

UKCEH Land Cover Plus: Hedgerows (2016-2021) England Dataset documentation

Richard K. Broughton, Rich Burkmar, Morag McCracken, Nadine Mitschunas, Lisa R. Norton, Denise W. Pallett, Justine Patton, John W. Redhead, Jo T. Staley, Claire M. Wood, Richard F. Pywell
UK Centre for Ecology & Hydrology
v.1.0; 11/12/2023

Contents	
Summary	1
Background	1
Product specification	2
Methodology	2
Validation	4
Model Limitations	5
Data access	6
Example output	6
Acknowledgements and funding	7
References	7

Summary

The dataset comprises a model of the extent and height class of woody linear features on field boundaries in England, including hedgerows, tree lines and semi-natural thickets of shrubs and trees. The model was derived from processing of the Environment Agency lidar product (National Lidar Programme), captured in 2016-2021. Due to high levels of uncertainty and predicted low density of woody linear features, areas excluded from the coverage include mountain/moor/heath, open water and the coastal zone, and urban/suburban areas. Woodland (i.e. areas of continuous non-linear woody cover) is also excluded. The model was validated by 'ground-truthing' against 38 test squares of 1 km² each, which were surveyed on the ground by trained fieldworkers in 2022 using Countryside Survey methods. The results showed good agreement between the lidar model and the ground-truthing, allowing for a spatial buffer tolerance around features to account for the different linear frameworks they were based upon. At the summary scale across all test squares, the total woody feature lengths showed a 96% agreement. At the scale of individual test squares, there was an average 76% agreement of woody feature lengths within a 20 m buffer tolerance. Height classifications of the model showed an exact agreement for 32% of the total feature lengths, and 63% agreement within a tolerance of one class above or below.

Background

Hedgerows and other linear woody features along field boundaries, such as tree lines and semi-natural thickets, are important habitats and cultural features in the UK's rural landscapes. Mapping and quantifying hedgerows has long been problematic due to the spatial scale involved in recording their three-dimensional structure and extent. Earlier methods of mapping have only achieved local coverage, as ground-based fieldwork methods are prohibitively labour-intensive on a large scale, while aerial photography can achieve greater coverage but lacks the all-important three-dimensional attribute of shrub and tree heights. The use of more advanced remote sensing technology, such as lidar and Interferometric Synthetic Aperture Radar, has allowed advances in the scale and detail of hedgerow mapping, but these were still limited by the spatial extent or resolution of data acquisition (Graham et al. 2019). Lidar is perhaps the most promising technique for mapping woody linear features, like hedgerows, as it can achieve a large spatial coverage (hundreds or thousands of km²) and full three-dimensionality (height measurements to cm scale). The availability of the output from the Environment Agency's National Lidar Programme (2016-2021) provided relatively high resolution (1 m) and near total coverage for England. We used

this opportunity to scale-up a methodology that we had developed in previous work that mapped woody linear features on a regional basis (e.g. Broughton et al. 2017), to now encompass all of England. The output is a national-scale model of linear woody field boundaries for rural England that depicts length and height (summarised into height class) and integrates spatially with other UKCEH Land Cover Map (LCM) products, and is also compatible with height classes used in the Countryside Survey. The output is therefore a representation of the presence and height of woody linear features associated with the boundaries of Land Cover Map polygons rather than a fully georeferenced map of the position of each hedgerow on the ground, although there is generally close agreement between the two.

Product specification

- UKCEH Land Cover Plus: Hedges (2016-2021) England is provided in the geopackage format as 5,881 individual tiles of 10 km resolution, totalling 9.51 Gb.
- The tiles are aligned to the British National Grid and the Ordnance Survey (OS) grid, and use the OS grid naming system.
- The data are classified polylines derived from airborne lidar collected by the Environment Agency's National Lidar Programme during 2016-2021, which are aligned to the UKCEH Land Cover linear framework of parcel boundaries.
- Polyline data are binned into 7 height classes (minimum 0.5 m high) for segments of a minimum 2 m long.
- Polyline segments are also attributed with length (in metres) and an interpretation of feature type (e.g. double hedge, single hedge).
- Values are based on underlying lidar data with a plus/minus 15 cm reported vertical accuracy, 1 cm vertical resolution and 1 m horizontal resolution.
- The product extent is England, but features are excluded from land covers of woodland, urban/suburban, coastal, mountain/moor/heath.

Methodology

Processing environment and source data

Processing was carried out in R on the LOTUS batch computing cluster, which is part of the JASMIN data analysis facility for environmental science (<https://help.jasmin.ac.uk/article/5004-lotus-overview>).

