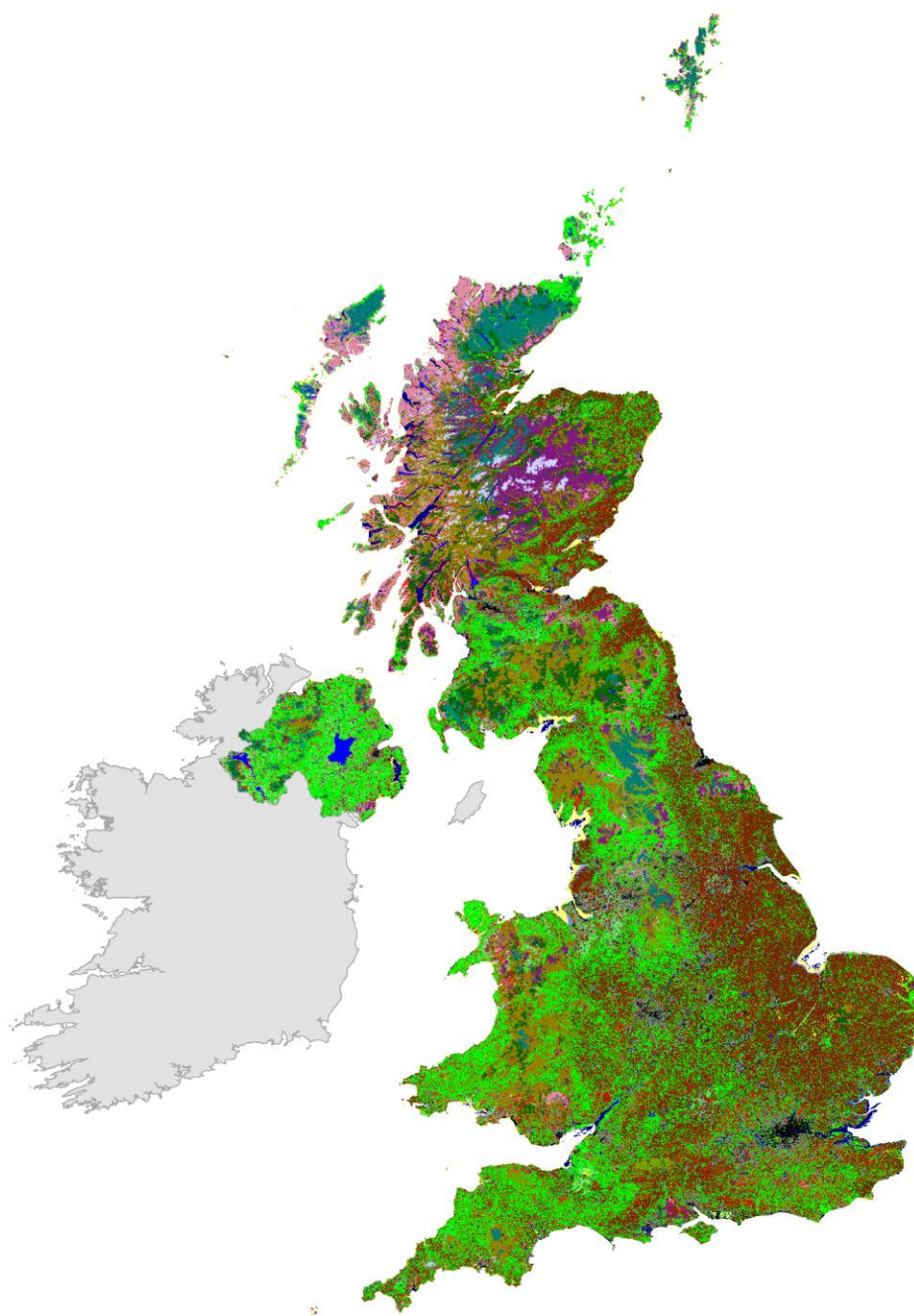


# The UKCEH Land Cover Map for 2021



Version	Date	Updates
1.0	20/07/2022	First version

## Purpose

This is the user guide for the new UKCEH Land Cover Map for 2021, LCM2021. The LCM2021 product consists of six datasets, three each for Great Britain and Northern Ireland comprising a 10 m dataset of classified pixels, a dataset of classified land parcels, and a 25 m pixel dataset or rasterised land parcels. Appropriate use of complex data requires a deeper understanding than structure and content. We therefore describe data production, validation and data accuracy. Our goal is to help users make informed decisions regarding the application UKCEH LCM data in their current and future work.

## Contents

Glossary.....	3
Introduction .....	4
UK CEH Land Cover Classes and BAP Broad Habitat Classes .....	4
Dataset descriptions .....	6
10 m Classified Pixel datasets .....	6
Land Parcel datasets .....	6
25 m Pixel Rasterised Land Parcel datasets .....	8
1km raster products.....	9
Material and Methods .....	10
Seasonal Composite Images .....	10
Context Rasters .....	10
Classification Scenes .....	11
The UKCEH Land Parcel Spatial Framework.....	12
Bootstrap Training .....	12
Random Forest classification .....	13
Product validation.....	14
References .....	14
Appendix 1. Notes on UKCEH Land Cover Classes .....	15
Appendix 2: Biodiversity Action Plan (BAP) Broad Habitats .....	20
Appendix 3. Full list of datasets for UKCEH LCM2021 .....	24
Appendix 4. Correspondence matrices for UKCEH LCM2021 .....	25
Appendix 5: Recommended colour recipe for displaying UKCEH Land Cover Classes .....	26
Appendix 6: Recommended colour recipe for displaying UKCEH Land Cover Classes colour blind friendly .....	27

## Glossary

**Bootstrap Training:** A method that automatically selects training observations from a historical land map. Bootstrap Training datasets are used by a Random Forest to classify Classification Scenes.

**Context Raster:** A multi-band raster with contextual information, such as terrain, coastal and urban proximity. Context Rasters are combined with Seasonal Composite Images to create Classification Scenes.

**Classification Scene:** A multi-layer raster comprising spectral information from a Seasonal Composite Image and Context Rasters. Classification Scenes are classified to yield a pixel land cover classification.

**Seasonal Composite Image:** A multi-band raster representing spectral observations across four seasons. These are designed to provide temporal, phenological information that a classifier can use to differentiate vegetation types.

**UKCEH Land Cover Classes:** A set of land cover classes defined by UKCEH, derived from the UK Biodiversity Action Plan Broad Habitats (Jackson *et. al.* 2000).

**UKCEH Land Parcel Spatial Framework:** A database of land parcel objects derived from generalising national cartography in order to remove unnecessary detail. All UKCEH LCM Land Parcel datasets use this spatial framework.

## Introduction

In this document we capitalise the first letters of some specific terminology. In most cases the meaning of these terms should be clear from the text, but for extra clarity a glossary is provided for these capitalised terms. Land cover types are also capitalised and/or italicised; these are described in appendices.

LCM2021 is the ninth UK land cover map produced by UKCEH (others include LCM1990, 2000, 2007, 2015, 2017, 2018, 2019, 2020). Land cover is given as 21 UKCEH Land Cover Classes based on Biodiversity Broad Habitats (BAP, see Jackson 2000). LCM2021 was created by classifying Classification Scenes comprising Sentinel-2 Seasonal Composite Images combined with ten Context Layers designed to reduce spectral confusion. The classification process is automatic, using a technique we call Bootstrap Training, combined with a Random Forest classifier.

Historically, because of high costs associated with labour intensive image pre-processing and gathering training data, there has been lag of approximately ten years between UKCEH LCMs. However, using new automated techniques we are able to produce new land cover maps annually. These methods are still evolving and we seek continuous improvements. If new methods yield significant improvements in accuracy, we will consider re-application across the entire catalogue. This will maximise temporal consistency, comparability and the potential for change detection, which is a fundamental goal for our land cover mapping. A benefit of annual production is that it helps to differentiate errors from real-world change. Most classification errors are random in space and time, so should not occur at exactly the same location year-after-year. With a rich time-series of land cover maps random errors will flicker on and off, whilst real land surface changes will persist. We therefore anticipate that as the annual series matures, UKCEH annual land cover maps will become an essential tool for monitoring the state and change of the UK countryside and support a wide range of environmental objectives. Formal validation of LCM2021 gives an overall accuracy of 82.6 %, at full thematic detail.

