NCEA Natural Capital and Ecosystem Assessment

# Living England 2022-23: Technical User Guide

## A satellite derived national classification of England’s broad habitats

September 2024

Natural England Research Report NERR141



NATURAL ENGLAND

## About Natural England

Natural England is here to secure a healthy natural environment for people to enjoy, where wildlife is protected and England's traditional landscapes are safeguarded for future generations.

## Further Information

This report can be downloaded from the Natural England Access to Evidence Catalogue. For information on Natural England publications or if you require an alternative format, please contact the Natural England Enquiry Service on 0300 060 3900 or email enquiries@naturalengland.org.uk.

## Copyright



Catalogue code: NERR141

# Report details

## Author(s)

Becky Trippier, Anne Stefaniak, Max Fancourt, Miles Clement, Chris Moore, Amy Woodget, Sophie Potter, Katherine Holmes, Bethany Hadfield, Alison Saunders, Rebecca Mein, Jade Bowling & Alexandra Kilcoyne.

## Project Managers

Anne Stefaniak and Becky Trippier

## Keywords

Satellite imagery, remote sensing, habitat mapping, object-based image classification, machine learning, natural capital.

## Acknowledgements

This project is funded by the UK Government through Defra's Natural Capital and Ecosystem Assessment programme and the Environmental Land Management schemes (ELMs).

We would like to thank the Natural England field surveyors and data providers for sharing datasets with us, and for their efforts in the field. This project would not be possible without the landowners who have kindly given permission for the project to access and survey their land, and we are extremely grateful for their continued contribution to the project. We would also like to thank all contractors and subcontractors for their continued field survey work alongside our field team collecting data for the project.

For further development of the methodologies, thank you to partners at JNCC, EA, and contractors from JBA Consulting Ltd and Terrabotics Ltd. Finally, we would like to thank and acknowledge contributions from colleagues across Natural England, including a special thanks to Scott Mitchell, Heather Thomas, Rachel Wade, Alison Goncalves, Marcelo Abreu, Lauren Attfield, Nick Tomline, Paul Ross, Nick Biddiscombe and the England Ecosystem Survey and England Peat Map Teams.

## Citation

TRIPPIER, B., STEFANIAK, A., FANCOURT, M., CLEMENT, M., MOORE, C., WOODGET, A., POTTER, S., HOLMES, K., HADFIELD, B., SAUNDERS, A., MEIN, R, BOWLING, J., & KILCOYNE, A. 2024. Living England 2022-23: Technical User Guide. A satellite derived national classification of England's broad habitats. [NERR141]. Natural England.

# Executive summary

Living England 2022-23 is a satellite-derived national habitat map showing the extent and distribution of England's diverse habitats. This dataset provides an assessment of the likely broad habitats present across England aligned to the UK Biodiversity Action Plan (UKBAP) habitat classification framework. This is the first publication using a standardised workflow, which is intended for use in future years enabling future assessment of habitat change. Future iterations of Living England will be published every two years informed by up-to-date datasets and updated habitat ground data.

The Living England 2022-23 dataset has been developed using satellite imagery, environmental datasets and machine learning techniques, informed by a programme of national-scale field data collection, designed to record habitats based on their appearance from above to align with satellite views. Living England uses a bespoke spatial framework derived from seasonal imagery mosaics using image segmentation and an object-based image classification approach to map habitats across England. Habitat data was collated through targeted field surveys as well as additional available ground data records to train a machine learning model, which assess the relationships of broad habitats against a range of environmental predictors including; Sentinel-1 (S1) SAR and Sentinel-2 (S2) Optical data, LiDAR (Light Detection and Ranging) topographic data, computed S1 coherence, backscatter and LiDAR texture indices, proximity distances to OS features, climatic data, and soil data. Where existing good quality data is already available for specific habitats (e.g. Rural Payments Agency (RPA)'s Crop Map of England (CROME) detailing arable habitats), these habitats were assigned directly from these datasets (specifically for urban and arable habitats). Where bespoke methods are more suitable (e.g. areas of bare ground and saltmarsh), these have been developed specifically for Living England. The habitat data layers were then combined with the modelled outputs to produce the final map. Alongside the predicted habitats, a novel reliability metric indicating the confidence in the habitat assignment for each segment is provided based on the underlying data sources and data validation.

Living England 2022-23 provides national coverage of broad habitats and now includes a mapped spatial extent of Solar Farms (0.1% of England's land cover). Living England 2022-23 reports an overall accuracy for modelled habitats of 87%, with a class-wise balanced accuracy of 90%. This varies between classes and regions across England. The reliability metric indicates 90% of England's area is mapped with a score of very high to medium reliability.

This technical user guide summarises the data and methodologies used to create the Living England 2022-23 product and highlights how data attributes can be used to inform use of the data. This guide accompanies the data release under an Open Government Licence (OGL).

# Contents

# 1. Introduction

## 1.1 Purpose and Scope

This Technical User Guide defines the product specification for the Living England 2022-23 national habitat map. It details the updated methodology used to derive the 2022-23 dataset, which will be applied for future iterations and replaces the methodology used for previous versions. The key updates to the methodology include new additional attributes such as the reliability metric, the removal of unclassified segments resulting in full national coverage with no unclassified segments, improved urban and water mapping, the inclusion of solar farms as a separate class, improvements to the model methodology resulting in improved accuracy especially for rarer habitats, and the integration of LiDAR data and summer mosaics. This guide also highlights key use cases and limitations to help users make informed decisions when applying this dataset to their own work. It provides a brief overview of the data inputs, modelling approach, output map, data formats and licensing information.

## 1.2 The Living England Project

High quality national-scale evidence is vital in assessing the state of our natural capital assets. National-scale habitat ground monitoring surveys are expensive, and data collection can be restricted due to accessibility and spatial coverage. Thus, there is a requirement for open, freely available data at scale to inform habitat assessments delivered in a reproducible, robust manner to inform policy and land management decisions. With policies such as the Environmental Land Management schemes (ELMs) and the England Tree Action Plan (ETAP) set to implement changes on our landscapes across England, it is more important than ever to have updatable, representative and reliable evidence of habitat extent and distribution over time.

The Living England project started in 2016 and is a multi-year project which produces a national habitat map of England using satellite imagery, field collected data and machine learning (Kilcoyne et al., 2017, 2022). The project is funded by the Department for the Environment, Farming and Rural Affairs (Defra) through the Natural Capital and Ecosystem Assessment (NCEA) programme, and Environmental Land Management schemes (ELMs). Living England is an integral part of the NCEA, delivering a nationwide survey of England's land, coast, and seas: mapping the location, extent and condition of our ecosystems and the benefits they provide. Through comprehensive monitoring, innovative measurement and the development of tools and guidance, the programme is providing insights on how and why our environment is changing and the impact of this, so that it can be better protected and managed.

The Living England project brings together expertise from ecology, Earth observation and data science to integrate knowledge of how habitats and dominant species appear on the ground. The project generates national-scale analysis-ready data, exploiting open-source European Space Agency (ESA) Sentinel imagery, used to produce Living England. Alongside in-house analysis and modelling, the project also runs national-scale ground data collection, trained and supported by field surveyors using bespoke tools for data collection. This will provide robust, reliable ground data throughout the longevity of the project for training and validating outputs.

The project plans to use the methodology set out in this guide in future years to produce updated iterations Living England every two years. These will be informed using the latest ground data collected and up-to-date satellite imagery. The project will continue to develop and optimise the approach to improve identification of habitats using Earth observation data, whilst limiting significant changes to the methodology.

## 1.3 Living England 2022-23

Living England 2022-23 (Figure 1) describes the extent and distribution of England's broad habitat types using input data from 2022-23. The resulting map describes England's broad habitats detected using 10 m Sentinel 2 (S2) satellite imagery. These categories are predominantly aligned with the UK Biodiversity Action Plan (UKBAP) broad habitats (JNCC, 2011). UKBAP is a terrestrial habitat classification system commonly used across England. The data represents predicted habitats from 2022-23 providing an assessment of natural capital assets.

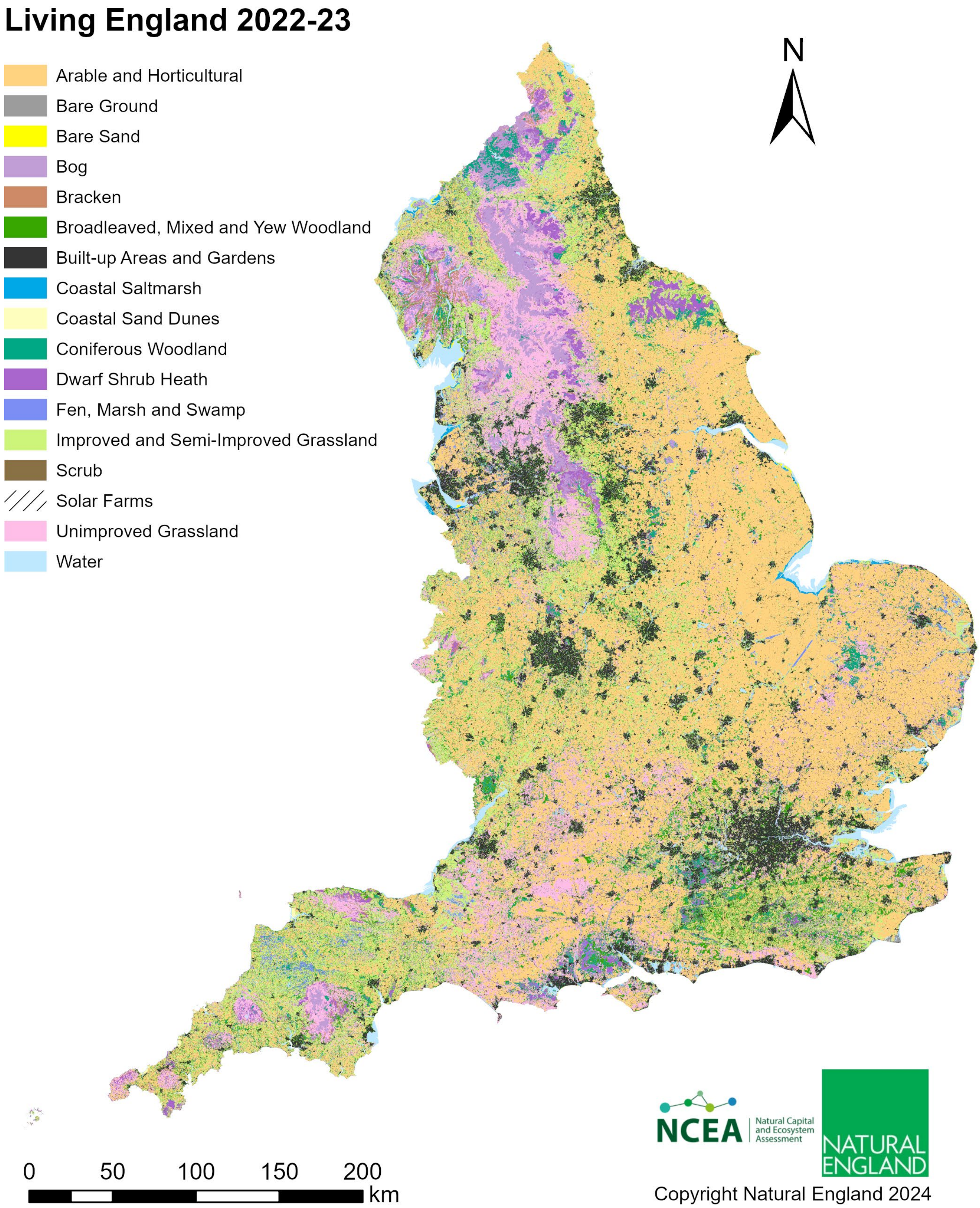


**Figure 1: Living England Habitat Map 2022-23**

Living England 2022-23 relates to imagery captured from Autumn 2022 and Spring and Summer 2023. Living England is projected using British National Grid (BNG) coordinate reference system (ESPG:27700) and covers England's terrestrial habitats up to Mean High-Water Springs (MHWS), as defined by the Ordnance Survey (Ordnance Survey, 2023a). The boundary of the whole Living England dataset extends up to the OS England boundary line (2023a) in order to capture the full extent of the coastal saltmarsh habitat, which extends beyond MHWS to the edge of the intertidal zone.

### 1.2.1 Data Format

Living England 2022-23 is freely available under an Open Government Licence (OGL).

The Living England dataset is provided as a geopackage representing spatial polygons, using a bespoke spatial framework, described in Section 3.3 Image Segmentation. The polygons are created using S2 imagery with a spatial resolution of 10 m, and a Minimum Mapping Unit (MMU) of 3 pixels (300 $m^2$).

The data is available to view on:

- Defra's MAGIC Platform (MAGIC, 2024)
- Natural England Open data portal (Natural England, 2024)
- Defra Data Services Platform (Defra, 2024)

The data are available to download via data.gov.uk as a geopackage (.gpkg) GIS file. Due to the size of the dataset, splitting of larger geographical areas may be required.

The dataset is also available as a 10 m raster layer, please contact livingenglandenquiries@naturalengland.org.uk for access to this data.

### 1.2.2 Data Attributes

For each segment within the Living England habitat map, a series of data attributes are provided. Table 1 shows the name, data type and a description of the attributes associated with each segment.

**Table 1: Descriptions of the data attributes in the Living England Habitat map 2022-23**

| Attribute short name | Attribute full name | Description | Data Type |
|---|---|---|---|
| **SegID** | **SegID** | Unique Living England segment identifier. Format is **LEZZZZ_BGZXX_YYYYYYY** where Z = release year (2223 for this version), X = BGZ and Y = Unique 7-digit number | String |
| **Prmry_H** | **Primary_Habitat** | Primary Living England Habitat | String |
| **Relblty** | **Reliability** | Reliability Metric Score | String |

| Attribute short name | Attribute full name | Description | Data Type |
|---|---|---|---|
| **Mdl_Hbs** | **Model_Habs** | List of likely habitats output by the Random Forest model. | List of Strings |
| **Mdl_Prb** | **Model_Probs** | List of probabilities for habitats listed in 'Model_Habs', calculated by the Random Forest model. | List of Integers |
| **Mixd_Sg** | **Mixed_Segment** | Indication of the likelihood a segment contains a mixture of dominant habitats. Either **Unlikely** or **Probable**. | Boolean |
| **Source** | **Source** | Description of how the habitat classification was derived. Options are:<br>**Random Forest**<br>**Vector OSMM Urban**<br>**Vector Classified OS Water**<br>**Vector EA saltmarsh, LE saltmarsh & QA**<br>**Vector RPA Crome, ALC grades 1-4**<br>**Vector LE Bare Ground Analysis**<br>**LE QA Adjusted** | String |
| **SorcRsn** | **Source_Reason** | Reasoning for habitat class adjustment if 'Source' equals 'LE QA Adjusted' | String |
| **Shap_Ar** | **Shape_Area** | Segment area ($m^2$) | Integer |

# 2. Data Sources

To create the Living England 2022-23 habitat map, data are collected, representing the habitat types and locations on the ground, during extensive field surveys. This is used alongside Earth observation data and a range of environmental data layers to indicate physical, chemical, and biological properties of the range of broad habitats across England.

Data sources include:

**(a) Ground data points**

- Living England national field survey data.
- Natural England active field data campaigns: including the Long-Term Monitoring Network, England Peat Map and Sites of Special Scientific Interest (SSSI) Protected Sites Monitoring.
- Other habitat surveys shared with the project, including from the National Plant Monitoring Scheme and Wetland annex 1 surveys.

**(b) Geospatial data**

- European Space Agency (ESA) Sentinel imagery: Sentinel-1 (S1) radar imagery & Sentinel-2 (S2) optical imagery.
- Topographic data: Environment Agencies (EA's) National LiDAR Programme collected between 2017 – 2022.
- Climatic data: Met Office's HadUK gridded climate product.
- Geology & soils datasets: BGS geology and soil parent material, Cranfield NATMAP soilscapes.
- Specific feature and habitat datasets: Ordnance Survey Mastermap (OSMM) Topographic layer, EA's Saltmarsh Extent and Zonation, Rural Payment Agencies (RPA) Crop Map of England (CROME).

A comprehensive list of all datasets used to create Living England 2022-23 and how they have been used can be found in Appendix 2.

# 3. Methodology

An overview of the processing steps used to create the Living England (LE) habitat map 2022-23 are outlined in Figure 2. Further details for each step are described throughout Section 3.

