England Peat Map Project

Final Report

May 2025

Natural England Research Report 149

www.gov.uk/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. Natural Capital and Ecosystem Assessment (NCEA) is Defra's largest Research and Development programme. It has set up a national environmental monitoring capability and is delivering a baseline assessment of England's land, freshwater, and coastal ecosystems.

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: NERR149

1. Report details

Authors

Christoph Kratz, Andrew Webb, Sam Dixon, Chris Miller, Alex Hamer, Anne Williams, Martha Tabor, Tom Hunt, Samuel Richardson, Sarah Lamb, Nick Tomline, Michelle Johnson, Jacob Podesta, Craig Dornan, Charlie Hart, Oliver Gutteridge, Mike Prince, Oliver Power, Phil Shea, Jessica Steward, Luisa DiLucchio, Daveron Smith, Tom Raven, Rachel Wade, Ruben Douglas & Elizabeth Cooper

Project Manager

Andrew Webb

Keywords

peaty, digital soil mapping, vegetation classification, satellite imagery, remote sensing, habitat mapping, classification, machine learning

Acknowledgements

Funded by the UK Government through the Nature for Climate Fund and Defra's Natural Capital and Ecosystem Assessment programme.

We would like to thank those who shared their peatland knowledge and provided their advice, steer, challenge and expertise - members of the Defra Soils and Peatland Science Team (particularly Chloe Smale, Vittoria Lauretano, Judith Stuart and David Smedley), NE Science Directorate and wider NE colleagues (Iain Diack, Ian Crosher, Alistair Crowle, Naomi Oakley, Matthew Shepherd, Catherine Weightman and members of the NE Peatlands team), Rebecca Artz, Rodney Burton, Ben Marchant, Matt Aitkenhead, Wahaj Habib, and members of the Defra England Peat Map Expert Group. Essential technical and/or administrative support of one kind or another was provided by colleagues in the NCEA Programme Management Office, England Ecosystem Survey, and Living England teams, particularly Mike Prince, Mark Rogers, Rachel Wade, Heather Thomas, and Lizzie Mitchell.

Thank you also to those external stakeholders who we worked with to deliver the project - particularly UKCEH (Jenny Williamson, Jonay Jovani and Chris Evans), University College Dublin (Alexis Hrysiewicz, Eoghan Holohan), TerraMotion Ltd (Andy Sowter), lowland peatland stakeholders Norfolk and Suffolk Broads Authority (Andrea Kelly), Fenland SOIL (Megan Hudson, Ian Holman, Elizabeth Stockdale), and the Fens East Partnership; Moors for the Future (Joe Margetts, Steve Maynard, Chris Dean) and other Great North Bog partnership member organisations (Yorkshire Peat Partnership, Cumbria Peat Partnership, Lancashire Peat Partnership, North Pennines National Landscape, Northumberland Peat Partnership), and The National Trust (Caroline Thoroughgood, Tia Crouch).

A heartfelt thank you also to the valuable contribution made by many hundreds of private landowners, tenants, businesses and others who kindly gave us permission to survey on their land. Without your support the EPM field survey would not have succeeded.

Finally, whether we have incorporated people's data into the model training process, used it to validate the outputs or used it for reference, we are very grateful to all the groups and individuals that provided data for use in the project. There are too many to mention here but they are all listed in a separate downloadable annex.

Citation

Kratz, C., Webb, A., Dixon, S., Miller, C.J., Hamer, A., Williams, A., Tabor, M., Hunt, T., Richardson, S., Lamb, S., Tomline, N., Johnson, M., Podesta, J., Dornan, C., Hart, C., Gutteridge, O., Prince, M., Power, O., Shea, P., Steward, J. … Cooper, E. 2025. England Peat Map Project Final Report. Natural England Research Report NERR149. Natural England.

2. Foreword

The twin crises of climate and nature find common cause in peatlands. Globally, peatlands store more than twice as much carbon as the world's forests and many are fantastic wild places, supporting characteristic biodiversity and providing quiet enjoyment to thousands of people. But these are far from the only services we receive from peatlands. When intact peat retains water, helping to reduce flooding and regulating water quality. In the lowlands, large areas are used for intensive agriculture, providing some of the most fertile soils for growing crops.

When degraded, damaged or dried, however, all these services are reduced, and peatland contributes significantly to greenhouse gas emissions as carbon oxidises and returns to the atmosphere. Balancing preservation of peatlands while continuing to benefit from the wide variety of services they provide creates a range of management and policy challenges which require a robust evidence base to address. Only by understanding where our peats are, how deep they are, and their environmental status can we assess how they contribute to the nation's carbon emissions, and how land use options alter these emissions. Good data is also essential to make informed decisions about planting trees, to assess opportunities for nature recovery, and to effectively restore peatland.

For too long we've made do with very limited evidence to help us. The publication of the England Peat Map combines observations and modelling to provide a comprehensive assessment of the location and status of our peatlands. The map now provides an up to date (and updateable) map of England's peat extent, along with a first-ever published national map of peat depth, and a picture of peatland condition derived from mapped vegetation and land cover, and from upland drainage and erosion patterns. Several things have made this possible. Recent advances in the field of data science provided the team with state-of-the-art machine learning and AI tools. These were applied to the huge quantities of existing survey data assembled, with thanks to the co-operation of many stakeholders. Additional commissioned peat surveys were made possible by the agreement of hundreds of private landowners. And the availability of plentiful, openly accessible earth-observation imagery and other data provided the predictors that made the models work.

The new England Peat Map, available to all, represents a step-change in the evidence we so badly need to help tackle the big challenges of restoring nature, maintaining our carbon stores, and continuing to benefit from the many services peatlands provide.

Gideon Henderson
Defra Chief Scientific Adviser

3. Summary

Defra commissioned the England Peat Map (EPM) Project to produce a new detailed set of map products describing England's peat resources, covering peat extent, depth and condition, and making extensive use of innovative modelling techniques, earth observation and field survey data. This new evidence was required because currently available national-scale peat mapping was out of date, of low resolution and limited scope (particularly lacking peat depth). The growing need to protect and restore peatlands, as reflected in the England Peat Action Plan (Defra, 2021), and the Net Zero strategy (HM Government, 2021) and the appreciation of their value in providing a wide range of ecosystem services (IUCN, 2025) provided the fundamental requirements for creating the new evidence.

Investment in EPM is expected to provide the supporting evidence for wide-ranging uses. Expected use-cases include the development of peat policy by Defra, future targeting of peatland restoration, the estimation of greenhouse gas emissions from peatlands, the comprehensive description of peatland condition, and the provision of opportunities for further research and development.

The main EPM products are:

- **Extent model**: a map of the predicted extent of peaty soils (see glossary), together with a map of the probability of occurrence of peaty soil, from which the predicted extent was generated
- **Depth model**: a map of the predicted depth, or thickness, of peaty soil, together with a map of the confidence we have in those depth predictions
- **Vegetation and land cover model**: a map of predicted vegetation and land cover classes on peaty soils, together with a map of the probabilities associated with the different classes at each location
- **Upland peat erosion and drainage features models**: maps of predicted surface features associated with peatland drainage and erosion in the uplands - grips, gullies, peat haggs and grip dams
- **Survey data**: a collection of the different types of survey data used to train and validate the models, including all the survey data commissioned directly by the Project, and as much of the survey data collated from other sources as our data licences allowed.

These products have been created using machine and deep learning modelling techniques which combine earth observation and other predictor data available at national scale with extensive field survey observations. We have demonstrated that these methods can be used effectively to map England's peatlands. The outputs represent a major improvement in our knowledge of England's peat resources, providing detailed new information of known quality, which is available to everyone. This new evidence will support progress in peatland protection and restoration in England for many years to come.

Following an initial phase of project definition with Defra in 2021, peat survey data collation from across the peatland community began and continued until the end of the project (in the expectation that data arriving too late for inclusion in this release can be considered for inclusion in future updates). The collation and use of the pre-existing peat survey data was a highly effective means of sourcing valuable data to train and validate the extent, depth and condition models. Some upland stakeholders had particularly valuable data holdings which we were able to access for the project. Because the distribution of these data was far from ideal from a modelling perspective we supplemented it with our own field surveys. A pilot survey was conducted in the first year of the project, informing the final design and delivery of commissioned field surveys which took place from 2022-24. We targeted surveys at areas of greatest need, aiming to ensure that model quality reached acceptable standards. Modelling approaches were researched and developed during 2021/22, and applied to the development of initial test outputs and national 'Beta products' from 2023.

Compared to the previous best available evidence, derived from the *England Peat Status GHG and Carbon Storage* layer (Natural England, 2010) these models indicate that peaty soil covers an area of 11,047km^2 in England, compared to 12,071km^2 previously. The 8.5% difference in these area estimates should not be interpreted as a reduction or loss of peat over time, as the methods used to calculate each were very different.

Despite incorporation of tens of thousands of peat survey records from across the peatland community, with supplementary extensive field survey, we recognise there remain uncertainties and limitations in the models. In particular, there is uncertainty in the lowlands, where there is less data (both survey data for training models, and good predictors of peaty soil). There will be some instances where areas of known peaty soils have not been predicted by the extent model, conversely areas where the extent model has confidently, but wrongly, predicted peaty soil occurrence. Where known peaty soil sites are not predicted this will normally reflect a lack of survey data for these areas and is something which we aim to address in a future update. Vegetation types and surface features were generally reliably detected, although here too, data limitations constrained the accuracies we were able to achieve. Some vegetation and surface feature types were inevitably more difficult to reliably identify, although in general terms (according to feedback from users, model accuracy metrics, and analysis of model outputs in comparison with other peat datasets) results were highly satisfactory.

Model uncertainties can be reduced with future iterations using additional data, and this is our intention for the extent and depth models. We do not consider it possible to use these models to map actual change in peaty soil depth and extent, because the model uncertainty is considerably greater than the anticipated change in peaty soil extent and depth over time. By contrast, iterating the condition models (vegetation and surface features) with new data is likely to start to pick up real change in condition, so these are more suited to the development of new approaches to monitoring.

We have worked closely with other peatland experts and modellers in the delivery of this project. Upland peat erosion and drainage features (also referred to as ‘surface features’) were modelled and delivered by Natural England and Defra colleagues in the AI4Peat Project. In addition, we worked closely with UK Centre for Hydrology (UKCEH) on a study of peat surface motion and its relationship with peat condition. This work established a new network of ground motion sensors ("peat cameras") and explored the ability of both peat cameras and satellite radar to describe peat condition. Further work is needed, as time-series data is currently insufficient to draw robust conclusions, although results are encouraging for raised bogs and some blanket bogs.

The model outputs are made available under the Open Government Licence (OGL), and accessible to view and/or download via a simple web viewer available at defra.maps.arcgis.com, together with survey data, summary statistics, and other project information. Models are published on Github (www.github.com/naturalengland) and peat camera data (also OGL licenced) can be accessed via an app at peatcam-shinyapp.datalabs.ceh.ac.uk.

A set of additional needs has been identified through user consultation and project discussions, and these are outlined as part of a set of potential next steps.

Contents

4. Introduction and Overview

4.1. Background

The primary reason for commissioning the England Peat Map (EPM) was a lack of up-to-date and detailed evidence on the state of England's peat resources, from which to develop future environmental policy. The development of a new peat map was therefore identified as a commitment in the England Peat Action Plan (Defra, 2021). Project funding was secured from the Nature for Climate Programme, to deliver the work as part of Defra's Natural Capital and Ecosystem Assessment (NCEA) Programme.

Project design and approach were strongly influenced by a range of factors. Given the large total area of peaty soils in England and their widespread distribution, a solely field-based approach to mapping was considered prohibitively expensive, and delivery would have been unachievable in the 3-year project timeframe due to limited numbers of surveyors. Consequently, a modelling approach was required. Also, because new peat survey data will continue to be produced from a variety of activities including restoration projects, monitoring, infrastructure development and planning, there are clearly opportunities to improve the models over time. However, whilst providing an excellent potential resource for model development, the data from these survey activities are not optimally distributed geographically - they are clustered unevenly in what amounts to a small portion of the overall peatland area. This leaves large areas of peat un-surveyed, and under-represented in the training data required for model development. So we recognised the need to commission new field survey as part of the project. In addition, the need to work collaboratively with peatland stakeholders necessitated the development of maps that could be readily shared, and used, by everyone - including information about their quality so that users could make informed decisions about appropriate uses. The EPM was envisaged as a new map that would not just enable the development of better environmental policy, but support a wide range of use cases such as better targeting of peatland restoration, improved reporting of greenhouse gas emissions from peatlands, identification of priority areas for nature recovery, future condition monitoring approaches and land use planning.