## UK CEH Land Cover Classes and BAP Broad Habitat Classes

At the turn of the century, conservation and regulatory agencies had reporting obligations under the UK Biodiversity Action Plan (BAP, Jackson *et al.* 2000) and were co-funding stakeholders of LCM2000 and LCM2007. Therefore, the 21 UKCEH Land Cover Classes are closely related to Biodiversity Action Plan (BAP) Broad Habitats, but they are not the same. BAP Broad Habitats were designed for field-based detection by botanists rather than via satellite remote sensing, therefore there are slight deviations. Agencies no longer have reporting obligations under the UK BAP, but we retain the land cover descriptions to preserve consistency and comparability with earlier products.

Describing complex land cover and habitat types that share similar nomenclatures and similar (but not equal) meanings precisely with words can be ambiguous. In most cases we could refer to UKCEH Land Cover Classes and BAP Broad Habitats synonymously without causing confusion, however we sometimes need to contrast them. To reduce ambiguity regarding land cover and habitat descriptions, we therefore italicise UK BAP Broad Habitats when explicitly referring to these. At all times when referring to a defined class (UKCEH or UK BAP) we will begin each element with a capital letter. For example: Improved Grassland refers to the UKCEH Land Cover Class, *Improved Grassland* the BAP habitat.

Some users do not require the full thematic detail of UKCEH Land Cover Classes so we also provide generalised land cover UKCEH Aggregate Classes in our Land Parcel product. Table 1 shows the relationship between UKCEH Land Cover Classes and Aggregate Classes, with the UK BAP Broad

Habitats. Usually there is a one-to-one relationship between UK BAP Broad Habitats and derived UKCEH Land Cover Classes. However, the UK BAP *Standing water and Canals* and the UK BAP *Rivers and Streams* are represented by a single UKCEH Land Cover Class, Freshwater. There is not a UK BAP for saltwater, but we separate saltwater and freshwater when possible, so include a UKCEH LCM Saltwater class. In two cases, we have found it appropriate to split UK BAP Broad Habitats. The UK BAP *Dwarf Shrub and Heath* class is split into UKCEH Land Cover Classes Heather and Heather Grassland. The UK BAP *Built up Areas and Gardens* class is split into UKCEH Land Cover Classes Urban and Suburban. Fuller details are given in Appendices 1 & 2.

*Table 1.* Relationship between UK CEH Aggregate classes, UK BAP Broad Habitats (Jackson 2000, highlighted in green) and UKCEH Land Cover class and associated integer identifiers. Additional notes on the on the relationships between UKCEH Land Cover Classes and UK BAP Broad Habitats are given in Appendices 1 and 2 together with satellite-based mapping considerations.

UKCEH Aggregate Class (AC)	AC Identifier	UK BAP Broad Habitat	UKCEH Land Cover Class	LC Identifier
Broadleaf woodland	1	<i>Broadleaved mixed and yew woodland</i>	Deciduous woodland	1
Coniferous woodland	2	<i>Coniferous woodland</i>	Coniferous woodland	2
Arable	3	<i>Arable and horticulture</i>	Arable	3
Improved grassland	4	<i>Improved grassland</i>	Improve grassland	4
Semi-natural grassland	5	<i>Neutral grassland</i>	Neutral grassland	5
		<i>Calcareous grassland</i>	Calcareous grassland	6
		<i>Acid grassland</i>	Acid grassland	7
		<i>Fen marsh and swamp</i>	Fen	8
Mountain, heath and bog	6	<i>Dwarf shrub and heath</i>	Heather	9
			Heather grassland	10
		<i>Bog</i>	Bog	11
		<i>Inland rock</i>	Inland rock	12
Saltwater	7		Saltwater	13
Freshwater	8	<i>Standing open water and canals</i>	Freshwater	14
		<i>Rivers and streams</i>		
Coastal	9	<i>Supralittoral rock</i>	Supralittoral rock	15
		<i>Supralittoral sediment</i>	Supralittoral sediment	16
		<i>Littoral rock</i>	Littoral rock	17
		<i>Littoral sediment</i>	Littoral sediment	18
			Saltmarsh	19
Built-up areas and gardens	10	<i>Built-up areas and gardens</i>	Urban	20
			Suburban	21



## Dataset descriptions

LCM2021 includes three geospatial datasets (Table 2) for Great Britain and Northern Ireland, six in total. Appendix 3 gives the official dataset name for each of these digital holdings.

### 10 m Classified Pixel datasets

The 10 m pixel datasets are produced from multiple classified Classification Scenes that are compiled to create a single mosaic of national cover. The Random Forest classifier derives a probability of membership for each class. The class with the highest is the most likely land cover. The first band of the 10 m pixel dataset is an integer representing the most likely UKCEH Land Cover Class. The second band is the probability that was associated with this result; this gives an indication of the confidence of the classification (Figure 1).

Unlike pixels of the 25 m Rasterised Land Parcel dataset, the 10 m Classified Pixels have not been generalised by combination with the UKCEH Land Parcel Spatial Framework. This preserves intricate features of the landscape such as narrow linear features and small patches of habitat that fall below the 0.5 hectare minimum mappable unit (MMU) of the UKCEH Land Parcel Spatial Framework. We anticipate that this extra detail will be of interest and useful but please note the formal validation exercises were performed against the Land Parcel dataset, not the 10 m Classified Pixels. We therefore recommend these datasets to specialist users who understand the consequence of this statement.

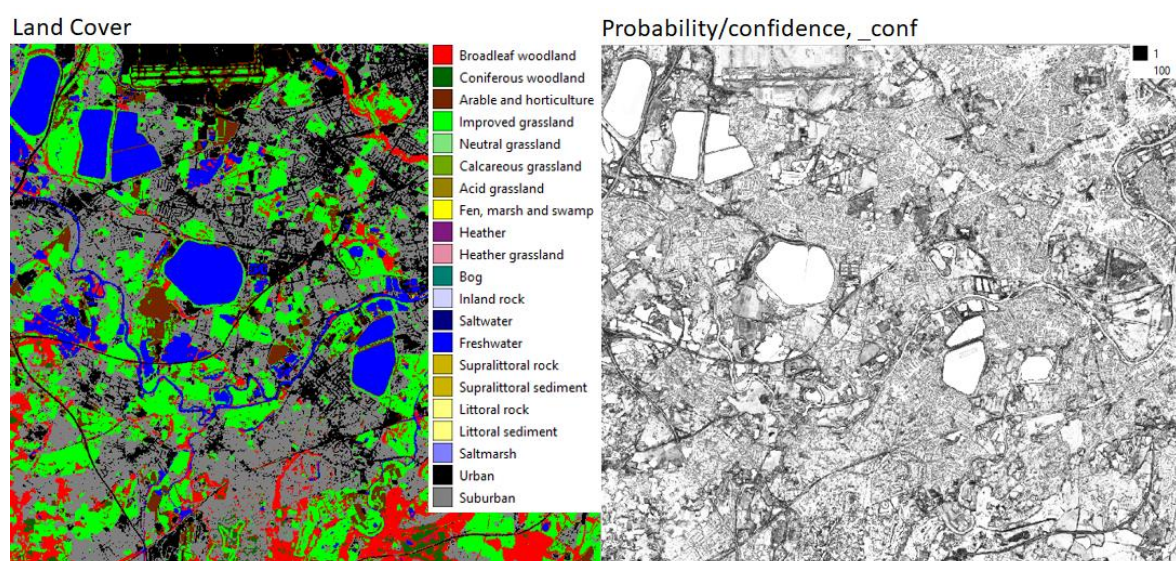


Figure 1. Left; 10 m classified pixels. Right; the probability associated with each classified pixel; this is an indicator of classification confidence.

### Land Parcel datasets

The Land Parcel datasets are the result of intersecting the 10 m Classified Pixel datasets with the UKCEH Land Parcel Spatial Framework to generate land parcel attributes (Figure 2, Table 3).