A 1 m x-y resolution (1 cm vertical z-resolution) Canopy Height Model (CHM) for England was supplied by the Environment Agency, pre-processed by the EA by subtracting the Digital Terrain Model (DTM) from the first-return Digital Surface Model (DSM), which were captured during the 2016-2021 National LiDAR Programme. Data were collected during leaf-off conditions (autumn/winter) and supplied as raster 5 km tiles.

Other datasets used in the processing (primarily for masking) included the UKCEH Land Cover Map 2021 (land parcels, GB) (<https://doi.org/10.5285/398dd41e-3c08-47f5-811f-da990007643f>), and Ordnance Survey (OS) open data products: DistrictMap Vector, OS OpenMap and OS Open Road.

Output data were generated on a 5 by 5 km tile basis and later combined into 10 by 10 km tiles.

Data processing

The LCM polygon boundaries were used as the spatial linear framework (polylines) on which the final product is generated. Essentially, the field boundaries were defined by the LCM data, and if woody features were present then these were populated from the lidar CHM data values.

For the CHM raster, height values under 0.5 m were filtered and removed. The remaining CHM raster values were reclassified into height classes which are compatible with those used in the Countryside Survey:

0.50-0.99 m = height class 1a
1.00-1.49 m = height class 1b
1.50-1.99 m = height class 1c
2.00-2.99 m = height class 2
3.00-3.99 m = height class 3
4.00-5.99 m = height class 4
6.00-maximum = height class 6

Modal filtering (11 pixel width) was applied to smooth the data, producing raster clumps of contiguous pixels with the same height class. Pixels belonging to a clump of ≤ 20 pixels and less than 5 m from a pixel in a clump of size 20 pixels or more were reclassified to that of the pixel in the larger clump. This further smoothed the raster data. The CHM raster was then converted to vector polygons.

Filtering field boundaries

The linear framework derived from the LCM polygons were filtered to exclude those field boundaries that abutted an excluded land cover class, which were: broadleaf woodland, coniferous woodland, mountain/moor/heath, saltwater, freshwater, coastal, urban and suburban areas. The reason for this is that each of these land classes include many CHM features that could be misinterpreted as hedgerows, such as walls and buildings, but would contain few actual hedgerows or other distinct woody linear features. As such, excluding these classes from analyses would remove a substantial number of type I errors, while sacrificing few genuine woody field boundaries. The resulting coverage essentially included all farmed and managed open land. A buffer of 50 m was also applied within the high water mark, to exclude coastal topography such as rocks and cliffs. Also masked out of analyses were individual buildings not already included in the LCM classes, such as farmsteads, whose footprint was derived from the OS data; a 5 m buffer was applied to capture garden walls, fences, sheds and associated infrastructure.

The vectorised CHM was overlaid with the field boundary polylines, and field boundaries were retained if they coincided at least 0.9 of their length within 20 m of CHM features.

Attributing woody feature types

A filter was run over a mask created from the raster CHM (\Rightarrow 0.5 m height values) which counted the number of pixels occurring within a radius of 10 m of each pixel (filter10).

A similar filter was run to count the number of pixels within 20 m of each pixel (filter20).

These two rasters are divided (filter20/filter10) to give a ratio raster, then a rule applied whereby ratio values of 2.5-4 were considered indicative of double features, i.e. parallel hedgerows or tree lines on either side of a lane or road. Values between 2.5 and 4 were given an attribute value of 2 and all other values 1.

The field boundary vector was buffered to 15 m and each resulting buffer polygon was attributed as being a single or double woody feature based on the modal value (either 2 or 1) of the overlaid ratio raster pixels in each polygon. An attribute label of 'isDouble' was transferred back to the field boundary vector line.

The OS vector data of the road network (OS Open Road) was buffered to 20 m, and if any polyline from the field boundary layer had at least 0.75 of its length within the road buffer then it was attributed to indicate a roadside hedge by applying the attribute label 'isRoadside'.

Final steps

The vectorised CHM polygon layer was intersected with the field boundary polylines within a tolerance buffer of 20 m, and where these intersected within the buffer then the resulting field boundary line segments were attributed with the height class from the vectorised CHM polygon. At this point, each field boundary line could be split into several line segments with different height class values. The field boundary lines were then clipped to the extent of the EA 5 km tiles, and then merged into 10 km tiles.

All 5 km field boundary tiles within a single 10 km tile (anything between 1 and 4 tiles) are combined into a single 10 km field boundary tile.

