NCEA
Natural Capital and Ecosystem Assessment

England Peat Map

Technical Supplement to the Final Report

May 2025

Natural England Research Report NERR 149 Annex 6

Authors: Christoph Kratz, Sam Dixon, Alex Hamer, Oliver Gutteridge, Chris Miller, Anne Williams, Michelle Johnson, Martha Tabor, Samuel Richardson, Nick Tomline and Phil Shea

www.gov.uk/natural-england

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

Acknowledgements

This project is funded by the UK Government through Defra's Natural Capital and Ecosystem Assessment programme.

Citation

Kratz, C., Dixon, S., Hamer, A., Gutteridge, O, Miller, C., Williams, A., Johnson, M, Tabor, M, Richardson, S., Tomline, N. and Shea, P. 2025. 'England Peat Map Technical Supplement to the Final Report' in Kratz, C. and others. 2025. England Peat Map Project Final Report. Natural England Research Report NERR149. Natural England.

Contents

1. Overview

This technical supplement is an Annex (referred to below as 'this Annex') to the England Peat Map Final Report (Natural England 2025) ('the main report'). It provides additional technical details which more fully document the development of outputs from the England Peat Map project. It is likely to be useful for people interested in the science underlying the England Peat Map, to better understand it or to attempt to replicate and improve on it.

2. Survey quality assurance & management

The field survey programme delivered by England Peat Map is described in chapter 5 of the main report. This Annex provides further information about how quality of the survey was assured and managed. The main methods were:

- Detailed training in the field to contractors and in-house surveyors
- Extensive protocols and documentation
- In-house quality assurance surveys with feedback to contracted surveyors
- Laboratory analysis of soil samples

2.1. Documentation and Training

Detailed field protocols were produced for both the vegetation and soil surveys which are available in a separate annex to this report. Additional guidance documentation was produced covering soil sample collection, peat probing, soil texture identification, and how to use the data collection application. DEFRA group biosecurity guidance and a GPS user manual were also provided to surveyors.

To support field survey delivery between November 2021 and August 2024 twelve training sessions were run on how to use the field data collection app, and how to undertake the England Peat Map Survey. A total of 92 people were trained. The training was developed and delivered by Dr Chris Miller supported by other Natural England specialists and colleagues.

2.2. Quality Assurance Surveys

The objectives of the quality assurance process were to:

- Identify errors in the survey data.
- Ensure contracted surveyors are following the agreed processes and procedures.
- Ensure remedial measures are put in place to reduce potential errors e.g. additional training.
- Improve survey accuracy using post survey data correction procedures.

Based on recommendations made by UKCEH it was decided to undertake quality assurance (QA) checks on 10% of the contractor's fieldwork. This approach was refined after the field survey pilot to encompass 10% of each field surveyor's work. Ideally to avoid bias each QA survey should be randomly selected, however to make best use of available QA surveyor resource we checked clusters of each surveyor's work, and in some cases the 10% requirement was exceeded. It is acknowledged that this may have inadvertently introduced some bias into the QA surveys as the selection was not random between survey sites.

A part blind survey approach was adopted where survey points were re-visited, and the survey was repeated without the QA surveyor referring to the original survey. Once the second survey was complete both sets of results were compared, and any differences noted. Whilst presenting coordination challenges, this approach allowed rapid identification of quality issues relating to specific surveyors as well as more general issues. This facilitated rapid feedback allowing quality issues to be addressed at an early stage. To improve the quality of the data collected, changes were made to the survey design, and automated and manual checks of the collected data were undertaken. Following the survey pilot, changes were made to the survey design to minimise quality issues. This included simplification of the field survey protocol to reduce the skill level required to deliver the survey. The requirement for soil samples to be taken and analysed for organic content helped reduce misidentification of mineral vs peaty soils. In addition, switching from using the Domin scale (see glossary) for cover to the nearest 5% helped to more precisely quantify vegetation cover reducing potential error.

Coding scripts were developed to undertake automated checks on the collected field survey data and provide the results in a user-friendly form. These included:

- checks on GPS position,
- identifying peat depth outliers,
- checking whether soil samples were collected,
- whether assigned vegetation class matched dominant vegetation found, and
- checking whether percentage cover adds up to 100 percent.

Automatic checks were compared the original survey and QA survey to rapidly identify discrepancies that had not been picked up by the QA surveyors during their comparisons.

Manual comparison of the original survey against the QA survey was also undertaken by staff who had not carried out the QA survey. Using professional judgement, thresholds were established for when there was a substantial difference between the two surveys and further investigation required. A 5-10% difference in vegetation percentage cover was not investigated as it was within the natural range expected between different surveyors. More substantial variations were investigated primarily using photographs of the quadrat and of the core, and a decision made whether to adjust the survey record, reject the original survey in favour of the QA survey or reject both surveys. All such decisions were recorded in the record.

The soil organic content identified through laboratory analysis was compared with the soil texture class which the surveyor identified in the field. Where there was a significant difference (i.e. more than one soil texture class difference e.g. Mineral vs Peaty Loam) the survey was investigated. Photos and field data were examined to identify misidentified soil samples, and in extreme cases resulted in the rejection of survey data.

Peaty soil depth measurement data was initially cleaned to remove data entry errors e.g. peaty depth recorded where mineral soil present at the surface. Data outliers/extreme values were also investigated and either corrected or removed from the dataset. Comparison between original and QA surveys was also undertaken and a decision made whether to reject the original survey or not. However, a detailed examination of the surveyor error associated with soil probing (i.e. whether surveyors had correctly detected point at which peaty soil transitions into non-peaty substrate) was not undertaken as it was outside of the scope of this project.

Feedback from the quality assurance process was presented to the contractor in regular meetings throughout the course of the contract. Whilst required actions from these meetings varied, they can be distilled into one of three actions: Individual feedback/re-training; reminder messages to survey group via a live messaging platform; change to field protocol.

2.3. Quality Assurance Analysis

Soil Survey

A total of 13.8% (542) of soil quadrats were rejected on quality grounds. Reasons for rejection included: QA survey better quality than original due to a variety of issues (46.3%); survey undertaken without landowner permission (16%); survey protocol not fully followed e.g. soil sample not taken when required (12.9%); all 5 depth points identical (10.3%); poor GPS accuracy (9.8%); Missed peaty soil at surface (1.7%); miscellaneous (3%).

The data collected from 24.2% (949) of soil surveys was changed as part of the Quality Assurance process. 76.7% (728) of changes were made as a result of laboratory data with the remainder being for data entry errors, soil horizon thickness incorrectly input, and corrections to position due to GPS issues.

Vegetation Survey

A total of 12.3% (368) of vegetation quadrats were rejected on quality grounds. Reasons for rejection included: QA survey better quality than original due to a variety of issues (26.4%); Species misidentification or cover over/under estimation (25.5%); data entry error (23.9%); Survey protocol not followed e.g. incorrect orientation/position (14.1%); Photo does not match survey (6.3%); poor GPS accuracy (3.8%).

Data collected from 81 (2.7%) vegetation surveys were changed mainly due to data entry errors e.g. incorrect vegetation class selected or because of QA surveys highlighting species misidentification.

Impact

The impact of the Quality Assurance process is difficult to judge accurately as the process has not been objectively tested. However, 38.0% of soil surveys and 15.0% of vegetation surveys were either rejected or changed as a result of the QA process. Misconceptions in survey process were also identified and remedial action taken to prevent their reoccurrence. Whilst the QA process has clearly reduced the number of errors in the data set it cannot be guaranteed to be error free.

2.4. Laboratory test and results

To confirm the accuracy of the field data soil samples were sent off for laboratory analysis to determine their organic content using Loss on Ignition (375^{o}C for 16 hours, also see glossary) so that they could be assigned to the correct soil texture class. For the main survey budget constraints limited soil sample collection to soil horizons where peaty soil had been identified or where identification was uncertain, and only one sample was taken per unique soil texture class. In May 2024 the requirement of when to take a soil sample was expanded to include organo-mineral soils following analysis of the loss on ignition data collected to date. Sufficient budget was allocated to ensure that any soil horizons encountered meeting these requirements could be analysed. All laboratory analysis was undertaken by the James Hutton Institute.

Table 2-1 Field Survey Loss on Ignition Results

Soil Texture Class	Soil Organic Matter Content	Number of samples	Percentage of horizons sampled
Mineral	<8%	206	11.8%
Organo-mineral	8-20%	392	22.4%
Peaty Loam or Peaty Sand	>20-35%	273	15.6%
Loamy Peat or Sandy Peat	>35-50%	225	12.9%
Peat	>50%	707	40.4%

A summary of the loss on ignition data can be found in Table 2-1, including the breakdown by soil texture class. The majority of soil samples (68.1%) analysed were Peaty (>20% organic matter content). Analysis of the loss on ignition data has revealed that 953 out of the 1747 (54.5%) samples taken corrected a misidentified soil profile. This resulted in changes to 853 soil surveys, 727 of which were used in the final model.

3. Modelling extent, depth and vegetation

The modelling process is described in detail in chapter 7 of the main report. The following sections provide detailed results and some more detailed description of methods to support understanding and replication of the work.

3.1. Model Evaluation Metrics for England Peat Map

As set out in section 7.3 of the main report, all models were trained with only part of the data, with a portion set aside for calculating validation metrics. Some of the models (extent and depth) were then retrained with all data, but no metrics calculated from this final model.

The models which each metric is used for are tabulated in the main report Table 7-3. The definition of each metric is set out in Table 3-1 below.

