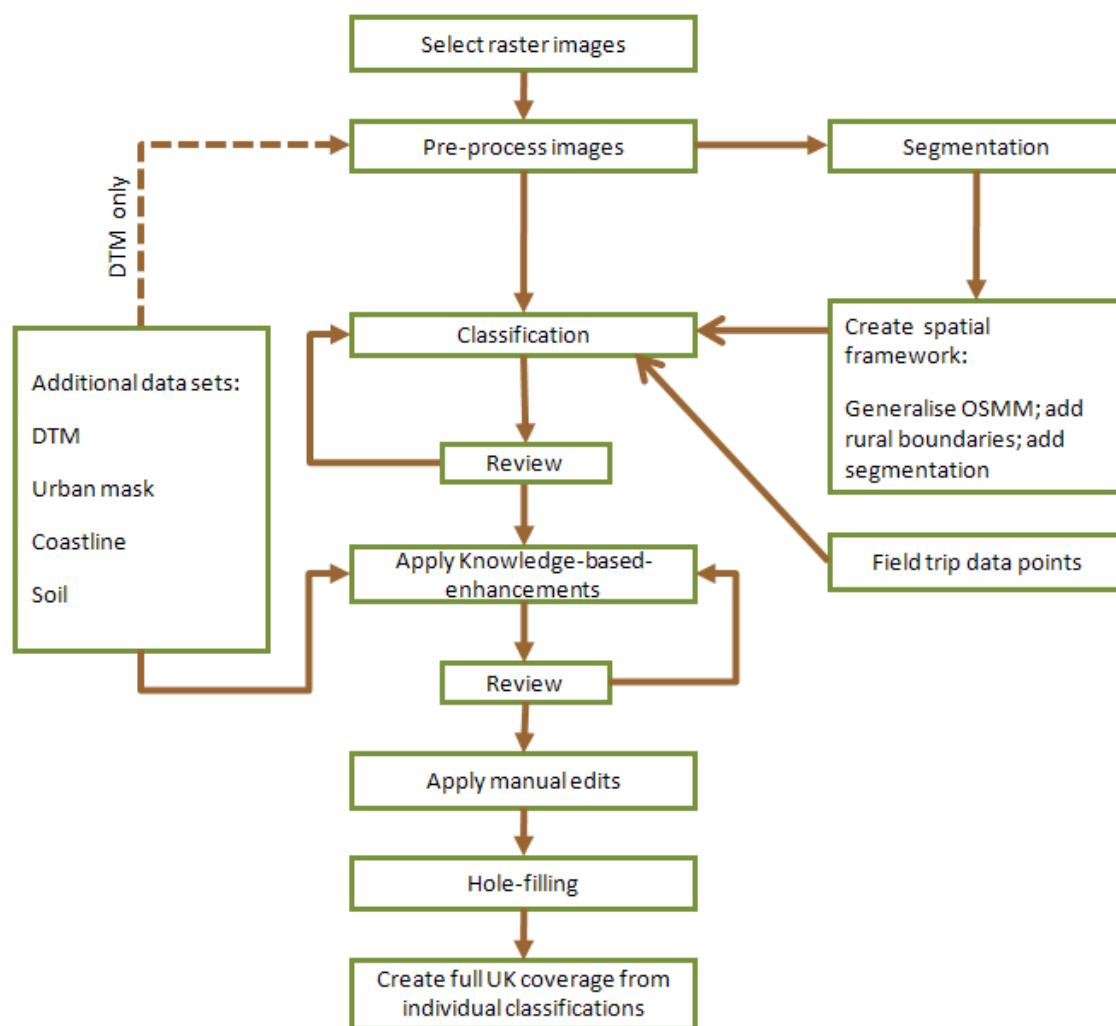

Figure 2. Flowchart showing the key stages in the image processing of LCM2007.