The new EPM Project was therefore designed with the following features:

- The use of state-of-the-art statistical modelling techniques to create the core products
- Strong reliance on the use of extensive, but scattered, pre-existing field survey data held by Defra organisations and collated from external stakeholders

- Supplementation of the pre-existing field survey data with new data from commissioned field survey, to achieve better representation of the different types of peat and peatlands across England in the models
- Incorporation of detailed, plentiful and openly accessible earth observation imagery - particularly from the EU Copernicus programme, with its high cadence and spatial resolution
- Providing a platform on which to build future work, particularly the development of new condition monitoring approaches using earth observation
- Publication of map outputs under the Open Government Licence, for everyone to use

Prior to EPM, the most commonly requested and used evidence of England's peat resources was the 'Peaty Soils Location' (PSL) map created by Natural England in 2008 (and currently available under a non-commercial government licence) as an accompaniment to the NE257 report, "England's Peatlands: Carbon Storage and Greenhous Gases" (Natural England, 2010). This dataset compiled mapped peat data from several existing sources:

- National Soils Map (digital version) © National Soils Resources Institute, Cranfield University. NSRI (2005).
- Natural England (2008). Biodiversity Action Plan Priority Habitat Inventory Mapping, Natural England, Sheffield.
- British Geological Survey. 1:50,000 scale BGS digital data (Superficial Geology), © Natural Environment Research Council, Licence 2006/072.

Peaty soils were classified in the PSL dataset into three types based on the source of the information but assumed to have the following general characteristics:

- Deep peaty soils: Areas covered mostly with peaty soils >40cm deep
- Shallow peaty soils: Areas covered mostly with peaty soils 10–40cm deep
- Soils with peaty pockets: Areas of mostly non-peaty soils, supporting smaller pockets of peat (such as flushes or exposures of buried peat) too small to map at a national scale.

This data was subsequently improved in 2012 with more detailed peatland condition and usage information designed to address a lack of evidence on the extent of moorland drainage, burning, erosion, and other condition features in upland England. This new data was derived from air photo interpretation and other sources (Penny Anderson Associates, 2012). Ground truthing of a sample of sites identified limited accuracy (61%) of the assigned condition information.

Bringing these sources of data into one place to give a coherent picture of peaty soils in all of England made the PSL map a valuable product. There are, however, a number of limitations:

(i) Old source data - Much of the field survey data underlying the source maps dated from the Lowland Peat Survey 1987 (Burton and Hodgson, 1987).

(ii) A small number of broad classes – deep peaty soils, shallow peaty soils, soils with peaty pockets. Deep peaty soils covers everything from 40cm to over 10m deep peats.

(iii) Limited usable information at local scale - 'Soils with peaty pockets' describes sometimes large areas within which some peat would be found, but without saying where or how much.

(iv) A lack of information about quality or accuracy of the data, either as separate sources, or combined in the final product.

(v) Unreliable information about condition.

(vi) Licence terms and conditions that restricted sharing and use. From release until 2021 users had to request a copy from Natural England. It was eventually made available on the Defra Data Sharing Portal (DSP) under the Non-commercial Government Licence in 2021.

These datasets have been the basis of various analyses of peat and peatlands in England such as:

- Towards an assessment of the state of UK Peatlands - JNCC Report No 445 (Higgins, 2011)
- IUCN UK Commission of Inquiry on Peatlands (Bain and others, 2011)
- Lowland peat systems in England and Wales - evaluating greenhouse gas fluxes and carbon balances (Evans and others, 2016)
- Implementation of an Emissions Inventory for UK Peatlands (Evans and others, 2017)
- The role of earth observation in an integrated framework for assessing peatland habitat condition and its impact on greenhouse gas accounting (Williamson and others, 2017)
- Commission of Inquiry on Peatlands : The State of UK Peatlands - an update (Artz and others, 2019)
- The UK Natural Capital: Peatlands (Office for National Statistics, 2019)

A common thread in these reports is the lack of robust, consistent and up to date evidence on peat resources, although the evidence used was the best available at the time. This reduces confidence in report conclusions, creating uncertainty and obstacles to progress in policy development and implementation.

In addition to realising the limitations of existing evidence sources, there was a growing appreciation of the power of machine learning, deep learning (Artificial Intelligence) tools, earth observation data and modelling approaches in mapping natural resources. In academia, the development of Digital Soils Mapping as a discipline was also growing, leading to new studies describing national peat maps in a variety of countries (Minasny and others, 2019).

4.2. Initiation and Delivery

It is in this context that the need for a new England Peat Map was identified. Following exploratory discussions between Defra Peat team and Natural England representatives in 2019, Natural England were commissioned to produce the England Peat Map by Defra Soils & Peatlands Science Team. Original aims for the Project were identified as follows:

1. Support delivery of the England Peat Action Plan by developing improved baseline evidence of the extent, depth and condition of peat across England, leading to:
2. Improved spatial prioritisation and targeting of peat restoration activities
3. Improved reporting of greenhouse gas emissions from peat
4. Better support for a range of other policy and delivery areas (e.g. Net Zero, Local Nature Recovery Strategies, natural capital accounting, appropriate tree planting, future land management schemes)
5. Provision of open data and methods which can be scrutinised (and therefore trusted), improved, and updated
6. Development of better methods for future peatland monitoring

Benefits from the approach taken, as described previously, would include:

- The first published England-wide map of predicted peat depths, extent and conditions
- Provision of up-to-date evidence whose accuracy can be described and shared
- A means of updating and improving models and outputs in future years
- Provision of more detailed (higher spatial resolution) mapping of peat than previously available
- Cost-efficiency from a reduced survey requirement
- Development of new approaches to describing and understanding peatlands and their condition (e.g. surface motion studies)
- Application of state of the art modelling approaches
- Exploitation of the value of existing survey data
- Integration with other NCEA Projects, particularly Living England

- Development of a range of technical capabilities in NE, and capacity to deliver: Earth Observation science, modelling, peatland field survey techniques, and improved data licencing / management
- Development of productive stakeholder relationships facilitating improved partnership and collaboration in future
- The ability to share EPM methods and outputs widely, fostering greater trust, greater data sharing and use, and helping avoid 'multiple versions of the truth'.

The EPM Project was formally established in January 2020, delivered by a project team in Natural England, with scientific steer provided by an expert panel, and governance provided via the NCEA Programme and Defra Soils and Peat Evidence Team.

Scope

The scope of the project included is wider than not just soils consisting purely of **peat**, for which there is no universally accepted definition (Lourenco and others, 2023), but encompasses all **peaty soils** where organic content may be 20% (see Glossary). Surface deposits of peaty soils, and deposits of buried peat within 1m of the ground surface were included in the field survey and modelling. Buried peat was not identified as a separate type or category in any of the models, and was incorporated within the extent and depth models along with surface deposits. Similarly, wasted peat (see Glossary) was not identified or mapped separately – wasted peat will be incorporated within the extent and depth models (but could perhaps be estimated from areas of lowland peaty soil under intensive agricultural land cover where depth is shallow).

Semi-natural and natural bog and fen vegetation classes were combined with land cover classes in a single model. This reflected both the expectation that we could resolve the desired semi-natural and natural classes in our modelling, and the availability of existing mapped sources for the other main land cover types we required.

Intertidal peat deposits were not surveyed or modelled – this would be particularly challenging given the modelling approach used. Peat deposits whose top is more than 1m below the ground surface were not mapped – the 1m threshold is an arbitrary but pragmatic depth limit which allowed field survey to proceed efficiently and focussed the modelling on surface peat deposits most at risk of further degradation.

AI4Peat

In December 2021, a project team led by Natural England, including members from across the Civil Service, competed in and won the Civil Service Data Challenge with a bid to use deep learning (a method of machine learning or artificial intelligence) to identify and map upland peatland artificial drains ('grips'). This became the 'AI4Peat' Project, managed and delivered by staff in NE's

Analysis Directorate and Defra Soils and Peatland Science team. AI4Peat undertook the technical modelling work, with support from EPM Project, providing the EPM Project with the core outputs for upland surface drainage and erosion features. The two projects co-ordinated activities and worked collaboratively to deliver the models and map layers for upland moorland grips and grip blocks, gullies and peat haggs (see Glossary for definitions). These features were chosen because they are the most prevalent erosion and drainage features in the uplands and there was a wealth of available field data. Lowland drainage features such as field drains are harder to map as they are often buried and more difficult to detect using image recognition approaches, especially with limited field data.

Ground Surface Motion Pilot Study

An original objective of the project was to explore the potential use of ground surface motion (the phenomenon where the peatland ground surface rises and falls due to changing water and gas storage within the peat, also known as 'bog breathing') as an indicator of peatland condition. If clear patterns of surface motion could be established, and correlated with satellite data and condition metrics, this could lead to the development of a new low-cost method of monitoring peatland condition consistently over wide geographical areas. A pilot study was commissioned during 2021/22 with the UK Centre for Ecology and Hydrology (UKCEH). This work established time series data from a set of ground surface motion sensors (aka "peat cameras") and explored patterns of ground motion behaviour in peatlands of different types, in relation to peat condition metrics and satellite interferometry. This work will be published separately and data from the peat cameras made openly available.

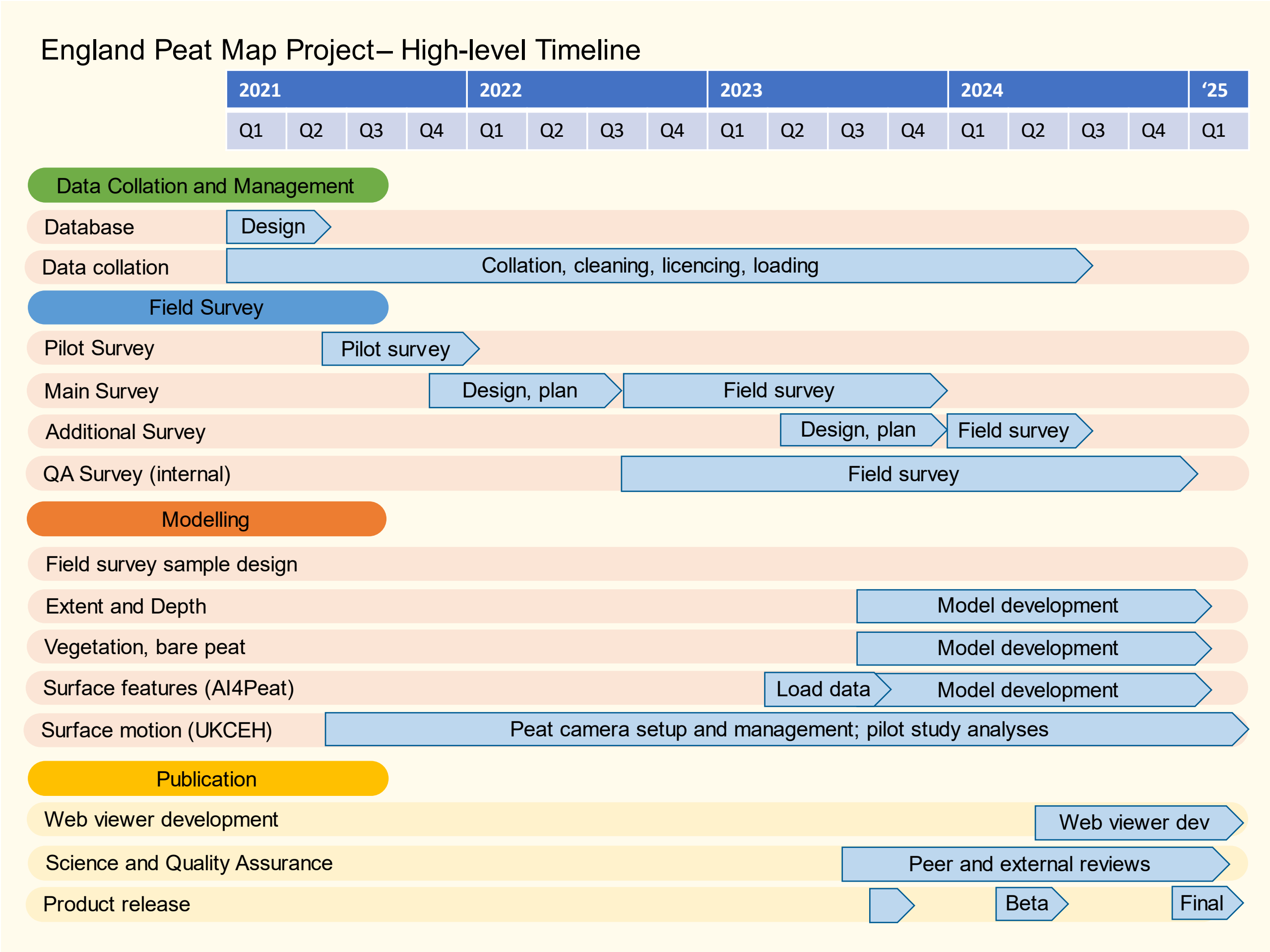


Figure 4-1 High-level timeline for the England Peat Map project

4.3. Outputs

Published Reports

- England Peat Map Project: Final Report (this report), and supplementary Annexes:
  - Annex 1: Field survey protocol – soils
  - Annex 2: Field survey protocol – vegetation
  - Annex 3: Peat probing guidance
  - Annex 4: Loss on ignition protocol
  - Annex 5: Data Supplement
  - Annex 6: Technical Supplement
- England Peat Map User Guide.
- A report by UKCEH detailing findings from analysis of peat camera time-series data, and its relationships to satellite radar data and peat condition.