## Classified land parcels

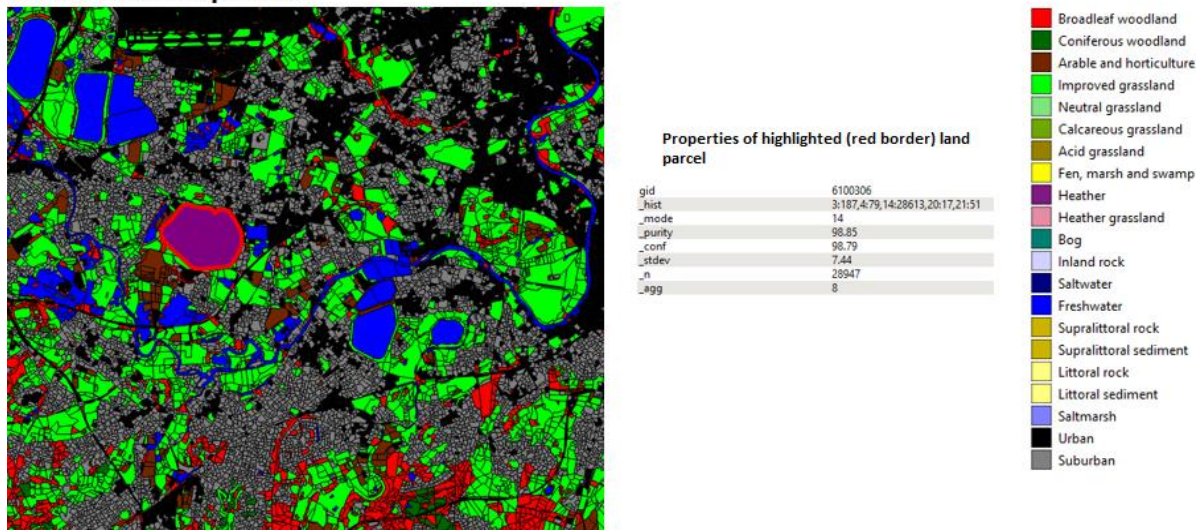


Figure 2. Classified land parcels and properties of the highlighted (red border) land parcel.

Table 2. UKCEH LCM product suite details (metadata).

	Great Britain, GB		Northern Ireland, NI	
Coordinate System	British National Grid. EPSG: 27700		TM 75 Irish Grid. EPSG: 29903	
Land Parcel Products				
Number of parcels	6741288		902248	
8-bit Raster Products				
Extent, metres. (East min, East max, North min, North max)	(0, 700000, 0, 1300000)		Truncated Irish Grid to: (180000, 400000, 300000, 500000)	
Product Name	Classified Pixels	Rasterised Pixels	Classified Pixels	Rasterise Pixels
Pixel resolution	10 m	25 m	10 m	25 m
Number of bands	2	3	2	3

Table 3. UKCEH LCM Land Parcel product attributes for LCM2021.

UKCEH LCM attribute name	Example	Description
gid (ogc_fid)	10079	This is a unique identifier for the land parcel

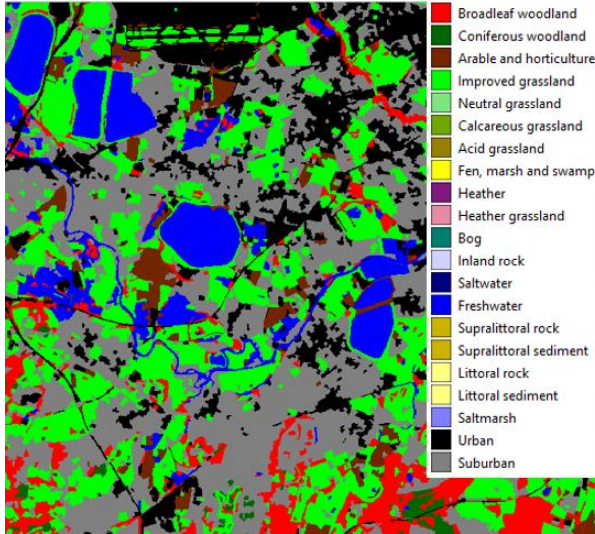
<i>_hist</i>	2:3, 11:361	This describes the frequency of 10 m pixels per UKCEH land cover class encountered in a land parcel given as a list of tuples <left>:<right> separated by commas. The figure to the left of the colon per tuple represents the UKCEH Land Cover Class identifier, to the right the frequency of occurrence.
<i>_mode</i>	11	This is the most frequently occurring (modal) land cover type.
<i>_purity</i>	99.06	This is the percentage of the modal land cover class over the total number of pixels ( <i>_n</i> ) encountered.
<i>_conf</i>	93.88	This when combined with <i>_purity</i> helps to understand classification confidence. It is the mean polygon value of the class membership probability for each pixel, rescaled to the range of 0 to 100.
<i>_stdev</i>	9.51	This is the standard deviation of <i>_conf</i> .
<i>_agg</i>	6	This is the aggregate class (see Table 1)
<i>_n</i>	319	This is the total number of 10 m pixels intersecting the land parcel

### 25 m Pixel Rasterised Land Parcel datasets

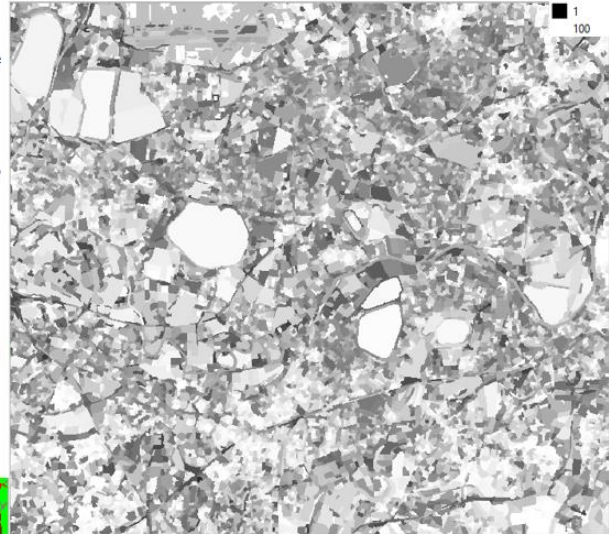
These are the result of rasterising the Land Parcel dataset into 25 m pixels, with the pixel origin matching the origin of the British National Grid (for Great Britain) or the Irish National Grid (for Northern Ireland). Three attributes are carried from the Land Parcel datasets, giving a 3-band raster. Band 1 is the dominant land cover *\_mode*; band 2 is *\_conf*; and band 3 *\_purity* (see Figure 3).



Per-parcel Land Cover, rasterised 25m



Per-parcel mean probability/confidence, \_conf



Per-parcel pixel purity, \_purity

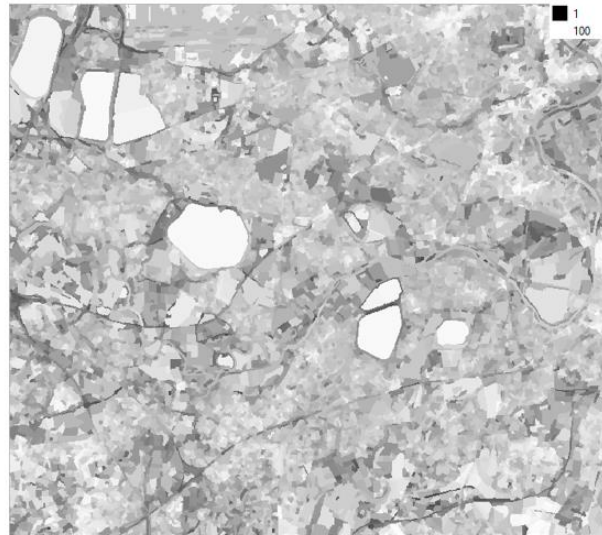


Figure 3. Left top, rasterised land parcels at 25 m pixel resolution. Right-top mean parcel most likely land cover probability (\_conf). Bottom per parcel purity (\_purity).