3.1 Spatial framework development

The LCM2007 spatial framework requirements were that it should represent real-world spatial objects, such as fields, that contain homogeneous groups of pixels. This was achieved

by generalising the cartographic data first, then adding agricultural boundaries and finally adding segments derived from the satellite-data.

3.2 Generalisation of OS MasterMap (OS MM) topography layer

Spatial generalisation is the process of simplifying a complex vector, so that unnecessary detail is removed, until the generalised vector reaches the required level of detail. It was necessary to generalise the OSMM topography layer so that it was suitable for the object-based classification of LCM2007. The generalisation applied to the OSMM topography layer was an iterative process, which repeatedly split and merged polygons. Determining whether to split or merge a polygon was based on the spatial attributes of the polygon (size and complexity) (Fig.3). Smaller polygons (red and cyan) were merged with larger polygons to create suitably sized polygons (Figure 4). Similar methods were applied for NI (Morton *et al.,* 2011).
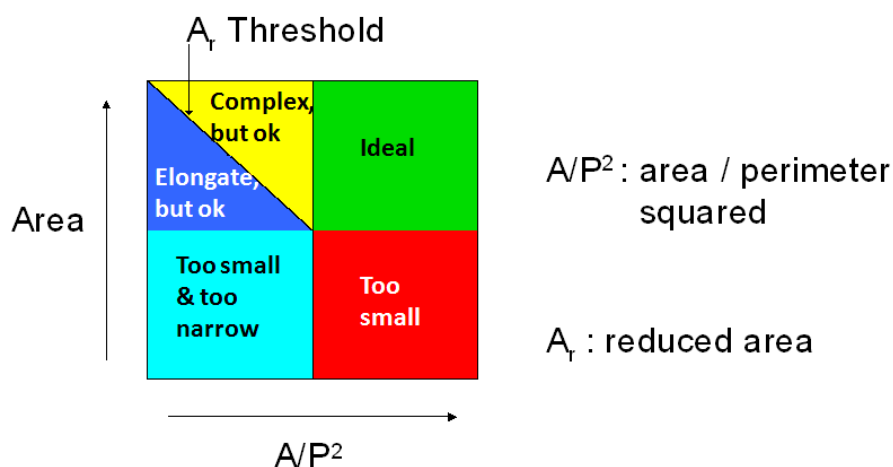

Figure 3. Scheme for classification of polygons based on area and perimeter of polygon. Note, polygons in Figure 4 are classified using this scheme.
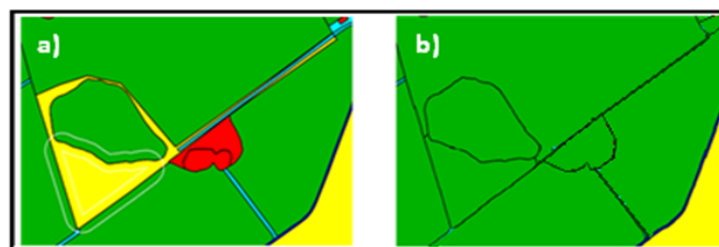
Figure 4: Polygons classified according to area:perimeter relationship (as in Figure 3), with a) polygons early in generalisation and b) polygons at end of spatial generalisation.

3.3 Raster pre-processing

The pre-processing steps carried out for LCM2007 are similar to those carried out for LCM2000 (Fuller *et al*., 2002) and to those used in other large-scale land cover mapping. The pre-processing flow line for LCM2007 was:

- Import data
- Apply cloud and cloud shadow masking
- Atmospheric correction

- Geo-registration
- Topographic correction
- Creation of composite images

To ensure consistency a handbook (Bradley *et al.,* 2009) was created to define the pre-processing methods and a log was kept for each image. After pre-processing, a segmentation algorithm was run to create segments to merge with the spatial framework produced by generalising the cartographic data.