Table 3-1 Evaluation metrics and definitions

Metric	Definition	Equation	Source
Overall accuracy	The proportion of correctly classified pixels in comparison to the reference data > 80% considered 'good'	$\frac{TP+TN}{TP+TN+FP+FN}$	van Rijsbergen, 1979
Precision	The proportion of true positives across all positive predictions > 80% considered 'good'	$\frac{TP}{TP+FP}$	van Rijsbergen, 1979
Recall	The proportion of true positives across all correctly predicted samples > 80% considered 'good'	$\frac{TP}{TP+FN}$	van Rijsbergen, 1979
F1 score	A balanced mean of precision and recall > 80% considered 'good'	$2 \times \frac{\mathrm{Precision} \times Recall}{\mathrm{Precision} + Recall}$	Hand, 2012
Root Mean Square Error (RMSE)	The average difference between predicted and actual values. Lower is better	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2}$	Hastie, Tibshirani and Friedman, 2009
Matthew's Correlation Coefficient (MCC)	Measuring the quality of binary classifications +1 perfect; – 1 bad; 0 random	$\frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP) \times (TP+FN) \times (TN+FP)}}$	Matthews, 1975
Frequency Weighted Intersection	The balanced similarity between the predicted region and reference region by identifying overlapping regions	$\frac{\sum_{i=1}^{N} n_i \left(\frac{TP_i}{TP_i+FP_i+FN_i}\right)}{\sum_{i=1}^{N} n_i}$	Russakovsky *et al.*, 2015

Metric	Definition	Equation	Source
over Union (fwIoU)	50% for basic acceptability, 70% for higher precision and above 75% very good		
Kappa	The statistical significance in comparison to random performance +1 perfect; – 1 bad; 0 random	$\frac{p_o - p_e}{1 - p_e}$	Cohen, 1960

3.2. Predictors

Section 7.2 of the main report summarises the predictors used in EPM, and Table 7-2 of the main report tabulates the models each predictor is used for. The following provides further information and in particular describes the creation of the **Bare Soil Sentinel-2 Mosaic** and the **Detrended LiDAR** predictor. Table 3-2 of this Annex summarises sources, temporal and spatial scales and other information about each predictor group.

The primary sources of predictors used in EPM modelling are derived from Earth observation imagery and sensors, aerial imagery and sensing as well as some mapped products. In addition, indices were derived from these primary sources.

Seasonal Sentinel-2 composites were created for Autumn 2022, Spring 2023 and Summer 2023. A winter composite was not feasible due to excessive cloud cover. These composites use solar day matching of Level-2A imagery to select the least cloudy imagery, based on the S2 Cloudless cloud mask, within each time period and infill until all pixels have been assigned a value, where no cloud-free imagery was available it is left as no data. The Sentinel-1 imagery underwent median filtering for the same time periods as the Sentinel-2 to reduce radar speckle in both VV and VH polarisation for Ascending and Descending acquisition modes. These Sentinel-1 and Sentinel-2 mosaics are the same used by the 2022-23 Living England habitat probability map (Trippier *and others*, 2024).

The national airborne LiDAR composite Digital Terrain Model (DTM) and Digital Surface Model (DSM) from the Environment Agency is used in the extent, depth and vegetation modelling which is a composite of imagery captured between 2017 and 2023. Further processing of these layers is undertaken to calculate the Canopy Height Model (CHM) from the DTM and DSM and slope derived from the DTM to provide further information for vegetation presence.

Additional predictors were used specifically for peat extent and depth modelling. The Land Utilisation Survey (1933-1949), Dudley Stamp, categorised Great Britain into 8 land use classifications at a 1km resolution. The bedrock and superficial geology predictors are derived from the British Geological Survey 1:625,000 scale digital geological map. Bedrock is defined as deposits laid down prior to the quaternary period (2.588 million years ago) and superficial deposits as laid down during the quaternary period. The OS

Open Rivers dataset was used to calculate the Distance to River and the OS Boundary High Water Mark dataset was used to calculate the Distance to Sea.

The bare peat and surface feature mapping used high resolution aerial photography of Great Britain (APGB) in upland (as defined by the Moorland Line) areas. A composite of upland areas was created using the latest available imagery from BlueSky Mapping at the time of the project (January 2023) and includes imagery from 2018 until 2023 for national coverage at 12.5 cm and 25 cm resolution for surface features and bare peat, respectively. In addition, detrended LiDAR is used alongside aerial photography for improved classification of grips, gullies and haggs.

Bare Soil Sentinel-2 Mosaic

A new mosaic was created by the England Peat Map team because notable spectral differences between peaty and non-peaty soils were observed in areas of the seasonal Sentinel-2 mosaics in bare arable-dominated regions. The mosaic enhances these spectral properties (Figure 3-1). The Bare Soil Index (BSI) was implemented to identify areas when they were bare by calculating the index for all cloud-free atmospherically corrected imagery captured between August 2022 and July 2023 to match the timings of the other Sentinel-2 mosaics. Cloud masks were applied through the S2 Cloudless implementation to remove no data areas into the mosaic creation. The 95% percentile of the BSI was taken for each pixel to remove extreme values for identifying bare soil image timings and these dates were selected to create the bare soil composite image across all lowland peatlands for use within the model.

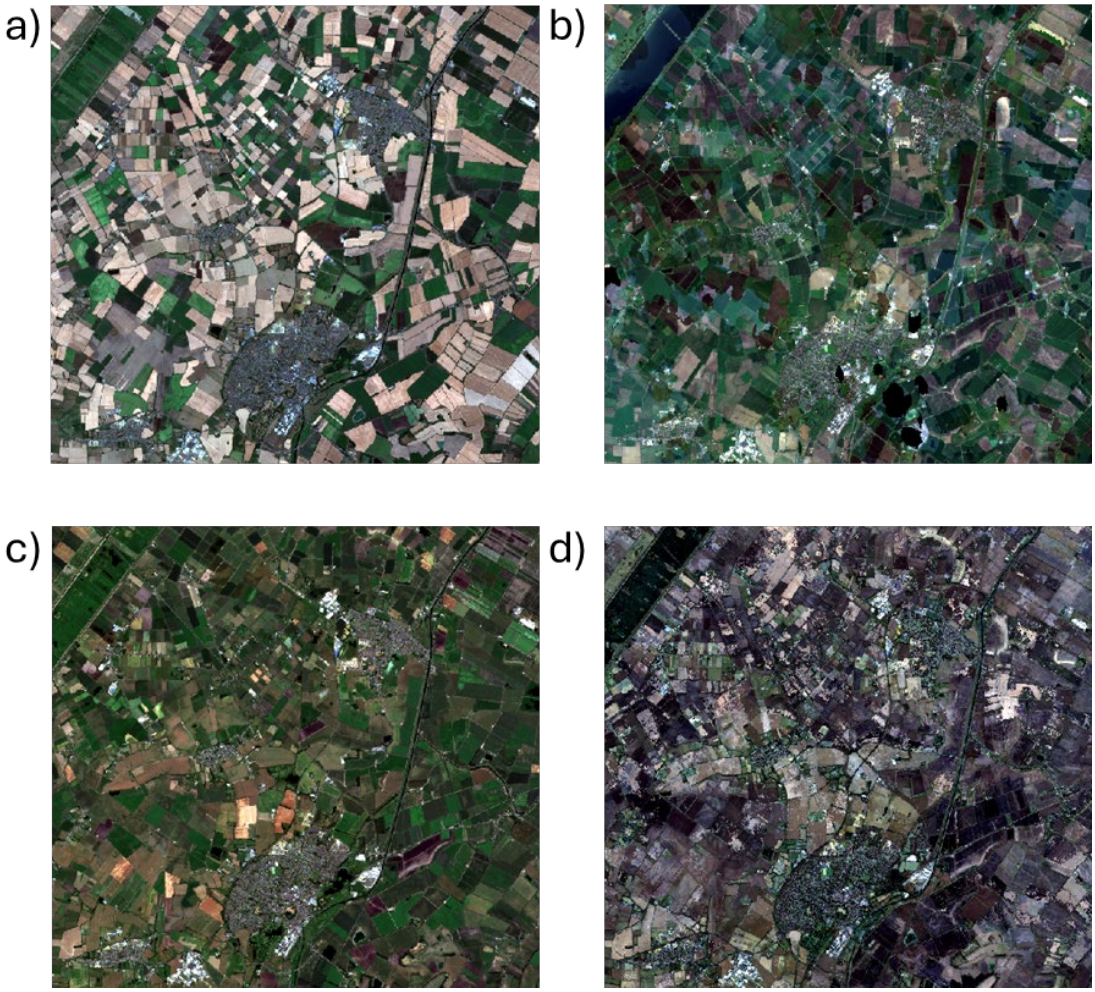

Figure 3-1 Sentinel-2 imagery for Autumn 2022 - Summer 2023 across the Cambridgeshire Fens with (a) autumn mosaic, (b) spring mosaic, (c) summer mosaic and (d) median bare soil mosaic. Subtle colour variations between mosaics are caused by different stretch implementations for visualisation. Contains modified Copernicus Sentinel data 2022-2023.

Detrended LiDAR for surface feature mapping

LiDAR is used alongside the aerial imagery to enhance the detection of grips, gullies and haggs and to calculate the depth and slope of modelled features. The 1 m resolution LiDAR derived DTM was detrended to remove large-scale elevation changes and highlight features relative to the local terrain. Detrending is achieved by subtracting the median elevation of a moving 100 m radius window from each 1 m pixel (Figure 3-2). After de-trending, the relatively small-scale features of grips and gullies can be clearly seen in the data. A range window sizes were trialled from 2 to 200 metres.