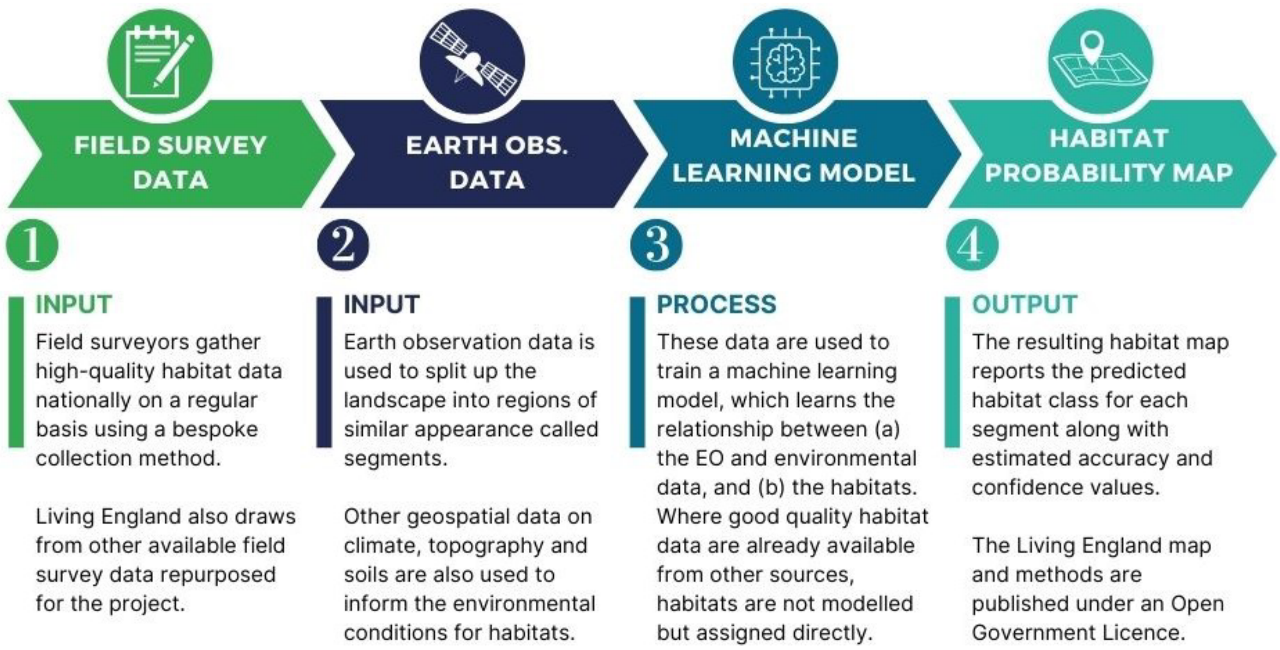

**Figure 2: A summary of the key stages of the Living England workflow: field survey data, earth observation data, machine learning modelling, and production of the habitat probability map.**

# 3.1 Field Data collection and preparation

## 3.1.1 Field data collection

Living England is supported by a national-scale field data collection programme, collecting data specifically for the purpose of informing Earth observation models. This allows for direct targeting of data collection, where field data is lacking, or model confidence is low and warrants further investigation in the field. This targeted collection has vastly improved the modelled approach and the data available to validate the Living England maps.

The Living England field survey data are collected via externally contracted field surveys as well as training of internal Natural England field surveyors according to the Living England field protocols.

Field data collection consists of targeting specific habitats and gaining access permission to survey a site. Once obtained, the surveyor will locate a targeted segment and collect a central data point, recording the percentage cover of habitats present. The habitat within the segment should be homogenous, covering >60% of the segment, to be recorded as a distinct broad habitat class. To facilitate easy and consistent data recording, a bespoke ESRI Field Maps application (ESRI, 2024) has been specifically developed for collecting data aligned to the Living England spatial framework.

## 3.1.2 Additional habitat survey data

To supplement field data collection for the project, survey data from ongoing and past national surveys within Natural England is also collated, as well as habitat survey data collated from partner organisations, including Forest Research's National Forest Inventory (NFI, 2023), data from the Environment Agency's coastal surveys (EA, 2018, 2022), and the National Plant Monitoring Scheme (NPMS, 2023). The data go through several cleaning and quality assurance (QA) steps and have been translated into the Living England habitat classification framework (Holmes et al., 2024) through an automated reproducible process, utilising a Living England habitat translation database. This database brings together knowledge from the correspondences between different habitat classifications (JNCC, 2008; Mountford and Strachan, 2022) and is supplemented by specialist knowledge from the team ecologists, to automate the translation of survey data collected using different habitat classifications in the field, into the Living England classes consistently (see Appendix 1).

### 3.1.3 Field data quality assurance

The combined field dataset consisting of the Living England surveys and existing surveys were collated, totalling 116,915 data points. The data went through an automated quality assurance process in R (version 4.2.3, R Core Team, 2023), to ensure the data are representative of the habitats identified.

The process included the following two checks:

1. **Data duplicates -** removal of duplicate data points where the same class has been identified multiple times for the same segment.
2. **Spectral outliers -** removal of data points which do not meet expected spectral reflectance value ranges, indicating this may no longer be representative of the habitat present in the latest imagery mosaics.

**Data duplicates** are removed through a 'one-per-segment process' which involves selecting unique Segment ID and broad habitat class combinations, retaining the data point with the latest survey date recorded. This process retains data points where multiple habitat classes have been identified in the same segment; however, these must meet the filter for spectral outliers for both habitats. A total of 17, duplicates were removed at this stage.

**Spectral outliers** are identified through comparisons with a subset of data where high confidence suggest it is representative of the current habitat state. The data subset is created with the current ongoing field surveys from Natural England, shown in Table 2 where data has been collected in the last 3 years (between 1$^{st}$ January 2021 - 12$^{th}$ December 2023). The subset of high confidence data also underwent a quality assurance step to flag any outliers in the data. Where data points were flagged as an outlier in any of the S2 bands minimum or maximum ranges across all three seasons, the points were removed. This resulted in 110 points being removed from the "high confidence" data subset.

**Table 2: Descriptions of the "high confidence" surveys and the number of points of each of these collected in the last 3 years, used to filter out spectral outliers in the Living England (LE) field dataset.**

| Survey | Description | Number of points in spectral filter |
|---|---|---|
| **Living England Data Collection** | Bespoke data collection coordinated by the LE field surveyors and contractors. Data is recorded based on the habitat segments to the LE habitat classification descriptions. | 9,921 |
| **NE Protected Sites Monitoring** | Monitoring surveys from protected sites collected by NE field surveyors recorded via the NESS CSMi application. Data collection is at a finer classification than LE at UKBAP priority habitat level. | 2,298 |

| | | |
|---|---|---|
| **England Peat Map (EPM) Vegetation Surveys** | Vegetation surveys as part of NE's England Peat Map project, aiming to map the depth, extent and condition of peatlands on a national scale. Data capture is at a finer classification than LE but aligned with LE broad classes. These points have passed quality assurance steps as part of EPM. | 583 |
| **Long-Term Monitoring Network** | Long-term monitoring of plant communities to identify change and possible drivers. Data collection is aligned to UKBAP broad and priority habitat classes. | 450 |

For each habitat class in the data subset, the minimum and maximum S2 spectral reflectance values and S1 backscatter values are calculated for each image band to give an expected habitat spectral range (refer to Section 3.2 Geospatial Data Processing). The spectral ranges are used to filter the remaining habitat data, with a data point removed if it is flagged as an outlier in any image band. Outliers within the high confidence datasets were flagged based on statistical inter quartile range thresholds, whereas outliers for other datasets were flagged where points fell above or below the values set by the high confidence data subset. Where the data references a segment covered by cloud in all three Spring, Summer and Autumn S2 mosaics, the data point is removed, however, no points were identified during this step. If it only has cloud cover across one or two of the seasonal image mosaics, but still sits within the habitat spectral range, the data point is retained. There were 818 points flagged with cloud in either one or two seasonal mosaics which were retained.

After removing any duplicates (17,294), cloud (0) and spectral outliers (27,692) of which the majority were from old data which were no longer representative of the habitat present, a total of 71,929 habitat data points remained and were used in the workflow for training and validation. Of this total, 69,069 data points were used to inform the habitat classification. In addition to these, an additional 418 hedgerow points were digitised by the Living England team based on the seasonal mosaics and segmentation, totalling 69,549 ground points to inform the modelled habitats (3.4 Model-Based Classification). This excludes points relating to non-modelled habitats where bespoke methods were used to map extent (Section 3.5 Specific Habitat Classification), with points used to evaluate and validate the bespoke map outputs.

# 3.2 Geospatial Data Processing

## 3.2.1 Biogeographic Zones

To accommodate the acquisition of cloud-free seasonal image mosaics, and accounting for phenological and habitat variations across England in the classification model, England was split into 13 Biogeographic Zones (Biogeographic Zones Living England, 2019) (BGZs). These are pictured in Figure 3 with a generalised description of the zones. The BGZs were derived by merging the National Character Areas (Natural England, 2021), which are distinct areas each with a unique ‘sense of place’, in accordance with satellite orbits.

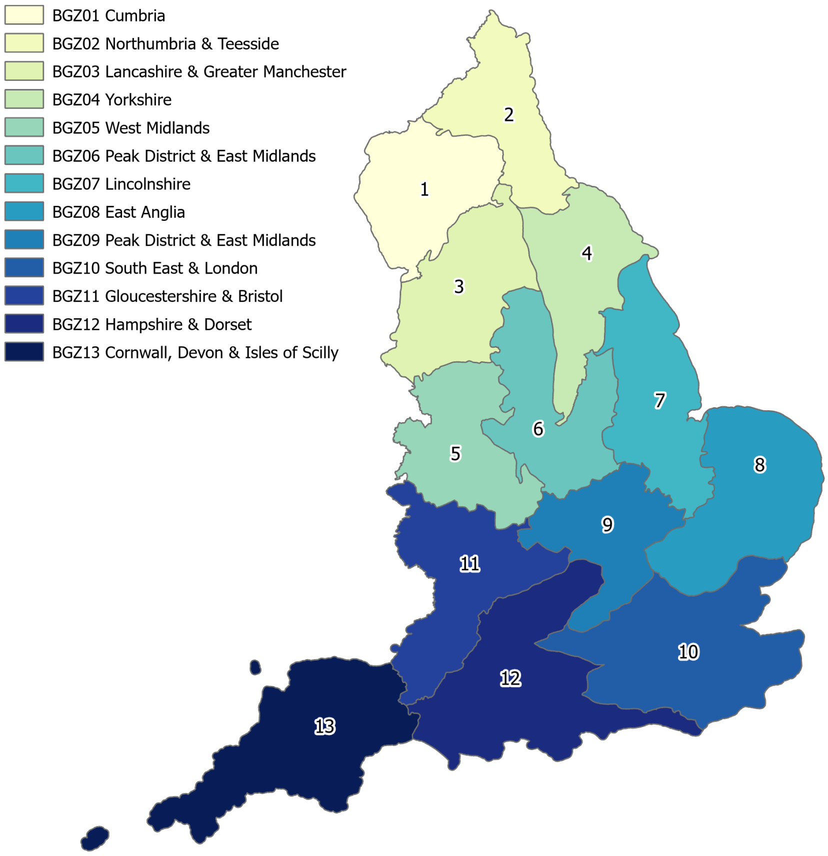


**Figure 3: Living England Biogeographical Zones (BGZs): generalised regions describing areas of shared phenological and habitat characteristics**

### 3.2.2 Sentinel-2 Cloud and Cloud Shadow Masking

Sentinel-2 (S2) is a wide swath, high-resolution, multispectral imager with 13 spectral bands including visible, near-infrared and short-wave infrared satellite part of the European Space Agency (ESA) and European Union (EU) Copernicus programme (ESA, 2012). S2 provides imagery of the Earth's surface at a spatial resolution varying from 10 m to 60 m. As an optical satellite, the land surface is often obscured by cloud, and modified in appearance by cloud shadow, both of which reduce the quality of the imagery used as input for habitat modelling. Cloud and cloud shadow must therefore be removed prior to modelling to produce cloud free mosaic images of the land surface for use in modelling.

The methodology for the identification of clouds, and subsequently cloud shadow, is based on the S2 cloudless cloud probability product (Gorelick et al., 2017, Copernicus Sentinel data, 2023) The S2 cloudless cloud probability layer is used to identify cloudy pixels, using a threshold cloud probability of greater or equal to 50%, following the testing of a range of potential threshold values, to ensure that as much cloud as possible was removed from imagery. Higher values often resulted in thin wispy clouds remaining and so a lower threshold was selected to ensure these clouds were removed as the inclusion of cloud was deemed more harmful to modelling, than the exclusion of valid imagery. Additional cloud was added to this mask using pixels identified as cirrus clouds by the S2 Scene Classification Layer (SCL).

Imagery, and accompanying cloud masks, captured on the same day were subsequently mosaicked together before the identification of cloud shadow. This was carried out to account for clouds on the edges of tiles, which depending on the solar azimuth may cast shadows onto different tiles. If images on the same day were not mosaicked together, such shadows would be undetected as they rest on a different tile to the parent cloud.

Cloud shadow was identified using a directional transform methodology (Braaten, 2024). Potential cloud shadow was identified by projecting areas detected as cloud along the solar azimuth for 10 km, a distance which was found to be sufficient in capturing cloud shadows across the latitudinal range being covered. This was repeated four times: at solar azimuth (AZ), AZ minus 10 degrees, AZ plus 10 degrees, and AZ plus 20 degrees, to account for variation in cloud height which modifies the projection angle of the cloud shadow in respect to the solar azimuth. Within this potential cloud zone, true cloud shadow was then identified by selecting pixels with a Near Infrared Red (NIR) value below a threshold of 0.15, to identify true cloud shadow from all potential cloud shadow areas. This threshold value was determined through iteratively testing a range of values and looking across multiple scenes to balance cloud shadow omission and commission errors. Water bodies including rivers and lakes are identified using a combination of the rivers layer as provided by OSMM, and the Normalised Difference Water Index (NDWI), and added back after cloud shadow masking, which often misidentifies water as cloud shadow. The S2 SCL was used to identify and mask oversaturated, ice and snow covered pixels. This was achieved by creating binary pixelwise masks using the flags within the SCL bitmask for oversaturated, ice and snow-covered pixels and combining these masks with the cloud, and cloud shadow mask. Small cloud and cloud shadow fragments less than 400 $m^2$ in area were removed as visual assessments highlighted that these were often erroneous. The final cloud and cloud shadow mask were then buffered using a 40 m circular kernel to smooth out the masks.

### 3.2.3 Sentinel-2 Image Mosaicking

As the width of S2's swath (290 km) (ESA, 2012) is less than the width of England, the whole area cannot be captured in a single pass by the satellite and so multiple scenes, which may have been captured at different times, need to be combined to create continuous national image cover. In addition, the masking of cloud and cloud shadow results in gaps in images, which require filling where possible, with cloud free imagery of the same area captured as close in time as possible.

Spatially continuous image mosaics were created for Autumn 2022, and Spring and Summer 2023. Images for each season were created by first collecting all available imagery within the date ranges listed in Table 3. The capture window for the Autumn and Spring seasons were one month greater than the two-month capture window for the Summer season, to enable greater cloud-free image availability.

**Table 3: Sentinel-2 Imagery capture dates for Living England 2022-23 and the number of scenes captured within each date range**

| Image Mosaic | Date range | Year | Number of scenes within date range |
|---|---|---|---|
| Autumn | 1st August – 31st October | 2022 | 832 |
| Spring | 1st March – 30th May | 2023 | 578 |
| Summer | 1st June – 31st July | 2023 | 465 |

All cloud and cloud shadow within these images were removed, and any satellite scenes collected on the same day were spatially joined as previously detailed. The image within each season with the lowest percentage of cloud cover was identified using the S2 metadata (averaged across each image mosaic captured on the same day) and used as the primary image for the purposes of mosaicking. The time difference between the primary image capture and all other images within the season was calculated and used to stack these in order of those closest in date to the primary image.

The image stack was mosaicked together using a "Swiss Cheese" methodology (Kilcoyne et al., 2022), with areas not covered by the primary image, or holes in the primary image being filled with the next image in the image stack with imagery in that area. The "Swiss Cheese" methodology is summarised visually in Figure 4.

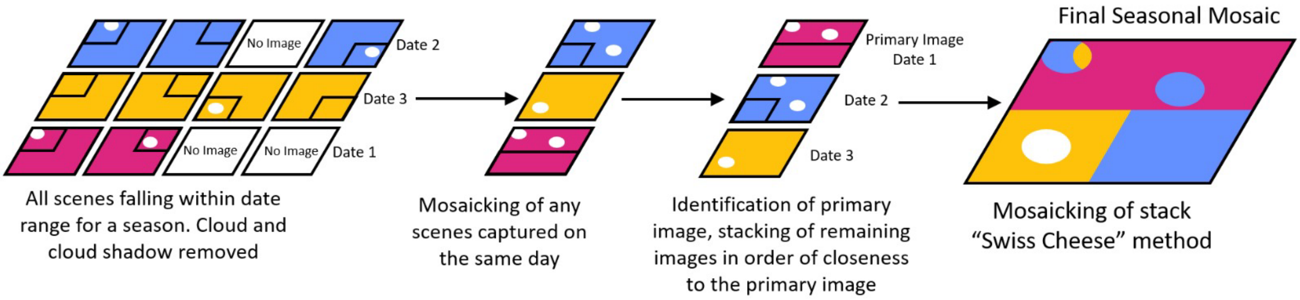


**Figure 4: A summary of the "Swiss Cheese" methodology for creating the Sentinel-2 seasonal image mosaics. This describes how imagery dates are stacked within a season and used to gap-fill over areas of cloud and cloud shadow.**

### 3.2.4 Sentinel-1 Backscatter and Coherence

S1 is also part of the ESA/Copernicus initiative equipped with a C-band Synthetic Aperture Radar (SAR) sensor. SAR operates at wavelengths which are not impeded by cloud cover or low illumination, and therefore can be acquired day or night and all-weather conditions providing additional information to the S2 data. As well as S1 backscatter, the Living England workflow also uses coherence maps derived from the S1 data, which provide a comparison of both phase and amplitude between a group of pixels from two images. This indicates the consistency of the Earth's surface and can provide valuable information on changes in surface properties such as height and type of vegetation, aiding identification of habitat type.