3.4 Classification

LCM2007 is a parcel-based supervised maximum likelihood classification of the UK predominantly based on medium resolution (20-30m) satellite data. It was produced by classifying 34 composite images and 21 single-date images according to Broad Habitat-based classes (Table 1). Training areas for the classification were chosen based on the ground reference points collected during the field trips and the suitability of the pixels within the polygon for use as a spectral reference. The field trip data points were supplemented, as necessary, by points based on OS maps and web-based mapping services, such as Google Earth. Once the training areas had been selected, they were reviewed and similar spectral signatures, corresponding to the same land cover type, were grouped into spectral classes after which the classification was run (Fuller *et al.,* 2002). The probabilities of the top five spectral classes (i.e. those that matched the observed spectral signature for the polygon most closely) were recorded for each polygon. Each classification was reviewed and re-classified until the accuracy reached the required level, or until the classification could no longer be improved (Figure 2).

3.5 Knowledge-based enhancements (KBEs)

The DEM, soil and urban extent data were used to inform a series of knowledge-based enhancement (KBEs) rules, which were used post-classification to refine the thematic resolution and increase the accuracy of the classification (see Morton et al., (2011) for details). KBEs were especially important in refining the grassland categories.

4   Validation

Validation is critical for informed use of any data product. The field trip points were used to identify suitable training areas for the classification (i.e. a training data set), and also to identify suitable points for a ground reference data set against which to validate the product (i.e. a testing data set). The training and testing data sets used separate data.

4.1 Results

The comparison of LCM2007 with the 9127 ground reference points shows (Table 1) shows:

- 83% correspondence for LCM2007 classes.
- Correspondence varies depending on class.
- Correspondence is generally good for intensively managed areas, e.g. Coniferous Woodland, Arable, Improved grassland, and poorer for semi-natural habitats that typically occur in mosaics and where habitat changes are more transitional. For these areas ecotone approaches giving percentage of different habitats per pixel or per land parcel might be more appropriate.
- The lowest correspondence is for Rough, Neutral, Calcareous and Acid grassland. LCM2007 is a land cover map, but based on Broad Habitats. In the case of grassland,

this creates complications, as grassland is the land cover for the Improved, Neutral, Acid and Calcareous grassland habitats, plus some types of '*Bog*'. The refinement from grassland land cover to grassland Broad Habitats is achieved via the KBEs.

Table 1. Accuracy of LCM2007 based on 9197 LCM2007 ground reference polygons. **Green squares** correspond at LCM2007 class-level.