Any isolated sections under 2 m in length were deleted as extraneous. Other sections under 2 m in length that were connected to others were merged with an adjacent larger segment and given its attributes. Touching segments that had the same attributes of height class, isDouble and isRoadside were combined into a single feature. A 'label' attribute was added to aid interpretation for users, based on the value of the isDouble and isRoadside attributes. Where isDouble and isRoadside is true then the label is 'Double hedge' (i.e. probably two parallel hedgerows either side of a road); where isDouble is true & isRoadside is false then the label is 'Probable wide single hedge' (e.g. a treeline or wide hedgerow); where isDouble is false & isRoadside is true then the label is 'Single hedge'; where isDouble & isRoadside are false then the label is 'Single hedge'. The full list of attributes is given in Table 1. The output was the UKCEH Land Cover Plus: Hedgerows (2016-2021) England model (LCPHE).

Table 1. Interpretation of the attributes of the model output.

Attribute	Meaning
section	Unique number of each segment
hgghtcls	Height class (see above)
isDb	'isDouble': interpreted as double (parallel) features
isRd	'isRoad': interpreted as a feature lying alongside a road
hlength	Length (metres) of the feature segment
label	Interpreted feature type, derived from Boolean terms of isDb and isRd attributes

Validation

The output was validated using 38 test squares of 1 km², distributed across England, which were surveyed by trained fieldworkers in 2022 according to Countryside Survey (CS) standard methods (Maskell et al. 2008). The surveys included mapping and classification of woody linear features (WLF) into height classes, based on features at least 20 m long and a maximum of 5 m wide. Woodland strips wider than 5 m were also recorded, and these were extracted from the CS data alongside the WLFs. The spatial data framework used in mapping during CS surveys was ultimately based on OS base data, which is also the root of the LCM framework underlying the LCPHE model, although in both frameworks the surveyors or processing could reshape features, allowing some divergence in the placing of boundary lines.

The LCPHE model was compared with CS WLFs in several ways. Most simply, the total length of 1436 linear features derived from the field survey in 38 CS test squares was compared with 6926 linear features from the LCPHE model in the same squares. These were compared on a like-for-like basis, excluding areas within the masked land cover areas of woodland, urban, mountain/moor/heath etc. The results of this comparison found a 96% agreement in summary total feature lengths between the two coverages, meaning that the total length of features from the LCPHE model was within 96% of that from the CS field survey across all 38 squares.

Secondly, a 20 m buffer was applied around CS WLF and woodland strips. The linear features in the LCPHE model were clipped within this buffer, and the total lengths of all LCPHE features within these buffers were compared with the CS WLF from which they were derived. As such, this comparison was a more detailed test of whether individual linear features within the CS data were replicated by the LCPHE model, allowing for a 20 m tolerance along the feature alignment. This analysis was performed individually for each of the 38 CS squares. The results showed an overall mean of 76% (15% SD) agreement

between feature lengths per square, with mean agreement ranging over 27-98% for individual CS squares.

Finally, height class of the linear woody features was compared by taking the midpoint of each LCPHE linear line segment and comparing its height class to the nearest CS WLF height class, within the 20 m buffer. Because the LCPHE model had a higher precision than the CS WLF, with a greater degree of segmentation, the comparison was typically comparing multiple segment midpoints against the single linear feature that they were aligned alongside. The percentage of the total length of 6926 LCPHE segments that tallied with the adjacent CS WLF height class was then calculated. The exact agreement between adjacent features with the same height class was 32% across all 38 CS squares, meaning that just under a third of the length of LCPHE features had the same height class as the CS WLF that they were adjacent to. However, 63% of the LCPHE feature length was within a single height class of the CS WLF, which was probably more realistic in allowing for variation in height class due to cutting or growth between the timing of the field survey and lidar data capture, which could be a disparity of up to 6 years. Further disparity or inaccuracy of the ground survey could be due to surveyor error or generalization (low precision/accuracy) that could introduce an unknown degree of error in feature location, length or height, as the ground survey was unlikely to provide a definitively accurate dataset.

Model limitations

The LCPHE model is a snapshot of the time of lidar data capture, which for a location can fall between 2016 and 2021. Hedgerows and other woody linear features may have changed in the interim, due to growth, management, toppling or removal, and new features may have appeared through planting or growth. The model gave a reasonably good agreement with field surveys in the areas analysed, but these excluded moorland, heathland and urbanized regions, and also woodland, although hedgerows and other linear woody boundary features are unlikely to be common in such environments. An area of approximately 24 x 25 km on the North Yorkshire Moors was also excluded from the model, as this area was not mapped by the EA lidar.