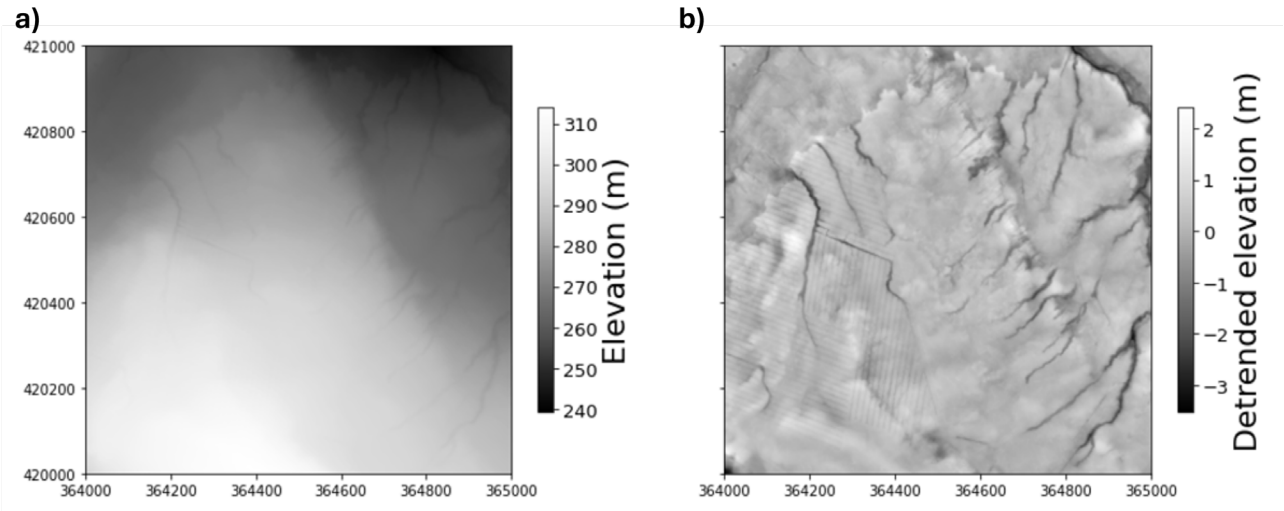

Figure 3-2 Example of original 1 km^2 LiDAR image (a) and corresponding detrended image (b)

The goal of the median layer conceptually is to put a line of best fit through the surface model, forming a straight line across the grips and gullies as though they were not present. The vertical difference between this median surface and the terrain model forms the detrended LiDAR layer. If the moving window size is too small, this line of best fit follows the features instead of cutting across the top.

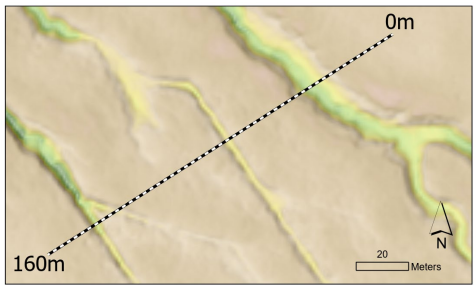

Figure 3-3 sample of gullies and grips in the West Pennines, with a transect line used for creating an elevation profile. Lidar data from Environment Agency 2020.

The difference in moving window sizes is illustrated in Figures 3-4 and 3-5 using a transect across a range of gully and grip features. Figure 3-4 shows the elevation profile for the transect shown in Figure 3-3. The black profile shows the actual topography as recorded in the Environment Agency's LiDAR digital terrain model. The various coloured lines show the smoothing effects for different moving median window sizes ranging from 2 m up to 200 m. The smallest 2 m median follows the terrain model very closely, so is not appropriate for detrending. At 20 m, the orange median line still drops down as it passes across the gully features. At the 100 m median in green, the line draws a fairly straight line across all of the depressions. This level line is important, as the detrended LiDAR is created by measuring the depth to the original terrain model (black line) below the chosen median line (in this case the green line). The 200 m median window applies too much smoothing in most areas. In the example below it is seen to pass above the high spots, meaning that they would show up as depressions in the detrended LiDAR, and overestimate the depth of features. There will always be an element of compromise here.

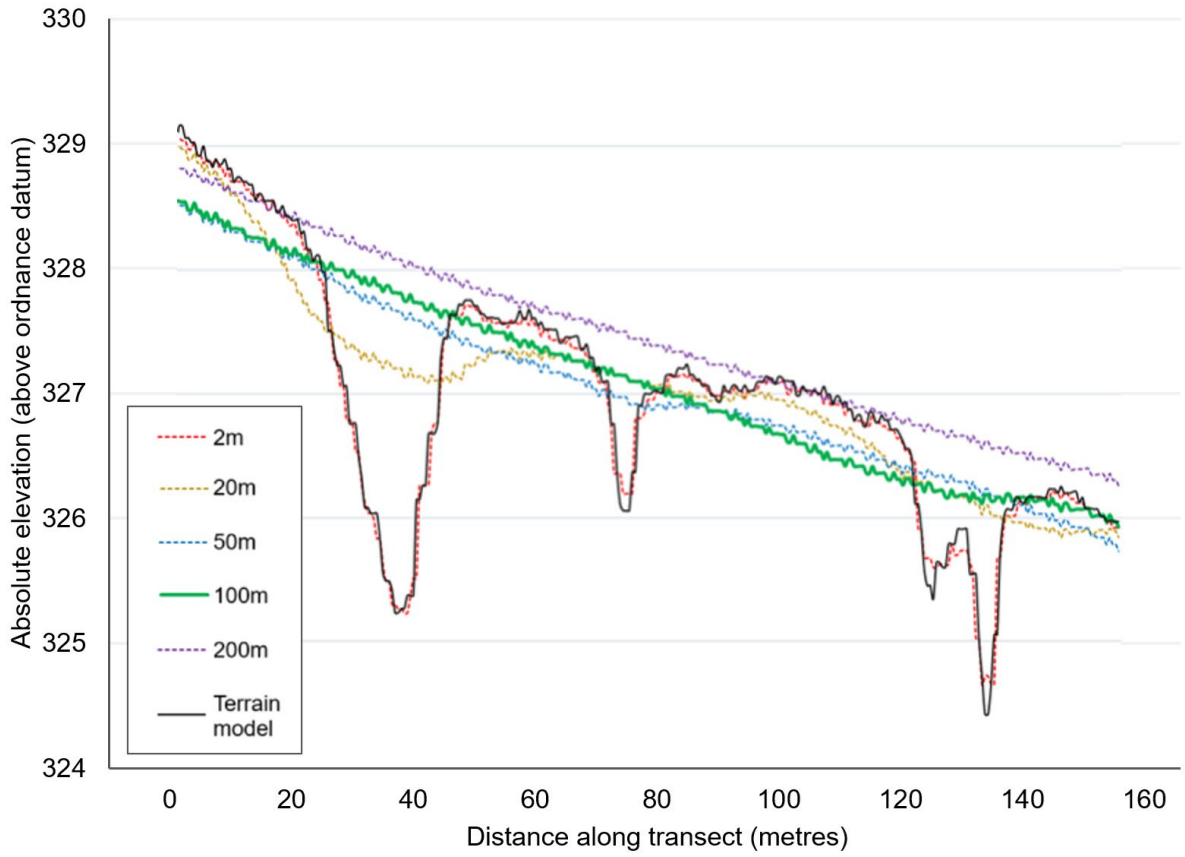

Figure 3-4 Elevation profiles for different median average moving window sizes, compared to the LiDAR digital terrain model from Environment Agency LiDAR

Figure 3-5 shows the resulting detrended LiDAR models for the transect shown above. This illustrates the issue with the smaller moving window sizes. This is obtained by subtracting the original Digital Terrain Model from the smoothed median surface. The red 2 m moving window follows the black 0 m axis line through the middle of the graph. If we used this as our detrended terrain model, we would measure the depth across the gullies to be near 0 m depth. This is clearly not helpful or desirable. By contrast, the largest 200 m window floats above many of the high points, which is also undesirable as it will cause us to overestimate the depths of the low points. The 100 m window size provides a better indication of the shape of the terrain, whilst normalising the elevation values to be relative to the average surface as shown in Figure 3-4.

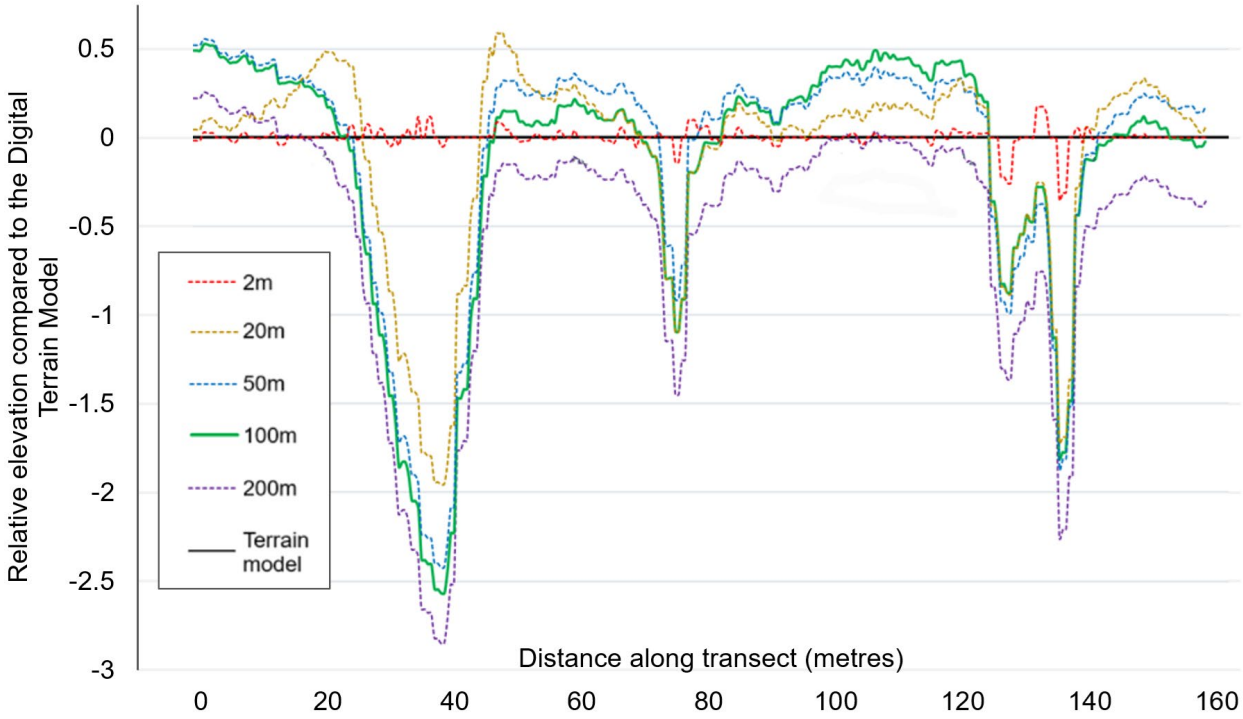

Figure 3-5 Detrended elevation profiles for different median average moving window sizes, derived from Environment Agency LiDAR

This gives us a normalised terrain model which clearly shows features which extend above and importantly for this work, below the surface. It also enables us to more accurately measure the depth along these features. To this end it is important for the median surface to pass as close as possible to the tops of the gully and grip features. In practice this is tricky as large scale and small scale topographic variations affect the different moving window sizes in different ways, but visual inspection suggests that the 100 m window size strikes a decent compromise for this purpose.