Published Maps

GIS Layers showing results of models for

- Peaty soil extent and probability

- Peaty soil depth and confidence
- Vegetation and land cover type on peaty soil
- Surface features on peaty soil in the uplands
- Bare peat in the uplands

GIS layers showing peat survey data used in model training and validation

- EPM-commissioned field survey (soils and vegetation)
- Pre-existing survey data collated from third parties

Database

- Soil and vegetation survey data collected by NE as part of the EPM project as well as other survey data collated by the project, along with licenses.
- Data from the network of c.49 peat camera sensors deployed on peatlands across England. Provided by UKCEH.

Models, Code and Tools

- Trained machine learning models used to predict the mapped outputs.
- Code and methodologies used to create the models.
- Tools and instructions to enable future revisions to be made to the extent and depth models by incorporation of new data.

Data Sharing Portal

Web viewer providing an interactive map of the outputs, as well as the tools to download published GIS data.

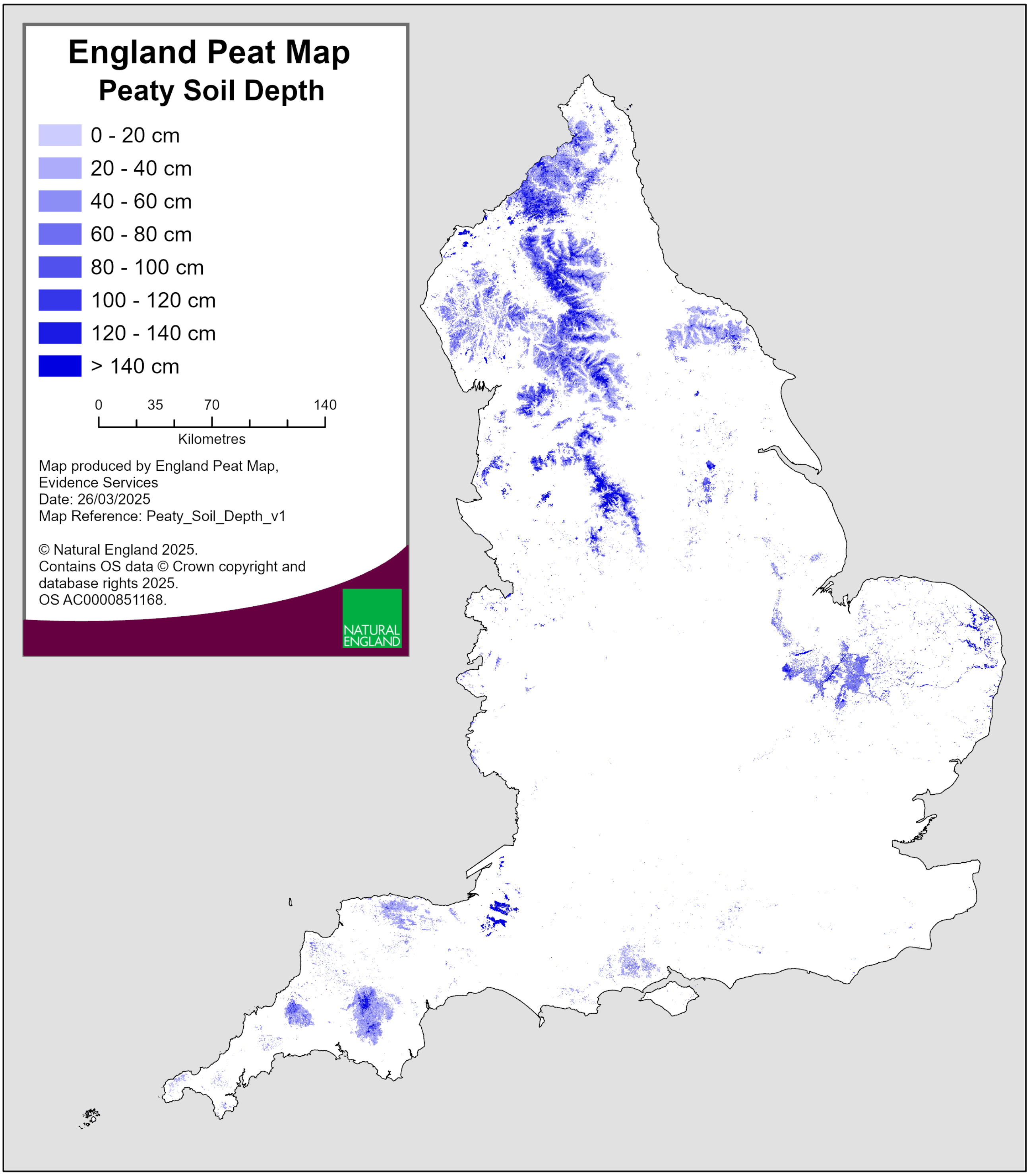


Figure 4-2 England Peat Map: peaty soil depth

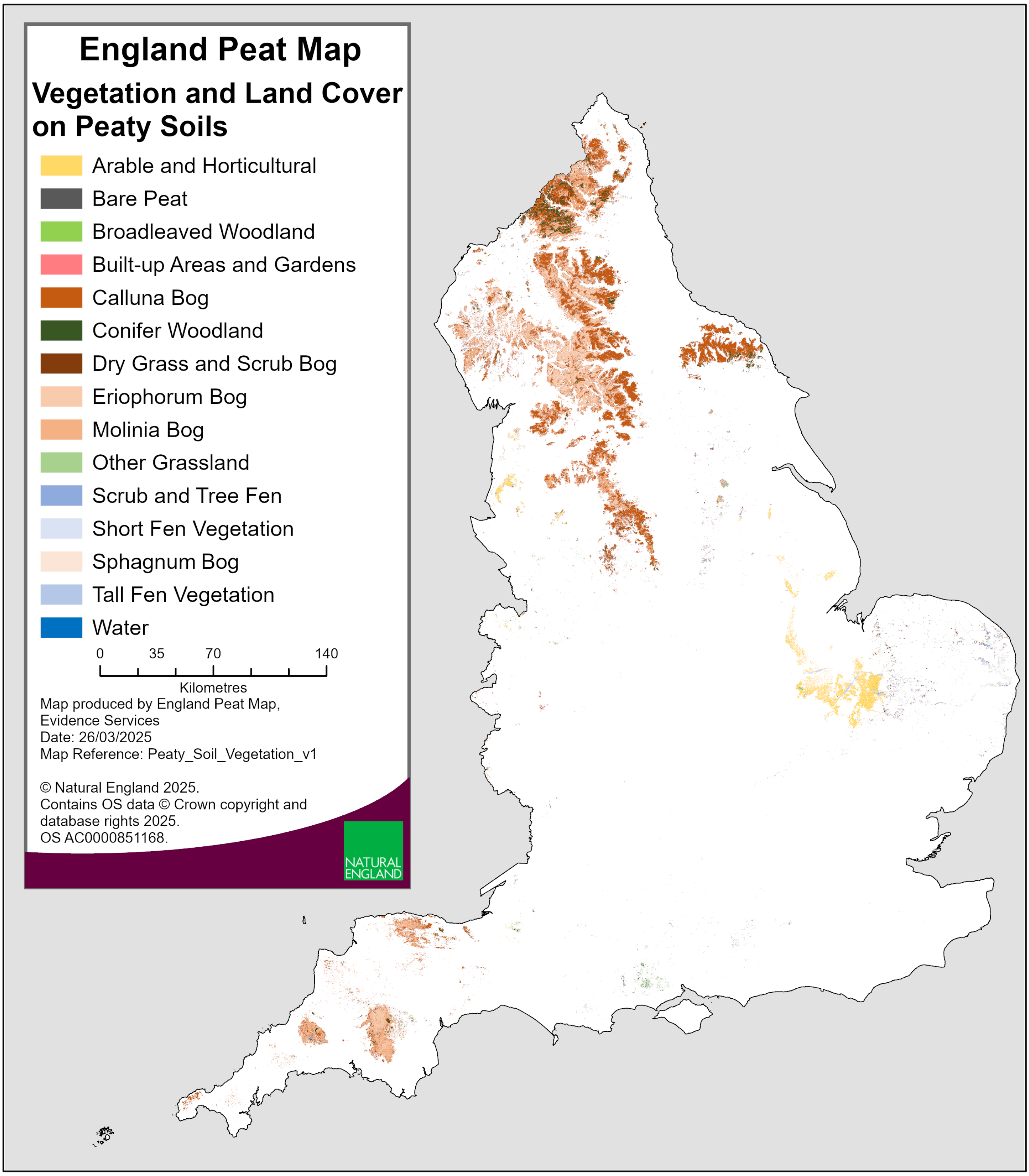


Figure 4-3 England Peat Map: vegetation and land cover on peaty soil

5. Field Survey

5.1. Survey data requirements

The purpose of the field survey was to collect high quality field data which could be used to support modelling of the extent, depth and condition of peaty soils in England. To ensure value for money, we targeted the field survey at areas where previous data collection had not occurred. However, it was not possible to obtain and licence all 3rd party data in advance of undertaking the field survey. Assumptions were therefore made about where gaps in third-party data were likely to be, and the field survey was targeted on these areas.

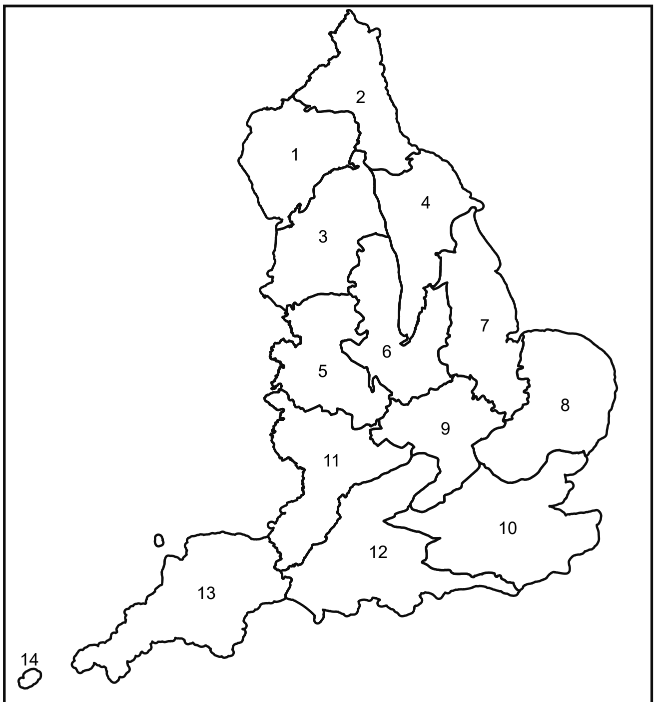


Figure 5-1 Map of England showing the 14 Biogeographic Zones (BGZs) used in the Living England Project

To reduce the potential impact of regional variations on the data models, the field survey was divided into the 14 Biogeographic Zones (BGZs) used by the Living England project (Trippier and others, 2024). BGZ 14 (Isles of Scilly) was not surveyed due to the high cost and small land area. Survey targets were established for each of the 13 BGZs addressing key evidence gaps (Figure 6-1).

Where possible, survey sites were randomly distributed within each BGZ to avoid sampling bias. Due to challenges in securing access permissions, we were restricted in which sites could be surveyed, resulting in a degree of selection bias.

To enable effective modelling the field survey was aligned with Sentinel 2 satellite data. We used a 10 by 10 metre survey quadrats orientated on the British National Grid. We used sub-metre accurate GPS units to locate these. We took multiple measurements of peat thickness in each quadrat to allow an average thickness to be used for modelling.

5.2. Field Survey Pilot

To help develop the England Peat Map field survey, a pilot survey was run at the start of 2022 to test the field survey method and associated/supporting tools, processes, data quality, and data management. We conducted and commissioned field surveys at a range of pre-selected sites broadly representative of the peat types and conditions that were expected to encounter in years 2 and 3 of the Project.

The pilot field survey was undertaken in Biogeographic zones 3 & 7 as these zones were considered to contain a good representation of the peatland types likely to be encountered by the full field survey. This includes lowland agricultural, upland, intact, degraded, deep, shallow and pockets of organic soil. The pilot resulted in changes to the field survey design and alignment with the UK Peatland Code, as well as improvements to access permissions, training, equipment, and the field survey app.

5.3. Sampling requirements and approach

The sampling requirements were developed from discussion with the England Peat Map soil & vegetation modelling team, members of the Peat Map Expert Group (an advisory group set up to provide strategic direction and oversight), and drawing on experience gained from running the field survey pilot. Due to differing needs the field survey was split into two distinct surveys: vegetation and soils.