Acquisition dates for the S1 datasets were chosen to coincide with likely cloud-free S2 imagery and were screened to reduce the chance of heavy rainfall or snow; both of which influence the dielectric properties of the ground surface modifying coherence (Zwieback et al., 2015, Palmisano et al., 2019, Souza et al., 2022). These were screened using ERA-5 atmospheric model data from the ESA Climate Data Store (CDS) accessed via the API (Hersbach et al., 2023). This applies threshold ranges for rainfall, snowfall and wind speed across multiple time periods, to remove any image dates likely to be substantially impacted by weather effects. The thresholds were established through testing across a range of image years, and then applied across 6-day intervals. The resulting thresholds were 10 m/s for windspeed, 15 mm/day for rainfall and 1 mm/day for snowfall. The screened dates used are described in Table 4.

S1 backscatter mosaics were created using Google Earth Engine (GEE) (ESA, 2023, Gorelick et al., 2017), with the image capture dates shown in Table 3, aligned to the date ranges for the S2 mosaics. The median pixel value was extracted from 4-week collections to reduce radar speckle. Both vertical-vertical (VV) and vertical-horizontal (VH) polarisations were extracted from the ascending and descending acquisition modes to produce temporally averaged S1 backscatter layers with reduced image speckle.

**Table 4 Sentinel-1 SLC Imagery capture dates for Living England 2022-23 and the number of scenes captured within each date range, after extreme weather screening was applied.**

| Sentinel-1 Coherence/Backscatter seasonal composite | SLC Imagery Dates Range | Year | Number of Dates After Removing Days with Extreme Weather |
|---|---|---|---|
| **Autumn 2022** | 30th August – 30th October | 2022 | 40 |
| **Spring 2023** | 1st March – 30th April | 2023 | 41 |
| **Summer 2023** | 9th June – 30th July | 2023 | 35 |

The average seasonal coherence maps were created using a processing chain in Python, version 3.7 (2018) using the ESA Sentinel Applications Platform (SNAP) software version 9.0 (2022). Coherence images were created using both ascending and descending orbit directions and VV and VH polarisations, utilising a seamless 30 m Digital Terrain Model (DTM) derived from merging the EA LiDAR data, Integrated Height Model (IHM) DTM, and SRTM DTM to capture the extent required (Environment Agency, 2023a, Bluesky International Ltd 2022, NASA SRTM 2013) for back geocoding and range-doppler terrain correction. Coherence image pairs were then geo-coded, and the mean average taken, such that each temporal baseline (6-day, 12-day, 18-day, 24-day) were equally weighted.

### 3.2.5 LiDAR National Programme

The EA's National LiDAR Programme provides high spatial resolution (1 m) topographic data in the form of Digital Surface Models (DSMs) and DTMs (Environment Agency, 2023a). The DSMs are derived from the first return LiDAR pulses and therefore provide valuable height data for objects above ground level such as buildings, trees and other vegetation. DTMs have these elements stripped out to give a smoother model of the bare surface of the Earth. Complete national LiDAR coverage was made available for the first time in 2023. For Living England 2022-23, LiDAR data captured between 2016 and 2023 was used to provide topographic characteristics of habitats and land surface features.

The LiDAR products were resampled to 10 m resolutions and have been used to derive slope and aspect layers and Topographic Wetness Index (TWI) layers for use in Living England, generated using tools within ArcGIS Pro (Esri, 2022). Canopy Height Models (CHM) were also generated by subtracting the DTMs from the DSMs for the whole of England. Once generated, the CHMs were subjected to spatial analysis using the 'Focal Statistics' tool in ArcGIS, producing a layer of spatial variability in canopy height (Esri, 2022).

### 3.2.6 Proximity to feature layers

Several other geospatial layers were used within the Living England methodology, including those characterising climatic, topographic and soil characteristics. Alongside these, three proximity layers were created to provide spatial constraints and describe three relationships:

- Proximity to water bodies, having removed those with areas <300 $m^2$ to align with the Minimum Mapping Unit (MMU),
- Proximity to urban habitats,
- Proximity to the coast.

The OSMM Topography Layer (2023b), contains over 500 million features representing objects in the physical environment, such as buildings, roads and water bodies, at a scale between 1:1250 to 1:10,000. For this iteration of Living England, the March 2023 release of OSMM was used. This was used to inform the image segmentation (section 3.3 Image Segmentation and to derive the proximity to water bodies layer, as well as informing the 'Built-up areas and gardens' class (section 3.5.1 Urban and Water. The OS Urban Area Extent (Ordnance Survey, 2023c) was used to derive the proximity to urban habitats layer was not resampled to the MMU, as the dataset is more generalised and reduces the computational demand for national-scale analysis. The proximity to the coastline layer was generated using the broader MHWS boundaries (Ordnance Survey, 2023a) and did not require resampling to the MMU. The proximity layers were all generated in ArcGIS Pro (Esri, 2022) calculating the accumulative distance to the closest source feature.

## 3.3 Image Segmentation

The Living England habitat map is generated using an object-based classification. Objects or 'segments' describe discrete spatial areas which make up the spatial framework used to classify and map different habitat types. Object-based Image Analysis (OBIA) has been adopted by Living England as although there are advantages and disadvantages with both OBIA and pixel-based approaches, the use of classifying objects allows our analysis to reflect meaningful features in the landscape by grouping spectrally similar pixels (Smith and Morton, 2010, Barber and Robinson, 2023). The use of objects also allows the model to see complementary contextual information to improve accuracy and greatly reduces within-class variation or 'salt-and-pepper' effects typically found with pixel-based classifications (Liu and Xia, 2010, Whiteside et al., 2011, Tewkesbury et al., 2015). This is particularly important in designing a methodology where advances in the resolutions of satellite-derived imagery are leading to smaller pixel sizes where this effect is likely to increase. However, OBIA approaches also have limitations and resulting segments often represent mixed habitats and can result in fuzzy boundaries for example around field edges. An OBIA approach was used for Living England as pixel-based approaches using S2 data with 10 m also consist of mixed habitats and testing of the segmentation parameters was undertaken through desktop approaches and with field surveyors to ensure as close alignment of the segmentation with ground features as possible.

The Living England segmentation process is conducted using a series of algorithms within the Trimble Geospatial (2023) eCognition software (Version 10.3) including multiresolution, chessboard, and spectral difference segmentation. This process identifies and groups image pixels into areas of similar spectral reflectance characteristics based on how homogenous their characteristics are. This uses multiple homogeneity criteria to generate segments, including colour, shape, smoothness and compactness. For Living England 2022-23, the segmentation has been generated using:

(a) S2 seasonal image mosaics from Autumn 2022, Spring 2023 and Summer 2023.
(b) LiDAR-derived CHM and focal texture layers.
(c) OSMM Topography urban and water features (see section 3.5.1 Urban and Water).

All raster layer inputs are 10 m resolution, either natively or resampled using bilinear interpolation. Separate processing loops are first run for urban and water regions, using the OSMM features. This allows for refined segmentation parameters enabling better representation of these habitat feature boundaries. The remaining segments are then run through a final segmentation to fully segment all regions. To ensure there are sufficient pixels within each segment to inform meaningful classification, a MMU of 3 pixels (300 $m^2$) is established within the segmentation, however, this can result in urban fringes being removed and potential undermapping. This results in the Living England 2022-23 product containing approximately 13.1 million segments.

## 3.4 Model-Based Classification

### 3.4.1 Model variables

To inform the model classification, training data was compiled from habitat labels as described in Section 3.1 Field Data collection and preparation and combined with variables used to define the relationship between environmental and contextual characteristics, and the presence of habitats. The predictor variables were extracted from the data using a reproducible automated process deployed in R (R Core Team, 2023). This included calculating the zonal statistics within each segment for each variable (minimum, maximum, mean and standard deviation). The predictors at this stage included:

- 10 S2 bands for each of the Autumn, Spring, and Summer mosaics (excluding band 1 aerosols and band 9 water vapor from the harmonised level 2A product),
- 2 band indices derived from the S2 seasonal mosaics: Normalised Difference Vegetation Index (NDVI) and Normalised Difference Wetness Index (NDWI),
- 2 bands for S1 backscatter for each of the Autumn, Spring, and Summer mosaics,
- 4 bands from S1 coherence product for each of the Autumn, Spring, and Summer mosaics,
- LiDAR topographic data layers: DTM, Slope, Aspect, Topographic Wetness Index, CHM and focal statistics,
- Climatic annual and seasonal average rainfall, and minimum, maximum and mean temperature for 2-year and 20-year periods,

- Proximity layers to coastal, urban and water OS features,
- Soil and geology datasets including; soil parent material carbonate content, grain size, European Soil Bureau (ESB) group, soil texture, soil thickness; soilscapes and wetness; and bedrock geology.

A total of 386 variables were fed into the modelling stage with 69,549 quality assured ground data points.

### 3.4.2 Model approach

Living England predicts most habitat classes using a Random Forest (RF) modelling framework deployed in R (R core team, 2023). The model is trained on the field habitat data, satellite imagery, and predictor variables as described in section 3.4.1.

Spectral data from remote sensing sources is represented as the radiance measured between different wavelength ranges called bands. These bands are often highly correlated with each other due to atmospheric and patterns in the objects, which reflect light and can influence multiple bands in a similar fashion. Although RF models are often assumed to be resilient to multicollinearity, removal of very highly correlated variables prior to training reduces redundancy in the model and is considered good practice in machine learning (ML), (Darst et al., 2018). Highly collinear variables were identified by calculating the correlation between all variables. For each variable pair with a correlation coefficient greater than 0.9, the average correlation of each variable in the pair across all other pairs was calculated and the variable with the highest value removed. Correlations between all pairs were then recalculated, and this process repeated until all pairwise correlation coefficients were less than 0.9. A total of 214 highly colinear variables were identified and removed by this process, leaving 148 variables for model building.

Following removal of the highly correlated variables, the RF model was trained on 80% of the field dataset while the remaining 20% was retained separately for validation and not used during the modelling process to ensure independent evaluation of model performance. Both training and validation subsets were balanced, so they each contained the same proportions of each habitat class.

Models can be influenced by key data points, that substantially influence the resulting model when included in either the training or validation set. To account for this, a k-fold approach was implemented whereby the 80/20 split of data was systematically rotated to ensure that all field data points were used to both train the model, and asses its performance (Hastie et al., 2009). A visual representation of the data splitting is shown in Figure 5.

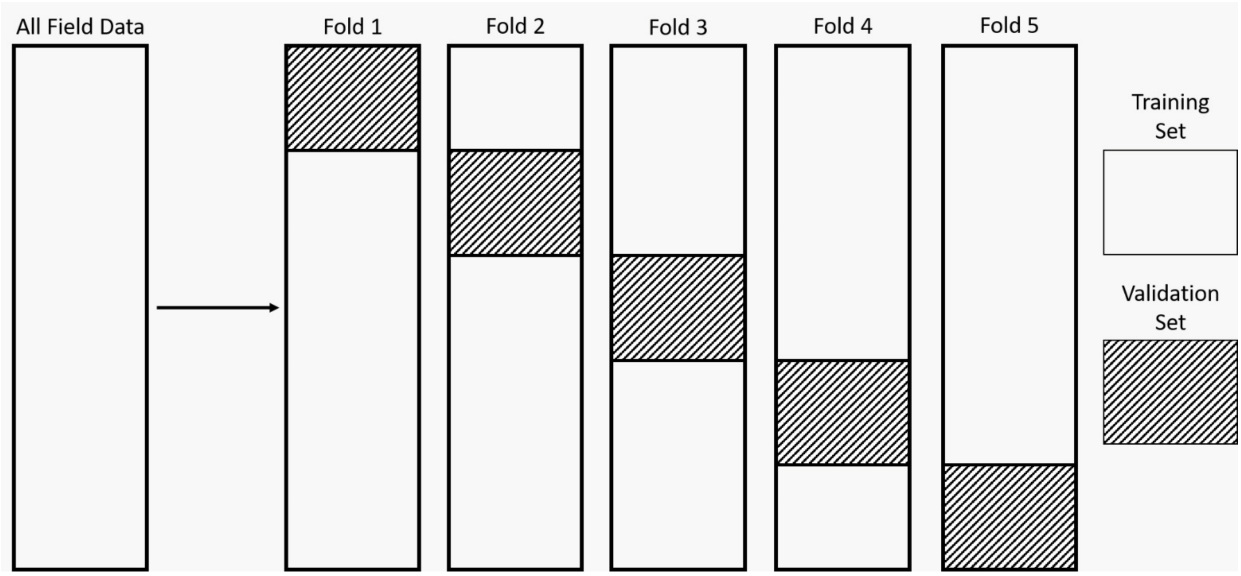


**Figure 5: Visual representation of k-fold data partitioning used for training the distributed random forest model. By splitting the data in this manner, each k-fold had an independent training and testing set, allowing for all the training data points to be in both the training and validation datasets. The best fold model was then selected and evaluated against the independent test dataset.**

The presence of highly influential field data points having a detrimental impact on model performance can be detected by the presence of large variations in model performance across the folds. The performance on each of the fold-models has been reported (Appendix 4) and the variation across all fold models analysed to make inferences regarding the generalisability of the model.

For each fold, a Distributed Random Forest (DRF) model was trained, and validated using the training and validation sets, respectively, for that fold. The model hyperparameters and the number of trees were optimised using the entire dataset, and subsequently used across all folds. The DRF was implemented using the H20ai library (H20.ai, 2022) due to its capacity to cope with segments with missing data, which allows the generation of predictions where data gaps exist due to cloud/cloud shadow. Missing values are interpreted as containing information, rather than being missing at random, which allows decision trees to be built that optimise performance in the case where data is missing, treating the absence of data as a decision point itself. This consequently allows predictions to be made in the absence of data.

Following training, each fold model was evaluated on the independent validation dataset and performance metrics are calculated including the confusion matrix, accuracy, balanced accuracy, and F1 score. Each of these are extracted at the class level (definitions in Table 5). The fold model with the highest balanced accuracy averaged across all classes was selected as the final model. Balanced accuracy provides a reliable measure of performance, considering both positive and negative, or presence and absence, prediction ability, and is independent of the relative frequency of each class due to it being standardised in its formulation. Thus, the use of balanced accuracy removes potential bias towards the largest classes and provides a better comparative metric across all classes regardless of extent or rarity for the purpose of selecting the best model. This model was then used to calculate validation and predict across all segments.

**Table 5: Metrics and definitions used to evaluate model performance.**

| Term | Definition |
|---|---|
| **One vs all** | To access the performance of a multiclass classification system, we access performance for each class individually using a one-vs-all approach where the current class is defined as the positive (or presence) class, and all other classes as the negative (or absence) class. |
| **True Positive** | After applying a one-vs-all approach, the model predicts positive (presence), and the ground data also shows positive (presence) for the class being evaluated. |
| **True Negative** | After applying a one-vs-all approach, the model predicts negative (absence), and the ground data also shows negative (absence) for the class being evaluated. |
| **False Positive** | After applying a one-vs-all approach, the model predicts positive (presence), but the ground data shows negative (absence) for the class being evaluated. |
| **False Negative** | After applying a one-vs-all approach, the model predicts negative (absence), but the ground data shows positive (presence) for the class being evaluated. |
| **Accuracy** | A measure of the model's overall performance across all classes. Calculated as the number of times the model correctly predicts the class (as determined by the ground data for each segment), divided by the total number of predictions made |
| **Balanced accuracy** | Accuracy due to its formulation is weighted more heavily towards the larger classes. Balanced accuracy normalises the contribution of each class to the average, so each class contributes equally to the final average. |
| **F1 Score** | Similar to balanced accuracy but gives more weight to whether segments are correctly identified. This contrasts with balanced accuracy, which gives equal weight to correct and incorrectly predicted classes. |

| Term | Definition |
|---|---|
| **Precision/ User accuracy** | After applying a one-vs-all approach, precision/user accuracy is calculated as the number of True Positives divided by the combined number of True Positives and False Positives made by the model. This statistic tells us when the model predicts the presence of a particular class, how often is it correct. |
| **Recall/ Producer accuracy** | After applying a one-vs-all approach recall/ producer accuracy is calculated as the number of True Positives, divided by the combined number of True Positives and False Negatives. This statistic tells us what proportion of all instances of a class did our model capture. |
| **Specificity** | After applying a one-vs-all approach what proportion of negative instances which were truly negative. |

### 3.4.3 Model results and validation

The final model had an overall accuracy of 0.87 or 87%. Examining balanced accuracy between classes (a measure of performance independent of the class frequency), reveals that performance across all the classes is good with the highest score being 0.96 (Broadleaved Mixed and Yew Woodlands) and the lowest 0.80 (Scrub). Broadleaved Mixed and Yew Woodlands, Coastal Sand Dunes and Conifer Woodlands have the highest balanced accuracy, while Scrub, and Improved and Semi-Improved Grasslands are the lowest performing classes. Table 6 details model performance for each class.