### 1km raster products

The 1km products for LCM2021 were created by summarising the LCM2021 25 m raster data set to produce percentage cover for each class and to identify the dominant (most widespread) class in each 1 km pixel. The products at 1 km are:

- Dominant cover at 1 km for LCM2021 classes (1-band)
- Dominant cover at 1 km for LCM2021 Aggregate classes (1-band)
- Percentage cover at 1 km for LCM2021 classes (21 bands)
- Percentage cover at 1 km for LCM2021 Aggregate classes (10 bands)

For the percentage cover products, integer values are reported, so the rounding of each value means that in some cases the sum of all the layers may be slightly above or below 100. Around the coast the values will be sum to less than 100 reflecting the proportion of mapped area in the 1 km square. One 1 km x 1 km pixel covers the area of 1600 25 m pixels, consequently, a land cover type has to cover more than sixteen 25 m-pixels to exceed 1% of the 1 km pixel area, and hence have a percentage cover value in the 1 km percentage cover data sets.

## Material and Methods

### Seasonal Composite Images

Seasonal Composite Images for classification were derived from Google Earth Engine. Sentinel-2 surface reflectance values were resampled to 10 m pixel resolution and median reflectance was computed for four time periods: January-March, April-June, July-September, and October-December, using ten Sentinel-2 bands 2, 3, 4, 5, 6, 7, 8, 8a, 11, and 12 (after Carrasco *et al.*, 2019), see Table 4. There were occasional seasonal gaps in the Seasonal Composite Images due to persistent cloud in an image compositing period which were represented by null data. Fortunately, these data gaps never occurred at a single location in all four of the Seasonal Composites, ensuring that at least some Sentinel-2 satellite data were available for classification. Our classification algorithm will tolerate partially complete spectral information, so we were able to produce land cover for the whole of the UK without the need to manually fill gaps.

*Table 4. Sentinel-2 spectral bands and spatial resolutions.*

Sentinel-2 Band	Central wavelength (µm)	Resolution (m)
Band 1 - Coastal aerosol	0.443	60
Band 2 - Blue	0.490	10
Band 3 - Green	0.560	10
Band 4 - Red	0.665	10
Band 5 - Vegetation red edge	0.705	20
Band 6 - Vegetation red edge	0.740	20
Band 7 - Vegetation red edge	0.783	20
Band 8 - Near infra-red	0.842	10
Band 8A - Vegetation red edge	0.865	20
Band 9 - Water vapour	0.945	60
Band 10 - Short wave infrared - cirrus	1.375	60
Band 11 - Short wave infrared	1.610	20
Band 12 - Short wave infrared	2.190	20

### Context Rasters

Spectral confusion can occur between different land cover types that have similar spectral properties. For example, bare rocks in the littoral coastal zone lack significant vegetation, so too do exposed mountain rocks and sealed urban surfaces. Spectrally these surfaces can appear very similar when viewed from space and extra detail is required to differentiate them. We used 10 m Context Rasters to resolve a range of confusion types. For Great Britain, the 10 m Context Rasters were:

1. Height, derived from the NEXTMap® terrain product from Intermap® Solutions.
2. Aspect, derived from the NEXTMap® terrain product from Intermap® Solutions.
3. Slope, derived from the NEXTMap® terrain product from Intermap® Solutions.

4. Distance from the nearest building, derived from Ordnance Survey open data.
5. Distance from road, derived from Ordnance Survey open data.
6. Distance from sea, derived from Ordnance Survey open data.
7. Distance from freshwater, derived from Ordnance Survey open data.
8. A foreshore binary mask, derived from Ordnance Survey open data.
9. A tidal water binary mask, derived from Ordnance Survey open data.
10. A woodland binary mask, derived from Ordnance Survey open data.

For Northern Ireland 10 m Context Rasters were:

1. Height, derived from the NEXTMap® terrain product from Intermap® Solutions.
2. Aspect, derived from the NEXTMap® terrain product from Intermap® Solutions.
3. Slope, derived from the NEXTMap® terrain product from Intermap® Solutions.
4. An urban binary mask derived from open data of the Northern Ireland Statistics and Research Agency.
5. A distance to coast layer derived from the Ordnance Survey or Northern Ireland open data.
6. A distance to freshwater layer derived from the Ordnance Survey or Northern Ireland open data.
7. A distance to road layer derived from the Ordnance Survey or Northern Ireland open data.
8. A foreshore binary mask, derived from Ordnance Survey or Northern Ireland open data.
9. A tidal water binary mask, derived from Ordnance Survey or Northern Ireland open data.

## Classification Scenes

For Great Britain, a grid of tiles based on a modified version of the Ordnance Survey 100 x 100 km tile grid was created (Figure 4). These were used to select and extract 50-band Seasonal Composite Images from the Google Earth Engine comprising the Seasonal Composite images and context raster layers. In total, we classified 32 Classification Scenes comprising full coverage of the GB land surface. Each Classification Scene was trained and classified independently. The approximate 100 x 100 km tile size was chosen as this provides a manageable size for processing. Moreover, if regions were much larger phenological variation due to climatic difference across a Classification Scene could begin to degrade results. Some tiles such as those encompassing the Western Isles, Orkney and Shetland, and Cornwall and the Scilly Isles are larger. These are intentionally enlarged to avoid a sparsity of training data due to the extensive presence of sea in these tiles. Occasionally where tile extents are modified to include specific areas overlap between adjacent tiles do occur.

For Northern Ireland, because of its much smaller area, it was possible to use a single Classification Scene determined by the minimum bounding rectangle of the Northern Ireland land mass (Figure 2). The 40-band Sentinel-2 Seasonal Composite Image was combined with Northern Ireland Context Rasters to produce a 49-band single Classification Scene.

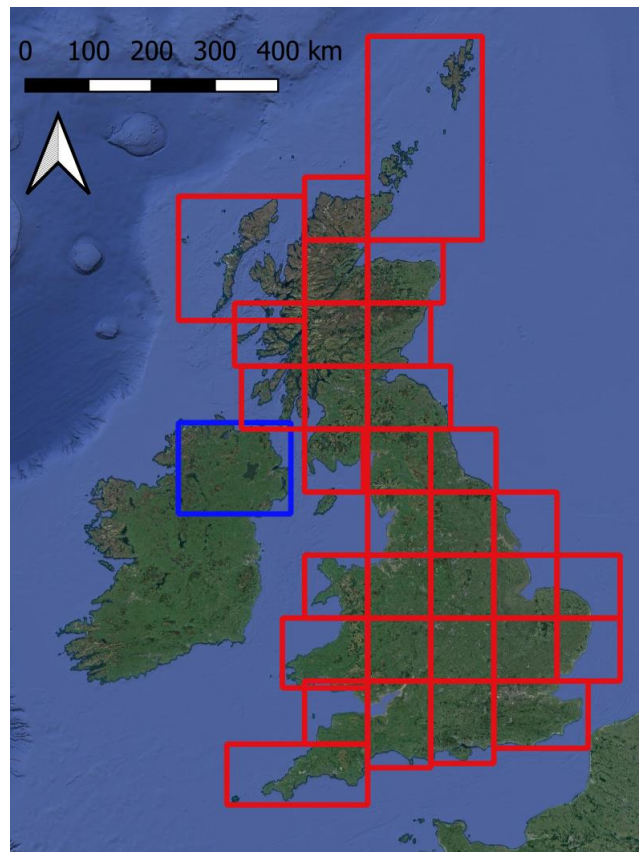


Figure 4. Tiles used for selecting Sentinel-2 Seasonal Composite Images for Great Britain (red) and Northern Ireland (blue).

### The UKCEH Land Parcel Spatial Framework

The UK Land Parcel Spatial Framework used for LCM2021 was originally developed for LCM2007 (Morton *et. al.*, 2011) and a modified version was used for LCM2015. We have had to make some minor changes since LCM2015. The total number of land parcels and their geometries are unchanged, but we have re-ordered database storage and defined new indices to enable faster processing. The consequence of this is that the unique identifiers (the *gid* attribute) for each land parcel do not match those provided in the LCM2015 Land Parcel dataset. For most users this will have no significance, but for users who wish to compare the new land cover with LCM2015 using parcel identifiers, instead of spatial overlap, this will not be immediately possible.