Vegetation Survey Development

Following discussions with UKCEH we adopted a vegetation classification framework aligned to land cover hierarchy for UK peatlands, proposed for use by the UK Greenhouse Gas Inventory (Table 5-1). It retains the classes of peatland vegetation required for GHG emissions reporting while also accommodating peatland restoration. We also sought to ensure that the new classes have the potential for identification using remote sensing. Whilst there is existing data for the spatial distribution for several of the vegetation classes in the framework (Table 5-1 Other Categories) there were some notable evidence gaps, particularly semi-natural bog and semi-natural fen. The EPM field survey therefore focussed survey activity on these classes in particular.

The EPM survey collected quadrats where the following vegetation classes were dominant: *Sphagnum sp.* dominated bog; *Eriophorum sp.* dominated bog; *Molinia caerulea* dominated bog; *Calluna vulgaris* dominated bog; Dry grass / scrub dominated bog; Short vegetation dominated fen; Tall vegetation dominated fen; Scrub/tree dominated fen. ‘Dominant’ was defined as having 60% or more cover of the target vegetation to allow for natural variation in cover estimates between surveyors. Although some of the vegetation classes occur on non-peaty soils (e.g. *Calluna vulgaris*), vegetation data was only collected where a peaty soil was present.

Table 5-1 Adapted vegetation framework for the England Peat Map based on the UK peatlands emissions inventory (Evans et al., 2017) with additional land cover classes commonly found on UK peatlands. Collated datasets used in our framework are listed in Other Categories.

Semi-natural bog	Semi-natural fen	Other categories
Sphagnum sp Dominated	Short vegetation dominated	Broadleaved Woodland (NFI)
Eriophorum sp. Dominated	Tall vegetation dominated	Coniferous Woodland (NFI)
Molinia caerulea dominated	Scrub/tree dominated	Water (OS)
Calluna vulgaris dominated	-	Built-up areas and gardens (OS)
Dry grass/scrub dominated	-	Arable and Horticultural (CROME & ALC)
Bare peat	-	Other Grasslands (LE)

Some aspects of the survey protocol were introduced specifically to allow field survey data to be effectively used for remote sensing applications. All vegetation data was collected from 10x10 metre quadrats aligned with Sentinel-2 data. A 20-metre buffer was placed around each survey point to avoid confusion between data points. Aerial foliar cover was used instead of the more usual total percentage cover (understorey of Sphagnum sp. was added mid-survey due to the challenges of locating Sphagnum sp. dominated areas). Cover was estimated to the nearest 5% because the large quadrat size made estimating to the nearest 1% inaccurate. Species groups (e.g. *grasses*) were used because identification down to species level was impractical over such a large quadrat area. The exception to this approach was for key habitat condition indicator species; for example *Molinia caerulea* was separated from the grass group.

Soil Survey Development

The approach to the EPM soil survey has been shaped by the need to understand the extent and depth of organic soils across England. In line with previous work on mapping peatlands (Natural England, 2010), the field survey has been focussed on locating **peaty soils** rather than being restricted to peat alone to ensure carbon rich shallow soils are not overlooked. Peaty soils are a group of soil texture classes which includes peat, loamy peat, sandy peat, peaty loam and peaty sand textures (Natural England, 2008). “Soils are classified 'peaty' if they contain more than 20% organic matter (OM) (25% OM for soils with more than 50% clay content)” (Natural England, 2008). For the peat soil texture class, a layer of peaty soil with a minimum thickness of 40cm in the top 80 cm of soil is also required for it to be defined as peat. However, the minimum thickness used to define peat varies significantly (10-70cm) between countries (Bord na Mona, 1985; Food and Agriculture Organization, 1988; Joosten & Clarke, 2002; Lourenco and others, 2023). Due to the lack of a consistent definition, and to increase the range of uses of the mapping products, we did not use depth to define peaty soil. Instead, percentage organic matter was used to differentiate between the soil texture classes, with *peaty* being defined as more than 20% organic matter (Natural England, 2008).

The soil survey fieldwork was targeted to take place on types of peatlands and peaty soils that we had least data for, primarily lowland peatlands, shallow peaty soils, small peatlands and the margins of larger peatlands. We created two types of sampling pattern, clusters and transects of candidate points, either of which were designed to take one workday to survey. A **cluster** contains 15 candidate survey points within an approximate circle of roughly 1 km diameter. Surveyors were asked to survey at least eight candidate points using selection rules described in the field survey protocols. The candidate points were generated before the survey and placed at random within a cluster area, with some buffering of the margins of the area and between the points. Clusters were used where we had prior knowledge of a gradient in the landscape. A **transect** comprises eight quadrats, five of which are along a 1 km long line and the remainder slightly offset near the start, middle and end of the line. There are also 36 depth-only measurement locations. Surveyors were asked where possible to survey all quadrats and depths. Transect lines were placed at random in areas of likely transition (e.g. between shallow and deep peaty soils), with their direction aligned with the likely gradient of transition. This approach helped gather evidence on where the characteristics of peaty soils change across the landscape.

We adopted a method for soil survey which combined limited peat coring with full peat probing, aiming to strike a balance between accuracy and efficiency. Peat probing, whereby a rigid or semi rigid pole is inserted into a peaty soil, is a rapid and cheap method for determining the thickness of organic rich soils but can be inaccurate (Parry and others, 2014). In contrast using a soil corer to extract peat cores down to the underlying mineral soil is very reliable. However, peaty soil deposits can be over 10 metres in depth (Lindsay, 2010) which makes soil coring extremely time consuming and costly. In the EPM soil survey protocol, at each quadrat surveyors take a single soil core down to 40 cm, and if peaty soil is present, use peat probes at five locations within the quadrat to estimate thickness. Soil core depth was limited to 40cm to allow identification of peat soil using current definitions, but minimised time spent coring, increasing the amount of data points that were collected. Where buried peaty soil was expected, the depth maximum was increased to 100 cm to aid better understanding of its spatial distribution. To reduce the impact of peat probing inaccuracy five measurements were collected within a 10x10 metre area and averaged.

Due to a lack of suitably trained soil surveyors available in the market, detailed characterisation of the peaty soil was replaced with soil texture identification. Bespoke soil texture classification guidance was produced to improve field identification. In addition, soil sample collection and laboratory analysis (loss on ignition) was added to the protocol to ensure accurate identification of peaty soil.

5.4. Survey delivery

The EPM field survey was delivered by contractors in two phases: a Main Survey in 2022/23 and an Additional Survey in 2024. The Main Survey contract included an equipment purchase component and contracts for both survey phases included securing access permissions. In addition, Natural England carried out a programme of surveyor training and quality assurance surveys.

Main Survey

The contract for the main EPM field survey was awarded to Fera Scientific Ltd. through a long-term service agreement and was undertaken between October 2022 and March 2024. Vegetation survey training for the Main Survey was delivered in December 2022 on Thorne Moor, South Yorkshire and Wicken Fen, Cambridgeshire. The vegetation survey took place between 25 January 2023 and 28 March 2024. Quality assurance checks were undertaken on this work between 07 February 2023 and 24 May 2024. The soil survey started late due to delays in equipment and access permissions. Soil survey training took place in January 2023 on Engine Farm & Darlow's Farm, Cambridgeshire. The soil survey took place between 18 April 2023 and 28 March 2024. Quality assurance checks were undertaken on this work between 26 June 2023 and the 23 May 2024.

Both the soil and vegetation surveys were distributed across biogeographical zones 1-13 (Trippier *et al.*, 2024). BGZ 14 (Isles of Scilly) was not surveyed due to cost and small land area. Also, the lack of known peatland habitat in most of BGZ 9 precluded the vegetation survey from this area.

Sites for vegetation survey were selected based on expert knowledge and examination of aerial photography, on the basis that they should contain at least one of the required vegetation classes. Additional sites were added to target specific shortfalls in data collection.

Sites for soil survey were selected based on known gaps in the evidence base and targeted at lowland peatlands, shallow peaty soils, small peatlands and the margins of larger peatlands. Sites were distributed across BGZs 1-13 to ensure a good geographic spread of survey points. We generated more potential soil survey sites than were required to allow for a proportion of landowners refusing access.

Delivery of the main EPM field survey was slower than expected due to a number of issues. In May 2023 a revised delivery target was developed with the contractor to more accurately reflect how may surveys could be completed. However, ongoing issues prevented this revised target from being delivered. In total, 342 soil surveys (8 quadrats per survey) were completed and 2,330 vegetation surveys (one quadrat per survey) - significantly below both the original and revised targets (Table 5-2). In addition, the scope of soil sample collection had to be reduced significantly from the contract specification, which had an impact on the quality of data being collected.

Table 5-2 Main Survey: Target versus delivery

Quadrats surveyed:	Original Target	Revised Target	Delivered	Percent Delivered Original	Percent Delivered Revised
Soil Survey	935	550	342	36.5%	62%
Vegetation Survey	2600	2,525	2,300	88.4%	91%

Additional survey

Because the Main Survey did not provide sufficient data for both the vegetation and soil models a contract for an Additional Survey was awarded to ICF via the Monitoring, Evaluation & Learning framework, and was undertaken between July and October 2024.

The soil survey requirement was split into 3 groups of locations: BGZs 7 and 8 had a significant shortage of data; BGZs 5,10,11 and 12 had a moderate shortage of data; and the remaining areas had a minimal data shortage. Increased delivery costs led to only groups 1 and 2 being surveyed with survey work prioritised for BGZs 7 and 8.

The vegetation survey was targeted at collecting data from under-represented vegetation classes within each BGZ as summarised in Table 5-3. Additional data was not required for Biogeographical zone (BGZ) 9, and BGZ 14 (Isles of Scilly) was not surveyed.

Table 5-3 Vegetation Survey Priorities for Additional Survey

BGZ	Vegetation Class
1	Scrub/Tree Bog
2	Scrub/Tree Bog
3	Sphagnum sp.; Eriophorum sp.; *Molinia caerulea*; Calluna vulgaris; Scrub/Tree Bog
4	Sphagnum sp.; Scrub/Tree Bog
5	Sphagnum sp.; Scrub/Tree Bog. Short & Tall Fen
6	Calluna vulgaris; Eriophorum sp. Bog
7	Short Fen; Tall Fen; Scrub Fen
8	Sphagnum sp.; Eriophorum sp.; Molinia caerulea; Calluna vulgaris; Scrub/Tree bog; Short Fen; Tall Fen; Scrub Fen.
10	Sphagnum sp.; Eriophorum sp.; Calluna vulgaris; Scrub/Tree bog
11	Sphagnum sp.; Eriophorum sp.; Molinia caerulea; Calluna vulgaris; Scrub/Tree Bog
12	Sphagnum sp.; Eriophorum sp.; Calluna vulgaris Bog
13	Eriophorum sp.; Calluna vulgaris Bog

A total of 68 soil surveys (8 quadrats per survey) and 376 vegetation surveys (one quadrat per survey) were successfully completed. Significant delivery issues were experienced which despite contract extensions only resulted in 41% of the required soil surveys being delivered (Table 5-4).

Table 5-4 Additional Survey: Target versus delivery

	Original Target	Delivered	Percent Delivered
Soil Survey	164	68	41.4%
Vegetation Survey	376	376	100%

Field Survey Results

Overall, 4,475 soil surveys (3,918 main & 557 quality assurance) were undertaken between April 2023 and October 2024. Due to quality issues only 3,537 soil surveys were used by the project (see section 6.4 for further details). The locations of these surveys is shown in Figure 5-2 and a summary of the soil texture class observed at the surface can be found in Table 5-5.