The LCPHE model could not distinguish well between woody linear features and other solid field boundaries, such as drystone walls, and this is a source of error, albeit one that was mitigated by excluding moorland and mountain land covers. Similarly, field boundaries composed of tall non-woody vegetation, such as nettles, reeds and bracken, could not be distinguished from hedgerows or shrubs, and this is another source of potential error. Most of these misclassified features will be captured in the lower height classes of 1a and 1b (below 1.5 m tall), and these can be filtered for removal if required. Misclassification of taller features potentially includes bridges, solar panels and other large infrastructure, although these are less likely to be coincident with polygon boundaries. Manual data cleaning to correct such errors was not attempted, and users are recommended to inspect the model data for their area of interest to remove known error features if possible or appropriate. Woody features not associated with field boundaries in the LCM linear framework were not considered, as the vast majority of woody features outside of woodlands were considered likely to occur along field boundaries. However, the alignment and location of the LCM framework, and the woody features derived from it in the LCPHE model, may differ from those on the ground or in other datasets, due to generalization of the polygon framework, and digitizing or positioning errors in the model or other spatial data. Due to the aforementioned potential for confusion with other features, and to increase interoperability with the LCM and other UKCEH data products, the aim of the LCPHE model is to produce, as accurately as possible, a representation of the presence and height of woody linear features associated with the LCM linear framework rather than a georeferenced map of the position of each hedgerow with high absolute accuracy. For example, the midline of a hedgerow is likely to fall to one side of a mapped field boundary, and not exactly on top of it, which instead may sit along the midline of a ditch at the field edge. Similarly, a hedgerow or tree canopy will occupy a substantial width along its length, and the mapped polyline of the linear feature may not fall along its midline. Further, a linear position may be correctly

mapped in the model, but this may not align with other spatial datasets due to georeferencing or digitizing error and variation that occur in all spatial data. As such, a spatial tolerance of some metres, and possibly a few tens of metres, should be expected between the location of modeled features and their position in other georeferenced datasets. The cumulative limitations mean that the model is most appropriate for larger scale summaries (e.g. national, regional or farm-estate scale) of total hedgerow lengths per height class, rather than at the smaller scale of individual fields and features, where errors along single features will be more obvious and significant. However, as with many such data products, at the smaller scale the model may also offer a valuable baseline for enhancement and amendment with a field survey that reduces the need for field surveyors to create hedgerow maps entirely from scratch.

Data access

Bespoke licensing conditions apply to these data. If you choose to download the data, the UKCEH Data Licensing team will contact you to negotiate a licence. Licensing will generally be free for academic and research purposes.

Example output

Figure 1 shows a sample area of the polyline model output, and Figure 2 shows an example of the tabulated attribute data attached to the polyline features.