Land parcels in the UKCEH Land Parcel Spatial Framework have a minimum area of approximately 0.5 hectares, the MMU. The land parcels were derived by generalising national cartography (Morton *et. al.*, 2011; Smith *et. al.*, 2007) and are designed to represent discrete real-world units of land such as fields, parks, urban areas, woodlands, lakes and so forth. It is usual, but not universal, that the land parcels are dominated by a single land cover type. Organising the 10 m Classified Pixels into land parcels helps to reduce classification noise to provide a clean, easier to use product. It also provides a convenient fixed structure for comparing land cover over time for change detection.

### Bootstrap Training

Bootstrapping is used to refer to a self-starting process that proceeds without external input. UKCEH have developed an automatic training process for land cover/habitat classification that does not require a fresh collection of (expensive) field-gathered data for classifier training, and we have named



the process Bootstrap Training. Land cover and habitat change is usually gradual. Transitions from one land cover or habitat to another typically occur over a number of years. Therefore recent habitat/land cover maps can be a valuable source of training data for a new map if the original maps are accurate and the update interval of the new map is short relative to target dynamics. When this is true, land cover observations from the historic maps can be used to sample the current satellite image to produce training observations. These can then be used by a RF classifier to yield a classification result, which contributes to the bootstrap for the next map and so forth. Because the historic maps give wall-to-wall coverage they provide a very large number of training observations, which is the key to learning success. Machine-learning algorithms, such as RF, rely on the majority signal to assign class membership, so when the Bootstrap Training set is very large if a minor proportion have changed class (are incorrect) over the refresh interval since these will have little influence on the dominant signal.

The Bootstrap Training dataset for UKCEH LCM2021 came from UKCEH LCM2018, LCM2019 and LCM2020. We filtered these land cover products retaining only land parcels with >80% purity and which were classified as the same land cover class across all three years. Figure 5 gives an example of a Bootstrap Training dataset and resulting classification result.

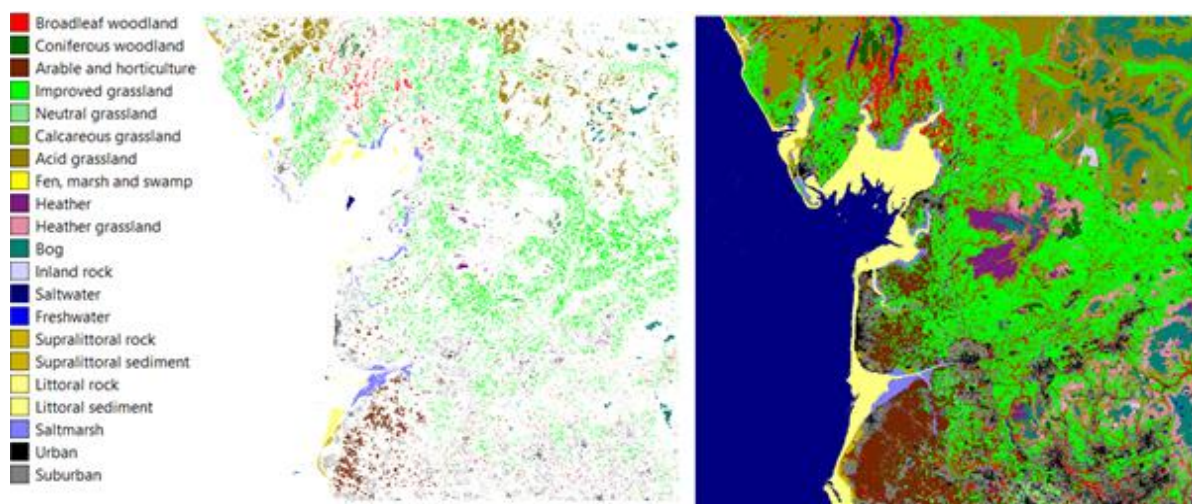


Figure 5. UKCEH Land Cover classes (left), a Bootstrap Training set (centre) and the resultant RF land cover classification (right).

### Random Forest classification

Random Forest (RF) classification (Breiman 2001) is a supervised learning technique that uses a training set of known observations to derive an empirical relationship which is then used to predict the membership of unknown observations. Bootstrap Training data represent land parcels; these we used to sample all underlying pixels from the Classification Scenes for 2021. Pixels were placed into labelled bags and from each bag 10,000 samples, with replacement were drawn to train the RF classifier. The RF classifier subsequently yields the 10 m Classified Pixel product (for example, Figure 3, right). Sampling with replacement ensures that all land cover classes have an equal number of pixel observations for training the RF, this balances learning. Without balanced learning the signal of rarer classes will be weak and susceptible to domination from commoner classes, causing misclassification.

The classification software used for UKCEH LCMs is bespoke and was developed by UKCEH scientific staff. It integrates the Weka (Frank *et al.* 2016) machine learning suite with a PostGIS geospatial database and gdal tools (<https://gdal.org/>). These are all open source technologies.



## Product validation

LCM2021 was validated using 35,182 validation reference points collected from across the UK and covering all 21 LCM classes. The reference data is a composite data set, derived from the GB countryside survey, open source National Forest Inventory data, Rural Payment Agency data and a set of bespoke LCM validation points generated from manual image interpretation and field collection. These were intersected with the UKCEH LCM 25 m rasterised polygon datasets to determine correspondence. The overall accuracy of LCM2021 is 82.6 %, with a 95 % confidence interval of 82.19 % - 82.99 % (Appendix 4).

## References

Breiman, L., 2001. Random forests. *Machine learning*, 45(1), pp.5-32.

Carrasco, L., O'Neil, A.W., Morton, R.D. and Rowland, C.S., 2019. Evaluating combinations of temporally aggregated Sentinel-1, Sentinel-2 and Landsat 8 for land cover mapping with Google Earth Engine. *Remote Sensing*, 11(3), p.288.

Frank, E., Hall, M. A., and Witten I. H. (2016). The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition, 2016.

Jackson D.L., (2000), Guidance on the interpretation of the Biodiversity Broad Habitat Classification (terrestrial and freshwater types): Definitions and the relationship with other classifications, JNCC Report 307, 73 pages, ISSN 0963 8091(<http://www.jncc.gov.uk/page-2433>)

Morton, D., Rowland, C., Wood, C., Meek, L., Marston, C., Smith, G., Wadsworth, R. and Simpson, I., 2011. Final Report for LCM2007-the new UK land cover map. Countryside Survey Technical Report No 11/07.

Smith, G., Beare, M., Boyd, M., Downs, T., Gregory, M., Morton, D., Brown, N. and Thomson, A., 2007. UK land cover map production through the generalisation of OS MasterMap®. *The Cartographic Journal*, 44(3), pp.276-283.

## Appendix 1. Notes on UKCEH Land Cover Classes

(see also Appendix 2 for a summary of UK BAP Broad Habitat definitions)

UKCEH Land Cover Class	Notes
Broadleaved woodland	<p>In the UK BAP <i>Broadleaved, mixed and yew woodland</i> the broadleaved woodlands are characterised by stands &gt;5 m high with tree cover &gt;20%. Scrub (&lt;5 m) requires a cover &gt;30 % for inclusion. Such fine distinctions cannot be made through remote sensing. Open-canopy woodland (stands with trees &lt;50 %) is a particular problem, albeit occurring relatively rarely. These are likely to be confused with other classes due to the dominance of the non-woodland understory.</p> <p>Broadleaved evergreen trees rarely occur in stands &gt;0.5 hectares; an area large enough to create suitable training areas appropriate for classification. Consequently the classifier would struggle with this land cover. It is likely these stands will be classified as Conifer because of the full-year chlorophyll signal.</p> <p>Mixed woodland 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.</p> <p>Stands with near-closed canopies can be interpreted easily in the field and pure examples can normally be found for training the classifier.</p>
Coniferous woodland	<p>The UK BAP <i>Coniferous Woodland</i> class includes semi-natural stands and plantations, with cover &gt;20 %. Classification of coniferous woodland is generally straightforward, but rare examples of open canopy semi-natural pinewoods are likely to be classified according to the dominant understorey class.</p> <p>The UK BAP includes new plantation and recently felled areas. These are land use, not land cover. Newly felled areas are often dominated by grass, heather and encroaching vegetation and more likely to be classified as these, instead of coniferous woodland.</p> <p>Deciduous larch has potential for confusion with broadleaved deciduous woodland but is generally correctly identified.</p>
Arable and Horticulture	<p>The BAP Broad Habitat <i>Arable and Horticulture</i> includes annual crops, perennial crops such as berries and orchards and freshly ploughed land. This is a very broad class and as a consequence has large potential for spectral confusion with non-arable surfaces. The main confusion between arable and other classes occurs between arable land and improved grassland. This is especially likely when grassland is managed by cutting, followed by periods of low growth and reflectance from chlorophyll. When this happens the observed seasonal reflectance pattern can be similar to graminid crops, such as wheat and barley. Indeed grass managed in this way is technically a crop, so an arable classification is not necessarily wrong.</p>