Table 5-5 Soil Survey: Soil Texture Class at Surface

Soil Texture Class at Surface	Number of Survey Points
Unsure	14
Mineral	1,273
Organo-mineral	1,043
Peaty Loam or Peaty Sand	227
Loamy Peat or Sandy Peat	233
Peat	747

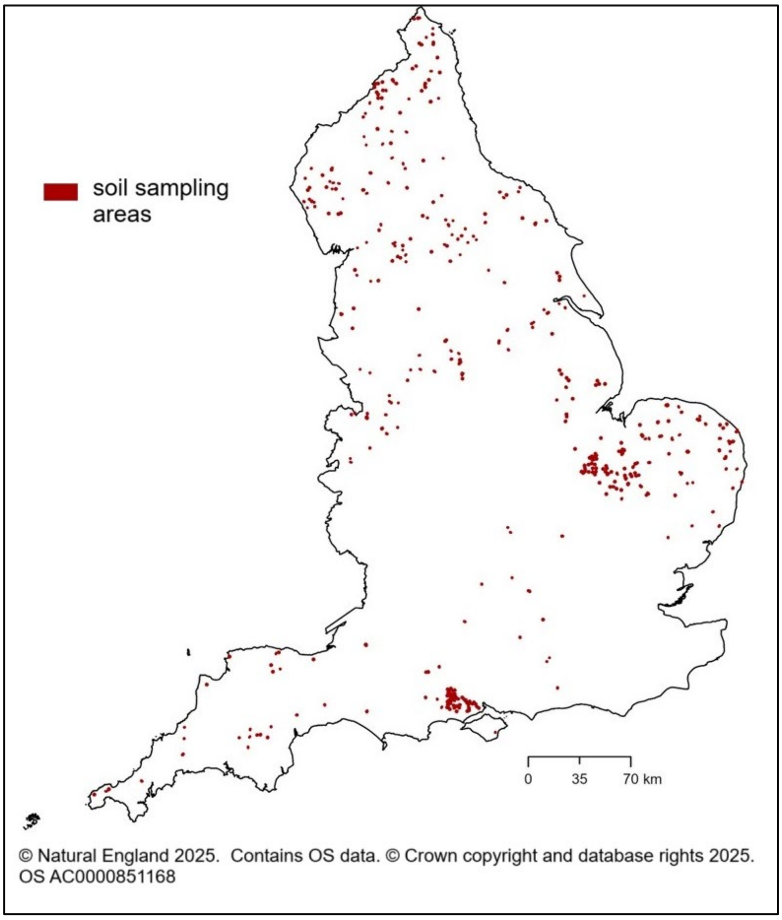


Figure 5-2 Soil survey locations

A total of 3,361 vegetation surveys (3,015 main & 346 quality assurance) were undertaken between January 2023 and October 2024. Due to quality issues only

2,730 vegetation surveys were used by the project (see section 6.4 for further details). The locations of these surveys are shown in Figure 5-3, and a summary of the number of quadrats collected for each vegetation class can be found in Table 5-6.

Table 5-6 Number of Vegetation Quadrats

Vegetation Class	Number of Quadrats Recorded
Sphagnum sp.	193
Eriophorum sp.	413
Molinia caerulea	368
Calluna vulgaris	388
Dry Grass/Scrub bog	362
Short Fen	340
Tall Fen	398
Scrub/Tree Fen	268

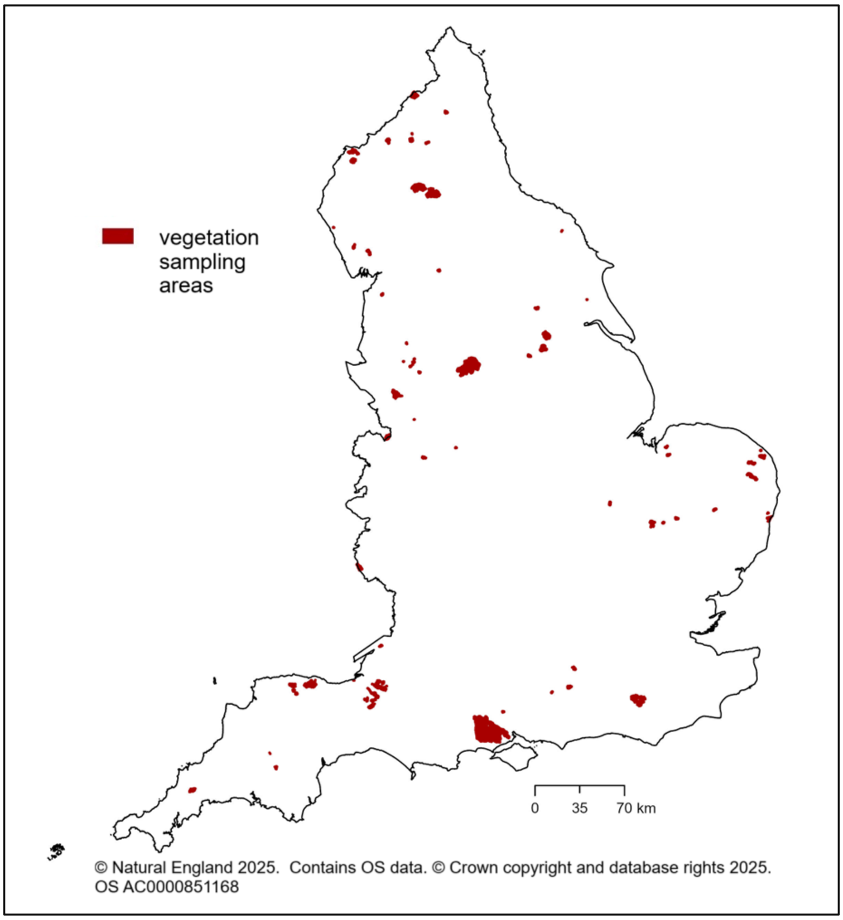


Figure 5-3 Vegetation survey locations

6. Peat Survey Data Collation and Management

6.1. Acquiring and Managing Peat Survey Data

Peatland restoration drives the production of large amounts of data continuously. Restoration funds require initial baseline surveys, whilst research into the effectiveness of restoration and site management require periodic measurement of peatland condition. Data from these activities, sourced from a wide range of projects and very variable in nature, augmented the field survey data substantially.

One of the major funds for peatland restoration, the Nature for Climate Peatland Grant Scheme (NCPGS), is administered by Natural England and presented the opportunity to regularly and reliably acquire data in a standardised format from many different places. This source constituted a significant proportion of the data points that were added to the project database. The rest of the database was drawn from varied organisations, providing diverse content, formats and conditions of use. This presented a challenge as inconsistent data takes time to manage and combine into a single database.

Data from different owners with differing interests posed problems in terms of ensuring intellectual property rights were fully taken into account. For a large amount of the data in the north of England, this was tackled with the help of members of the Great North Bog Project. For the rest of England, these problems had to be tackled by project officers on a case-by-case basis.

To facilitate future data sharing, the project developed a draft data exchange standard (Natural England, 2023), designed as a common way of organising peat depth data for sharing and a starting point for further development of data standards in this field.

6.2. Data sources, availability and engagement

Most of the collated data from the north of England was administered via a memorandum of agreement (MoA) with the Peak District National Park which hosts Moors for the Future (MFF), an organisation working towards the restoration of upland environments in the Peak District and south Pennines. MFF are part of the Great North Bog (GNB), a wider partnership of peat partnerships for Lancashire, Cumbria, Northumberland, Yorkshire and the North Pennines National Landscape. Through this arrangement, the project was able to gain access to 1,714 datasets, of which 1,553 were submitted to the England Peat Map. Priorities for collating data were the size of the dataset and the recency of the data, or in the case of surface feature data only the most recent data was selected. An important and time consuming element of the work was to coordinate signing of licences by data owners. See below for the licensing of this data.

Outside of the Great North Bog area there is considerably less regional coordination of peatland survey activities and the England Peat Map project team sourced data directly from many different stakeholders. Natural England, the Environment Agency, Forestry England, Forest Research and other Government bodies all collect data as part of their activities, e.g. during SSSI monitoring surveys and Long Term Monitoring Network surveys and an extensive trawl identified many suitable datasets for inclusion. We approached partnerships working on peatland projects across England, Protected Landscape bodies and academics working in the field. We identified a particular data gap in the lowlands, and worked with stakeholders in those areas, particularly Fenland SOIL, who greatly helped to identify and secure data. In some cases, even relatively small amounts of data were particularly valuable because the filled in gaps in very poorly surveyed areas, such as Wishmoor Bottom in Surrey.

This process of sourcing data allowed us to also engage with stakeholders about the England Peat Map in general and to foster a sense of community ownership and collaboration. We reached out through different channels, beginning with Natural England area teams and peat-oriented colleagues, generating interest which was followed up by some regular meetings. This highlighted the existence of likely peat surveys and suggested external stakeholders to contact. We then approached larger organisations and presented and attended events such as the IUCN UK Peatland Programme conference. Particular emphasis was placed on stakeholders who could supply data on which to train the models, those who could provide information and data with which to validate the models (including feedback on the draft outputs) and those who were potential users of the final outputs. The result was to substantially augment the volume of data available for developing the England Peat Map outputs (see Table 6-1).

Some data was only available in formats such as images or pdf documents where we had to weigh the amount of potentially useful data against the work required to extract it. For instance, we received pdf documents from the Woodland Creation Planning Grant Scheme which contained some soil data. Whilst their design was useful for the original purpose (steering tree planting towards the best places), it was difficult to extract soil data. As each document only contained a small amount of data, we decided the time investment would be disproportionate. Similarly, the original lowland peat survey from the 1980s could have been a very useful source of data but was only partially digitally scanned, the rest being on paper. Some digitised data relating to the lowland peat survey was acquired through UKCEH and used in the project.

In other cases the data, or the permission to use it, did not arrive in time. This was the case with the Environment Agency's 'Lowland agricultural peat: water for peat pilots'. Finally, in some cases we were never sufficiently confident that we were licenced to use data. Licensing is discussed in greater detail in the next section.

Table 6-1 Number of different observations used by the project collected from the field survey and collated from external sources (excludes peat absence records received from the British Geological Survey)

Data Type	Field Survey	Collated	TOTAL
Extent (presence/absence)	13240	90164	103404
Depth	8797	80214	89011
Vegetation	2384	18831	21215
Bare Peat	0	65820	65820
Surface Features	0	257900	257900

6.3. Data licensing

For the England Peat Map, Natural England sought to use the most comprehensive data available, as well as publish as much of the underlying data as possible. However, Natural England cannot necessarily use or share all data that it receives – the decision as to what can be done with data usually rests with the person who creates or commissions it. Ideally data would be accompanied by a licence describing the terms and conditions of its use, but often this is implicit or missing. We therefore gave considerable attention to determining the source, ownership, and terms of use for all received data, including acquiring explicit new licences wherever possible.

There are two main groups of standardised licenses: governmental licences and creative commons licences. We also encountered many non-standard licences, some of which allowed for the project to use and publish the data itself whilst some only facilitated use of the data in creation of the project's final outputs. Table 6-2 provides an overview.

Table 6-2 Common licence types applied to environmental data

Licence	Permits publication of project outputs	Permits publication of data itself	Details
Open Government Licence (OGL)	Yes	Yes	Allows Government organisations (local and national) to publish data and information anyone can use for any purpose. Users must attribute the data to the government organisation that published it.
Non-Commercial Government Licence (NCGL)	Yes	No	As above excludes uses where the primary purpose is for commercial gain.
Creative Commons Zero (CC-0)	Yes	Yes	A licence designed by the Creative Commons organisation for anyone to use to

Licence	Permits publication of project outputs	Permits publication of data itself	Details
			allow their data or information to be used by anyone for any purpose.
Creative Commons BY (CC-BY)	Yes	Yes	As above but users must attribute the data to the person or organisation that published it.
Creative Commons Non-Commercial (CC-BY-NC)	Yes	No	As above but data cannot be used for commercial purposes.
Natural England template – publishing data and derivative works	Yes	Yes	Allows for NE to derive products from data and publish both those products and the data itself with an open licence (OGL).
Natural England template – derivative works only	Yes	No	As above but the data itself cannot be published.
Other bespoke licences	Possibly	Unlikely	These are non-standardised licences, usually produced by the organisation issuing the data, that may or may not be worded in a way that secures NE's right to use the data and publish the outputs.

The project acquired more than 1,700 datasets using all the licence types described above (apart from NCGL). The *NE template – derivative works only* licence was the most frequent due it being the licence of choice for a large number of datasets from the GNB. The number of datasets and number of different licences presented an administrative challenge in terms of cataloguing, metadata, recording data ownership and storing licences appropriately.

6.4. Types of data and a peat data standard

Peatland surveys differ according to the organisations commissioning them, the people conducting them and reasons why they are needed. Even where data was collected for very similar purposes it may still be formatted differently, making it harder to use beyond its original purpose without human intervention to modify it in some way, e.g. renaming column headings. To facilitate greater sharing and the ease of use of existing data in decision making and research – the "collect once, use many times" principle (IGGI, 2005) – the project developed a draft data exchange standard for peat depth measurements. The publication of this standard, in January 2023, is the first step in enabling organisations to share data in a universally accepted format. It defines words, data formats and mandates the inclusion of attributes that are very important for research and model development, e.g. whether the bottom of a peat layer was reached by peat probing.

Using the standard will help the community of peatland organisations in England to adhere to the FAIR data principles. These principles are guidelines for making data Findable, Accessible, Interoperable and Reusable (Wilkinson and others, 2016) as part of a wider movement to guide data managers towards best practice in the interests of greater data use. Standardised data is easier to discover, manage, archive and share and can therefore be accessed by future projects and studies, many of which require combined datasets covering large areas or time periods (Horsburgh and others, 2009; Campbell and others, 2016). Whilst the peat data standard is specifically targeted at depth surveys, we used many different data types summarised in Table 6-3.