**Table 6: Detailed performance statistics for the final Random Forest model informing Living England 2022-23.**

| Class | Precision/ User accuracy | Recall/ Producer accuracy | Balanced accuracy |
|---|---|---|---|
| **Broadleaved, mixed and yew woodland** | 0.97 | 0.94 | 0.96 |
| **Bog** | 0.77 | 0.88 | 0.93 |
| **Bracken** | 0.69 | 0.80 | 0.90 |
| **Bare sand** | 0.93 | 0.90 | 0.95 |
| **Coastal sand dunes** | 0.80 | 0.97 | 0.98 |
| **Coniferous woodland** | 0.82 | 0.92 | 0.95 |
| **Dwarf shrub heath** | 0.84 | 0.80 | 0.89 |
| **Fen marsh swamp** | 0.67 | 0.69 | 0.84 |

| | | | |
|---|---|---|---|
| **Inland rock** | 0.73 | 0.74 | 0.87 |
| **Improved and semi-improved grassland** | 0.60 | 0.78 | 0.88 |
| **Scrub** | 0.65 | 0.60 | 0.80 |
| **Unimproved grassland** | 0.83 | 0.70 | 0.84 |

## 3.5 Specific Habitat Classification

Several habitats are classified outside of the model-based classification, where there is reliable data detailing the extent of the habitat class already available. This is referred to as the vector approach in Table 1 and ensures consistency with other habitat data products published by the Defra Group and partners, such as the EA Saltmarsh Extent and Zonation dataset and the RPA's Crop Map of England (CROME). Automated workflows have been developed to ensure consistency with the modelled habitats and allow for quality checks to be undertaken when transposing the data into the Living England spatial framework.

Table 7 provides a summary of the broad habitats and features classified in this way and the datasets and methods applied. The table also lists the layer hierarchy, describing the order in which the derived vector layers are then added in when merging the final Living England classification dataset.

**Table 7. Summary of Living England Habitat classes/features not assigned through the modelled approach and how these have been derived.**

| Layer hierarchy | LE Habitat Class | Datasets | Approach Summary |
|---|---|---|---|
| **1** | Water | OSMM Topography Layer (2023b) | Areas within OSMM identified as 'Water'. Segments are assigned where they have any overlap with these features. |
| **2** | Built-up Areas and Gardens | OSMM Topography Layer (2023b), OpenStreetMap 2024 | Areas within OSMM identified as 'ManMade', supplemented by additional areas along urban fringes in the OS Open Built-Up Areas dataset. Segments are assigned where they have any overlap with OSMM features. Allotments were noted as a feature missing from OSMM. Location data for allotments were downloaded from using the OpenStreetMap API using the tags landuse: allotment. These features were then overlaid to assign segments as Built-up areas and Gardens where segments had over 1% overlap with the allotment polygons. |
| **3** | Coastal Saltmarsh | EA Saltmarsh Extent & Zonation (2023b), EA LiDAR DTM (2023a), EA Coastal Flood Boundary and AIMS Spatial Sea | A combination of two saltmarsh mapping methods informed by the EA Saltmarsh Extent and Zonation dataset, and a bespoke method developed with JBA Consulting Ltd. Segments are assigned |

| | | | |
|---|---|---|---|
| | | Defences (2023c), OS VectorMap District (2023d) | where they have 50% overlap. These are assessed for inclusion through manual visual assessments against the S2 mosaics and Aerial Photography GB (APGB) imagery. |
| **4** | Solar Farms | S2 seasonal mosaics | Digitised in Esri ArcGIS through comparisons with the 2022-23 S2 seasonal mosaics. |
| **5** | Arable and Horticultural | RPA CROME (2022), Agricultural Land Classification (Natural England, 2023) | Subset of classes taken from CROME aligning to LE arable and horticultural class. Segments are assigned where they have >1% overlap with these features, and then sense checked against the ALC grade, removing any grade 5 - very poor-quality agricultural land. |
| **6** | Bare Ground | OSMM Topography Layer (2023b), LE S2 seasonal mosaics | Areas within OSMM identified as 'Mineral Workings' or 'Spoil Heaps'. Segments assigned with 50% overlap and NDVI is <0.2 in all seasons. Additional areas identified using a bare soil index analysis where segments below a threshold value of 0.1 in Spring and 0.2 in Summer and Autumn were included. |

### 3.5.1 Urban and Water

The following feature attributes within OSMM Topography Layer were selected to identify the Living England classes of 'Built-up Areas and Gardens' and 'Water' (Ordnance Survey, 2023b):

- Built-up Areas and Gardens:
    - Class identified where OSMM features are classed as "Manmade" or "Multiple" in the MADE attribute field.
    - Removed from this class are OSMM features that contain "Water" or "Landform" in the DESCRIPTIVE GROUP column, or "Mineral Workings" or "Spoil Heap" within the DESCRIPTIVE TERM attribute field.
- Water:
    - Class identified where OSMM features as classed as "Water" in the THEME attribute.

OSMM provides a higher level of spatial detail than is visible from the S2 imagery. As a result, it is necessary to pre-process the OSMM features to better align them with the Living England spatial framework. This alignment comprises the following steps, with a before/after alignment comparison provided in Figure 6:

- Feature aggregation to generalise and disaggregate any multipolygon features,

- Removal of any features below the LE MMU (300 $m^2$) and any holes within features,
- Dilute and expand features to help remove any long thin features where the width is unlikely to be visible at the S2 image resolution. This is carried out through iterative buffering to remove empty features where a negative buffer is applied – 3 m for water and –7 m for urban features. The positive buffer is then applied to return features to their original shape (3 m for urban and 7 m for water).
- Alignment of features to S2 pixels (10 m)
- Further removal of features below MMU (300 $m^2$) and any holes within features.

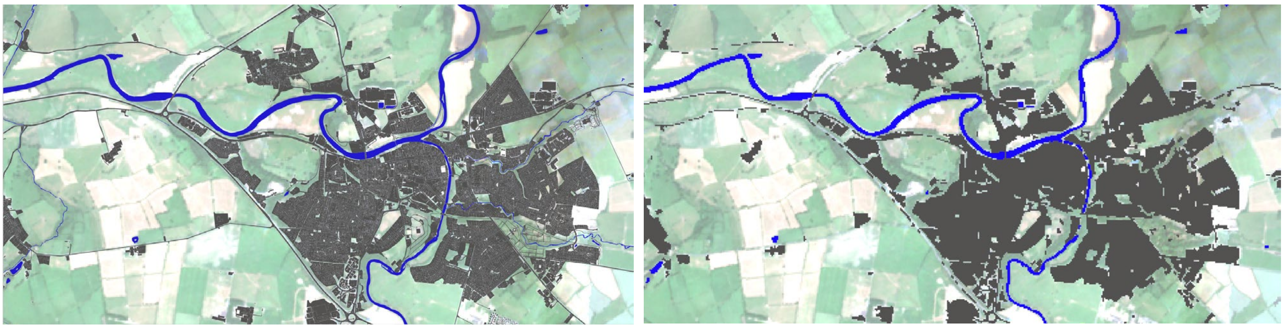

**Figure 6: Comparison between raw OSMM data (left) and the processed segments used in Living England 2022-23 (right) to inform 'Built-up Areas and Gardens' in black and the 'Water' in blue.**

Through assessment of the urban data, it was noted that allotments and caravan parks were notable features missing from the dataset. Caravan parks were manually selected through later quality assurance of the final draft dataset, to ensure these developed features were incorporated into the Built-up Areas and Gardens class. Data showing the location of allotments were obtained from OpenStreetMap using an API call through the R package 'osmdata', version 0.2.5 (Padgham et al., 2023). The allotment data was overlaid with the segmentation with a 1% overlap rule intended to be a lenient threshold to capture all the areas intersecting with allotment features and disturbed land use.

### 3.5.2 Coastal Saltmarsh

The EA Saltmarsh Extent and Zonation dataset (2023b) is a widely used product mapping the extent of saltmarsh from interpreted aerial imagery and ground surveys. The previous use of this dataset in Living England found some saltmarsh areas experiencing change due to expanding or declining extent, were not fully captured when compared with our latest satellite imagery. As a result, a new methodology was developed alongside JBA Consulting Ltd to ensure saltmarsh detection is aligned to the S2 mosaics used in Living England 2022-23, as well as still allowing interoperability with the existing EA data product.

The JBA approach (Cunningham and Cutts, 2023) uses elevation data and a rule-based approach to identify tidal stages around the coastline and constrain saltmarsh to the intertidal zone from Mean High-Water Neap (MHWN) to MHWS tidal levels. This utilised open datasets including the EA Coastal Flood Boundary dataset and LiDAR DTM resampled to 2 m and 10 m resolutions (EA, 2023a, 2023c), OS VectorMap District (2023d) and openly available tools in GDAL/OGL (v.3.5.1, 2024) and SAGA (v7.8.2, Conrad et al., 2015), Once the intertidal zone was delineated, a threshold approach was applied based on the NDVI (see Equations), Median VH backscatter, slope and Normalised Difference Infrared Index (NDII). Thresholds were iterated to ensure the best possible capture of saltmarsh extent and assessed for each zone. The thresholds used were slope <5°, tidal stages >1, autumn NDVI >0.2, spring NDVI >0.15, autumn S1 ascending <26 and autumn NDII <-0.18. This resulted in a single binary output of saltmarsh presence or absence. The method was run for each zone separately prior to collating on a national scale. A summary of the workflow is presented in Figure 7 below.

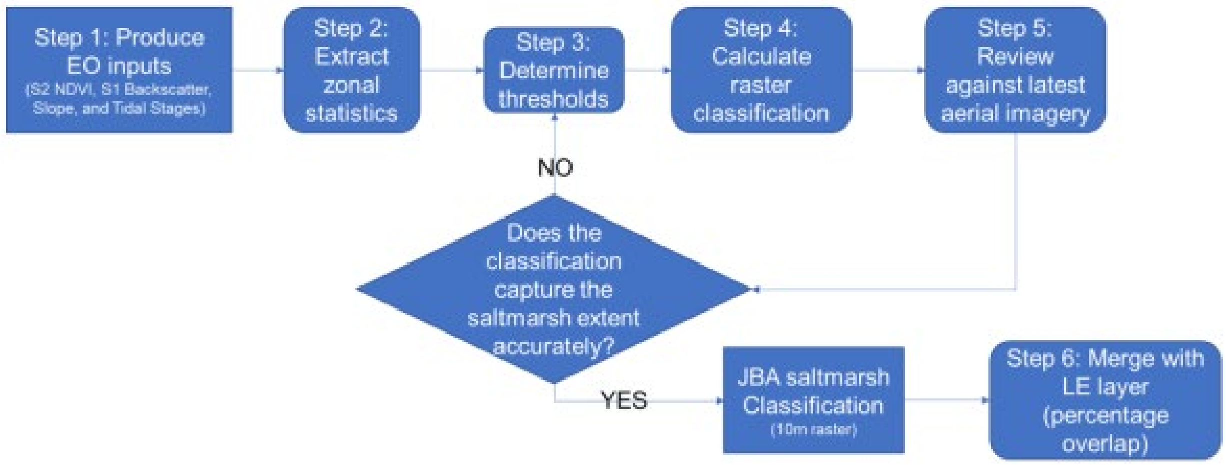


**Figure 7: A diagram summarising the 'JBA approach' workflow developed by JBA Consulting Ltd and the Living England team to map England's saltmarsh extent and integrate results with the Living England spatial framework. Source: Cunningham and Cutts, 2023.**

The outputs of the two saltmarsh extent datasets were then overlaid with the Living England spatial framework and where segments met an overlapping threshold of >50% these segments were assigned to the 'Coastal Saltmarsh' class. The results were then combined and then visually examined to quality assure the data for inclusion. Segments were checked through comparisons with the Living England S2 seasonal mosaics, Living England saltmarsh ground data, and high resolution APGB imagery (Bluesky International Ltd., 2023). Through this assessment, additional areas of saltmarsh not detected in either dataset were also flagged for inclusion, totalling an area of 27.93 $km^2$. Where two or more assessors agreed on the presence of saltmarsh in an identified segment, these included within the saltmarsh class.

### 3.5.3 Arable and Horticultural

The Crop Map of England (CROME) 2022, produced by the Rural Payments Agency (RPA) (2023), provides a modelled output of detailed crop types across England. This product maps multiple types of agriculture, including areas of pasture and grazing and permanent grassland, which don't directly align with Living England's definition of arable and horticultural. Therefore, the data is subset to selected classes to meet the definition, described in Appendix 3. These classes were overlaid with the Living England segments and where >1% area overlaps, the segments are mapped as arable and horticultural. The threshold for this class is more lenient than others due to the hexagonal spatial framework used in CROME and is therefore smaller and less representative of arable fields compared to the Living England segments. This helped reduce overmapping of the arable classes.

During the quality checks, some segments were flagged in upland areas, which did not align with the Living England class definition of arable. Therefore, an additional step was included to extract the Provisional Agricultural Land Classification (ALC) grade for each segment, which describes the possible agriculture activity given climatic, site and soil characteristics present (Natural England, 2023). Where a segment fell within a ALC grade 5 area, this was removed from the Living England arable and horticultural class, as this indicated poor quality agricultural land more likely to contain rough grazing and pasture activity more closely aligned to Living England's improved and semi-improved grassland class.

### 3.5.2 Bare ground

Bare ground encompasses a variety of features, which can have variable appearances in the satellite imagery. For Living England, bare ground refers to bare soil, peat and rock, large enough to meet the MMU of 300 $m^2$, with consistent unvegetated cover across the seasonal mosaics. This includes features such as quarries, with large areas of aggregate rock and soil persistent on the land, as well as degraded peatland areas with large patches of bare peat, and shingle beaches. Where ground data on inland rock is available, these have been included within the bare ground mapping; however, some bare ground sites are not captured well by the ground data and a hybrid method was developed using OSMM features and spectral analysis to identify additional segments.

Quarry features were identified using the OSMM Topography Layer (2023b) where the **DESCRIPTIVE TERM** attribute field is recorded as active "Mineral Workings" or "Spoil Heap". These were aligned to the Living England spatial framework, and where a segment met a 50% overlap in addition to NDVI values <0.2 indicating segments were bare in all three seasons (Remote Sensing Phenology, 2018), these were classed as Bare Ground.

To capture additional areas of bare soil and peat, a Bare Soil Index (BSI) was calculated from the Autumn, Spring, and Summer imagery mosaics (see Equations). The thresholds were selected based on visual assessments against the images. Where segments met threshold values of 0.1 for the Spring BSI and 0.2 for the summer and autumn BSI, these were also assigned to the Bare Ground class.

### 3.5.4 Solar Farms

Although not a broad habitat, solar farms were mapped as a separate class due to their distinct appearance in the landscape, impeding assessments of the underlying habitat and to remove confusion with other habitat classes.

The Department for Energy Security and Net Zero collate the Renewable Energy Planning Database (REPD, 2024) to track renewable electricity projects across the UK. This point location data from January 2024 was used to inform a visual assessment and manual digitisation of solar farms. The solar farms were manually digitised across England based on the S2 seasonal mosaics, where solar farms were present in any image. These were overlapped with the LE segments and were assigned to the “Solar Farm” class where they met a 40% overlap threshold.

## 3.6 Habitat Map Collation

### 3.6.1 Combining habitat mapping approaches

The habitats classified outside of the model were first assigned to the spatial framework, combining these in the hierarchical order shown in Table 7, with water being merged in first, followed by built-up areas and gardens, coastal saltmarsh, solar farms, arable and horticultural, and bare ground respectively. The modelled habitats were then assigned to the remaining segments, taking the habitat class with the highest probability (A prob) for each segment as the ‘Primary Habitat’ (A pred). Where secondary probabilities (B prob) for a segment were within 30% of the primary habitat predicted probability, these were flagged as ‘Probable’ mixed segments.

The modelled hedgerow and scrub classes were combined for the final habitat map due to the uncertainty associated with distinguishing hedgerow features at the S2 spatial resolution. The modelled bare ground and individual habitat mapping layers were also combined into a single ‘Bare Ground’ class, with the source attribute denoting where this has been derived from for each segment.

The Living England spatial framework extends to the OS England boundary (Ordnance Survey, 2023a). All habitat classifications were clipped to the MHWS boundary line, with the exceptions of bare sand where this intersects with the MHWS boundary, or coastal saltmarsh where the whole intertidal extent is mapped.

### 3.6.2 Quality Assurance

Several quality assurance checks were carried out throughout the workflow to ensure that data going into each stage of the workflow met a high-quality standard. The modelled habitats and specific habitat classification vector layers were visually interrogated before combining into the final habitat map. After the habitat map layers were combined, a series of desktop assessments were carried out to identify any data quality issues and assess how well particular habitats classes were being mapped. This included:

1. A qualitative assessment of each BGZ and each habitat class by multiple assessors comparing to the S2 seasonal mosaics.
2. Targeted assessments of selected habitat features and habitats of concern using ecological knowledge of habitat feasibility, e.g. coastal habitats, bracken and bog on specific soil types, Isles of Scilly and Isle of Wight, solar farms and segments where mixed ground data is available.
3. Peer review of a draft data version with Defra group stakeholders and collation of feedback.