Improved Grassland	Improved grassland is distinguished from semi-natural grasslands based on its higher productivity, lack of winter senescence, location and/or context. Grasslands lie on a continuum, so some confusion with other grassland types is inevitable. Confusion with grass-like crops will also occur.
Neutral Grassland	The UK BAP Broad Habitat <i>Neutral Grassland</i> is expected to be challenging for satellite-based classification. BAP <i>Neutral Grassland</i> is defined by botanical composition and includes semi-improved grasslands managed for silage, hay or pasture (Jackson, 2000). There is not generally an obvious spectral difference between these and other productive grass types. However, the inclusion of Context Rasters for slope and distance to rivers appear to have helped greatly with Neutral Grassland detection.
Calcareous Grassland	Calcareous Grassland class is mapped spectrally. However, the inclusion of ancillary layers for slope is expected to improve results. UKCEH does not have free access to a highly resolved soil PH/soil type layer, which we would expect to help further. For regions known to contain substantial coverage of Calcareous Grassland, for example Limestone Dales of Derbyshire and North Yorkshire, the South Downs and Salisbury Plain our results match expectations.
Acid Grassland	<p>The UK BAP <i>Acid Grassland</i> can be spectrally variable, depending on dominant species composition. Deciduous Acid grassland, dominated by <i>Molinia caerulea</i> has a distinct signal from acid grasslands dominated by mixtures of other grasses, rushes, mosses, herbs and sedges. In other work we have been able to refine this class successfully. However, we did not make this separation in historical maps, so we are not able to retrieve suitable observations from Bootstrap Training.</p> <p>Bracken has a very distinctive spectral signal, but only at certain times of the year when its foliage begins to dominate its grassland understory. Historically, with restricted availability of satellite images we could not reliably separate the UK BAP <i>Bracken</i> class from <i>Acid Grassland</i> so we combined these into a single UKCEH Land Cover Class. With the greater image frequency and therefore better access to seasonal signals it may now be possible to overcome this historic limitation, but to do this we will need novel training data as we will not be able to retrieve a signal from Bootstrap Training.</p>
Heather; and Heather grassland	<p>For LCM2007 we refined the BAP <i>Dwarf Shrub and Heath</i> into two classes, depending on the density of Heather, producing the UKCEH Heather and Heather grassland classes (it is heather when there is greater than 25 % Heather Cover). This was to retain some consistency between the LCM1990 and LCM2000 classes Open Shrub Heath and Dense Shrub Heath. In some parts of the UK, significant areas of low-lying non-heather shrubs occur. For example, Gorse can form a dominant shrub layer.</p> <p>Note: the Land Cover Maps typically show confusion over Heather, Heather grassland and Bog. However, they are often difficult to separate in the field. It is</p>

	challenging to accurately estimate coverage above and below the defining threshold.
Fen, Marsh and Swamp	The UK BAP <i>Fen, Marsh and Swamp</i> includes fen, fen meadows, rush pasture, swamp, flushes and springs. From a remote sensing perspective Fen, Marsh and Swamp is problematic as it can be comprised of a wide range of vegetation types and many patches are below the MMU of the UKCEH Land Parcel Spatial Framework. The small size of many Fen, Marsh and Swamp patches, plus their typically mosaic nature make it difficult to find reliable training data. Consequently, Fen, Marsh and Swamp is likely to be underestimated in some regions. However, substantial areas of contiguous reed dominated Fenland appear to be well detected.
Bog	<p>The UK BAP <i>Bog</i> includes ericaceous, herbaceous and mossy swards in areas with a peat depth &gt;0.5 m. We cannot detect peat depth from satellites.</p> <p>Vegetation on deep peat soils represent a continuum involving acid grassland, dwarf shrub heath and some types of fen, marsh and swamp and the separation of continuously varying land cover into discrete types can be difficult, especially when they exist in a complex small patch mosaic and their definitions are vague.</p> <p>We retain the Bog class to maintain consistency with historical UKCEH LCM products and the RF classifier learns Bog presence based on training data automatically generated from these. The predicted distribution occurs in regions where it is expected, so is a good indicator of where Bog is likely to be occurring. However, Bog and the range of upland vegetation classes expected to occur on peaty soils (Acid grassland, Fen marsh and swamp, Heather, and Heather grassland) show a cluster of interclass confusion (Appendix 4). This is partly due fine-scale variation but largely an effect of ambiguous definitions. UK BAP Broad Habitats (on which UKCEH Land Cover Classes are based) were not defined with satellite remote sensing in mind.</p>
Saltwater	Saltwater is rarely different spectrally from freshwater, and the saltwater distribution predicted by the RF classifier is determined by coastal Context Rasters in Classification Scenes. There will be some confusion between Saltwater and Freshwater in tidal rivers, but not substantial. Occasionally, Saltwater is confused with non-vegetated surfaces close to the coast and this happens because the automatically generated Saltwater training classes coincide with the tide being out in the satellite view. The effect has so far been trivial but the result is that we predict Saltwater with slightly lower accuracy than Freshwater. Our main goal is to map land cover so coastal water and intertidal regions are not high priority.

Freshwater	<p>The UKCEH Freshwater class comes from merging two BAP BHs (<i>Standing Open Water and Canals</i>, and <i>Rivers and Streams</i>) since they cannot be separated by spectra. In many cases, small and/or narrow water bodies fall below the MMU of the UKCEH Land Parcel Spatial Framework so effectively disappear into the dominant surrounding vegetation. Where these features are appropriately aligned and sufficiently wide, pixels they may be detected and if so will be available in the Raster Classification datasets.</p> <p>Water bodies &gt;0.5 ha and wider than 40 m are mapped with very high accuracy. The exceptions are temporary water bodies and quarries. Water in some quarries is strongly affected by the minerals in the rock and can result in atypical colours and misclassification.</p>
Inland Rock	<p>The BAP Broad Habitat <i>Inland Rock</i> covers both natural and artificial exposed rock surfaces which are &gt;0.25 ha, such as inland cliffs, caves, screes and limestone pavements, as well as various forms of excavations and waste tips such as quarries and quarry waste. Opportunistic vegetation is common amongst rocky landscapes. We classify UKCEH Inland Rock if rock has the dominant signature.</p>
Urban; and Suburban	<p>Within the <i>Built-up Areas and Gardens</i> BAP Broad Habitat we can reliably separate two UKCEH categories: Urban and Suburban. Urban includes dense urban, such as town and city centres, where there is little, if any, vegetation. Urban also includes areas such as dock sides, car parks and industrial estates. It is sometimes confused with other non-vegetated surfaces; for example open cast quarries or more rarely coastal rocks or ploughed fields.</p> <p>Suburban includes suburban areas where the spectral signature is a mix of urban and vegetation signatures. Suburban and Urban lie on a continuum and confusion is expected.</p>
Supralittoral Rock	<p>Features that may be present in this coastal class include vertical rock, boulders, gullies, ledges and pools generally forming a narrow band when viewed from above. Only limited areas can be mapped using satellite remote sensing.</p>
Supralittoral Sediment	<p>This class includes sand dunes, which are reliably mapped. Areas of coastal sand may be confused between this class and the littoral sediment class. Supralittoral sediments can stabilise and from increasing volumes of vegetation. Heavily vegetated littoral sediment is likely to be classified as a vegetation class.</p>
Littoral Rock	<p>These classes are those in the maritime zone on a rocky coastline. They are generally more extensive than supralittoral rock and thus more readily detected using satellite images.</p>