Table 6-3 Data types used by the project

Data type	Description	Comments
Depth probes	Observations consist of measurements of the depth of peaty soil taken by probing the ground coupled with a spatial reference.	This method allows for a wide area to be covered quickly. A large proportion of the external data collated by the project is of this type.
Soil cores	Observations consist of soil cores removed from the ground for analysis in the field and / or by photograph, coupled with a spatial reference.	This method produces data that are a very reliable record of soil composition but there are a lot fewer of them as they take time to record.
Soil organic content analysis	Observations consist of the result of various tests on soils samples to determine organic content, performed in the field or a laboratory, coupled with a spatial reference	Laboratory tests such as loss on ignition are the most reliable way of testing for the presence of peat. It takes many more resources to generate however and consequently there is a lot less of this data available.
Other forms of depth measurement	Some of the more technological methodology, e.g. using electromagnetism to penetrate the ground such as penetrating.	These are very rare forms of data and many are experimental, available only as a result of academic research.
Standardised habitat classifications	Phase One and the National Vegetation Classification are two very common examples of this.	Vegetation surveys are common on for sites designated for their ecology.
Other classifications	Sometimes bespoke classifications are used to describe vegetation. At other times single species (e.g. Molinia caerulea) are mapped, or vegetation records appear as target note to peat surveys.	The usefulness of these kinds of data depends on whether the classification system can be fitted within the system being used in the models. Where individual species are recorded then this can happen easily.

Data type	Description	Comments
On the ground surveys	Observations consist of individual surface features drawn on a map by people on the ground looking directly at grips, gullies, haggs and bare peat etc.	This group includes data of many different physical aspects of peat. Where restoration has taken place, this is often easy to come by as people monitor the effects of work such as grip damming.
Digitisation developed from aerial imagery	Observations consist of surface features digitised by tracing them from aerial maps into a GIS.	Largely as above but relies on features that can be seen from the sky.
Data on water levels and other aspects of the environment	Dipwells, both automatically read and manually read. Observations consist of the spatial reference for the dipwell (or other measuring tool), the reading and the date.	Often in designated sites, dipwells are installed to measure water levels. This data is hard to find outside of designated or restoration sites however.

6.5. Feedback on interim and beta outputs

Having incorporated data submissions from the peat community in the modelling process, further input was sought at two key stages: 'interim outputs' (covering the north of England and the Fens) and 'beta outputs' – an initial version of outputs covering all of England. People with extensive knowledge of soils in specific areas were approached for comment. Through the MoA with MFF, peat partnerships in the north were approached for their input. In the rest of the country, the network of stakeholder contacts that had been built up through the early stages of the project was employed. Feedback from both internal and external stakeholders was sought. Eleven interviews were conducted, eight with Natural England and three with external stakeholders. All regions of England were covered except the South East.

Consultees were asked about their overall impression of the accuracy of the outputs as well as any specific points where they could say that the models had made accurate or inaccurate predictions. A tool to gather feedback was developed in the form of a web portal which presented the outputs alongside various other useful map layers, including the currently published Peaty Soils Location, and allowed users to place points where they could then input their feedback. To facilitate deeper feedback on the beta outputs, semi structured interviews were conducted. These asked standard questions but gave respondents a lot of freedom to move between geographical areas and topics, providing feedback where it would be most useful. Feedback points were submitted in a systematic manner, coupled with extensive notes. This process greatly increased the overall depth of feedback received. It allowed it to be centred on specific places to ensure greater geographic coverage and a focus on areas where the models appeared to be performing less well. Consultees were also able to submit comments in writing and on a web portal.

One of the clearest themes was the extent of peaty soils in the lowlands, which was seen as an over-prediction. This contrasted with the depth of peaty soils in the lowlands which was seen generally as an under-prediction, especially at certain sites where the depth had been measured but the data was not available to train the model. But there were also some contradictory comments suggesting that peaty soils in the lowlands were under predicted for extent and over predicted for depth. These tended to be site specific rather than general. For the uplands, there were a lot fewer comments overall but were similar to the lowlands, i.e. an over estimation of extent and an under estimation of depth.

Comments on the vegetation model covered the majority of vegetation classes, but some classes attracted more comments than others. *Molina sp.,* in particular, attracted a lot of criticism for being confused for other classes. Both types of fen classes (short and tall) drew comments and in several interviews it was suggested they had become confused with grasslands or pasture. Arable, *Sphagnum sp.* and *Calluna vulgaris* all drew more than one comment in the interviews whilst in submissions via the web portal, *Calluna vulgaris* attracted the greatest number of points by a large margin with patches of it having been identified from desk-based studies by feedback providers. An additional area of vegetation feedback was the misclassification of bracken as other vegetation types.

There were fewer feedback comments on surface features throughout the whole process, although this would be expected as surface features modelling was restricted to upland areas more data was available. The feedback that that did come in was spread across the different types of surface features, with grips receiving a little more focus than the other types. It was spread evenly in terms of whether features had been incorrectly predicted, missed or whether the wrong type of feature had been identified.

Feedback comments were used to improve models and outputs in the remaining time available, for example by focussing on optimisation of the extent model in the lowlands, or on the balance of Molinia and Calluna dominated bog vegetation classes in the uplands. These changes affected the models in their entirety, so the difference at individual feedback sites may be smaller than improvements to the models overall.

7. Modelling

7.1. Data preparation

Peat presence, absence and thickness data

Any survey which collected information about the presence or absence of peaty soils was considered for model training and validation data. The highest priority was given to potentially large datasets and datasets where a methodology was available. High priority was also given to datasets in areas where data was sparse (e.g. lowland areas, shallow soils, small peatlands). Survey data with little or no methodology were omitted from the modelling data. Sources of derived data (such as maps created from field data) were noted and in some cases used for planning or validation, but no location data was extracted from derived data sources. Once usage rights over the data were established, data was examined for suitability, attributes of interest were extracted and exported to a spatial database.

Biogeographic Zone (BGZ)	Presence Observations	Absence Observations
1	19,509	13,728
2	16,939	28,655
3	49,635	48,622
4	11,057	15,767
5	2,863	50,367
6	10,935	29,153
7	660	14,393
8	1,751	20,543
9	145	19,077
10	301	44,159
11	371	16,281
12	987	26,676
13	6,926	15,201
14	-	-

Table 7-1 Presence and Absence Observations in each Biogeographic Zone (BGZ)

The attributes extracted are: the presence or absence of peaty soil (without differentiating which type of peaty soil), depth and thickness of the first and second peat horizon if present, coordinates, date, surveyor information, probe type and whether the probe reached the bottom of the peaty horizon. Metadata for each source dataset includes title and abstract of each dataset, a concise description of survey methods, and geographic metadata based on the INSPIRE data standard. Duplicate data and missing values were removed if they shared the same location and attribute values as another point.

Vegetation data

Field survey data outside of the EPM field survey has been collected as a georeferenced point dataset to improve our vegetation coverage across England's peatlands. This data has been collected from previous surveys across various sectors including, government, peat partnerships, Non-Governmental Organisations (NGOs) and academia. These data come in various formats and collection was undertaken with various methods, therefore a quality assurance process was undertaken before their use in the vegetation modelling work.

Standardisation was achieved by selecting samples with percentage cover recorded within quadrats and assigning each quadrat to the vegetation condition classes outlined in Table 5-1. External data were included in the labelled datasets to reduce the likelihood of changed vegetation types being introduced into the modelling workflow. The EPM is driven by Earth observation imagery therefore we are only able to capture information about the aerial coverage of vegetation. This was considered when assigning vegetation classes and to align collated data with the EPM survey. A 60% threshold was applied to all quadrats for dominant aerial cover and where a dominant 60% threshold was not achieved, they were removed from our dataset. These collated field surveys were not consistent in their collection; therefore, this data is only used as training data for the modelling approach. Vegetation data was considered outliers if it exceeded 1.5 times the interquartile range (IQR) above the third quartile or below the first quartile for each vegetation class across all three seasonal cloud-free Sentinel-2 mosaics.

Upland bare peat data

Desk-based surveys are routinely carried out by external partners to identify sites for peatland restoration. Part of these surveys include digitising bare peat from aerial photography which provide a large, labelled dataset appropriate for training and validating bare peat models. These labels were rasterised to a 25 cm x 25 cm grid to match the resolution of our bare peat product and bare peat image chips were created where there was a 50% or greater presence of bare peat. A subsequent quality assurance process was carried out for all external data through visual inspection of each dataset to account for seasonal differences. Additional quality assurance of manually labelled data from desk-based studies provides less uncertainty and representative accuracy statistics, particularly as non-bare peat areas are the dominant land cover class.

Upland peat drainage and erosion data

Grips, gullies and haggs ground truth data (Figure 7-1, left) were obtained from the following external organisations – Yorkshire Peat Partnership, Lancashire Peat Partnership, Northumberland Peat Partnership, Cumbria Peat Partnership, Environment Agency and the National Trust. Features were digitised as lines and done so primarily for the purposes of planning restoration. Due to this, the spatial accuracy of digitisation is not perfect, and we expect that each organisation and user will have a slightly different definition of each feature (Figure 7-1, right). Grip lines were buffered by 0.75m, gully lines were buffered by 1.5m and hagg lines were buffered by 1m to create polygons. In total 111,792 grip, 78,075 gully and 118,423 hagg lines were available. The buffer sizes were chosen after visual inspection of grips and gullies. The vast array of linear data from different sites from around England made it impractical to tailor the buffer sizes by site, and so a single buffer size was chosen which best represented the majority of observed cases. For grips this works reasonably well due to their uniform and regular size. Gullies and Haggs are far more irregularly shaped, so a one-size-fits-all approach is less ideal for these features. The alternative approach would be to manually digitise, and some manual digitising was carried out over the West Pennines to create more accurate training data. However, this was found to be impractical over larger scales, as a site the size of the West Pennine Moors would take up to a week to complete.

Peat dams were digitised by the AI4Peat team along a selection of the grip lines identified. Dams were digitised as point features and then converted into 6 m x 6 m bounding boxes ready for model training.

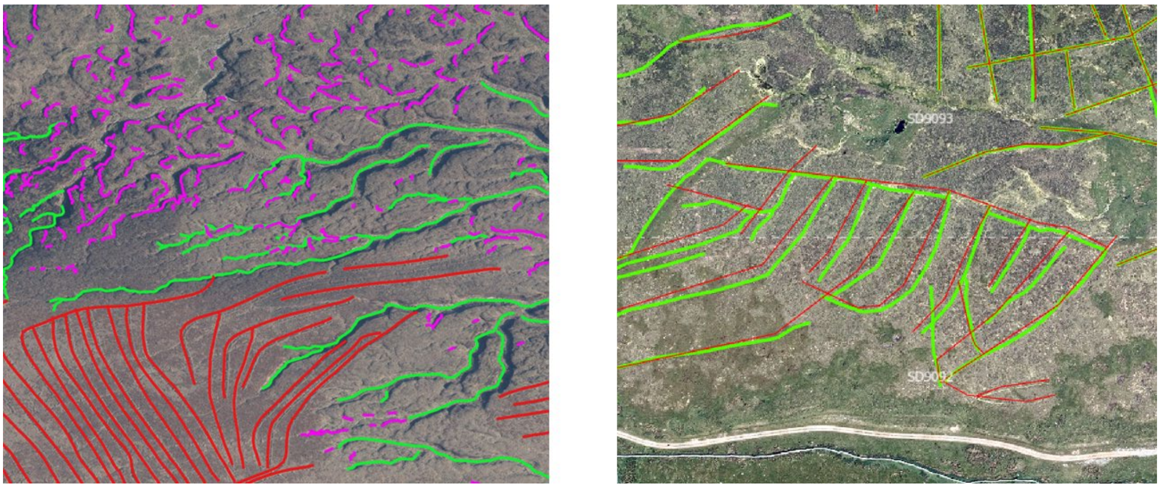


Figure 7-1 **left**: Example area showing ground truth labels in the Yorkshire Dales. Grips are in red, haggs are in purple and gullies are in green; **right**: example of spatial inaccuracy of feature digitisation. Red lines are as received from partners, green lines are manually edited using aerial photography and LiDAR. Aerial imagery © 2015 Getmapping plc and Bluesky International Ltd.

7.2. Predictors

Predictors are datasets which, when compared to observation data (e.g. presence of peaty soil), enable us to estimate what the observation would be in places where we have no direct survey data. They cover the whole of England and combining them helps improve the accuracy of the machine learning outputs. We have drawn predictors from satellite imagery, airborne surveys, and mapped products. The predictors used across all the England Peat Map products are listed in Table 7-2 and more information is provided in Annex 6.

Satellite data comes in the form of radar (Sentinel-1), photographic imagery (Sentinel-2, both European Space Agency) and thermal imagery (Landsat-8, NASA). All satellite data was compiled from data recorded between 1 Aug 2022 and 31 Jul 2023. The Sentinel-1 and Sentinel-2 mosaics are the same as those used by the 2022-23 Living England habitat probability map (Trippier and others, 2024). Sentinel-2 imagery is captured whenever a satellite passes overhead. However, these images do not cover all of England at once and are often obscured by clouds. To address this, we create mosaics by combining images from different times, minimising cloud interference and ensuring complete coverage. These mosaics are organised by season, producing separate mosaics for spring, summer, and autumn. We also created a new mosaic focusing on bare soil, particularly in lowland areas, see Annex 6 for more details. The Landsat thermal imagery was processed to create an annual median.