As a result, the classes of some segments were manually adjusted where they were found to be incorrect or it was noted certain features were missing in third party data. These are detailed within the data attributes noted in the 'SourceReason' attribute for transparency (Table 1). The attribute identifies the quality assurance check where the Living England team or external partners identified an adjustment was required. Model accuracies presented in Table 6 were calculated before manual adjustment took place.

## 3.7 Reliability Metric

Each segment in Living England is attributed with a primary habitat prediction, a list of modelled habitats (Model probs, >0.1 threshold) and the model probability scores (primary habitat probability) for each segment, giving an indication of the accuracy of these predictions. Alongside this, a reliability score has been developed to provide an indicative measure of how reliable the prediction is. This method provides an easy-to-interpret metric sitting alongside the habitat map, describing the confidence in the habitat class predictions based on the model quality and data sources these have been derived from. The reliability scores are described in Table 8 below:

**Table 8: A table describing the Living England 2022-23 Reliability score descriptions. This is a categorised measure of how reliable the primary habitat prediction is considered to be. A trusted source indicates the reliability of the specific habitat classifications rather than**

**RF modelled habitat and level of alignment with the LE segmentation and classification framework compared to the original dataset.**

| Reliability Score | Description |
|---|---|
| Very High | Primary habitat prediction is likely to be highly reliable. Data is from a trusted source or the habitat is distinctive, and sufficient ground data has been collected, which is representative of the habitat class in that location. |
| High | Primary habitat prediction is likely to be reliable. Data is from a high confidence source, or the ground data is representative and in sufficient quantities to reliably predict the habitat class. |
| Medium | Primary habitat prediction is of mixed reliability. Data is from a trusted source, however interpretation to fit the LE spatial framework or classification may be limited. The habitat class may show mixed responses in the data compared with the quality assured ground data collected. |
| Low | Primary habitat prediction is unlikely to be reliable. Data is from a trusted source, however interpretation to fit the LE spatial framework or classification may be limited or quality assessments have shown the results are uncertain. The habitat may not be distinctive enough with the representative ground data collected. |
| Very Low | Primary habitat prediction is of poor reliability. Data is from a trusted source, however interpretation to fit the LE spatial framework or classification may be limited or quality assessments have shown the results are uncertain. The habitat may not be distinctive enough with the representative ground data collected. |

Communicating uncertainty in the data is particularly important where data has been sourced from modelled outputs and third-party data, where these have inherent biases from the approaches used to derive these, which are often overlooked when using the data in practice to inform land management and policy decisions.

### 3.7.1 Reliability for modelled habitats

The reliability of predictive models is seldom considered beyond the overall accuracy metrics generated by the model, and without calibration. To aid data users and policy makers in using the dataset, an additional metric has been derived to describe the level of reliability for each modelled habitat prediction.

The reliability scores for the modelled habitats, were calculated by first extracting the modelled probability for each segment (Model_probs) output by the RF model. However, the modelled probabilities are only relative to each other and do not indicate the probability of a segment being correctly predicted. To align the modelled probabilities to the likelihood of a segment being correctly identified, the model probabilities are calibrated against the event rate (how often the ground data is correctly being predicted by the model). For example, if a modelled probability score of 0.7 was only being correctly predicted 40% of the time based on the ground data, the new calibrated score would be 0.4. This better represents the level of confidence the model has in the prediction made being correct. Model probabilities were calibrated separately for each broad habitat class.

This method of calibration improved the interpretability of the RF model probabilities and was found to be statistically significant for all habitat classes through the calculation of Briers scores. The calibrated model scores were then converted to a categorical reliability score (very low to very high) using Table 9. The reliability score relates to the primary habitat being predicted. A higher reliability score thus indicates higher confidence in the prediction being correct based on the ground data, while a lower indicates a lower confidence.

**Table 9: A table describing the calibrated model score ranges used to convert data into categorical reliability scores for modelled habitats in the Living England 2022-23 map.**

| Calibrated Model Score Range | Reliability Score |
|---|---|
| 0.00 - 0.20 | Very Low |
| 0.21 - 0.40 | Low |
| 0.41 - 0.60 | Medium |
| 0.61 - 0.80 | High |
| 0.81 - 1.00 | Very High |

### 3.7.3 Reliability for specific habitat classification

Reliability was developed for each of the specific habitat classifications based on the data sources and methodological approaches outlined in Section 3.5 Specific Habitat Classification. These were assigned by:

- **Water** – OSMM is a highly accurate data and up-to-date source for water bodies and as this was directly used to create the LE spatial framework, this was deemed to be "Very High" reliability across all water classified segments.
- **Urban** – OSMM is a highly accurate data source for built-up areas and relatively up-to-date and was directly used to create the LE spatial framework. As this is a highly dynamic habitat experiencing a lot of change, this was deemed to be "High" reliability, acknowledging that timing of updates to the dataset and our imagery capture dates may cause some slight inaccuracies.

- **Arable –** Arable segments were selected from the RPA CROME dataset using a subset of the crop classes and assessed against the ALC classification map. As the original data format is a modelled output using a hexagonal spatial framework, the segments are assigned “Medium” reliability, due to inherent limitations of the underlying data and conversion into the LE spatial framework and habitat classes.
- **Bare ground –** This was deemed “Medium” reliability due to the combination of a mix of data sources and a highly dynamic habitat class. Bare ground is difficult to interpret over the 2-year LE capture period in part due to the seasonal variability.
- **Solar farms** – These have been manually digitised and assessed by multiple assessors against the S2 seasonal mosaics, and so were deemed “High" reliability across all segments.
- **Coastal Saltmarsh** - The reliability score was assigned based on the agreed presence of saltmarsh between both the EA saltmarsh dataset and the JBA methodology derived dataset and how many assessors' agreed saltmarsh was present when assessing. This was assigned the classes as indicated in Table 10.

**Table 10. A table describing the reliability scoring applied to the Coastal Saltmarsh segments, where data were compared from two separate datasets (EA saltmarsh dataset and JBA approach dataset), and then visually quality assured by three assessors.**

| Number of saltmarsh datasets where segment is present (2 datasets) | Number of assessors in agreement (3 assessors) | Reliability Score |
|---|---|---|
| **0** | >= 2 | Very High |
| **2** | 3 | Very High |
| **2** | 2 | High |
| **1** | 3 | High |
| **1** | 2 | Medium |
| **1** | 1 | Low |
| **2** | 1 | Low |
| **1** | 0 | Very Low |

### 3.7.4 Reliability Adjustment

The reliability scores were further assessed through comparisons with our quality assured ground data. Two rules were applied to adjust the reliability scores:

- Where a ground data point was captured within a segment and the habitat recorded correctly matched the Primary Habitat assigned, this is given a “Very High” reliability score,
- Where a ground data point was captured within a segment and the habitat recorded indicates a different habitat to the Primary Habitat assigned, this is given a “Very Low” reliability score and is flagged for further investigation.

### 3.7.5 Reliability Results

Figure 8 shows the mapped reliability scores across England. At a national level, 47% of England’s area cover is predicted with high to very high reliability. Only 10% of cover, was regarded as having low to very low confidence in the results and have been highlighted as areas to focus on future improvement and to target further field investigations.

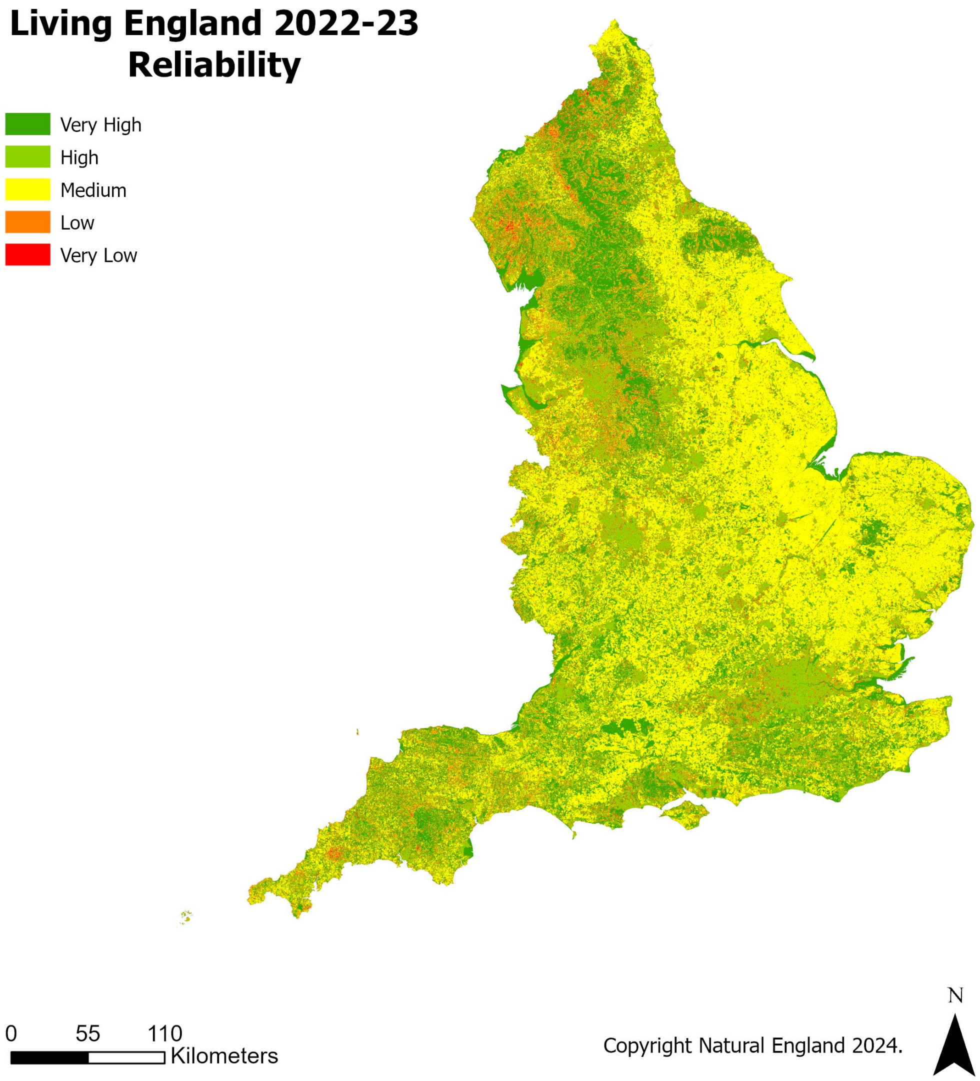


**Figure 8 A map showing the reliability scores of the Living England 2022-23 predicted habitats, ranging from very high to very low.**

**Table 11: A table of the computed number of segments and total area coverage of each reliability score category in Living England 2022-23.**

| | Very Low | Low | Medium | High | Very High |
|---|---|---|---|---|---|
| **Number of segments** | 1088643 | 684969 | 4730381 | 3168924 | 3485776 |
| **Total area cover (km$^2$)** | 7791.49 | 5810.49 | 57238.08 | 19859.35 | 42125.03 |
| **Total area cover (%)** | 5.87 | 4.37 | 43.09 | 14.95 | 31.71 |

The reliability scores are highly variable between habitat classes, with the lowest reliability scores reflecting the issues identified through visual assessments of multiple assessors. For example, the classes with the greatest proportion of very low reliability scores are bare sand (42.49% total bare sand area mapped), bare ground (41.16% total bare ground mapped) and coastal sand dunes (39.37% of total CSD area mapped). This reflects known issues identified through difficulty surveying these habitats and collecting sufficient ground data within each BGZ, and separability with other habitat classes. Of the modelled habitats, bog and coniferous woodland have the highest proportion of area classed as very high reliability, with 72.01% of bog and 69.20% of coniferous woodland mapped with very high reliability. This reflects the model's ability to predict these habitats with confidence and correlates to assessor checks.

# 4. Known Data Issues and Limitations

**Disclaimer:**
Living England is derived from a machine learning model; thus, it represents the most likely habitat present in a particular area based on the input variables. As with any modelling process, it is not 100% accurate and should be used with caution alongside other data sources when informing further environmental applications or land management decision making. With the 2022-23 map, a reliability metric has been provided (see Section 3.7 Reliability Metric) to indicate the level of confidence in the data present across the national-scale map.

For Living England 2022-23, the method has been standardised and will be used for creating future iterations of Living England, subject to technological advances and significant developments. As a result, there have been numerous changes made to the methodology compared to the previously published Living England Phase IV dataset (Kilcoyne et al., 2022). **Direct comparisons therefore cannot be drawn between the two products due to methodological differences** and will not provide a true assessment of habitat change.

Further considerations when using Living England 2022-23 include:

- Living England is an object-based classification and thus will group similar pixels together to represent real-world objects. This approach may have some resultant over- or under-segmenting of features on the ground, as well as generalises segments to a single habitat class where habitats in reality are often mixed mosaics or in transition. A new attribute has been included titled 'Mixed_Seg' to flag where a segment is more likely to be impure with a mix of broad habitat classes present.
- Living England is based on the UKBAP broad habitat classification but modified for the habitats detectable using an Earth observation approach. This collates priority habitats into broad classes as described in the habitat descriptions (Appendix 1).Some habitat sub-classes within a broad class may differ spectrally and in their characteristic relationships to others, and therefore lead to confusion when mapped through the modelling approach. Where it is not possible to reliably and robustly classify to UKBAP broad level these have been aggregated, as is the case with unimproved grassland combining acid, calcareous and neutral grasslands, and bare ground inclusive of inland rock and bare soil/silt/peat. These will be targeted by the Living England team for future development and ground data collection.

- As the Living England approach is based on satellite-derived data, this is restricted by what is visible from above. Therefore, overhanging vegetation will restrict the view of the habitats below and cause some underestimation, for example in riparian habitats.
- Living England is informed by ground data collected through national field surveys and repurposing of other national ground survey data. This covers a range of capture dates and therefore will not necessarily represent data captured during our reporting window; however, this has gone through our quality assurance processes outlined in Section 3.1.3 Field data quality assurance

Through quality assurance of the Living England 2022-23 dataset, several known habitat mapping issues have been identified. Many of these issues are identified within segments with a reported low or very low reliability score. These include:

- Confusion between improved & semi-improved grassland and unimproved grassland due to the similar spectral responses and appearance in the satellite imagery. This is an area where the team will continually improve and target field data collection to further refine in future.
- Hedgerow features are not consistently detected through the segmentation model due to the spatial resolution of the imagery used in Living England; however, due to refinement of this process some large hedgerow objects are being detected. Hedgerow segments are included as a separate class in the random forest model and later merged to the scrub class due to the uncertainty associated with this class definition. As open data becomes available through the Defra group for hedgerows, this will be further refined.
- Scrub habitats are noting as overmapping across particularly arable fields where these have been missed in the bespoke mapping methodology. This is particularly visible in the Northwest of England in BGZ01 around Carlisle, Silloth and Egremont.
- Urban fringes and small urban segments within developed areas are sometimes missed where these were missing in the OSMM dataset. Similarly due to the spatial resolution of the data and visibility in the satellite imagery, road networks are often not fully connected due to overhanging vegetation along verges and segment size limitations.
- Specific features where land is heavily modified such as golf courses, caravan parks and allotments often are misclassified and flagged as a mixed segment. Similarly, polytunnel features are not all detected as arable habitats.
- There is some misclassification of coastal sand dunes particularly with higher successional dune segments, for example in BGZ03. Some further refinement has looked to correct this; however, there is still some confusion particularly with scrub and improved grassland habitats with similar appearances. Bare ground is a highly generalised class and is currently undermapping, further work will investigate refining the approach used and the ground data feeding into this class in future.
- Arable habitat mapping in some locations demonstrate some misclassification, with some inherent inaccuracy brought in from the use of third-party data and transferred to the Living England spatial framework. As a result, there is some misclassification where arable fields and hedgerows are underrepresented. This is particularly prominent in BGZ13 where these habitats are misclassified as Fen, Marsh and Swamp habitats. The Living England team are working closely with the RPA to improve this for future iterations and improve the accuracy of arable habitat capture.

# 5. Use cases

Living England should be carefully considered if it is the most suitable data for individual use cases. Generally, Living England can be used for:

- Broad-scale habitat (national/regional) extent metrics,
- National habitat extent and connectivity assessments for targeting nature recovery,
- Environmental policy decision making,
- Assessment of large-scale natural capital assets,
- Ecosystem service modelling,
- Updating the evidence base for key policy areas such as ELM.

Living England should not be used for:

- Living England is a broad habitat map rather than a land cover or land use map and therefore should only be used to determine extent and presence of broad habitats and not to inform on presence/absence of specific species.
- Change-detection or comparisons of habitat change between Phase IV and LE 2022-23 due to methodological differences. Separate outputs are planned to publish change detection using the LE workflow in the future taking the various methodological implications into account.
- Living England is a predictive map. Note that accuracy is variable across habitats and biogeographic zones.
- Living England is a broad habitat map with a minimum mapping unit of 300 $m^2$ and should not be used to assess higher level details within segments.
- Condition assessments – Living England is not designed to inform on the condition of the mapped habitats and is only designed to map extent and presence/absence.