Littoral Sediment; and Saltmarsh	<p>The BAP Broad Habitat <i>Littoral Sediment</i> has a subclass, the BAP Priority Habitat <i>Saltmarsh</i>. Saltmarsh is generally distinct from nearby vegetation and only occurs near the coast. As a consequence we can map this well with remote sensing. The Saltmarsh class is occasionally subject to commission error, when we mistake other vegetation in the coastal zone (mainly Arable) as Saltmarsh.</p> <p>The Littoral Sediment is sometimes confused with the Supralittoral Sediment class.</p>
-------------------------------------	--

## Appendix 2. Biodiversity Action Plan (BAP) Broad Habitats

This appendix provides a brief summary of the JNCC definitions of the Broad Habitats and is based on Jackson (2000). There is some duplication of material Appendix 2 with Appendix 1. However in combination they provide users with key information to understand UKCEH Land Cover Classes and the Broad Habitat definitions they are based on. Note: class numbers here are JNCC's Broad Habitat class numbers, not UKCEH Land Cover Class numbers.

The text in this Appendix has been copied from Morton *et. al.*, 2011, with very minor updates to remove obsolete cross-references to historical projects and documents, plus some corrections. Some new comments are included regarding BAP Broad Habitat *Linear Features* and the new 20 m Classified Pixel datasets.

### 1. Broadleaved, Mixed and Yew Woodland

This Broad Habitat is characterised by vegetation dominated by trees >5 m high when mature, with tree cover >20 %. Scrub (<5 m) requires cover >30 % for inclusion in this Broad Habitat. It includes stands of both native and non-native broadleaved trees and yew. Woodlands dominated by coniferous species but with >20 % cover by deciduous species are included in this category. Areas of fen woodland dominated by species such as willow (*Salix* spp.), alder (*Alnus glutinosa*) or birch (*Betula* spp.) are also included.

### 2. Coniferous Woodland

This Broad Habitat is characterised by vegetation dominated by trees >5 m high when mature, which forms a canopy having a cover of >20 %. 'Coniferous Woodland' includes semi-natural stands and plantations and includes both native and non-native coniferous trees.

### 3. Boundaries and Linear Features

This Broad Habitat type covers a range of linearly arranged landscape features such as hedgerows, lines of trees, walls, stone and earth banks, grass strips and dry ditches. These features are not included as a UKCEH Land Cover Class as they are generally too narrow to be reliably captured from the satellite images. However, linear features are of general interest and an important component of the landscape. Some linear features will be visible in the 10 m Classified Pixel dataset.

However, regardless of whether or not we can see linear features, we cannot classify them using our current methods. Linear features are in fact just thin fragments of land cover types that we already map. Membership in this class is defined by shape, not spectra. We do not have tools for classifying shapes. New tools for line-finding within the 10 m Classified Pixels products could help with their detection.

### 4. Arable and Horticulture

This Broad Habitat includes annual crops, perennial crops, woody crops, intensively managed commercial orchards, commercial horticultural land (such as nurseries, commercial vegetable plots and commercial flower growing areas), freshly-ploughed land, annual leys, rotational set-aside and fallow.

### 5. Improved Grassland

'Improved Grassland' is characterised by vegetation dominated by a few fast-growing grasses such as *Lolium* spp., and also white clover (*Trifolium repens*), on fertile, neutral soils. Improved grasslands are

typically either managed as pasture, or mown regularly for silage production or in non-agricultural contexts for recreation and amenity purposes.

## **6. Neutral Grassland**

This Broad Habitat type is characterised by vegetation dominated by grasses and herbs on a range of neutral soils usually with a pH of between 4.5 and 6.5. It includes enclosed dry hay meadows and pastures, together with a range of grasslands which are periodically inundated with water or permanently moist.

## **7. Calcareous Grassland**

*Calcareous Grassland* is characterised by vegetation dominated by grasses and herbs on shallow, well-drained soils which are rich in bases (principally calcium carbonate) formed by the weathering of chalk and other types of limestone or base-rich rock. Soil pH tends to be high (>6) although it may be as low as 5.

## **8. Acid Grassland**

*Acid Grassland* is characterised by vegetation dominated by grasses and herbs on a range of lime-deficient soils, which have been derived from acidic bedrock or from superficial deposits such as sands and gravels. Such soils usually have a low base status, with a pH of <5.5.

## **9. Bracken**

This Broad Habitat type covers areas dominated by a continuous canopy cover of bracken (*Pteridium aquilinum*) at the height of the growing season. It does not include areas with scattered patches of bracken or areas of bracken, which are less than 0.25 ha. These are included in the Broad Habitat type with which they are associated.

## **10. Dwarf Shrub Heath**

'*Dwarf Shrub Heath*' is characterised by vegetation that has >25 % cover of plant species from the heath family (ericoids) or dwarf gorse *Ulex minor*. It generally occurs on well-drained, nutrient-poor, acid soils. This habitat type does not include dwarf shrub dominated vegetation in which species characteristic of peat-forming vegetation such as cotton grass *Eriophorum* spp. and peat-building sphagna are abundant, or that occurs on deep peat (>0.5 m) as these are included in the 'Bog' Broad Habitat type.

## **11. Fen, Marsh and Swamp**

This habitat includes fen, flushes, springs, fen meadows, rush pasture and swamp. Fens are peatlands that receive water and nutrients from groundwater and surface runoff, as well as from rainfall. Flushes are associated with lateral water movement, and springs with localised upwelling of water. Marsh is a general term usually used to imply waterlogged soil; it is used more specifically here to refer to fen meadows and rush-pasture communities on mineral soils and shallow peats. Swamps are characterised by tall emergent vegetation. Reedbeds (i.e. swamps dominated by stands of common reed *Phragmites australis*) are also included in this type. Apart from rush pasture, examples of this Broad Habitat are relatively rare.

## **12. Bog**

This Broad Habitat type covers wetlands that support vegetation that is usually peat-forming and which receive mineral nutrients principally from precipitation rather than ground water. This is

referred to as ombrotrophic (rain-fed) mire. The Bog Broad Habitat includes ericaceous, herbaceous and mossy swards in areas with a peat depth >0.5 m.

### **13. Standing Open Water and Canals**

This Broad Habitat type includes natural systems such as lakes, meres and pools, as well as manmade waters such as reservoirs, canals, ponds and gravel pits.

### **14. Rivers and Streams**

The '*Rivers and Streams*' Broad Habitat type covers rivers and streams from bank top to bank top, or where there are no distinctive banks or banks are never overtopped, it includes the extent of the mean annual flood.

### **15. Montane**

The '*Montane Habitats*' category includes a range of vegetation types that occur exclusively in the montane zone such as prostrate dwarf shrub heath, snow-bed communities, sedge and rush heaths, and moss heaths. The distinction between the sub-montane and montane zone is often blurred and the two usually merge through a band of transitional vegetation.

### **16. Inland Rock**

This Broad Habitat type covers both natural and artificial exposed rock surfaces which are >0.25 ha, such as inland cliffs, caves, screes and limestone pavements, as well as various forms of excavations and waste tips such as quarries and quarry waste.

### **17. Built-Up Areas and Gardens**

This Broad Habitat type covers urban and rural settlements, farm buildings, caravan parks and other manmade built structures such as industrial estates, retail parks, waste and derelict ground, urban parkland and urban transport infrastructure. It also includes domestic gardens and allotments. This type does not include amenity grassland, which should be included in the '*Improved Grassland*' category.

### **18. Supralittoral Rock**

This habitat occurs above the high water mark, in areas influenced by wave splash and sea spray. Features that may be present include vertical rock, boulders, gullies, ledges and pools, depending on the wave exposure of the site and its geology.

### **19. Supralittoral Sediment**

This habitat occurs above the high water mark, but in areas influenced by wave splash and sea-spray. It includes shingle beaches, sand dunes and machair.