Sentinel-1 radar imagery can be processed in different ways. We used the 'backscatter' portion of the radar signal to provide information about the characteristics of the terrain such as its roughness. Unlike optical imagery, radar imagery is unaffected by cloud cover. We also calculated the 'InSAR coherence' of images taken at different points in time, which can use phase and amplitude between image pairs to indicate small changes in the elevation of the surface. To ensure consistency, we create a median mosaic over the same period as each of the Sentinel-2 mosaics.

The Environment Agency's national airborne LiDAR programme uses aircraft-mounted lasers to generate Digital Terrain Models (DTM) and Digital Surface Models (DSM) across England. For our extent, depth and vegetation modelling, we used a composite of imagery captured between 2017 and 2023. A Canopy Height Model (CHM) was derived by subtracting the DTM from the DSM. The DTM was also used to calculate slope (for vegetation, extent, and depth), elevation, and aspect (for extent and depth).

For the peaty soil extent and depth models, we derived several indices from the DTM: terrain roughness index (TRI), topographic wetness index (TWI) and ‘geomorphons’ (at two spatial scales). Geomorphons (Jasiewicz & Stepinski, 2013) categorise terrain into landforms such as slopes, ridges, and valleys. Additionally, we generated detrended LiDAR, which removes large-scale elevation changes in the DTM to highlight finer topographic variations (see Annex 6). This was combined with aerial photography to enhance the classification of grips, gullies and haggs.

Table 7-2 Summary of predictors used in the England Peat Map products. Sources: ESA = European Space Agency, NASA = National Aeronautics and Space Administration, EA = Environment Agency, NE = Natural England, BGS = British Geological Survey, OS = Ordnance Survey.

Predictors	Raw Data Source	Extent	Depth	Vege-tation	Bare Peat	Surface Features
Sentinel-1 backscatter and InSAR	ESA	+	+	+	-	-
Sentinel-2 imagery	ESA	+	+	+	-	-
Sentinel-2 bare soil mosaic (annual composite)	ESA	+	+	-	-	-
Landsat-8 thermal	NASA	+	+	-	-	-
LiDAR-derived slope	EA	+	+	+		
LiDAR-derived elevation (DTM), aspect, TWI, TRI and geomorphons	EA	+	+	-	-	-
LiDAR-derived DTM (detrended)	EA	-	-	-	-	+
LiDAR-derived CHM	EA	-	-	+	-	-
Dudley Stamp Land Utilisation Survey	EA	+	+	-	-	-
Bedrock and superficial geology	BGS	+	+	-	-	-
Flood risk from tidal and river water (1%) and from surface water (3%)	EA	+	+	-	-	-
Broad Landscape Type	NE	+	+	-	-	-
Distance to sea, distance to river	OS	+	+	-	-	-
Aerial Photography	BlueSky Mapping	-	-	-	+	+

Additional predictors were used specifically for peat extent and depth modelling. The Land Utilisation Survey (1933-1949), Dudley Stamp, categorised Great Britain into eight land use classifications at a 1km resolution. Natural England’s 159 National Character Areas (NCA) Profiles (Natural England, 2021) classify each NCA into one of 19 Broad Landscape Types (BLT). 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 mostly unconsolidated sediments that accumulated during the quaternary period. The Environment Agency's Flood Map for Planning (Rivers and Seas) Flood Zone 3 (Environment Agency, 2024) and its Risk of Flooding from Surface Water Extent (Environment Agency, 2013) were used to classify areas which had a 1% or greater chance of flooding from rivers and seas, and a 3% or greater chance of flooding from surface water respectively.

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. We used imagery from 2018 until 2023 for upland features and bare peat.

Most of these predictors comprise many layers of data. For instance, the Sentinel mosaics are each 13 layers representing different ranges of the light spectrum and categorical predictors such as Dudley Stamp are broken out into eight layers, one for each category of land use. The number of predictors reported for each model is therefore much greater than the number of rows in Table 7-2.

7.3. Model Validation and Performance Metrics

While models were being developed, we set aside a portion of the training data as 'validation data'. This was then used to test to see how good a model is at predicting for locations it has no information about by calculating a range of metrics (see Table 7-3, and Appendix 6).

In some cases, validation data was then used to help train a final model, but we do not report metrics for these because they would be artificially good. Instead, we report the metrics of the precursor model, trained without validation data.

Table 7-3 Metrics used to evaluate model performance for the different models

Metric	Extent	Depth	Vege-tation	Bare Peat	Surface Features
Overall Accuracy *> 0.8 (80%) considered 'good'*	+	+	+	+	+
Precision *> 0.8 considered 'good'*	-	-	+	+	+
Recall or Sensitivity *> 0.8 considered 'good'*	+	+	+	+	+

Metric	Extent	Depth	Vege-tation	Bare Peat	Surface Features
Specificity *> 0.8 considered 'good'*	+	+	+	+	+
F1 Score *> 0.8 considered 'good'*	+	+	+	+	+
Mean Absolute Error (MAE) *Lower is better*	-	+	-	-	-
Root Mean Square Error (RMSE) *Lower is better*	-	+	-	-	-
Matthew's Correlation Coefficient (MCC) *+1 perfect; 0 random; – 1 bad*	+	-	-	+	+
Intersection over Union (IoU) *0.5 acceptable, 0.7 higher precision, > 0.75 very good*	-	-	-	+	+
Kappa *+1 perfect; 0 random; - 1 bad*	-	-	+	+	+

7.4. Extent and Depth Modelling

Methodology

EPM used machine learning models to create a prediction of the extent of peaty soils, and their thickness, in England. The output products are illustrated in section 4.3 above. Further technical details are provided in Annex 6 to this report.

Data Preparation

Models were trained on the presence and depth attributes of the soils field data described in section 5.36. The 'presence' data is highly clustered and also imbalanced. We addressed clustering by thinning all data to a density of no more than five observations per 100 meter square (500 per square km). This still leaves large differences in sampling density, so we included spatial predictors in the model (BGZ and Broad Landscape Type) to allow the model to distinguish between higher and lower density areas. This contrasts with our 'beta' model approach, where we created separate models for different regions.

The data is also extremely imbalanced, with a ~3:1 ratio between absences and presences nationally. This is spatially variable and in some areas of England it is reversed. We addressed this by adding a parameter to the models which accounts for imbalanced training data.

For the depth model, we excluded any depth measurements of less than 10 cm but did not balance, thin or weight our data.

A total of 82 predictors were assessed as potential predictors for determining the extent and depth of peaty soil in the models (Table 7-2). We reduced the number of predictors used by removing variables that had significant correlation with other variables; made a minimal contribution to the models; or lead to substantial inaccuracies in the predicted maps (e.g. due to missing data).

Defining and fitting models

The modelling was carried out using the R statistical programming language (version 4.3.1) (R Core Team, 2024) using the tidymodels framework (Kuhn & Wickham, 2020). The XGBoost library was used to create ensemble models predicting extent (binary classification) and depth (regression) (Chen & Guestrin, 2016).

We fitted binary classification models to predict the probability of peaty soil for all of England. An initial split randomly retained 75% of the observations as the training dataset. Spatial 10-fold cross-validation assessed ensemble model performance (Roberts and others, 2017). The fitted models were evaluated with the remaining 25% testing subset against the metrics described in Table 7-3. We chose the best performing model and used the terra package (Hijmans, 2024) to predict the probability of peaty soil presence for every 10m cell in England. We tested a number of probability thresholds to determine which performed best at correctly predicting peaty soil presence in our validation data. We also tested using different thresholds for different parts of the country. The best performing threshold was used to create a prediction boundary for peaty soil presence.

We fitted regression models to predict the depth of peaty soil for all of England and cropped it to the area predicted by the extent model. We used the same methods for splitting and cross validating our models, and evaluated models using 25% of the data held back from training.

Post-processing

To account for a coarser bedrock geology layer and localised mosaic habitats, we identified areas where bedrock was at the surface (e.g. limestone pavement) using Sentinel-1 data and removed this from our prediction. We also removed areas of standing and open water where the use of satellite imagery would be inappropriate and reduce accuracy (Li and others, 2022), and constrained our prediction to the mean high water line.

Interpolation of residuals

Although peat thickness can vary a lot over very short distances, spatial correlation has been found in peaty soils, meaning that, on average, areas close to one another have similar thickness than more distant areas. This spatial structure can be modelled and interpolation techniques can improve predictions of extent and depth of peaty soils (Young et al., 2018; Poggio, Lassauce and Gimona, 2019; Finlayson et al., 2020).

The difference between the predicted and observed peaty soil thickness is called the ‘residual error’. It has been shown that machine learning models based on earth observation data can be improved by estimating the spatially autocorrelated residual error with kriging (see glossary) and adding this to the model predictions (Poggio and others, 2019). We calculated the residual error and examined its spatial structure by creating semi-variograms and fitted models to them. Kriging was used to interpolate estimates of the residual at each predicted depth location. The sum of the predicted depth and the predicted residual are the final depth prediction for each cell.

In addition, variance is estimated for the residual at each cell, which can be used as a guide for overall uncertainty of the spatially correlated component of the depth prediction. We averaged the variance over 1km by 1km cells to obscure the location of individual training points. This last step was required to comply with licensing constraints on some of the data we used (the data that is not subject to those restrictions has been published separately).

Results

Predictor importance

Having trained the models, it is possible to determine the contribution each predictor makes to the final prediction. This is expressed as ‘Importance’ and given as a percentage. Figure 7-2 shows the 20 most important predictors for the depth model and the extent model and also shows how a predictor that ranks highly in one of the models scores in the other model. The most important variables predicting peaty soil extent are topographic (elevation), historical (Dudley Stamp Land Utilisation Survey) and autumn satellite imagery, with smaller contributions from summer and spring imagery. Broad landscape type also makes a substantial contribution, as do superficial and bedrock geology, suggesting that the landscape context is important in determining the presence of peaty soils. Other aspects of geography (wetness and roughness indices, geomorphons, see glossary) and radar backscatter are individually less important but collectively provide the model information about small scale variability. The presence of InSAR predictors in the top 20 most important predictors suggests that surface movement might be informative of peaty soil presence. Biogeographic zone is also in the top 20, suggesting that the model takes into account regional differences in peaty soil distribution. River and tidal flood risk is the least important of the top 20.

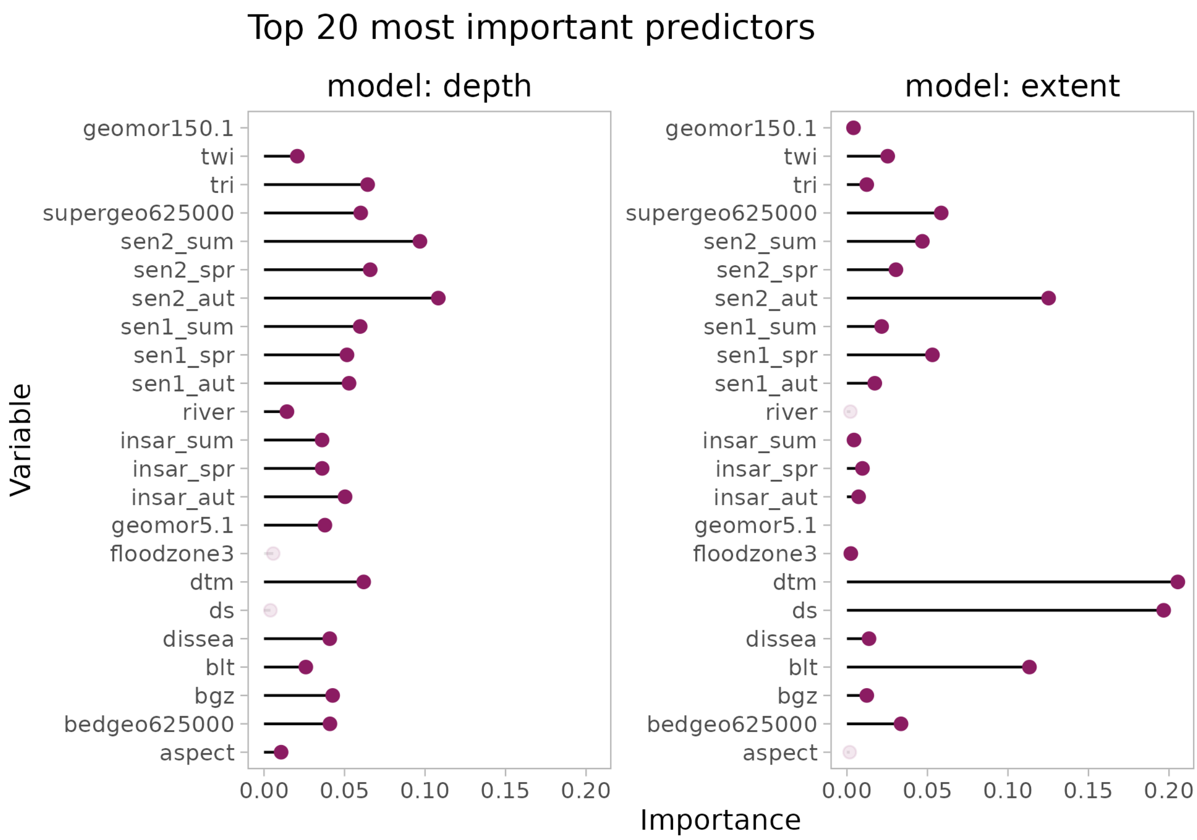


Figure 7-2 Model predictor importance. The 20 most important predictors for depth and extent models respectively are shown in bold. If a predictor does not have an importance value, it is not present in that model.
Predictors: geomor150.1 = geomorphons at 1,500 m; twi = Topographical Wetness Index; tri = Terrain Roughness Index; supergeo6250000 = superficial geology 1:625000; sen2_… = Sentinel 2 imagery for summer, spring or autumn; sen1_... = Sentinel 1 backscatter for summer, spring or autumn; river = distance to river; insar_... = Sentinel 1 coherence for spring or autumn; geomor5.1 = geomorphons at 50 m; floodzone3 = 3% flood risk from rivers and sea; dtm = elevation from digital terrain model; ds = Dudley Stamp Land Utilisation; dissea = distance to sea; blt = Broad Landscape Type; bgz = Biogeographic Zone; bedgeo625000 = bedrock geology 1:625000; aspect = aspect from digital terrain model.