Table 3-2 Summary of predictor sources used for England Peat Map

See table 7-2 of the main report for a different overview of predictors.

Source	Description	Spatial Scale	Temporal Scale	License	Reference	How are these data used?
Sentinel-1 backscatter	Radar imagery from Sentinel-1 satellite mission, capturing C-band synthetic aperture radar images.	10 m	1 Aug 2022 – 31 Jul 2023	OGL	(Gorelick *and others*, 2017)	Created 10m multitemporal mosaics using Living England (LE) workflow in Google Earth Engine (GEE), used in extent, depth and vegetation models.
Sentinel-1 InSAR coherence	Single Look Complex (SLC) images containing amplitude and phase information	15 m	1 Aug 2022 – 31 Jul 2023	OGL	(Gorelick *and others*, 2017)	Created 10m multitemporal mosaics as above, used in extent, depth and vegetation models.
Sentinel-2	Optical imagery from Sentinel-2 satellites capturing 13 spectral bands	10 m	1 Aug 2022 – 31 Jul 2023	OGL	(Gorelick *and others*, 2017)	Created 10m multitemporal mosaics as above, used in extent, depth and vegetation models. Created 10m bare soil mosaic using EPM workflow in GEE, used in extent and depth models.
Landsat-8	Optical imagery from the Landsat-8 satellite capturing	30 m	1 Aug 2022 – 31 Jul 2023	OGL	(Gorelick *and others*, 2017)	Created 10m annual median lowland thermal mosaics in GEE used in extend and depth models.

Source	Description	Spatial Scale	Temporal Scale	License	Reference	How are these data used?
LiDAR	EA's National LIDAR programme capturing accurate elevation data of the height of the terrain and surface objects on the ground	1 m	2016 - 2023	OGL	(Environment Agency, 2023)	Derived 10m rasters of: - slope used in extent, depth and vegetation models. - aspect, Digital Terrain Model (DTM), Terrain Roughness Index (TRI), Topographic Wetness Index (TWI) and Geomorphons used in extent and depth models. - Canopy Height Model (CHM) used in vegetation model. - detrended DTM used in surface feature model.
Geology (Bedrock and Superficial) 1:625,000	National geological map data	1:625,000	2016	License agreement for use	(British Geological Survey, 2016)	Rasterised to 10m. Used in the extent and depth models
Dudley Stamp Land Utilisation Survey	Land use survey across England	1 km	1933-1949	OGL	(Environment Agency, 2011)	Rasterised to 10m. Used in the extent and depth model.
OS Open Rivers	GB river network generalised from OS large-scale data.	1:25,000	2024	OGL	(Ordnance Survey, 2023)	Derived 10m raster of distance to river. Used in the extent and depth model
OS Boundary Line	Mean high water mark (England and Wales)	1:10,000	2024	OGL	(Ordnance Survey, 2024)	Derived 10m raster of distance to the sea. Used in the extent and depth model
Aerial Photography	Aerial survey imagery across Britain	12.5 cm & 25 cm	2018 - 2023	License agreement for use	(Bluesky International Ltd, 2023)	Used as the only input to bare peat and dam models and alongside LIDAR for grip, gully and hagg models

Source	Description	Spatial Scale	Temporal Scale	License	Reference	How are these data used?
Flood Map for Planning (Rivers and Seas) Flood Zone 3	Areas of land at risk of flooding, when the presence of flood defences are ignored and covers land with a 1 in 100 (1%) or greater chance of flooding each year from Rivers; or with a 1 in 200 (0.5%) or greater chance of flooding each year from the Sea.	Not known	2004-2024	OGL	(Environment Agency, 2024)	Rasterised to 10m. Used in the extent and depth model.
Risk of Flooding from Surface Water	Extent of flooding from surface water that could result from a flood with a 3.3% chance of happening in any given year. Previously known as the updated Flood Map for Surface Water (uFMfSW).	Variable	2013	OGL	(Environment Agency, 2013)	Rasterised to 10m. Used in the extent and depth model.

Table 3-3 Indices used for England Peat Map

Metric	How is this index used?	Equation	Source
Topographic Wetness Index (TWI)	Used to inform the extent and depth model	$Ln((FlowAcc + 1) / Tan(SLOPE * (\pi/180)))$	(Kopecký *and others*, 2021)
Bare Soil Index (BSI)	Used to inform the imagery selected for the Sentinel-2 bare soil mosaic	$BSI = \frac{(SWIR2 + RED) - (NIR + BLUE)}{(SWIR2 + RED) + (NIR + BLUE)}$	(Nguyen *and others*, 2021)

3.3. Model Performance

Peaty Soils Extent Probability Thresholds

The **extent model** estimates the probability that peaty soil is present. To make a final prediction we used the evaluation metrics to determine the optimal threshold at which to predict that peaty soil is present or not present (see Table 3-4). The final extent model uses a probability threshold of 0.4 because it has good performance across all metrics, scored highest for MCC, and in particular it balances sensitivity and specificity, with both scoring above 0.9. We also gave consideration to using a threshold of 0.5, which also scores above 0.9 in both sensitivity and specificity and scores best in accuracy and F1 Score. However we gave preference to the MCC metric which is better at balancing false positives and false negatives (Chicco & Jurman, 2020). The 0.4 threshold also has a higher sensitivity, at the cost of a slight reduction in specificity.

Table 3-4 Extent model performance metrics at different probability thresholds

prob. threshold	accuracy	f1_score	mcc	sensitivity	specificity
0.1	0.942	0.895	0.862	0.982	0.929
0.2	0.950	0.907	0.877	0.975	0.941
0.3	0.953	0.912	0.883	0.970	0.947
0.4	0.954	0.913	**0.885**	0.964	0.951
0.5	**0.955**	**0.914**	0.884	0.956	0.954
0.6	0.954	0.912	0.882	0.946	0.957
0.7	0.953	0.909	0.878	0.929	0.961
0.8	0.949	0.899	0.865	0.895	0.968
0.9	0.934	0.859	0.820	0.800	0.979

Confusion Matrix for Peaty Soils Extent

	Predicted Absent	Predicted Present
Observed Absent	78,740	4,083
Observed Present	1,002	26,825

Full Confusion Matrix for National Vegetation Classification

Predicted (columns) Observed (rows)	Sphag-num sp.	Erio-phorum sp. bog	Molinia caerulea bog	Calluna vulgaris bog	Dry grass & scrub bog	Short fen veget'n	Tall fen veget'n	Scrub & tree fen
Sphagnum sp.	23	5	2	9	0	0	1	0
Eriophorum sp. bog	5	393	8	97	2	1	4	1
Molinia caerulea bog	0	18	127	18	2	0	6	0
Calluna vulgaris bog	2	43	5	3,519	1	1	0	1
Dry grass & scrub bog	0	4	5	7	42	1	1	2
Short fen vegetation	0	1	1	0	0	51	8	1
Tall fen vegetation	0	1	1	0	0	12	63	1
Scrub & tree fen	0	3	3	0	2	2	2	34

Full Confusion Matrix for Upland Bare Peat Classification

	Predicted Absent	Predicted Present
Observed Absent	23,644,541	305,282
Observed Present	479,771	450,406

3.4. Depth interpolation

Geostatistical analysis and predictions were carried out in R using the gstat package (Pebesma, 2004; Gräler *and others*, 2016). Variogram fitting was carried out using the automap package (Hiemstra *and others*, 2008).

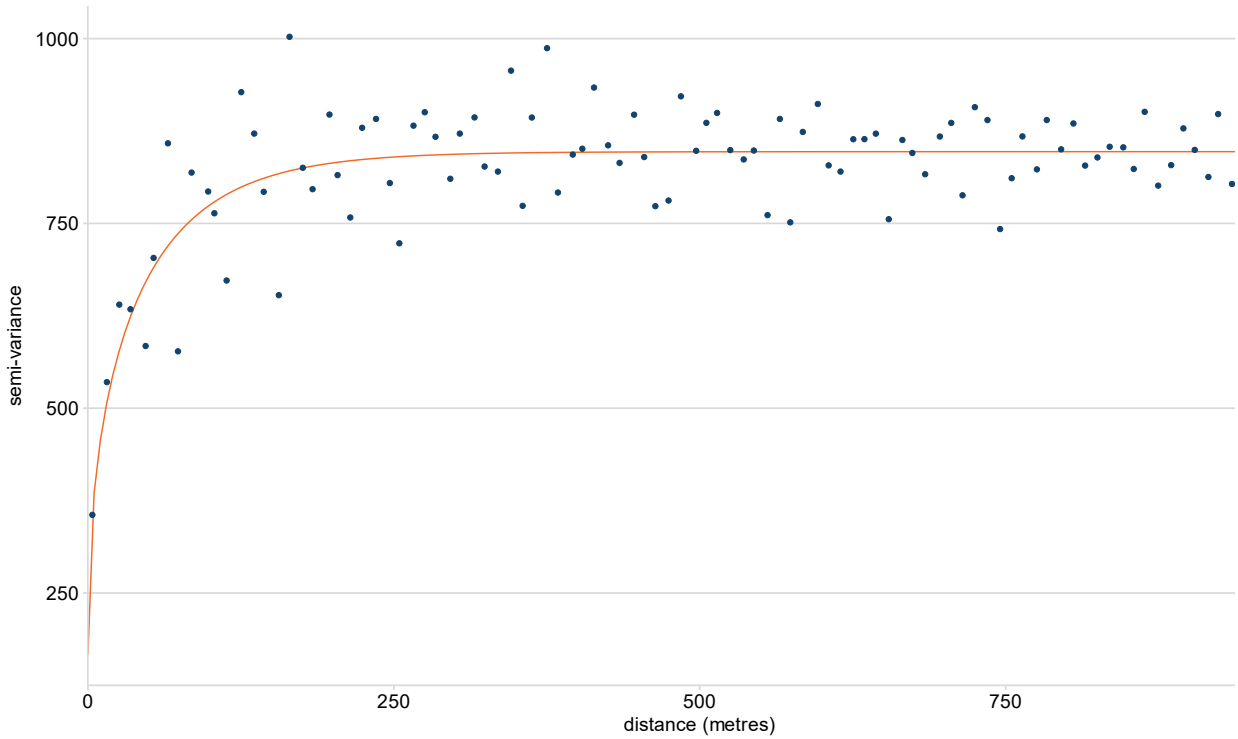

Figure 3-6 Empirical and fitted variograms for depth residual.
Model = Matern (M. Stein's parameterisation), nugget = 159.37, partial sill = 687.71, range = 65.53