### **20. Littoral Rock**

The geology and wave exposure of the shore influence the form of Littoral Rock habitats, which can be as varied as vertical rock, shore platforms, boulder shores, or rocky reefs surrounded by areas of sediment. In general, *Littoral Rock* tends to be colonised by algae in wave-sheltered conditions, and by limpets, barnacles and mussels as wave-exposure increases.

## **21. Littoral Sediment**

Areas of *Littoral Sediment* are widespread around the UK forming features such as beaches, sand banks, and intertidal mudflats. A large proportion of this habitat occurs in estuaries and inlets where it can cover extensive areas. Saltmarsh is included within this Broad Habitat.

## **22. Inshore Sublittoral Sediment**

All areas of sea and estuary class are assumed to be *Inshore Sublittoral Sediment*. It is defined as within six nautical miles of the shoreline by JNCC.



### Appendix 3. Full list of datasets for UKCEH LCM2021

	Dataset	Land mass	Full Citation
<b>UKCEH LCM2021</b>	10 m Classified Raster Product	Great Britain	Land Cover Map 2021 (10 m classified pixels, GB)
	Land Parcel Product	Great Britain	Land Cover Map 2021 (land parcels, GB)
	25 m Rasterised Land Parcel Product	Great Britain	Land Cover Map 2021 (25 m rasterised land parcels, GB)
	10 m Classified Raster Product	Northern Ireland	Land Cover Map 2021 (10 m classified pixels, N. Ireland)
	Land Parcel Product	Northern Ireland	Land Cover Map 2021 (land parcels, N. Ireland)
	25 m Rasterised Land Parcel Product	Northern Ireland	Land Cover Map 2021 (25 m rasterised land parcels, N. Ireland)
	1 km Summary Rasters	Northern Ireland and Great Britain	Land Cover Map 2021 (1km summary rasters, GB and N. Ireland)

## Appendix 4. Correspondence matrices for UKCEH LCM2021

	Broadleaved woodland	Coniferous woodland	Arable	Improved grassland	Neutral grassland	Calcareous grassland	Acid grassland	Fen, Marsh, Swamp	Heather	Heather grassland	Bog	Inland Rock	Saltwater	Freshwater	Supra-littoral rock	Supra-littoral sediment	Littoral rock	Littoral sediment	Saltmarsh	Urban	Suburban	Total	User's Accuracy
Broadleaved woodland	1704	218	19	73	24	0	10	0	5	26	2	4	0	9	1	0	0	1	1	2	13	2112	80.7
Coniferous woodland	55	649	3	1	1	1	4	0	5	16	3	0	0	0	0	0	0	0	0	1	0	739	87.8
Arable	22	3	10102	306	28	1	2	0	0	3	1	10	0	6	0	0	0	0	1	13	13	10511	96.1
Improved grassland	100	4	1027	4835	186	55	175	18	1	82	29	4	0	8	0	26	1	0	4	9	24	6588	73.4
Neutral grassland	18	11	39	230	503	0	13	19	2	4	0	0	0	3	0	0	0	1	0	2	0	845	59.5
Calcareous grassland	30	4	26	47	4	946	15	0	1	6	0	13	0	0	0	0	0	0	0	1	0	1093	86.6
Acid grassland	20	1	88	177	6	55	1245	0	29	228	39	1	0	1	0	0	0	0	0	0	4	1894	65.7
Fen, Marsh, Swamp	15	0	5	14	2	1	4	577	0	0	2	0	0	8	0	0	0	0	0	0	0	628	91.9
Heather	7	1	9	1	0	0	30	1	819	104	121	0	0	0	1	0	0	0	0	0	0	1094	74.9
Heather grassland	17	3	9	12	0	5	158	3	81	299	106	4	0	3	0	1	0	0	0	0	1	702	42.6
Bog	0	3	0	3	0	3	31	4	27	71	877	0	0	1	0	0	0	0	0	0	0	1020	86.0
Inland Rock	0	0	10	2	1	9	4	0	3	2	2	125	0	0	0	0	0	0	0	8	0	166	75.3
Saltwater	0	0	0	0	0	0	0	0	0	0	0	0	73	0	0	0	0	13	0	0	0	86	84.9
Freshwater	13	0	1	4	1	0	1	2	0	0	0	0	0	548	0	0	0	0	0	0	2	572	95.8
Supra-littoral rock	0	0	0	5	0	0	7	0	0	0	0	0	0	0	42	6	11	6	0	0	0	77	54.5
Supra-littoral sediment	1	0	3	2	4	0	1	0	0	0	0	0	0	0	1	178	0	7	1	0	0	198	89.9
Littoral rock	0	0	0	0	0	0	0	0	0	1	0	1	0	0	16	5	86	17	0	1	0	127	67.7
Littoral sediment	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	11	7	211	3	0	0	241	87.6
Saltmarsh	2	0	0	5	1	0	0	13	0	1	0	0	0	2	1	8	1	9	169	0	0	212	79.7
Urban	19	0	12	29	15	0	1	1	0	0	0	26	0	3	0	7	1	2	1	2343	303	2763	84.8
Suburban	151	1	17	223	44	0	2	0	1	1	2	10	0	3	0	3	0	0	0	329	2727	3514	77.6
Total	2174	898	11370	5969	820	1076	1703	638	974	844	1184	198	82	595	62	245	107	267	180	2709	3087	35182	
Producer's Accuracy	78.4	72.3	88.8	81.0	61.3	87.9	73.1	90.4	84.1	35.4	74.1	63.1	89.0	92.1	67.7	72.7	80.4	79.0	93.9	86.5	88.3	Accuracy	82.6

## Appendix 5. Recommended colour recipe for displaying UKCEH Land Cover Classes

UKCEH Land Cover Class	UKCEH Land Cover Class number	Red	Green	Blue
<b>Broadleaved woodland</b>	<b>1</b>	255	0	0
Coniferous Woodland	2	0	102	0
Arable and Horticulture	3	115	38	0
Improved Grassland	4	0	255	0
Neutral Grassland	5	127	229	127
Calcareous Grassland	6	112	168	0
Acid grassland	7	153	129	0
Fen, Marsh and Swamp	8	255	255	0
Heather	9	128	26	128
Heather grassland	10	230	140	166
Bog	11	0	128	115
Inland Rock	12	210	210	255
Saltwater	13	0	0	128
Freshwater	14	0	0	255
Supralittoral Rock	15	204	179	0
Supralittoral Sediment	16	204	179	0
Littoral Rock	17	255	255	128
Littoral sediment	18	255	255	128
Saltmarsh	19	128	128	255
Urban	20	0	0	0
Suburban	21	128	128	128

A QGIS symbology file, lcm\_raster.qml, which contains the recommended colour recipe for displaying UKCEH land cover classes as given in Appendix 5 (above), is provided alongside the respective data product.

## Appendix 6. Recommended colour recipe for displaying UKCEH Land Cover Classes colour blind friendly

UKCEH Land Cover Class	UKCEH Land Cover Class number	Red	Green	Blue
Broadleaved woodland	1	51	160	44
Coniferous Woodland	2	0	80	0
Arable and Horticulture	3	240	228	66
Improved Grassland	4	1	255	124
Neutral Grassland	5	220	153	9
Calcareous Grassland	6	255	192	55
Acid grassland	7	178	145	0
Fen, Marsh and Swamp	8	253	123	238
Heather	9	128	26	128
Heather grassland	10	230	140	166
Bog	11	205	59	181
Inland Rock	12	210	210	255
Saltwater	13	0	0	92
Freshwater	14	0	0	255
Supralittoral Rock	15	152	125	183
Supralittoral Sediment	16	204	179	0
Littoral Rock	17	255	255	128
Littoral sediment	18	255	255	128
Saltmarsh	19	128	128	255
Urban	20	0	0	0
Suburban	21	128	128	128

A QGIS symbology file, `lcm_raster_cb_friendly.qml`, which contains the recommended colour blind-friendly colour recipe for displaying UKCEH land cover classes as given in Appendix 6 (above), is provided alongside the respective data product.