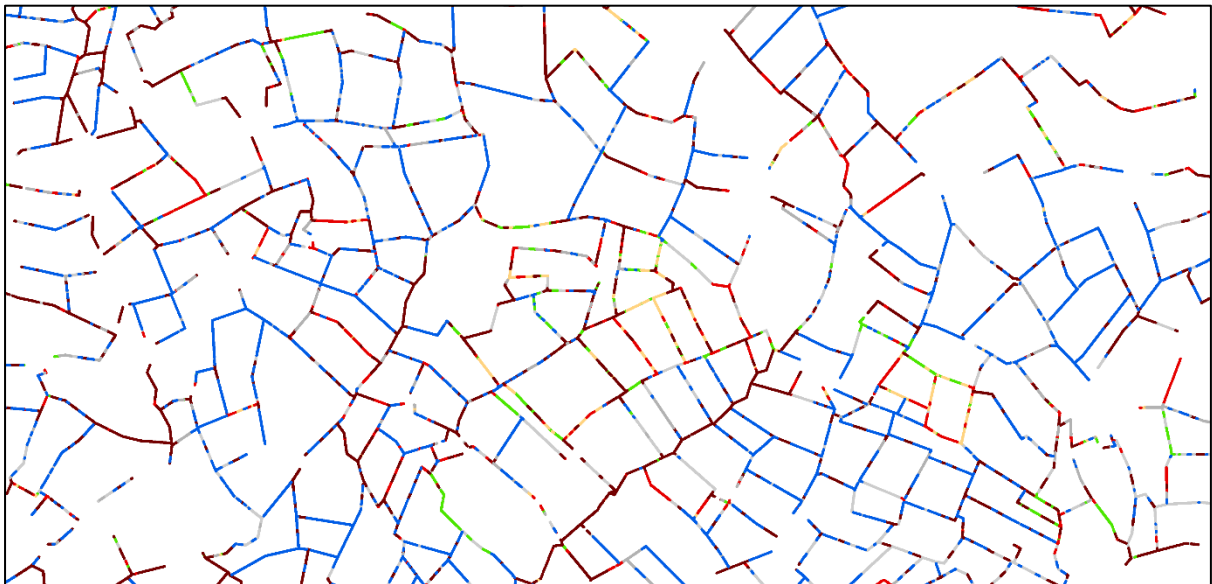


Figure 1. A 4 x 2 km sample area of the model output, showing polyline segments of woody boundaries of land parcels or fields, where the boundary features are colour-coded by height class (blue or green for shorter features and red for taller hedges and tree-lines).

FID	Shape *	section	hghtcls	isDb	isRd	hlength	label
11589	Polyline	85	1b	1	1	56.799736	Double hedge
11590	Polyline	86	1b	1	1	18.345127	Double hedge
11591	Polyline	87	1b	1	1	12.082161	Double hedge
11592	Polyline	88	1b	1	1	97.528616	Double hedge
11593	Polyline	89	1b	1	1	101.214233	Double hedge
11594	Polyline	9	1b	1	1	20.055224	Double hedge
11595	Polyline	90	1b	1	1	10.08757	Double hedge
11596	Polyline	91	1b	1	1	34.225675	Double hedge
11597	Polyline	92	1b	1	1	33.48043	Double hedge
11598	Polyline	93	1b	1	1	7.720321	Double hedge
11599	Polyline	94	1b	1	1	26.614812	Double hedge
11600	Polyline	95	1b	1	1	4.354241	Double hedge
11601	Polyline	96	1b	1	1	233.224178	Double hedge
11602	Polyline	97	1b	1	1	38.000866	Double hedge
11603	Polyline	98	1b	1	1	6.000137	Double hedge
11604	Polyline	99	1b	1	1	22.000995	Double hedge
11605	Polyline	1	1b	1	0	7.269761	Probable wide single hedge
11606	Polyline	10	1b	1	0	13.790113	Probable wide single hedge
11607	Polyline	100	1b	1	0	34.932477	Probable wide single hedge
11608	Polyline	101	1b	1	0	33.879578	Probable wide single hedge
11609	Polyline	102	1b	1	0	11.032557	Probable wide single hedge
11610	Polyline	103	1b	1	0	12.319846	Probable wide single hedge
11611	Polyline	104	1b	1	0	18.47977	Probable wide single hedge
11612	Polyline	105	1b	1	0	14.466509	Probable wide single hedge
11613	Polyline	106	1b	1	0	21.906171	Probable wide single hedge

Figure 2. A sample attribute table from the model output, where each record represents a classified line feature (see Table 1).

Acknowledgements and funding

This research was funded by the Natural Environment Research Council (NERC) under research programmes NE/N018125/1 ASSIST – Achieving Sustainable Agricultural Systems www.assist.ceh.ac.uk and NE/W005050/1 AgZero+ : Towards sustainable, climate-neutral farming.

ASSIST and AgZero+ are initiatives jointly supported by NERC and the Biotechnology and Biological Sciences Research Council (BBSRC). Claire Carvell (UKCEH) contributed to project management.

This work used JASMIN, the UK collaborative data analysis facility.

Hedgerow data collection was carried out under the Natural England supported project ITT9718 ‘An evaluation of Agri-Environment Scheme impact on hedgerows in England’.

Significant support was provided by Alastair Duncan and the Geomatics team at the Environment Agency. Contains Ordnance Survey data © Crown copyright and database right 2023.

References

Broughton, R.K., Gerard, F., Haslam, R. & Howard, A.S. (2017). Woody habitat corridor data in South West England. NERC Environmental Information Data Centre. (Dataset). <https://doi.org/10.5285/4b5680d9-fdbc-40c0-96a1-4c022185303f>

Graham, L., Broughton, R.K., Gerard, F. & Gaulton, R., 2019. Remote sensing applications for hedgerows. *The Ecology of Hedgerows and Field Margins*, pp.72-89.

Maskell, L.C., Norton, L.R., Smart, S.M., Carey, P.D., Murphy, J., Chamberlain, P.M., Wood, C.M., Bunce, R.G.H. & Barr, C.J. (2008) CS Technical Report No.1/07 Field Mapping Handbook. Centre for Ecology and Hydrology (Natural Environment Research Council).