4. Modelling surface drainage and erosion features

The upland peat drainage and erosion features models, delivered by the AI4Peat project team are described in section 7.6 of the main report. The following provides more details about the modelling process, parameterisation, post-processing, calculation of dimensions and metrics. It also addresses the issue of quality assurance and ethics in the context of this application of artificial intelligence, and considers potential future work.

4.1. Model Selection

The training datasets for each feature (grips, gullies, haggs and dams) were split into training (70 %), validation (20 %) and test(10 %) datasets. To ensure even spatial coverage in each of these sets, the full dataset was split using the third numerical value of the spatial index of the 50 m chip (e.g. value 3 is used for the chip with index SD123456NW). Chips with values 0-6 were assigned to the training dataset, values 7-8 to the testing dataset and values equal to 9 the validation dataset. Due to this method of selecting training datasets, the proportions don't line up exactly with the 70/20/10 split, though due to the even distribution of the spatial index in the 50m chips, the proportions are close enough to refer to them as such. This is shown in Figure 4-1.

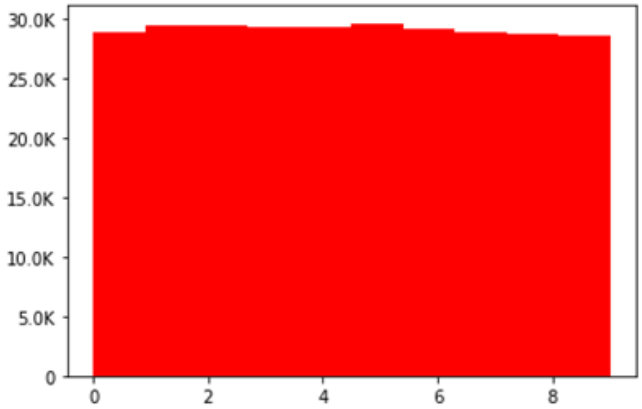

Figure 4-1 The distribution of the 3rd digit in the spatial index ID for the 291,049 chips in the full training dataset.

Due to the requirements of certain model architectures needing input sizes divisible by 32, such as the FPN model, the 400x400 pixel chips had to be resized to 416x416. This extra padding was done using the constant mode with a value of 0. The images were also normalised to have means and standard deviations closer to the means and standard deviations of the ImageNet dataset the models were pretrained on. These means and standard deviations are as follows: mean = [0.485, 0.456, 0.406, 0.42], std = [0.229, 0.224, 0.225, 0.22]. The first 3 bands are the RGB bands for APGB data, with the final band being the LiDAR.

Grips, Gullies and Haggs

All models went through an experimentation process that tested the following aspects of the model with the accompanying parameter sets:

- Model architecture (PSPNet, Linknet, Unet, FPN, Manet)
- Dropout (0, 0.1, 0,2, 0.3, 0.4, 0.5), backbone (ResNet18, ResNet32, ResNet50)
- Optimiser (Adam, NAdam, SGD with and without momentum (Nesterov))
- Loss function (Dice, Focal, Lovasz, Jaccard, SoftBCE)
- Learning rate (0.0001-0.1), momentum (0.7-0.99), beta2 (0.99-0.9999), epsilon (1e-8-1e-4), weight decay (1e-8-0.0001), momentum decay (1e-8-0.1)

Each feature was tested over 5 experiments, each one testing the scientific parameters, along with nuisance parameters (which tended to be the learning rate and momentum of the loss functions) and all remaining parameters fixed. Each experiment also has an average of 60 trials, this ranged from 30-100 trials. Most of the experiments used a random parameter selection algorithm to test the full width of the parameter space, however the final experiment of the numerical hyperparameters of the optimiser used a Tree-structured Parzen Estimator sampler. This sampler is a Bayesian-style sampler that homes in on an optimal set of hyperparameters.

The experiments were set up to explore trends in the effects of each hyperparameter on the performance of the models, as opposed to a brute force parameter-space search. This was done due to the oversized parameter space to search, and to gain meaningful insight into the purpose of the hyperparameters.

For the Grip and Gully models, there were two different datasets: A “raw” dataset and an “improved” one. The “raw” dataset comprised of a larger number of chips, made using line segments buffered by a constant width. The “improved” dataset was comprised of a smaller number of chips, made by hand digitising the features by drawing around their edges. A further experiment was therefore created to test the ideal split between training on the larger raw dataset and fine-tuning on the smaller improved dataset. This covered raw_epochs (1-5) and improved_epochs (5-15). After testing both grips and gullies for the ideal split between raw and improved data, it was determined that for grips, the inclusion of improved data did not have a big enough impact on the results to justify their use. Whilst the metrics did improve, the visual performance of the model appeared to suffer. This is believed to be due to the overfitting of the model to the - much smaller – improved dataset, though further testing should be undertaken to confirm and rectify. Whilst the gullies did show a similar effect in that the improvement to the metrics outstrips the visual improvement, the visual results did appear better once improved data was included, usually due to the ability of the improved data to better capture the irregular and variable width of the features.

Dams

To detect dams where restoration has already occurred, an object detection model was used. Unlike the semantic segmentation models used for the other features, these models

do not pick out the exact shape of a feature, but instead draw a bounding box around the feature of interest.

To find the best model for detecting the dams, three different model architecture-backbone combinations were tested.

- *Faster R-CNN with ResNet50 FPN*: Faster R-CNN (Region-based Convolutional Neural Network) with a ResNet50 backbone and Feature Pyramid Network (FPN) is a two-stage detector. It first generates region proposals (potential object locations) and then classifies them and refines their locations in a second stage. The addition of FPN helps the model capture objects of varying sizes by combining feature maps from multiple layers. It's accurate but typically slower than single-stage detectors because of the extra region proposal step.
- *RetinaNet with ResNet50 FPN*: RetinaNet is a single-stage detector, meaning it predicts objects and their locations in one step. It also uses a ResNet50 backbone with FPN, which helps detect objects of different sizes. RetinaNet is known for its "focal loss" function, which adjusts its focus to better detect harder (e.g., smaller) objects that can be overlooked by other detectors. It's faster than Faster R-CNN but tends to be slightly less precise.
- *YOLOv8 (You Only Look Once, Version 8):* YOLOv8 is the latest iteration in the YOLO family, which is known for its speed and efficiency in real-time object detection. YOLO models are single-stage detectors, which means they predict object classes and locations in a single forward pass through the network, allowing for very fast detection.

All these models were used pre-trained on the COCO (Common Objects in Context) dataset which includes around 200k labelled images with 80 different object classes. This gives the models the advantage of being able to already detect simple structures in images such as edges of objects. The first two models can be run using Pytorch, while YOLOv8 is most effectively used through the Ultralytics package.

A detailed comparison of the three models was carried out using the Optuna package which allows a Bayesian search over a hyperparameter grid. For each model three optimisers were tested: Adam, NAdam and SGD. For each model-optimiser combination 20 trials were run across the hyperparameter grid, resulting in 180 trials in total. The Faster R-CNN and RetinaNet models learned very quickly and usually reached maximum performance by around 10 epochs, so each trial was run for 10 epochs. The YOLOv8 model took longer to learn, so each trial was run for 100 epochs but was set to stop after 20 epochs if no improvements in the validation loss were made. The performance of each model on the validation dataset was logged using Mlflow.

Figure 4-2 shows how the validation mAP50 metric varies with training epochs for the best two YOLO models and the best Faster R-CNN and RetinaNet models.

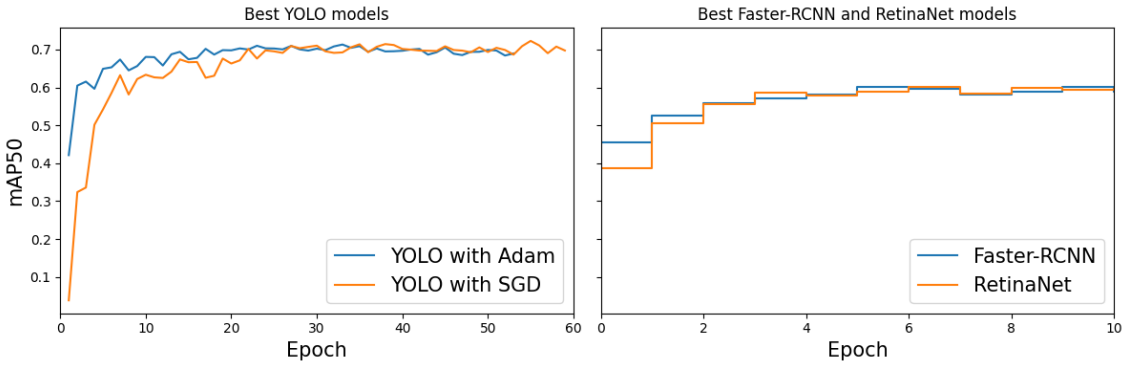

Figure 4-2: Validation mAP50 for each training epoch for best performing YOLO models on the left and Faster-RCNN and RetinaNet models on the right.