Since the release of LE Phase IV in March 2022, this has been used to inform:

- 25 Year Environment Plan D1 Indicator: Quantity, Quality and Connectivity of Habitats
- Environment Land Management schemes (ELMs) Data science and modelling e.g. Environmental value modelling
- NCEA projects including integration for analysis and validation with England Peat Map, Green Infrastructure, Priority Habitat Inventory, Local Nature Recovery Network portal etc.
- Commercial organisations assessing Biodiversity Net Gain and Net Zero.

Existing use cases should look to use the updated dataset in relation to the reliability metric to assess validity of existing analyses and to understand the caveats and limitations.

# 6. Contact

For any queries or further technical detail, please contact the Living England team at livingenglandenquiries@naturalengland.org.uk.

The code used to create Living England will be publicly available on the Natural England Github pages. The Living England code repository comprises of R, Python and Google Earth Engine scripts that have been used to create this product and are subject to licence conditions.

# Equations

| Band Retio | Equation | Reference |
|---|---|---|
| **Topographic Wetness Index (TWI)** | *Ln ((FlowAcc + 1)/ Tan (SLOPE * (π/180)))*<br>*where FlowAcc is the Flow Accumulation* | ESRI ArcGIS Pro 3.2, 2023 |
| **Normalised Difference Vegetation Index (NDVI)** | $NDVI = \frac{NIR - RED}{NIR + RED}$ | The IDB project, 2024 |
| **Normalised Bare Soil Index (BSI)** | $BSI = \frac{(SWIR2 + RED) - (NR + BLUE)}{(SWIR2 + RED) + (NIR + BLUE)}$ | Nguyen et al. 2021. |
| **Normalised Difference Water Index (NDWI)** | $NDWI = \frac{Green - NIR}{Green + NIR}$ | Gao et al, 1996 |
| **Normalised Difference Infrared Index (NDII)** | $NDII = \frac{SWIR2 - SWIR1}{SWIR2 + SWIR1}$ | Cunningham and Cutts, 2023. |
| **Accuracy Metric** | **Equation** | **Reference** |
| **Accuracy** | $\frac{TP + TN}{TP + TN + FP + FN}$ | van Rijsbergen, 1979 |
| **Precision** | $\frac{TP}{TP + FP}$ | van Rijsbergen, 1979 |

| Accuracy Metric | Equation | Reference |
|---|---|---|
| **Recall/Sensitivity** | $\frac{TP}{TP + FN}$ | van Rijsbergen, 1979 |
| **Specificity** | $\frac{TN}{TN + FP}$ | van Rijsbergen, 1979 |
| **Balanced accuracy** | $\frac{sensitivity + specificity}{2}$ | van Rijsbergen, 1979 |
| **F1 Score** | $\frac{2 \cdot precision \cdot recall}{precision + recall}$ | van Rijsbergen, 1979 |
| **In Multiclass assessment, performance is evaluated in a one-vs-all fashion**<br>**TP – True positives (model correctly positive)**<br>**TN – True negative (model correctly predicts negative)**<br>**FN – False negatives (model incorrectly negative)**<br>**FP – False positives (model incorrectly positive)** | | |

# References


Barber, R. and Robinson, P. 2023. Scoping Object-based Change Detection for Living England. NECR460, Natural England

Biogeographic Zones Living England (2019). Available from: [Biogeographic Zones Living England (2021) - data.gov.uk]

Bluesky International Ltd. 2023. Aerial Photography GB (APGB). Available from: https://www.blueskymapshop.com/products/aerial-photography .

Bluesky International Ltd. and Getmapping PLC., 2022. Aerial Photography and Height Data. Available from: https://apgb.blueskymapshop.com/

Braaten, J. 2024. Sentinel-2 Cloud Masking with s2cloudless. Available from: https://developers.google.com/earth-engine/tutorials/community/sentinel-2-s2cloudless

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., and Böhner, J. 2015: System for Automated Geoscientific Analyses (SAGA) v. 2.1.4, Geosci. Model Dev., 8, 1991-2007, doi:10.5194/gmd-8-1991-2015.

Cunningham, A. and Cutts, A. 2023. NCEA1.1_Saltmarshes: Mapping expansion and degradation of Saltmarsh. Final report. Version 2. JBA Consulting Prepared for Natural England. (unpublished).

Darst, B.F., Malecki, K.C. and Engelman, C.D. 2018. Using recursive feature elimination in random forest to account for correlated variables in high dimensional data. BMC Genet 19 (Suppl 1), 65. https://doi.org/10.1186/s12863-018-0633-8

Defra. 2024. Defra Data Services Platform. Available from: https://environment.data.gov.uk/.

Department for Energy Security and Net Zero. 2024. Renewable Energy Planning Database (REPD). Available from: https://www.data.gov.uk/dataset/a5b0ed13-c960-49ce-b1f6-3a6bbe0db1b7/renewable-energy-planning-database-repd .

Environment Agency 2023a. National LIDAR Programme. Available from: https://www.data.gov.uk/dataset/f0db0249-f17b-4036-9e65-309148c97ce4/national-lidar-programme Available from: https://environment.data.gov.uk/survey

Environment Agency. 2023b. Saltmarsh Extent & Zonation. Available from: https://www.data.gov.uk/dataset/0e9982d3-1fef-47de-9af0-4b1398330d88/saltmarsh-extent-zonation]

Environment Agency. 2023b. Coastal Design Sea Levels – Coastal Flood Boundary Extreme Sea Levels. Available from: https://www.data.gov.uk/dataset/73834283-7dc4-488a-9583-a920072d9a9d/coastal-design-sea-levels-coastal-flood-boundary-extreme-sea-levels-2018

Environment Agency. 2022. LIDAR Sand Dunes 2022 dataset. Correspondence colleagues at the EA.

Environment Agency. 2018. Coastal Dune Geomatics Mapping Ground Truthing dataset. Correspondence colleagues at the EA.

European Space Agency . 2023. Copernicus Sentinel data 2022-23' for Sentinel data

European Space Agency, 2022. Science Toolbox Exploitation Platform, SNAP 9.0. Available from: https://step.esa.int/main/

Esri. 2022. ArcGIS Pro version 3.0.1. Available from:https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview

European Space Agency . 2012. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services (ESA), SP-1322/2 (March 2012) SP-1322/2 Sentinel-2 (esa.int). Available from: https://sentinel.esa.int/documents/247904/349490/S2_SP-1322_2.pdf

ESRI 2024 ArcGIS Field Maps. Available from: https://www.esri.com/en-us/arcgis/products/arcgis-field-maps/resources .

Forestry Commission. 2023. National Forest Inventory Woodland England 2022. Available from: National Forest Inventory England 2022 (data.gov.uk).

Gao, B.C., 1996. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. Remote sensing of environment, 58(3), pp.257-266.

GDAL/OGR contributors (2024). GDAL/OGR Geospatial Data Abstraction software Library. Open Source Geospatial Foundation. URL https://gdal.org DOI: 10.5281/zenodo.5884351

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R. 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment.

H2O.ai. 2022. h2o: R Interface for H2O. R package version 3.42.0.2. https://github.com/h2oai/h2o-3.

Hastie, T., Tibshirani, R., Friedman, J.H. and Friedman, J.H., 2009. The elements of statistical learning: data mining, inference, and prediction (Vol. 2, pp. 1-758). New York: springer.

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horanyi, A., Munoz Sabater J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., Thepaut, J-N. 2023: ERA5 hourly data on single levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS).

Holmes, K., Stefaniak, A., Trippier, B., Potter, S., Woodget, A., Fancourt, M., Hadfield, B., Saunders, A., Mein, R., Bowling, J. 2024. Living England Habitat Classification Dashboard for Surveyors: Version 2.0. Natural England. Unpublished.

Kilcoyne, A.M., Clement, M., Moore, C., Picton Phillipps, G.P., Keane, R., Woodget, A., Potter, S., Stefaniak, A. and Trippier, B. 2022. Living England: Technical User Guide. NERR108. Natural England

Kilcoyne, A., Alexander, R., Cox, P. and Brownett, J. 2017. Living Maps: Satellite-based Habitat Classification. Evidence Project SD1705. Available online at: http://randd.defra.gov.uk/Document.aspx?Document=14198_SD1705_FinalReport.pdf

Liu, D., and Xia, F., 2010. Assessing object-based classification: advantages and limitations. Remote Sens. Lett. 1, 187–194. https://doi.org/10.1080/01431161003743173

JNCC 2011. UK Biodiversity Action Plan Priority Habitat Descriptions. Available from: https://jncc.gov.uk/our-work/uk-bap-priority-habitats/.

JNCC 2008. Spreadsheet of Habitat Correspondences. Available from: https://hub.jncc.gov.uk/assets/9e70531b-5467-4136-88f6-3b3dd905b56d#Habitat-correspondences-2008.xls.

MAGIC (Defra). 2024. MAGIC Map. Available from: https://magic.defra.gov.uk/.

Met Office, Hollis, D., McCarthy, M.P., Kendon, M., Legg, T., and Simpson, I. 2018. HadUK-Grid gridded and regional average climate observations for the UK. Centre for Environmental Data Analysis. http://catalogue.ceda.ac.uk/uuid/4dc8450d889a491ebb20e724debe2dfb

Mountford, E.P. and Strachan, I.M. 2022. JNCC UK Habitat Correspondence Tables. Excel spreadsheet: JNCC Habitat Correspondence Tables v8.2.xlsx. JNCC, Peterborough. (unpublished).

NASA Shuttle Radar Topography Mission (SRTM). 2013. Shuttle Radar Topography Mission (SRTM) Global. Distributed by OpenTopography. https://doi.org/10.5069/G9445JDF.

National Plant Monitoring Scheme. 2023. National Plant Monitoring Scheme survey data (2015-2022). NERC EDS Environmental Information Data Centre. https://doi.org/10.5285/f7ef2dc5-2bce-4436-8f65-90f7a99acff2

Natural England. 2024. Natural England Open Data Portal. Available from: https://naturalengland[1]defra.opendata.arcgis.com/.

Natural England. 2023. Provisional Agricultural Land Classification (ALC). Available from: https://www.data.gov.uk/dataset/952421ec-da63-4569-817d-4d6399df40a1/provisional-agricultural-land-classification-alc

Natural England, 2021. National Character Areas (England). Available from: https://naturalengland-defra.opendata.arcgis.com/datasets/9185e7efe65f4e47b4a722446c061e62_0/about

Nguyen, C.T., Chidthaisong, A., Diem,. P.K. and Huo, L.Z. 2021. A Modified Bare Soil Index to Identify Bare Land Features during Agricultural Fallow-Period in Southeast Asia Using Landsat 8. Land. 10, 231. https://doi.org/10.3390/land10030231.

Open Street Maps (2024). Available from: OpenStreetMap

Ordnance Survey. 2023a. Ordnance Survey Boundary Line. Available from: https://osdatahub.os.uk/downloads/open/BoundaryLine.

Ordnance Survey. 2023b. OS MasterMap Topography Layer. Available from: https://www.ordnancesurvey.co.uk/products/os-mastermap-topography-layer

Ordnance Survey. 2023c. OS Open Built Up Areas. Available from: https://www.ordnancesurvey.co.uk/products/os-open-built-up-areas#overview

Ordnance Survey. 2023d. OS VectorMap District. Available from: OS VectorMap District | Data Products | OS (ordnancesurvey.co.uk)

Padgham, M., Rudis, B., Lovelace, R., Salmon, M., Maspons, J., Smith, A., Smith, J., Gilardi, A., Spinielli, E., North, A., Machyna, M., Kalicinski, M., Pousson, E. 2023. 'Osmdata' R package, version 0.2.5. Available from: https://cran.r-project.org/web/packages/osmdata/index.html

Palmisano, D., Satalino, G., Balenzano, A., Bovenga, F., Mattia, F., Rinaldi, M., Ruggieri, S., Skriver, H., Davidson, M.W., Cartus, O. and Wegmuller, U., 2019. Sensitivity of Sentinel-1 interferometric coherence to crop structure and soil moisture. *In IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium (pp. 6219-6222). IEEE.*

Python. 2018. Python version 3.7. Available from: https://www.python.org/

R Core Team 2023. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Remote Sensing Phenology. 2018. NDVI, the Foundation for Remote Sensing Phenology. U.S. Geological Survey. November 27, 2018. Available from: https://www.usgs.gov/special-topics/remote-sensing-phenology/science/ndvi-foundation-remote-sensing-phenology#overview [Accessed 18/03/2024].

Rural Payments Agency. 2023. Crop Map of England 2022. Available from: https://environment.data.gov.uk/dataset/cc389fe9-f026-4b20-a80f-f424ee833ea6

Smith, G.M., and Morton, R.D., 2010. Real World Objects in GEOBIA through the Exploitation of Existing Digital Cartography and Image Segmentation.]

Souza, W.D.O., Reis, L.G.D.M., Ruiz-Armenteros, A.M., Veleda, D., Ribeiro Neto, A., Fragoso Jr, C.R., Cabral, J.J.D.S.P. and Montenegro, S.M.G.L., 2022. Analysis of environmental and atmospheric influences in the use of sar and optical imagery from sentinel-1, landsat-8, and sentinel-2 in the operational monitoring of reservoir water level. *Remote Sensing*, 14(9), p.2218.

Tewkesbury, A.P., Comber, A.J., Tate, N.J., Lamb, A., and Fisher, P.F., 2015. A critical synthesis of remotely sensed optical image change detection techniques. Remote Sens. Environ. 160, 1–14. https://doi.org/10.1016/j.rse.2015.01.006

The IDB Project. 2024. Index Database: a database for remote sensing indices. Available from: https://www.indexdatabase.de/

Trimble Geospatial 2023. eCognition (version 10.3). Available from: https://geospatial.trimble.com/products-and-solutions/trimble-ecognition.

van Rijsbergen C.J. 1979 Information Retrieval. Butterworths, London, UK

Whiteside, T.G., Boggs, G.S. and Maier, S.W. 2011. Comparing object-based and pixel-based classifications for mapping savannas. International Journal of Applied Earth Observation and Geoinformation. 13 (6), 884-893. https://doi.org/10.1016/j.jag.2011.06.008

Zwieback, S., Hensley, S. and Hajnsek, I., 2015. Assessment of soil moisture effects on L-band radar interferometry. *Remote Sensing of Environment*, 164, pp.77-89.

# Appendix 1 – Habitat description tables

The tables below describe what is included and excluded in each of the LE broad habitat classes and how these translate to UKBAP broad and priority habitats. These have been compiled specifically for the purpose of assessing habitats with Earth observation data and how data is collected on the ground, using knowledge from our ecological expertise and existing habitat classification sources (JNCC 2011). Please contact the Living England team if you have any questions. Note that some cells have been left deliberately blank.