LCM2007 \ LCM Ground reference polygons	Broadleaved woodland	Coniferous woodland	Arable and horticulture	Improved grassland	Rough grassland	Neutral grassland	Calcareous grassland	Acid grassland	Fen, marsh and swamp	Heather	Heather grassland	Bog	Montane habitats	Inland rock	Saltwater	Freshwater	Supralittoral rock	Supralittoral sediment	Littoral rock	Littoral sediment	Saltmarsh	Urban	Suburban	Sum of rows	User's accuracy (%)
Broadleaved woodland	593	18	6	3	1	1	0	2	12	10	1	0	0	0	0	1	0	1	0	0	1	0	0	649	91
Coniferous woodland	44	1035	3	6	0	0	0	1	1	1	5	0	0	1	0	1	0	0	1	0	0	0	0	1099	94
Arable and horticulture	5	2	1248	54	15	0	5	0	2	2	4	0	0	0	0	0	3	1	0	0	0	2	0	1353	92
Improved grassland	7	5	54	1047	79	19	13	12	3	3	0	0	0	0	0	0	0	0	0	3	0	1	0	1259	83
Rough grassland	15	10	17	40	177	18	11	26	4	7	19	4	0	0	0	1	0	4	0	2	4	1	0	364	49
Neutral grassland	1	0	0	3	9	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	35	63
Calcareous grassland	0	1	1	0	4	2	41	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	49	84
Acid grassland	2	6	0	8	117	0	6	155	0	2	14	0	0	4	0	0	0	0	0	0	0	0	0	321	48
Fen, marsh and swamp	2	0	0	0	8	0	0	0	98	0	0	0	7	0	0	1	0	0	0	0	1	0	0	109	90
Heather	5	7	0	1	3	0	1	11	1	352	29	3	0	0	0	1	0	0	0	0	0	0	1	413	85
Heather grassland	8	15	1	0	16	0	2	14	1	54	132	0	3	2	0	1	0	1	0	0	0	0	0	250	53
Bog	1	1	2	2	42	0	0	13	1	54	51	114	0	7	0	0	0	0	0	0	0	0	0	289	39
Montane habitats	0	0	0	0	8	0	0	0	0	4	2	0	14	6	0	1	0	0	0	0	0	0	0	34	41
Inland rock	0	0	0	0	2	0	0	0	0	0	0	0	0	177	0	1	0	1	1	1	0	5	0	193	92
Saltwater	0	0	5	0	0	0	0	0	0	0	0	0	0	0	153	56	0	0	1	4	1	0	0	223	69
Freshwater	1	1	0	0	0	0	0	0	1	1	0	0	1	2	3	409	0	0	0	2	1	0	0	419	98
Supralittoral rock	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	23	0	0	1	0	0	0	24	96
Supralittoral sediment	0	0	4	1	0	0	0	0	0	1	0	0	0	0	0	0	1	126	0	9	3	0	0	164	77
Littoral rock	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	3	3	0	0	108	93
Littoral sediment	0	0	0	0	0	0	0	0	0	1	0	0	0	1	4	2	1	9	3	256	11	2	0	298	86
Saltmarsh	0	0	0	0	4	0	0	1	1	0	0	1	0	0	14	1	0	2	0	11	179	0	0	188	95
Urban	1	0	11	6	1	0	0	0	0	0	0	0	0	12	0	1	0	0	0	2	0	559	84	680	82
Suburban	0	0	5	10	3	0	0	0	0	1	0	0	0	1	0	2	0	0	0	0	0	62	520	606	86
Sum of columns	685	1103	1367	1181	499	61	79	236	125	493	257	122	25	230	174	478	28	145	106	294	202	632	605		
LCM2007 correspondence: Producer's accuracy (%)	87	94	91	89	35	36	52	66	78	71	51	93	56	77	88	86	82	87	94	87	89	88	86		

Correspondence (LCM2007 classes): 83%

BH correspondence: 85%

5   Conclusions

LCM2007 contains almost 10 million objects, which present the land cover of the UK using 23 Broad Habitat-based land cover classes. LCM2007 represents a large-scale application of OBIA methods to produce a meta-data rich product, which is suitable for a wide range of science, policy and land management applications.

6   Accessing LCM2007 data sets

LCM2007 is available as a vector data set; a 25m raster data set and a series of 1km products (see www.ceh.ac.uk for further details). The 1km products can be downloaded freely from the CEH Information gateway, whilst the other data sets are available on request (a fee may be chargeable) from CEH (see CEH website or email: spatialdata@ceh.ac.uk for details).

7   Acknowledgements

References

BRADLEY, A., THOMSON, A., CUEVAS, M., and LOWE, B., *(2009) Raster pre-processing chain for LCM2007 (ENVI)*, Unpublished CEH internal document, pp.62

FULLER, R. M., SMITH, G. M., SANDERSON, J. M., HILL, R. A., THOMSON, A. G., COX, R., BROWN, N. J., CLARKE, R. T., ROTHERY, P., and GERARD, F. F., (2002) *Countryside Survey 2000 Module 7. Land Cover Map 2000. Final Report.* NERC/Centre for Ecology and Hydrology. (CEH: Project Report Number C00878) (Unpublished)

MORTON, D., ROWLAND, C., WOOD, C. MEEK, L., MARSTON, C., SMITH, G., WADSWORTH, R., SIMPSON, I.C. 2011. *Final Report for LCM2007 - the new UK land cover map. Countryside Survey Technical Report No 11/07.* NERC/Centre for Ecology & Hydrology 112pp. (CEH Project Number: C03259).