The ground truth data are not always accurate, and the bounding boxes will not always be centred precisely on the dam itself. In addition, the clearest indicator of where a dam is varies depending on the dam - sometimes the actual dam is obvious, but sometimes the overflow pools behind the dam are the clearer indicator. This means that the model predicted bounding boxes won't always overlap exactly with the ground truth bounding boxes. To assess the performance of the model a minimum acceptable IoU needs to be set to define when the model is successful. We have defined this as 50%.

The model predicts both the location of objects and the confidence in those predictions. The results and the performance of the model when comparing to the ground truth will depend on the minimum confidence threshold set.

Figure 4-3 shows how the model predictions vary with three different confidence thresholds. For lower thresholds (shown on the left) there are a higher number of model predictions (shown in blue). This means an increase in the number of correctly identified dams or "true positives" as almost all dams in the ground truth data (shown in red) have been captured, but it also means in increase in false positives – where the model thinks there is a dam where there isn't. For higher thresholds (on the right) the reverse is true: while there are fewer false positives, there are also fewer true positives. Setting an appropriate confidence threshold to balance these is important to maximise the quality of the output.

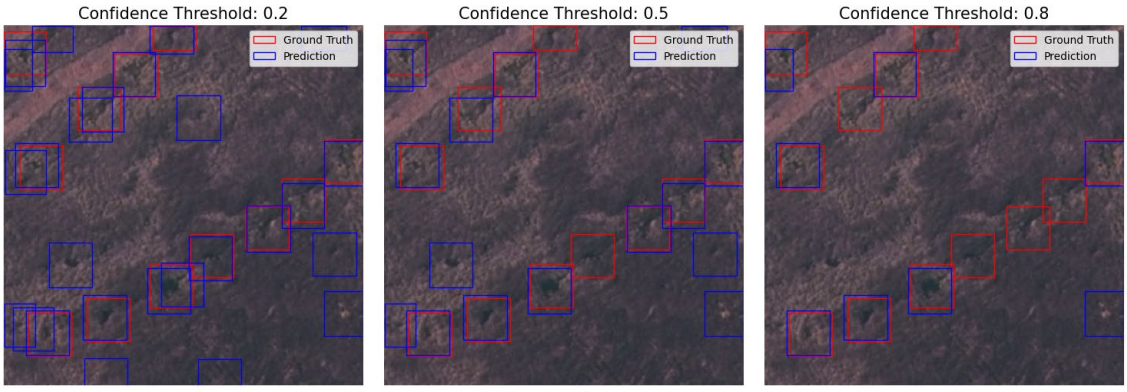

Figure 4-3: Ground truth and model inference for three different confidence thresholds. Aerial imagery © 2015 Getmapping plc and Bluesky International Ltd.

This balance is captured by the precision and recall metrics, where precision effectively measures what proportion of your predictions are correct, and recall measures the proportion of actual existing objects you model manages to capture. With a higher confidence threshold, the precision will generally increase, while recall will decrease.

Figure 4-4 shows how the recall, precision and MAP50 vary with confidence threshold for the best performing model for the holdout test dataset. This shows that the best mAP50 is achieved with a fairly low confidence threshold.

In the post processing we remove detections that are not on grips, so many remaining false positives will likely be removed. It was therefore decided that a low confidence threshold of 0.1 would maximise the number of dams detected without creating too many false positives. Any dams below this confidence threshold were removed.

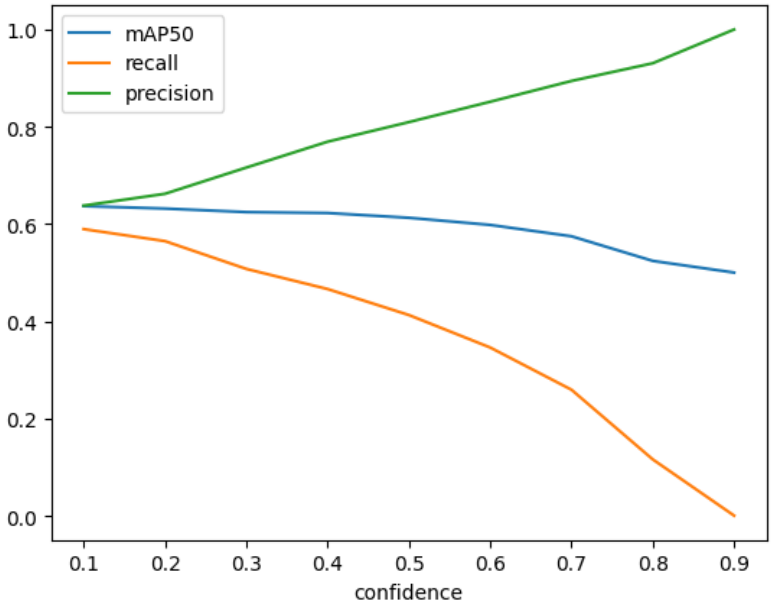

Figure 4-4: How precision, recall and mAP50 vary with confidence threshold.

4.2. Final model parameters

Grips

The Grip model selected by the experimentation process has the following hyperparameters:

Architecture: FPN, batch_size: 16, Beta2: 0.9943339059275593, Dropout: 0.3, encoder_depth: 5, encoder_name: resnet34, Eps: 2.1413053466815238e-08, Feature: Grip, learning_rate: 0.000765386594404144, Loss: Dice, Momentum: 0.8223237921224302, number_of_epochs: 20, Optimiser: Adam, Quality: Improved, weight_decay: 3.5076443639918004e-06

Gullies

The gully model selected by the experimentation process has the following hyperparameters:

Architecture: FPN, batch_size: 16, Dropout: 0.3, encoder_depth: 5, encoder_name: resnet34, Feature: Gully, improved_NUM_EPOCHS: 15, learning_rate: 0.001, Loss: Jaccard, number_of_epochs: 20, Optimiser: Adam, Quality: Improved, raw_NUM_EPOCHS: 5

Haggs

The hagg model selected by the experimentation process has the following hyperparameters:

parameters = {"learning_rate": 0.007314, "number_of_epochs": 20, "batch_size": 16, "architecture": FPN, "encoder_name": "resnet34", "loss": Dice, "encoder_depth": 5, "dropout": 0.0, "optimiser": "NAdam", "momentum": 0.924627, "beta2": 0.999016, "weight_decay": 1.24431e-8, "momentum_decay": 3.45674e-7, "eps": 1.21252e-7}.

Due to an error in inference, the hagg model was trained on images padded with a reflection, yet the inference was carried out with a padding of a constant value of 0. This lead to the hag outputs having a distinct edge effect. This has meant we have had to revert to using the August model outputs instead, though this could change if we have time to rerun the model correctly.

Dams

The best performing model based on the mAP50 metric was a YOLOv8 model, trained for 53 epochs with:

- A batch size of 8
- A learning rate l_0 of 0.00012
- Adam optimiser with weight decay =1.39e^{-3} and momentum=0.90969

4.3. Post processing

The post-processing steps involve first buffering all polygons outwards by 1 m (equivalent of 8 pixels) and unioning all polygons that overlap within a 1 km grid. This step merges polygons cut by the 50 m tile edges and any polygons separated by distances less than 2 m. The merged polygons are then buffered back in by 1 m. To smooth the pixelated edges, shapely's simplify function returns a subset of the polygon coordinates making sure all new points are located within 0.5 m of the original points (Figure 4-5). The resulting polygons were filtered by area to clean the datasets. Any haggs less than 1 m^2 any grips less than 50m^2 and any gullies less than 500m^2 were removed. The feature polygons were grouped based on the 100km British National Grid index and written out as GeoJSON before being converted to geopackage format. Features mapped outside of the peat extent were then removed from the dataset.

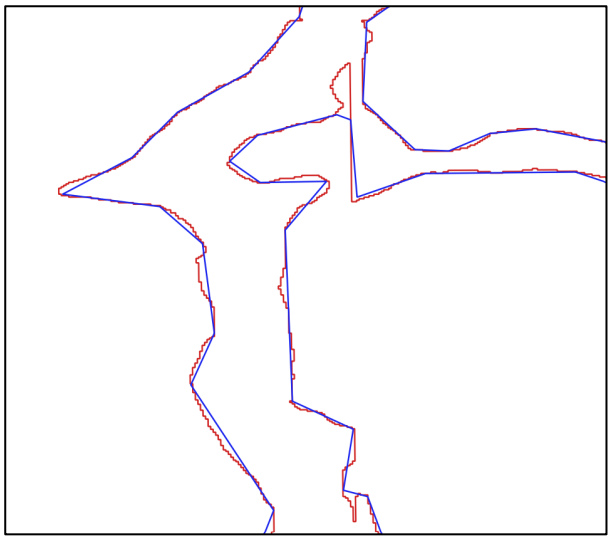

Figure 4-5 Example of a gully polygon before post-processing methods are applied (red) and after it has been buffered and unioned and then simplified (blue).

4.4. Dimensions

Width

For the width, we assume the polygon can be represented as a rectangle with width: W and length: L. The area of this rectangle will be $A = WL$, and the perimeter will be $P = 2W + 2L$. Since area and perimeter are values we can calculate for our polygons, we can use those to work out the width and length of this hypothetical rectangle. Solving the two equations for the two unknowns yields: $L = \frac{P + \sqrt{P^2 - 16A}}{4}$, $W = \frac{A}{L}$. These equations give the dimensions of a hypothetical rectangle with area and perimeter equal to that of the polygon in question.