| LE Broad Habitat | UKBAP Broad Habitat | Living England Description | Inclusions | Exclusions | LE Sub-class (inc. UKBAP Priority Habitats – PH) |
|---|---|---|---|---|---|
| **Arable and horticultural** | Arable and horticultural | This broad habitat type covers arable cropland (including perennial, woody crops, managed, commercial orchard), commercial horticultural land (such as nurseries and vegetable plots and commercial flower growing areas), freshly ploughed land, annual leys and rotational set-aside. | Cereal field margins<br>Un-seeded fallow | Field boundaries, domestic gardens, and allotments, fallow land, short rotation coppice and cover crops. | Arable field margins – PH<br>Arable<br>Horticultural |
| Bare ground (Includes **'Inland rock'** and **'Bare soil / silt / peat'**) | Inland rock<br>Supralittoral rock<br>Supralittoral sediment | Both natural and artificial exposed rock surfaces which are greater than 0.25 ha, such as inland cliffs, caves and screes and limestone pavements, as well as various forms of excavations and waste tips such as quarries and quarry waste. Maritime cliffs and slopes comprise sloping to vertical faces on the coastline where a break in slope is formed by | Calaminarian grassland<br>Hard cliffs, soft cliffs, ledges, steep rocky slopes, boulder fields.<br>Maritime grassland. | Cliff faces.<br>Open mosaic habitats.<br>Scrub on shingle.<br>Heavily vegetated (> 60% vegetation cover) coastal vegetated shingle. | Calaminarian grasslands – PH<br>Calaminarian grasslands (heavily vegetated but < 60% cover)<br>Inland rock outcrop and scree habitats - PH<br>Limestone pavement - PH |

| LE Broad Habitat | UKBAP Broad Habitat | Living England Description | Inclusions | Exclusions | LE Sub-class (inc. UKBAP Priority Habitats – PH) |
|---|---|---|---|---|---|
| | | slippage and/or coastal erosion. A cliff-top which extends landward at, at least the limit of maritime influence, which in some exposed situations may continue for up to 500 m inland.<br>Shingle is defined as sediment with particle sizes in the range 2 - 200mm. Shingle structures take the form either of spits, barriers or barrier islands formed by longshore drift, or of cuspate forelands where a series of parallel ridges piles up against the coastline<br>Expanses of bare soil, silt, mud or peat | Coastal shingle with pioneer species (e.g. sea kale, sea pea, sea beet, sea campion). More stable and mixed communities. | Freshly ploughed arable land and un-seeded fallow.<br>Bare sand. | Maritime cliffs and slopes – PH<br>Maritime cliffs and slopes (heavily vegetated but < 60% cover)<br>Coastal vegetated shingle – PH<br>Coastal vegetated shingle (heavily vegetated but < 60% cover) |
| **Bare sand** | N/A | Sand is defined as sediment with particle sizes in the range 0.06 to 2 mm. | | Coastal sand dunes | N/A |
| **Bog** | Bog | 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. | Raised bog and blanket bog, all intermediates in between.<br>Modified bog, including impoverished vegetation dominated by purple moor grass *Molinia caerulea* or | Lowland areas with predominantly acid substrata including valleys and basin mires that receive acid surface seepage, giving rise to vegetation similar to that of bogs.<br>Bog covered with dwarf shrub heath vegetation. | Lowland raised bog – PH<br>Lowland raised bog (rush dominated)<br>Lowland raised bog (purple moor grass dominated)<br>Blanket bog – PH<br>Blanket bog (rush dominated)<br>Blanket bog (purple moor grass dominated) |

| LE Broad Habitat | UKBAP Broad Habitat | Living England Description | Inclusions | Exclusions | LE Sub-class (inc. UKBAP Priority Habitats – PH) |
|---|---|---|---|---|---|
| | | | hare's-tail cotton grass *Eriophorum vaginatum* | | |
| **Bracken** | N/A | Areas dominated by a continuous canopy cover of bracken *Pteridium aquilinum* at the height of the growing season. | Areas covered by Bracken | Areas of bracken under forest or woodland canopy which are included in the 'Broadleaved, mixed and yew woodland' or the 'Coniferous woodland' broad habitat types.<br>Scattered patches of bracken. | Bracken |
| **Broadleaved, mixed and yew woodland** | Broadleaved, mixed and yew woodland | Vegetation dominated by trees that are more than 5 m high when mature, which form a distinct, although sometimes open canopy with a canopy cover of greater than 20 %. It includes stands of both native and non-native broadleaved tree species and yew *Taxus baccata,* where the percentage cover of these trees in the stand exceeds 20 % of the total cover of the trees present. | | Woodlands that are dominated by conifer trees with less than 20% cover provided by broadleaved or yew trees. Recently felled broadleaved, mixed and yew woodland. New planting. | Lowland mixed deciduous woodland - PH<br>Upland oakwood - PH<br>Traditional orchards - PH<br>Wet woodland - PH<br>Upland birchwoods - PH<br>Wood pasture and parkland - PH |

| LE Broad Habitat | UKBAP Broad Habitat | Living England Description | Inclusions | Exclusions | LE Sub-class (inc. UKBAP Priority Habitats – PH) |
|---|---|---|---|---|---|
| **Built-up areas and gardens** | Built-up areas and gardens | Urban and rural settlements, farm buildings and other man-made built structures. | Industrial estates, retail parks. It also includes domestic gardens, airport runways, urban transport infrastructure, docklands, waste and derelict ground, open mosaic habitats on previously developed land, allotments. | Amenity grassland, urban parkland, solar panels, polytunnels. | Built-up areas and gardens<br>Open mosaic habitats on previously developed land - PH |
| **Coastal Saltmarsh** | Littoral sediment | The upper, vegetated portions of intertidal mudflats, lying approximately between mean high water neap tides and mean high water spring tides. Saltmarsh vegetation consists of a limited number of halophytic (salt tolerant) species adapted to regular immersion by the tides. | Species-poor low-mid marsh, and the more diverse communities of the mid-upper marsh | Transitions to freshwater or grassland - doesn't go beyond where upper-most zone is populated by terrestrial plants. Some marsh types not included where created at recently breached managed realignment sites. | Coastal saltmarsh - PH |

| LE Broad Habitat | UKBAP Broad Habitat | Living England Description | Inclusions | Exclusions | LE Sub-class (inc. UKBAP Priority Habitats – PH) |
|---|---|---|---|---|---|
| **Coastal Sand Dunes** | Supralittoral sediment | Coastal sand dunes develop where there is an adequate supply of sand (sediment within the size range 0.2 - 2.0 mm) in the intertidal zone and where onshore winds are prevalent. The dry sand is blown landwards and deposited above high-water mark, where it is trapped by specialised dune-building grasses which grow up through successive layers of deposited sand. | Embryonic and mobile dunes with few plant species and Marram grass *Ammophila arenaria*.<br>Semi-fixed dunes with predominantly bare sand, marram still common with increasing number of plant species.<br>Fixed dune grassland.<br>Occasional areas of Coastal vegetated shingle may be included where it sits as a mosaic habitat within the Coastal Sand Dune complex | Dunes colonised by sallows *Salix* spp., birches *Betula* spp., or alder *Alnus glutinosa*.<br>Scrub on sand dunes (e.g. sea buckthorn dominated).<br>Unvegetated coastal shingle. | Coastal sand dunes - PH |

| LE Broad Habitat | UKBAP Broad Habitat | Living England Description | Inclusions | Exclusions | LE Sub-class (inc. UKBAP Priority Habitats – PH) |
|---|---|---|---|---|---|
| **Coniferous woodland** | Coniferous woodland | Vegetation dominated by trees that are more than 5 m high when mature, which form a distinct, although sometimes open canopy with a canopy cover of greater than 20 %. It includes stands of both native and non-native coniferous trees species (with the exception of yew *Taxus baccata*) where the percentage cover of these trees in the stands exceeds 80 % of the total cover of the trees present. | | Woodlands with less than 80 % cover provided by conifer trees. Recently felled coniferous woodland. New planting. | Native pine woodlands – PH<br>Non-native coniferous plantation |
| **Dwarf shrub heath** | Dwarf shrub heath<br>Montane habitats | Characterised by vegetation that has >25 % cover of plant species from the heath family (ericoids) or dwarf gorse. It generally occurs on well drained, nutrient-poor, acid soils. Heaths do occur on more basic soils but these are more limited in extent and can be recognised by the presence of herbs characteristic of calcareous grassland. Montane habitats include 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. | Dry and wet heath types occurring in the lowlands and the uplands. Dwarf shrub dominated vegetation in which species characteristic of peat-forming vegetation such as cotton-grass and sphagna are abundant. Montane heath. Exclusively | Heath types which are exclusively alpine in distribution.<br>Heath types on sand dunes or shingle.<br>Heath types on maritime cliffs and slopes that are influenced by salt spray.<br>Calcareous grassland, fens and springs, blanket bogs and rock habitats in the montane zone. Grasslands and dwarf shrub heath that straddle the montane boundary. | Lowland heathland (dry) – PH<br>Lowland heathland (wet) – PH<br>Lowland heathland (dominating bog)<br>Upland heathland (dry) – PH<br>Upland heathland (wet) – PH<br>Upland heathland (dominating bog)<br>Mountain heaths and Willow Scrub – PH |

| LE Broad Habitat | UKBAP Broad Habitat | Living England Description | Inclusions | Exclusions | LE Sub-class (inc. UKBAP Priority Habitats – PH) |
|---|---|---|---|---|---|
| | | | montane habitat types can be recognised by their floristic composition and their physiognomy (prostrate vegetation). Widespread arctic-alpine species. | | |
| **Fen, marsh and swamp** | Fen, marsh and swamp | Vegetation types found on minerotrophic (groundwater fed), permanently, seasonally, or periodically waterlogged peat, peaty soils, or mineral soils.<br>-Fens are peatlands which receive water and nutrients from groundwater and surface run-off, as well as from rainfall.<br>-Flushes are associated with lateral water movement, and springs with localised upwelling of water.<br>-Marsh is a general term usually used to imply waterlogged soil; it is most specifically here to refer to fen meadows and rush-pasture communities on mineral soils and shallow peats. | Reedbeds (i.e., swamps dominated by stands of common reed *Phragmites australis*). | Neutral and improved grasslands on floodplains and grazing marshes.<br>Flushes.<br>Ombrotrophic mires (blanket, raised and intermediate bogs).<br>More than 30 % cover areas of carr (fen woodland dominated by species such as willow *Salix* spp., alder *Alnus glutinosa* or birch *Betula* spp.) | Upland flushes, fens, and swamps – PH<br>Upland flushes, fens and swamps (rush dominated)<br>Lowland fens – PH<br>Lowland fens (rush dominated)<br>Reedbeds - PH<br>Purple moor grass and rush pastures - PH |

| LE Broad Habitat | UKBAP Broad Habitat | Living England Description | Inclusions | Exclusions | LE Sub-class (inc. UKBAP Priority Habitats – PH) |
|---|---|---|---|---|---|
| | | -Swamps are characterised by tall emergent vegetation. | | | |
| **Improved and semi-improved grassland** | Improved grassland | Characterised by vegetation dominated by a few fast-growing grasses on fertile, neutral soils. It is frequently characterised by an abundance of rye-grass and white clover. Improved grasslands are typically either managed as pasture of mowed regularly for silage production or in non-agricultural contexts for recreation and amenity purposes; they are often periodically resown and are maintained by fertiliser treatment and weed control.<br>Semi-improved grasslands, which have undergone some historic artificial improvement (e.g. by fertilisers), but may be more biodiverse / species-rich, are also included here. | Sown grasslands as part of an arable crop if more than one year old.<br>Fallow that has been seeded or otherwise developed a grassland sward. | Sown grasslands less than one year old.<br>Fixed grassland on sand dunes. | Improved grassland<br>Improved grassland (rush dominated)<br>Semi-improved grassland<br>Semi-improved grassland (rush dominated)<br>Species-rich semi-improved grassland<br>Species-rich semi-improved grassland (rush dominated)<br>Species-poor semi-improved grassland<br>Species-poor semi-improved grassland (rush dominated)<br>Coastal and floodplain grazing marsh – PH<br>Coastal and floodplain grazing marsh (rush dominated) |
| **Scrub** | N/A | Dense scrub – Patches of shrubs less than 5 metres tall with continuous (> 90%) cover. May be either mixed or single species scrub. Also includes | Patches with occasional trees more than 5m tall, tree species | Small hedgerow features, Coppice, Coppice with standards, recently felled | Mixed Scrub<br>Single species scrub (Alder buckthorn, Bay willow, Birch Blackthorn, Box, |

| LE Broad Habitat | UKBAP Broad Habitat | Living England Description | Inclusions | Exclusions | LE Sub-class (inc. UKBAP Priority Habitats – PH) |
|---|---|---|---|---|---|
| | | some hedgerows in this category where large enough to detect with the imagery resolution. | less than 5 m tall, coastal scrub, scattered sea buckthorn scrub on dunes. | woodland, scattered scrub where patches of scrub cover < 90%. | Bramble, Buckthorn, Dogwood, Eared willow, Elder, Goat willow, Gorse, Grey willow, Hawthorn, Hazel, Juniper, Montane, Osier, Purple willow, Rhododendron, Sea buckthorn, Spindle, Wayfaring tree) |
| **Solar Farms** | N/A | Ground-mounted collections of solar photovoltaic panels, or solar farms, greater than 300m$^2$ in total area. These may be situated on any underlying LE broad habitat except Built-up areas and gardens. | Collections of solar photovoltaic panels over-lying other LE broad habitats such as Improved and Semi-improved Grassland, or Arable and Horticultural. | Solar farms situated within areas designated as Built-up areas and gardens.<br>Very small numbers of solar photovoltaic panels (less than 300m$^2$ in total area) that may be associated with powering individual buildings.<br>Roof-mounted solar panels. | Solar Farms |
| **Unimproved grassland** | Acid grassland<br>Calcareous grassland | Unimproved (normally species-rich) grassland occurring on acid soil, circumneutral soil and shallow lime-rich soils normally underlain by chalk or limestone rocks. It includes a variety of | | Fixed grasslands on sand dunes. | Lowland dry acid grassland - PH<br>Lowland calcareous grassland – PH |

| LE Broad Habitat | UKBAP Broad Habitat | Living England Description | Inclusions | Exclusions | LE Sub-class (inc. UKBAP Priority Habitats – PH) |
|---|---|---|---|---|---|
| | Neutral grassland | grasslands such as pioneer annual-rich calcifuge communities on dry sandy soils, wet acidic grasslands typified by species such as heath rush, hay meadows and pastures, a range of grasslands which are inundated with water periodically and permanently moist or even waterlogged grassland where the vegetation is dominated by grasses. The grassland may be enclosed and managed, or tall and unmanaged.<br>At least two of the following are likely to apply:<br>-cover of rye-grasses and white clover is less than 10 %<br>- the sward is species-rich (more than 15 vascular plant species/m2 including grasses)<br>- there is a high cover of wildflowers and sedges (at least 30 %) excluding white clover, creeping buttercup and injurious weeds. | | | Lowland calcareous grassland (*Brachypodium pinnatum* dominated)<br>Lowland meadows – PH<br>Lowland meadows (rush dominated)<br>Upland calcareous grassland -PH<br>Upland calcareous grassland (*Brachypodium pinnatum* dominated)<br>Upland hay meadows – PH<br>Upland hay meadows (rush dominated)<br>Coastal and floodplain grazing marsh – PH<br>Coastal and floodplain grazing marsh (rush dominated)<br>Upland acid grassland / grassland moorland and rough grazing |
| **Water** | Rivers and streams | Natural and near-natural running waters. Natural systems such as lakes, meres, pools and saline lagoons, as well as | Open water zones with open water for at least the majority of | Marginal emergent vegetation that is greater than 5m wide. | Rivers - PH<br>Oligotrophic and dystrophic lakes - PH<br>Mesotrophic lakes - PH |

| LE Broad Habitat | UKBAP Broad Habitat | Living England Description | Inclusions | Exclusions | LE Sub-class (inc. UKBAP Priority Habitats – PH) |
|---|---|---|---|---|---|
| | Standing open water and canals<br>Sublittoral sediment | man-made waters such as reservoirs and gravel pits. | the year.<br>Standing waters including marl lakes and brackish water lakes. | Areas of wetland habitat adjacent to the waterbody that are greater than 0.25 ha.<br>Areas of wet woodland greater than 0.25 ha unless canopy cover is less than 30 %.<br>Unimproved floodplain grasslands, marshy grassland, wet heath, fens, bogs, flushes and swamps.<br>Flood waters/extreme events.<br>Small canals, ditches and ponds. | Eutrophic standing waters - PH<br>Aquifer fed naturally fluctuating water bodies - PH<br>Saline lagoons - PH<br>Ponds – PH<br>Canals |

# Appendix 2 – Data Sources for Living England 2022-23

| Dataset | Data Owner | Description | Spatial Scale | Temporal Scale | Licence | Reference | How is this data used |
|---|---|---|---|---|---|---|---|
| **Sentinel-2 Multispectral Imagery** | European Space Agency (ESA) | Optical imagery from Sentinel-2 satellites capturing 13 spectral bands. | 10 m pixels | 1st August 2022 – 30th April 2023 | Open Access | Google Earth Engine | Used to produce cloud-masked multitemporal mosaics used throughout the LE workflow, including to derive the object framework and to inform the habitat classification model. Where bands are captured at 20 m, these have been resampled to 10 m. |
| **Sentinel-1 Backscatter Imagery** | European Space Agency (ESA) | Radar imagery from Sentinel-1 satellite mission, capturing C-band synthetic aperture radar images. | 10 m pixels | 1st August 2022 – 30th April 2023 | Open Access | Google Earth Engine | Used to produce multitemporal mosaics. Current revisit time of 12 days due to failure of the Sentinel 1B system. |
| **Sentinel-1 Single Look Complex (SLC) Imagery** | European Space Agency (ESA) | SLC images containing amplitude and phase information. | 20 m pixels | 1st August 2022 – 30th April 2023 | Open Access | Copernicus Data Space Ecosystem | Processed to calculate an average seasonal coherence maps. |
| **EA LiDAR DSM & DTM** | Environment Agency (EA) | EA's National LIDAR programme capturing accurate elevation data of the height of the terrain and surface objects on the ground, at 1m spatial resolution. | 1 m pixels | 2017-20232 | OGL | National LIDAR Programme | Used to derive slope, aspect, Canopy Height Models (CHM), Focal statistics (texture proxy) and Topographic Wetness Index (TWI). These layers were resampled to 10 m and were used in both the segmentation and classification models. Focal statistics at 1 m resolution were also included to inform the classification model to provide greater definition of variability within segments. The LiDAR DTM was also used in the |

| Dataset | Data Owner | Description | Spatial Scale | Temporal Scale | Licence | Reference | How is this data used |
|---|---|---|---|---|---|---|---|
| | | | | | | | saltmarsh workflow at 2 m and 10 m resolutions. |
| **Met Office HadUK Gridded Climate Product** | Met Office | Collection of gridded climate variables interpolated from Met Office UK land surface observations. | 1 km pixels | 2 year – 2021-22, 20 year – 2003 - 22 | OGL | Metoffice.gov.uk | Variables used: temperature (min, mean, max), and rainfall variable, with annual averages over 2- and 20-year time periods as well as 2-year seasonal averages (DJF, MAM, JJA, SON) |
| **Geology** | British Geological Society (BGS) | Generalised digital geological map data from BGS. | 1:50,000 | 2016 | Licence agreement for use | BGS Geology 50k | Used to inform the habitat classification model |
| **Soils – BGS Soil Parent Material** | British Geological Society (BGS) | National map of the parent material from which a soil has formed. | 1km pixels | 2016 | OGL | Soil Parent Material Model | Used to inform the habitat classification model. Selected variables of: Carbonate content, grain size, soil group, soil texture, soil thickness. |
| **Cranfield NATMAP** | Cranfield | National map of soilscapes describing the soil types across England. | 1:250,000 | 2005 | Licence agreement for use | NATMAP Soilscapes | Used to inform the habitat classification model. Selected variables of: SoilScape class and Soil Wetness. |
| **OS Mastermap Topographic Layer (OSMM)** | Ordnance Survey (OS) | Detailed object layer identifying features in the landscape topography, such as roads, buildings, parks. | 1:1250 to 1:10,000 | March 2023 | PSGA Member licence | OS OpenData Hub | Used to derive the 'Built-up areas and gardens' and 'Water' classes, features such as quarries within the 'Bare ground' class, and to derive proximity layers to inform the classification model. |

| Dataset | Data Owner | Description | Spatial Scale | Temporal Scale | Licence | Reference | How is this data used |
|---|---|---|---|---|---|---|---|
| **OS Open Built-up Areas** | Ordnance Survey (OS) | Vector dataset representing the built-up areas of GB. | 200,000 m² | December 2022 | OGL | OS Data Hub | Used to create the generalised urban proximity layer. |
| **OS Boundaries dataset** | Ordnance Survey (OS) | Vector layer of authoritative and national boundaries. | N/A | 2023 | OGL | Boundary-Line™ | Used to obtain Mean High Water line and England boundary to inform land-sea cut off of mapping. |
| **OS VectorMap District** | Ordnance Survey (OS) | District level vector mapping data. | Polygon layer | 2023 | OGL | OS VectorMap® District | The Tidal Water and Foreshore datasets were used in the JBA saltmarsh workflow. |
| **Saltmarsh Extent and Zonation** | Environment Agency (EA) | Dataset of the extent of saltmarsh for England, mapping compiled from aerial photography and field surveys. | 20 cm pixel aerial imagery | 2016 - 2023 | OGL | data.gov.uk | Used to inform the 'Coastal Saltmarsh' LE class. |
| **Crop Map of England (CROME)** | Rural Payments Agency (RPA) | A vector dataset containing the main crop types of England, produced using a supervised classification from the RPA. | 0.41 ha hexagonal grid | 2022 | OGL | data.gov.uk | Used to inform the 'Arable and Horticulture' LE class. |
| **NASA SRT (Shuttle Radar Topography Mission) M30** | National Aeronautics and Space Administration, National Geospatial Intelligence Agency, Jet Propulsion Laboratory (NASA-NGA-JPL) | Interferometric radar data used to create a near-global set of land elevations. Further info: USGS Website | 30 m pixels | 2000 | Public domain | Earth Explorer | Used in the coherence workflow to supplement LIDAR DTM extent for areas immediately bordering England in Scotland and Wales (for areas where S1 tracks overlap borders and to avoid edge effects). Data were resampled to 10 m spatial resolution. |

| Dataset | Data Owner | Description | Spatial Scale | Temporal Scale | Licence | Reference | How is this data used |
|---|---|---|---|---|---|---|---|
| **EA IHM (Integrated Height Model) DTM** | Environment Agency (EA) | Digital Height data covering England. Combination of EA Aerial LiDAR and APGB height. Data natively 2 m resolution, resampled to 10 m using bilinear resampling. | 10 m pixels | 2016-2019 | Licence agreement for use | Home \| apgb | Used in the coherence workflow to supplement LIDAR DTM extent for a small area of North Yorkshire. |
| **ERA5** | Copernicus Climate Change Service at ECMWF | Atmospheric model data, including rainfall, snowfall and wind speed at hourly intervals. | Gaussian grid of N320 (about 31 km) | 2022-2023 | Copernicus Products Licence | Climate Data Store | Used to screen rainfall in S1 backscatter data acquisition. Data obtained for 6 days prior to Sentinel-1 backscatter acquisitions (26$^{th}$ July 2022 – 30$^{th}$ April 2023). |
| **EA Coastal Flood Boundary: Extreme Sea Levels** | Environment Agency (EA) | A GIS dataset and supporting information providing design / extreme sea level and typical surge information around the coastline of the UK. | Point data | 2018 | OGL, Copyright Associated British Ports 2010-2014 | Coastal Design Sea Levels - Coastal Flood Boundary Extreme Sea Levels (2018) - data.gov.uk | Used in the saltmarsh workflow. |
| **AIMS Spatial Flood Defences** | Environment Agency (EA) | The Environment Agency's (EA) Spatial Flood defences layer is the only comprehensive and up-to-date dataset in England that shows flood defences currently owned, managed or inspected by the EA | Polygon | 2020 | OGL | Defra Data Services Platform | Used in the saltmarsh workflow. |
| **National Forest Inventory** | Forest Research | National woodland map of GB covering woodland area over 0.5 hectares with minimum 20% canopy cover or potential to achieve it, with min width of 20 m. | 1:1250, 1:2500 and 1:10000 | 2022 | OGL | National Forest Inventory - Forest Research | Used to form ground dataset. |

| Dataset | Data Owner | Description | Spatial Scale | Temporal Scale | Licence | Reference | How is this data used |
|---|---|---|---|---|---|---|---|
| **Priority Habitat Inventory - B Button** | Natural England (NE) | Digitised polygons from NE surveyors submitted to the PHI project for priority habitat updates. | N/A | 2016 - 2018 | OGL | Priority Habitat Inventory | Used to form ground dataset. |
| **Coastal Dune Geomatics Mapping: Ground Truthing** | Environment Agency (EA) | Provided to the LE project by partners at the EA. | N/A | 2012 - 2018 | OGL | Unpublished | Used to form ground dataset. |
| **LIDAR Sand Dunes 2022** | Environment Agency (EA) | Provided to the LE project by partners at the EA. | N/A | 2022 | OGL | Unpublished | Used to form ground dataset. |
| **National Grassland Survey** | Natural England (NE) | Priority Grasslands Survey conducted by NE surveyors | N/A | 2017 - 2018 | OGL | Unpublished | Used to form ground dataset. |
| **Agri-Environment HLS Monitoring** | Natural England (NE) | Evaluation of HLS options using Common Standards Monitoring (CSM). Stands of Priority Habitat mapped in the field using handheld tablets. | N/A | 2019 - 2011 | Licence agreement for use | Unpublished | Used to form ground dataset. |
| **Lowland Heathland Survey** | Natural England (NE) | An evaluation of Environmental Stewardship options looking at the effectiveness of Lowland Heathland Higher Level Stewardship options | N/A | 2017 | OGL | Unpublished | Used to form ground dataset. |

| Dataset | Data Owner | Description | Spatial Scale | Temporal Scale | Licence | Reference | How is this data used |
|---|---|---|---|---|---|---|---|
| **Long Term Monitoring Network** | Natural England (NE) | Long-term monitoring of plant communities to identify change and the possible drivers of change | N/A | 2012 - 2022 | OGL | Unpublished | Used to form ground dataset. |
| **Northumberland Border Mires Survey** | Natural England (NE) | Wet Bog Quality Index (WBQI) Survey carried out by Helen Adamson or Newcastle University on behalf of Natural England. | N/A | 2017 | OGL | Unpublished | Used to form ground dataset. |
| **New Forest Mires Wetland Survey** | Natural England (NE) | Wetland survey across the New Forest Mires Wetland from 2017 | N/A | 2018 | OGL | Unpublished | Used to form ground dataset. |
| **NE Field Unit Surveys** | Natural England (NE) | Site surveys and SSSI monitoring commissioned by NE national and Area teams. | N/A | 2014 - 2019 | OGL | Unpublished | Used to form ground dataset. |
| **National Plant Monitoring Scheme** | NERC EDS Environmental Information Data Centre | Plot level plant occurrence data for the National Plant Monitoring Scheme | N/A | 2015 - 2022 | OGL | National Plant Monitoring Scheme survey data (2015-2022) | Used to form ground dataset. |
| **Space2 Eye Lens: Ainsdale NNR** | Natural England (NE) | Baseline habitat condition model pilot for monitoring change across Coastal Dunes of Ainsdale NNR & Long-Term Monitoring Network (LTMN) site using S2. | N/A | 2012 - 2019 | OGL | Unpublished | Used to form ground dataset. |
| **Space2 Eye Lens: State of the Bog collated surveys** | Natural England (NE), MMU, Carlos Benson | Selected surveys from the collated datasets in the State of the Bog project, collation of walkover plots of habitat condition. | N/A | 2012 - 2019 | OGL | Unpublished | Used to form ground dataset. |

| Dataset | Data Owner | Description | Spatial Scale | Temporal Scale | Licence | Reference | How is this data used |
|---|---|---|---|---|---|---|---|
| **West Pennines Designation NVC Survey** | Natural England (NE) | Vegetation surveys informing the boundary of the West Pennine Moors SSSI | N/A | 2012 | OGL | Unpublished | Used to form ground dataset. |
| **Uplands Inventory** | Natural England (NE) | Condition surveys of a representative national stratified random sample of PHI blanket bog upland (wet and dry) heath and upland calcareous grassland habitat polygons. | N/A | 2008 - 2010 | OGL | Priority Habitat Inventory | Used to form ground dataset. |
| **West Cumbria Mires Survey** | Natural England (NE) | Natural England Mires survey from 2014 | N/A | 2014 | OGL | Unpublished | Used to form ground dataset. |
| **Wetland Annex 1 Inventory** | Natural England (NE) | Inventories for Annex 1 wetland habitats - alkaline fens, transition mire and quaking bog, calcareous fen with Cladium, Molinia meadows (partial) and Depressions on peat (partial) | N/A | 1991 - 2014 | OGL | Unpublished | Used to form ground dataset. |
| **NE Protected Sites Monitoring** | Natural England (NE) | Data Collected across protected sites monitoring via NESS CSMi app | N/A | 2019 – 2023 | Some restrictions to licensing | Unpublished | Used to form ground dataset. |
| **England Peat Map Vegetation Surveys** | Natural England (NE) | Vegetation surveys as part of NE's England Peat Map project. | N/A | 2023 | OGL | Unpublished | Used to form ground dataset. |

| Dataset | Data Owner | Description | Spatial Scale | Temporal Scale | Licence | Reference | How is this data used |
|---|---|---|---|---|---|---|---|
| **EA coastal saltmarsh species surveys** | Environment Agency (EA) | Data on the presence and percentage cover of saltmarsh flowering plants and marine monitoring points. | N/A | 2017 | OGL | Saltmarsh Species - data.gov.uk | Visually assessed and selection used to form ground dataset. |
| **Living England desktop validation** | Natural England (NE) | Data points created by the LE team in Phase II through comparisons with S2 mosaics and ESRI world imagery. | N/A | 2021 - 2022 | OGL | Unpublished | Used to form ground dataset. |
| **Living England national field data collection programme** | Natural England (NE) | Bespoke data collection for the Living England project, collected through ESRI Field Maps. | N/A | 2021 - 2023 | OGL | Unpublished | Used to form ground dataset. |
| **Aerial Photography GB (APGB)** | Bluesky/ Getmapping Ltd | Aerial survey imagery across Britain | 12.5 cm and 25 cm pixels | Variable | APGB member Licence Agreement | Home \| apgb | Used for visual quality assessment only. |
| **Provisional Agricultural Land Classification** | Natural England | Provides a provisional agricultural land classification grade across England based on an assessment of climate, site and soil. | 1:250,000 | 2023 | OGL | Provisional Agricultural Land Classification (ALC) - data.gov.uk | Used to inform the LE arable class. For sense checking Crop Map of England data transposed onto Living England segments. |
| **Renewable Energy Planning Database (REPD)** | Department of Energy Security and Net Zero | A database tracking the UK renewable electivity projects over 150kw. | N/A | 2024 | OGL | Renewable Energy Planning Database: quarterly extract - GOV.UK (www.gov.uk) | Point locations used to inform areas for desktop assessment of solar farms and manual digitisation. |

| Dataset | Data Owner | Description | Spatial Scale | Temporal Scale | Licence | Reference | How is this data used |
|---|---|---|---|---|---|---|---|
| **Open Street Map 2024** | OpenStreetMap Foundation | A free, open geographic database updated and maintained by a community of volunteers via open collaboration. Contributors collect data from surveys, trace from aerial imagery and also import from other freely licensed geodata sources. | N/A | 2024 | Open Data Commons Open Database License | OpenStreetMap | Used to inform allotment locations. |

# Appendix 3 - Modelled metrics

This table shows the overall accuracy, balanced accuracy, and mean F1 scores for each of the k-folded datasets variants of the dataset. Each K-fold splits the overall dataset into a 20% validation set used to test the model trained on the remaining 80%. The splits are rotated systematically so that each data point has an opportunity to be used both in training and in testing. K-folding allows us to investigate the sensitivity of the model to the data that it is trained on by examining the variability in accuracy metrics across different training/validation subsets. If each of the k-folds shows highly variable performance, then it indicates that the model performance is dependent on the presence/absence of certain influential data points for training/validation, and thus is not a robust model. Looking below at the results of this we can see that the performance of the model for all 3 metrics, and across all folds is extremely similar. Similar performance across all models, despite differences in the data used to test and validate indicates that the performance of the model is not reliant on any particular data point and thus the performance can be considered robust.

| Metric | k-fold 1 | k-fold 2 | k-fold 3 | k-fold 4 | k-fold 5 |
|---|---|---|---|---|---|
| **Overall accuracy** | 0.87 | 0.87 | 0.87 | 0.87 | 0.86 |
| **Balanced accuracy** | 0.90 | 0.90 | 0.89 | 0.90 | 0.89 |
| **Mean F1 Score** | 0.79 | 0.79 | 0.78 | 0.79 | 0.77 |

# Appendix 4 - CROME arable class alignment with Living England 2022-23

This table describes the inclusion of classes from the Rural Payments Agency's Crop Map of England (CROME) to inform the arable class mapping in Living England. Some classes have not been included due to these classes not aligning to the Living England habitat classification description for Arable habitats. This includes the mapping of non-arable habitats and fallow, Ryegrass and cover crop classes, which closer align to Living England's proved grassland habitat description.

| CROME classes utilised for Living England 2022-23 'Arable' class | | CROME classes not included |
|---|---|---|
| AC01 Spring Barley | AC61 Mixed Crop-Group 4 | AC100 Italian Ryegrass |
| AC03 Beet | AC62 Mixed Crop-Group 5 | CA02 Cover Crop |
| AC04 Borage | AC63 Winter Barley | LG14 Clover |
| AC05 Buckwheat | AC64 Winter Linseed | SR01 Short Rotation Coppice |
| AC06 Canary Seed | AC65 Winter Oats | FA01 Fallow Land *[where vegetation is established]* |
| AC07 Carrot | AC66 Winter Wheat | HE02 Heathland and Bracken |
| AC09 Chicory | AC67 Winter Oilseed | PG01 Grass |
| AC10 Daffodil | AC68 Winter Rye | NA01 non-vegetated or sparsely vegetated land |
| AC14 Hemp | AC69 Winter Triticale | WA00 Water |
| AC15 Lettuce | AC70 Winter Cabbage | TC01 Perennial Crops and Isolated Trees |
| AC16 Spring Linseed | AC71 Coriander | NU01 Nursery Crops |
| AC17 Maize | AC72 Corn gromwell | WO12 Trees and Scrubs, short Woody plants, hedgerows |
| AC18 Millet | AC74 Phacelia | AC00 Unknown or Mixed Vegetation |
| AC19 Spring Oats | AC81 Poppy | |
| AC20 Onions | AC88 Sunflower | |
| AC22 Parsley | AC90 Gladioli | |
| AC23 Parsnips | AC92 Sorghum | |
| AC24 Spring Rye | AC94 Sweet William | |
| AC26 Spinach | LG01 Chickpea | |
| AC27 Strawberry | LG02 Fenugreek | |
| AC30 Spring Triticale | LG03 Spring Field beans | |
| AC32 Spring Wheat | LG04 Green Beans | |
| AC34 Spring Cabbage | LG06 Lupins | |
| AC35 Turnip | LG07 Spring Peas | |
| AC36 Spring Oilseed | LG08 Soya | |
| AC37 Brown Mustard | LG09 Cowpea | |
| AC38 Mustard | LG11 Lucerne | |
| AC41 Radish | LG13 Sainfoin | |
| AC44 Potato | LG15 Mixed Crops – Group 1 Leguminous | |
| AC45 Tomato | | |
| AC50 Squash | | |

| CROME classes utilised for Living England 2022-23 'Arable' class | | CROME classes not included |
|---|---|---|
| AC52 Siam Pumpkin<br>AC58 Mixed Crop-Group 1<br>AC59 Mixed Crop-Group 2<br>AC60 Mixed Crop-Group 3 | LG16 Mixed Crops – Group 2 Leguminous<br>LG20 Winter Field beans<br>LG21 Winter Peas | |

www.gov.uk/natural-england