One of the only obvious downsides to this method is that it fails when the section within the square root is negative since this produces an imaginary number. Solving $P^2 - 16A < 0$ gives $\frac{P^2}{A} < 16$. If the ratio of the square of the perimeter to the area is too small, then there is no rectangle that has the same area and perimeter. This appears to happen when the polygon is more circular, with a square being the cut-off point. Anything more circular than a square i.e. pentagon, octagon, circle etc. will fail. This can mean that the method fails for small, rounded features.

Depth

Depth attributes for each grip and gully polygon are assigned by extracting the detrended LiDAR pixels that overlap with the modelled feature polygon. The minimum, mean and maximum detrended elevation pixel values within each feature are assigned as an attribute. In the rare instances where a negative depth value is returned (i.e. the feature overlaps positive LiDAR elevation values) the depth attribute is assigned 0 to avoid confusion.

Slope

A slope value is assigned to all grip and gully polygons that have a perimeter greater than 100 m. Slope is calculated as the elevation difference between the top and bottom of a feature divided by the length (m/m or %). To calculate an elevation difference and length along a modelled feature polygon a minimum rotated rectangle is fit around each qualifying feature (magenta polygon, Figure 4-6). The length from the top to the bottom of the feature is the longest edge of this rectangle. The longest edge of the rectangle is divided into 15 equal segments (black points, Figure 4-6) and the first and last segment is used to clip the top and bottom of each feature (orange filled polygons, Figure 4-6). The original (un-detrended) LiDAR layer is clipped to these top and bottom segments and the minimum elevation pixel is selected. The absolute difference between these minimum elevation values is calculated and divided by the length.

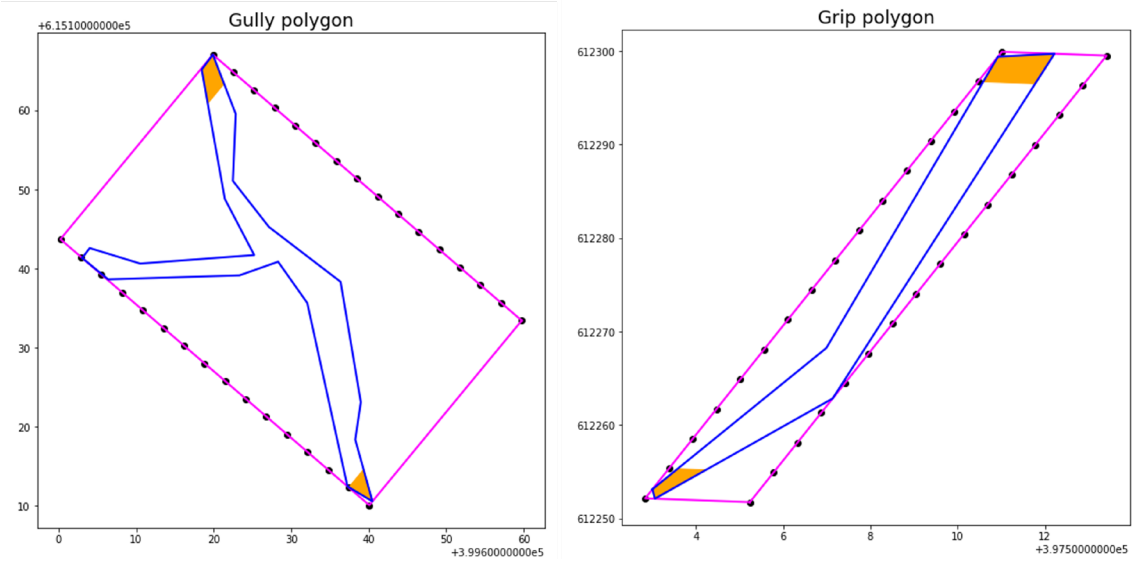

Figure 4-6 Example of the methodology used to calculate the slope of modelled grip and gully features (blue polygons). The magenta polygon is a minimum rotated rectangle fitted to individual features. Black points are the coordinates of equal divisions along the longest edge of the rectangle used to segment the feature and the yellow filled polygons are the top and bottom segments of the polygons where elevation values are extracted from to calculate elevation differences.

There are instances where the slope attribute assigned to individual polygons will not be meaningful. For example, features that have extensive branching (e.g. Figure 4-7) will have one slope value for the whole network rather than for individual linear sections. Merging polygons to avoid breaks at every 50 m image chip boundary also exacerbates this issue. We have attempted to remove the slope attribute from these features by filtering the polygons based on perimeter and removing the slope from those with a perimeter to bounding box area ratio greater than 0.5.

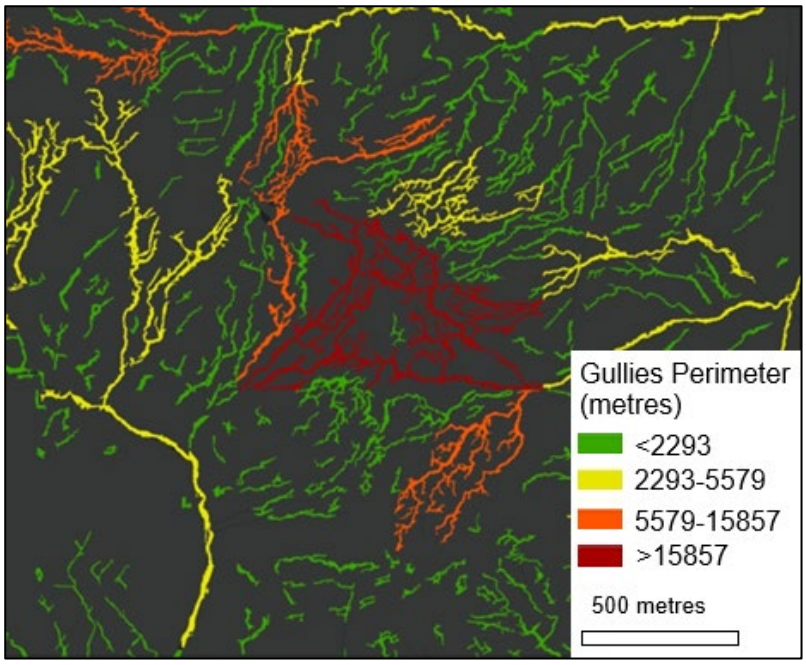

Figure 4-7 Example of complex gully polygons. Contains OS data © Crown Copyright and database right 2025. Contains data from OS Zoomstack.

In future, methods to split polygons should be explored, for example using tools aimed at hydrological analysis such as stream order

4.5. Grid based accuracy metrics

In an attempt to overcome the limitations imposed by the ground truth labels on accuracy metrics, a grid based approach was implemented. To do this we divide the region covered by the digitised ground-truth polygons into grids of a resolution that is within a tolerated distance of model-feature offset. In each grid the presence or absence of modelled and ground-truth polygons allows us to assign a score corresponding to either true positive (both model and ground-truth features exist in the grid), true negative (neither model or ground truth polygons are present in the grid), false positive (a modelled polygon is present but a ground-truth polygon is absent from the grid) and false negative (a ground-truth polygon is present in the grid but a modelled polygon is absent) (Figure 4-8). With these scores we calculate accuracy, precision, recall and the F1-score for the final outputs.

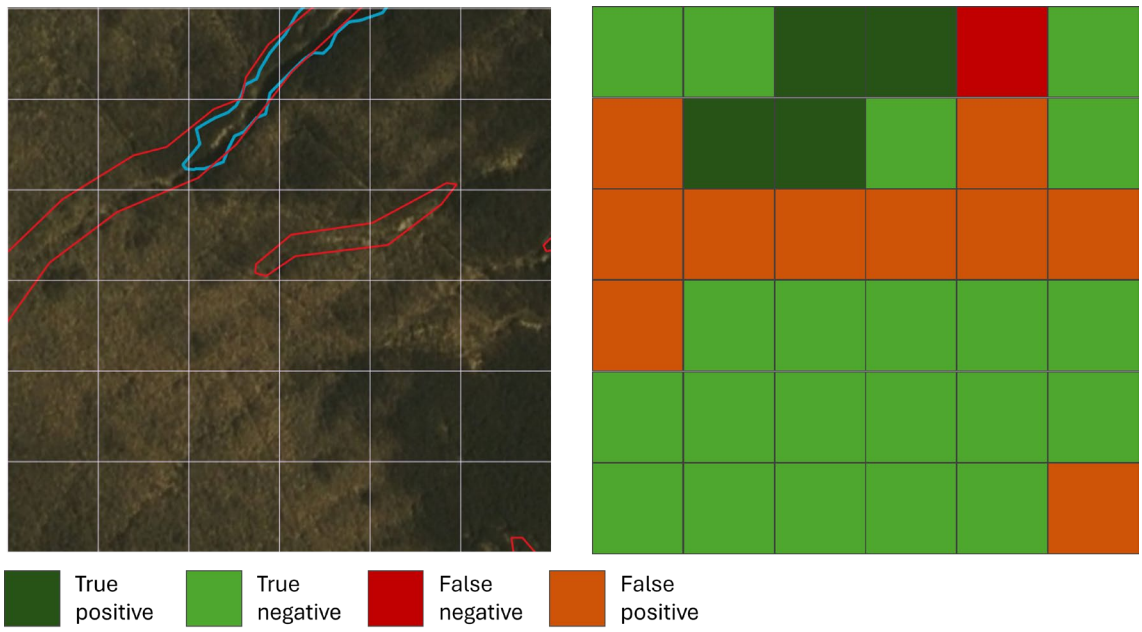

Figure 4-8 Example of the methodology used for grid-based model accuracy metrics. The blue polygon in the left image is a digitised ground-truth gully and the red polygons are modelled gully feature. The area covered by the left image is divided into 36 10 m x 10 m grids (this coarse resolution is for visualisation purposes only) and the score awarded to each grid is shown in the right image. Aerial imagery © 2015 Getmapping plc and Bluesky International Ltd.

Accuracy, precision, recall and the F1 scores are calculated for the grip and gully outputs using the improved ground-truth data located in the West Pennines (Table 4-1). The West Pennine dataset used both aerial imagery and LiDAR to manually digitise feature outlines. This differs from the line data provided by peat partnerships, which after buffering to convert to polygons are likely to be less faithful to the features on the ground (see section 7.6 of the main report). We calculate accuracy metrics for all 50 m BNG squares that include a ground-truth polygon in the West Pennine dataset (matching the training chip sizes). We then provide scores at both a 1 m and 5 m resolution for these regions.

The improved West Pennine ground-truth dataset does not include hagg data. For the hagg metrics we use the buffered line data provided by peat partnerships. The accuracy metrics are calculated using all hagg polygons in the SD 100km^2 BNG region (Table 4-1).

The metrics imply that the gully outputs are more reliable than the hagg and grip outputs with F1 scores of 79.86, 58.76 and 58.80 respectively for a 5m resolution. However, for the reasons explained above, and also highlighted by other studies (Dadap et. al., 2021, Robb et. al., 2023), we urge caution in using these metrics alone to determine the quality of the outputs.

Table 4-1 Performance metrics for the post-processed, surface feature model outputs (all in BNG 100km grid region SD) for different features. Ground truth datasets are as follows: West Pennine improved labels for gullies and grips, Peat partnership line labels buffered by 1 m for haggs.

Metric	Score (5m resolution)	Score (1m resolution)
Gullies		
Accuracy	90.06	92.36
Precision	73.45	68.56
Recall	87.48	83.52
F1 score	79.86	75.30
Grips		
Accuracy	82.77	91.67
Precision	80.87	58.53
Recall	46.14	35.55
F1 score	58.76	44.24
Haggs		
Accuracy	85.67	92.35
Precision	58.95	45.21
Recall	58.64	51.64
F1 score	58.80	48.21

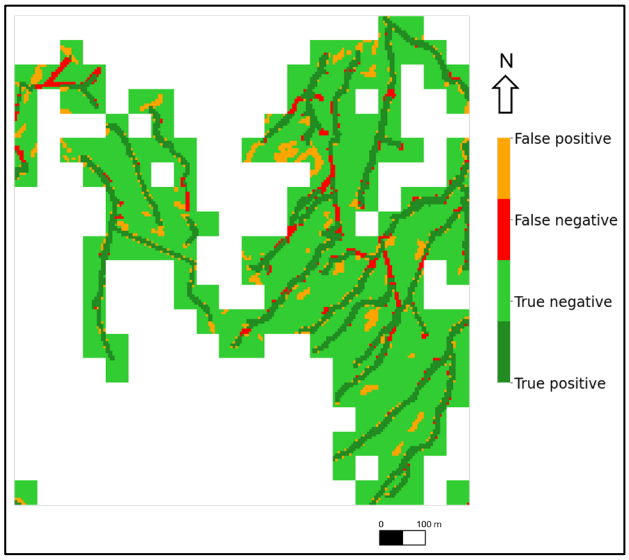

Figure 4-9 Example of the pixel-based method of calculating accuracy metrics for gullies at a 5 m resolution for a 1 km^2 region (SD6420) of the West Pennine ground-truth dataset.

4.6. AI and Quality Assurance

The use of AI and machine learning in the public sector is still in its early stages, and standardized guidance on how to apply these technologies effectively is evolving. This project has followed established scientific practices, such as thorough model validation, data quality assurance, and sound data management.

We also engaged with end users to ensure the outputs align with practical needs and expectations. However, there are inherent challenges when deploying AI in this context:

1. **Transparency and Explainability**: AI models, especially those based on machine learning, often operate as "black boxes," meaning it can be difficult to understand why a particular decision or classification was made. This lack of transparency can reduce trust in the outputs, especially for non-technical stakeholders.

2. **Ethical and Responsible Use**: The potential for misclassification or bias in model predictions raises ethical concerns. For instance, misidentified features could lead to inappropriate restoration actions or misallocation of resources. As clear guidelines for ethical AI use are still under development, we relied on general principles of fairness, accuracy, and accountability to guide our work.

3. **Uncertainty and Communication**: Despite best efforts, no AI model is perfect, and it is crucial to communicate the uncertainty in the results. Users should interpret outputs cautiously, especially in regions where training data is sparse or environmental conditions are atypical.

Recommendations

Guidance and Transparency
To maximise the usability of the mapping datasets, clear guidance that includes accountability should be provided to explain how end-users should interpret and apply the data. This should include practical use cases, limitations (e.g., accuracy issues), and instructions for restoration teams.

Accountability and Feedback
Establish robust feedback mechanisms, allowing users to report inaccuracies or concerns easily. A dedicated contact or role such as a change manager may look to oversee user queries and dataset updates after the tool is deployed. This would allow for further improvements to be made, enhancing the usability and user experience

Misuse Mitigation
Regularly updated maps can support environmental benefits, including restoration tracking, carbon offset estimation, and flood management. However, this may lead to potential misuse of the tool, such as a valuation tool in carbon offset markets.

4.7. Future Work

There are several areas in which more work could improve and expand the results presented here.

Accuracy improvements

As discussed previously, there is the potential to further improve on the surface feature datasets in several ways:

1. Including more high quality and more varied training data to improve performance for all features.
2. For dams, including additional training data with separate categories such as stone dams and bunding to allow the model to detect a wider range of dams and distinguish between them.
3. Including aerial imagery from previous years in the training data to allow the model to understand a wider variety of environmental conditions such as differing vegetation and light conditions.
4. Exploring other remote sensing datasets. For example, some initial exploratory work suggests that using data from Airbus may improve performance due to the more detailed colour information provided. In addition, using LIDAR point cloud data rather than 1m resolution gridded data may also improve the quality of the outputs.

Extending spatially and temporally

A key advantage of the modelling approach used is that the established methodology and quick run times mean new outputs can be generated relatively quickly. This opens up opportunities to extend spatially, for example to the Devolved Administrations, and also extend temporally by applying the models to previous years of aerial imagery and looking at changes over time.

Extending to lowland peat would also be beneficial to map lowland drainage. The different landscape may mean the models would need to be re-trained or the outputs filtered to remove grip-like linear infrastructure such as roads and paths which may confuse the model and are more common in the lowlands.

5. References

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

British Geological Survey 2016. *BGS Geology 50K*. Available at: https://www.bgs.ac.uk/datasets/bgs-geology-50k-digmapgb/.

Chicco, D. and Jurman, G. 2020. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation, *BMC Genomics*, 21(1), pp. 1–13. Available at: https://doi.org/10.1186/s12864-019-6413-7.

Cohen, J. 1960. A coefficient of agreement for nominal scales, *Educational and Psychological Measurement*, 20(1), pp. 37–46. Available at: https://doi.org/10.1177/001316446002000104.

Environment Agency 2011. Digital Land Utilisation Survey of Britain, 1933-1949. Available at: https://magic.defra.gov.uk/Metadata_for_MAGIC/Digital Land Utilisation Survey 1933.pdf.

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

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*, 202, pp. 18–27. Available at: https://doi.org/10.1016/j.rse.2017.06.031.

Gräler, B., Pebesma, E. and Heuvelink, G. 2016. Spatio-Temporal Interpolation using gstat, *The R Journal*, 8(1), pp. 204–218. Available at: https://journal.r-project.org/archive/2016/RJ-2016-014/index.html.

Hand, D.J. 2012. Assessing the Performance of Classification Methods, *International Statistical Review*, 80(3), pp. 400–414. Available at: https://doi.org/10.1111/j.1751-5823.2012.00183.x.

Hastie, T., Tibshirani, R. and Friedman, J. 2009. *The Elements of Statistical Learning*. 2nd editio. New York, NY: Springer (Springer Series in Statistics). Available at: https://doi.org/10.1007/978-0-387-84858-7.

Hiemstra, P.H., Pebesma, E.J., Twenh"ofel, C.J.W. and Heuvelink, G.B.M. 2008. Real-time automatic interpolation of ambient gamma dose rates from the Dutch Radioactivity Monitoring Network, *Computers & Geosciences* [Preprint]. Available at: https://doi.org/http://dx.doi.org/10.1016/j.cageo.2008.10.011.

Kopecký, M., Macek, M. and Wild, J. 2021. Topographic Wetness Index calculation guidelines based on measured soil moisture and plant species composition, *Science of The Total Environment*, 757(143785). Available at: https://doi.org/10.1016/j.scitotenv.2020.143785.

Matthews, B.W. 1975. Comparison of the predicted and observed secondary structure of T4 phage lysozyme, *Biochimica et Biophysica Acta (BBA) - Protein Structure*, 405(2), pp. 442–451. Available at: https://doi.org/10.1016/0005-2795(75)90109-9.

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(3), pp. 1–18. Available at: https://doi.org/10.3390/land10030231.

Ordnance Survey 2023. Open Rivers. Available at: https://www.ordnancesurvey.co.uk/products/os-open-rivers.

Ordnance Survey 2024. Boundary Line. Available at: https://www.ordnancesurvey.co.uk/products/boundary-line.

Pebesma, E.J. 2004. Multivariable geostatistics in {S}: the gstat package, *Computers & Geosciences*, 30, pp. 683–691.

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

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C. and Fei-Fei, L. 2015. ImageNet Large Scale Visual Recognition Challenge, *International Journal of Computer Vision*, 115(3), pp. 211–252. Available at: https://doi.org/10.1007/s11263-015-0816-y.

Trippier, B., Stefaniak, A., Fancourt, M., Clement, M., Moore, C., Woodget, A., S, P., Holmes, K., Hadfield, B., Saunders, A., Mein, R., Bowling, J. and Kilcoyne, A. 2024. *Living England 2022-23: Technical User Guide*. Available at: https://publications.naturalengland.org.uk/publication/5260859937652736.

www.gov.uk/natural-england

NATURAL ENGLAND