Depth predictions are driven by satellite imagery (all Sentinel 2 mosaics) and radar (all Sentinel 1 and InSAR mosaics). Elevation, aspect, terrain roughness, wetness and geomorphons are also important contributors, as are superficial and bedrock geology. Distance to sea and rivers also affect the prediction. As with the extent model, biogeographic zone is an important predictor, suggesting that the model takes into account regional differences in peaty soil thickness. In both models it is notable that the ‘slope’ predictor does not appear in the top 20, even though it is generally agreed to be a strong predictor of peat presence and thickness. It is likely that this is because slope is strongly correlated with other predictors in particular terrain wetness.

Model performance

A large number of extent and depth models were developed and evaluated, many performing well in some parts of the country, but poorly in others. It is likely that this is for two reasons: the different characteristics of peatlands, particularly between upland and lowland soils, and the substantial imbalance of available data. We initially addressed this by creating separate models for different geographical areas but later found that adding a weighting to training data according to its density was very effective and allowed us to provide the model with considerably more information. We also added predictors (Biogeographic Zone and Broad Landscape Type) which gave the models information about regional-scale differences and allowed it to adjust for different characteristics at a landscape scale.

The **extent model** estimates the probability that peaty soil is present. To make a final prediction we had to determine a threshold of probability that would be acceptable, again using the evaluation metrics. The final extent model uses a threshold of 40% probability, because it has good performance across all metrics (Table 7-4), scored highest for MCC, and in particular it balances sensitivity (the likelihood that peaty soil is predicted where it occurs) and specificity (the likelihood that a prediction of peaty soil is in fact correct), with both scoring above 95%. Results for all thresholds are provided in Annex 6.

Table 7-4 Summary accuracy metrics peaty soil **extent** model

Overall accuracy	F1 Score	MCC	Sensitivity	Specificity
0.954	0.913	0.885	0.964	0.951

The **depth model** estimates the thickness of peaty soil where it is predicted to be present by the extent model. We evaluated this both by how well the model fits the observations (R Squared) and by how much the model predictions and observations differ (RMSE and MAE). We found (Table 7-5) that modelling alone left a very large margin of prediction error to be accounted for. However, after interpolating the residuals, this decreased markedly. The predicted residuals were added to the predicted depth to create the final depth prediction. In addition, variance was estimated for the residual at each cell.

Table 7-5 Summary accuracy metrics peaty soil **depth** model

Model stage	RMSE	MAE	Huber Loss	R Squared
Model	67.5 cm	46.0 cm	45.5	0.47
Residual interpolation	28.6 cm	19.4 cm	18.9	0.92
Final prediction	30.9 cm	20.8 cm	20.3	0.91

7.5. Vegetation and Bare Peat Modelling

Methodology

Vegetation modelling approach

A pixel-based classification is used for mapping vegetation presence on England's peatlands, where the minimum mapping unit is 100m^2 (a single 10 m x 10 m pixel). The vegetation classification approach uses an XGBoost machine learning algorithm (Chen & Guestrin, 2016) to match the known characteristics of each vegetation type with specific characteristics from the model predictors (Table 7-2). The parameterisation of the model is achieved through bootstrapping various configurations, where each model holds a random 20% of the labelled data back to validate the model predictions using stratified random sampling and the best-performing parameters are selected for the final model run. During this step, predictors are removed with high collinearity (> 0.9) and recursive feature elimination to reduce model overfitting and ensure optimal feature extraction. The trained vegetation model used the following parameters after 5-fold cross validation: 200 estimators with a maximum depth of 4 with 0.2 learning rate. The model is applied to new imagery where a probability of each vegetation condition class for each pixel is provided, and the highest probability class is assigned. Ancillary layers from CROME, NFI, OS and LE are embedded into the map as a post-processing step to the classification. Each layer is rasterised to the 10 m x 10 m grid used for imagery and embedded into the classification where there is 50% or greater presence of the ancillary layer within each 10 m x 10 m pixel.

Upland bare peat modelling approach

The England Peat Map uses a pixel-based classification for modelling the bare peat condition class in upland areas across England defined by the RPA Moorland Line (Figure 7-3). Labelled bare peat data was provided by external partners after undertaking desk-based assessments of aerial photography. This data is used to train and test a Convolutional Neural Network (CNN) deep learning algorithm implemented through the FCN-8 architecture (Shelhamer and others, 2015) written in Tensorflow (Abadi and others, 2016).

The aerial photography and labelled data is split into 250 x 250 pixel subsets, also known as chips, to be fed into the model and normalised. The model holds 20% of the labelled data back for validation using systematic stratified sampling based on bare peat presence to ensure small and larger areas are included in both training and validation datasets. Data augmentation is applied to training chips by performing a set of horizonal and vertical flips to increase the number of training samples by a factor of 3 and improve performance (Shorten & Khoshgoftaar, 2019). A pre-trained model previously used for image classification of 3-band Earth observation imagery was used for transfer learning in this study to reduce the requirement for extensive model training (Hamer, 2021). The bare peat model used the model with the following parameters across 50 epochs with early stopping: 0.0001 learning rate, 32 batch size with an Adam optimizer. The model provides a bare peat presence probability for each pixel and bare peat presence is assigned to each pixel if the probability is above 50%. Post-processing steps include the removal of bare peat found in water bodies and urban areas, based on the same layers used in the vegetation model, and areas smaller than 4 m^2 (IUCN, 2023).

Results

The primary predictors for the vegetation mapping were (1) DTM, (2) Sentinel 2 red band and (3) slope and achieved an overall accuracy of 96.63 %. The bare peat modelling approach achieved an overall accuracy of 96.84 %. The full accuracy metric breakdown for these two approaches can be found in Table 7-6 and Table 7-7, respectively. The confusion matrix for these two approaches is also provided in Annex 6. The best-performing semi-natural bog class was *Calluna vulgaris*-dominated (with an F1 score of 99.18 %) which dominates much of the upland bog area in England. The main cause of confusion across all classes is *Eriophorum sp.*-dominated which is misclassified in mixed community areas of *Calluna vulgaris*. The national vegetation model was able to differentiate between semi-natural bog and fen well with misclassification occurring predominately within their own respective vegetation groups. Tall fen vegetation has areas of confusion with short fen vegetation, but this could also occur from seasonal differences between field survey and satellite imagery capture.

Bare peat classification performs well in most upland areas but there is an overprediction in *Calluna vulgaris*-dominated areas and areas in shadow (e.g. areas with greater hill shade), likely due to the spectral similarities between bare peat and darker areas of *Calluna* vulgaris and shadow in aerial photography. The model has performed well for larger areas of bare peat with less dense vegetation surrounding the area, but smaller areas were omitted from the classification, particularly in thinner surface feature channels across the upland landscape.

Table 7-6 Summary performance metrics for the national vegetation model.

Vegetation class	Overall accuracy	Kappa	F1 Score	Precision	Recall
Overall	0.94	0.82	0.93	0.93	0.94
Semi-natural bog					
Sphagnum sp. bog	-	-	0.66	0.77	0.57
Eriophorum sp. bog	-	-	0.80	0.84	0.77
Molinia caerulea bog	-	-	0.79	0.84	0.74
Calluna vulgaris bog	-	-	0.97	0.96	0.99
Dry grass and scrub bog	-	-	0.76	0.86	0.68
Semi-natural fen					
Short fen vegetation	-	-	0.78	0.75	0.82
Tall fen vegetation	-	-	0.77	0.74	0.81
Scrub and tree fen	-	-	0.79	0.85	0.74

Table 7-7 Summary performance metrics for the national upland bare peat model

Overall accuracy	Precision	Recall	F1 Score	MCC	Kappa	IoU
0.97	0.79	0.74	0.76	0.52	0.52	0.67

7.6. Upland Peat Erosion and Drainage

This component of the England Peat Map was delivered by the AI4Peat project (see section 4.2 above4.2 above).

Modelling Approach

Peatland erosion and drainage features (grips, gullies, haggs and dams, also referred to collectively as ‘surface features’ in this report) were mapped by interpreting high resolution (12.5 cm per pixel) aerial photography and Lidar (1m per pixel) using a type of artificial intelligence known as ‘deep learning’. The majority of analysis and data processing was done using Microsoft Azure’s cloud computing platform. This allowed the use of distributed computing which significantly sped up processing times and storage of very large datasets required for analysis.

Study area

Surface features are mapped for the upland peaty soils of England only. We define upland as the land within The Rural Payments Agency’s (RPA) moorland line (Figure 7-3). The moorland line was buffered outwards by 100 m to allow image tiles to overlap with the boundary. The imagery and lidar datasets that fall within this extent were extracted.

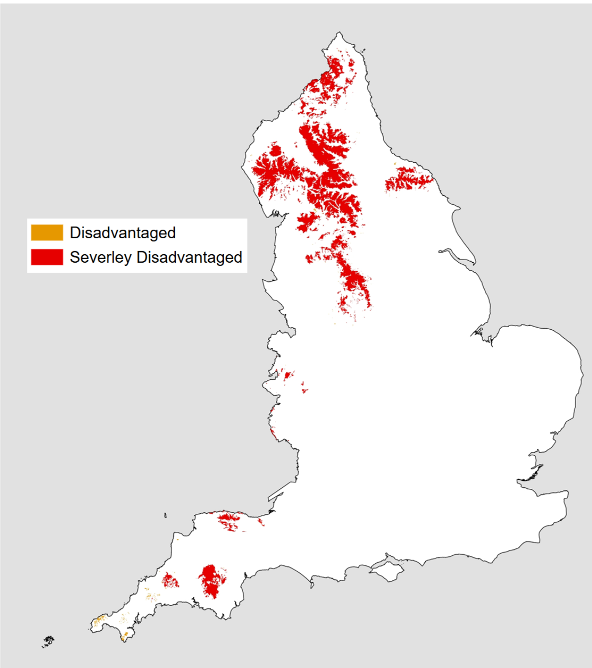


Figure 7-3 Area enclosed by the moorland line Contains OS data © Crown copyright and database rights 2025. OS AC0000851168

Pre-processing

To process the aerial imagery, feature labels and lidar into a suitable format for model training and inference, we used Databricks’ Mosaic geospatial library. This is an extension for Spark and enabled us to distribute the operations of indexing and geoprocessing across multiple compute cores.

To prepare the data for training and inference, the labelled features and detrended lidar datasets need to first match the pixel resolution and format of the aerial imagery data. For grips, gullies and haggs the digitised polygons labels that are used to train models were converted from vector format into raster datasets at the same 12.5 cm resolution as the aerial imagery. A binary map of features was created by giving a value of 0 or 1 to each 12.5 cm pixel indicating the absence or presence of a feature respectively. The 1 m resolution detrended lidar used to supplement the aerial imagery was resampled to the same 12.5 cm resolution using a nearest neighbour algorithm.

For the dams, the point locations of dams were converted into 6m bounding boxes in the format suitable for model training. For the YOLOv8 model (see below) this was a list of bounding boxes for each image in the format (x_{center}, y_{center}, width, height), while for the other models tested this was a table of bounding boxes in the format (x_{min}, y_{min}, x_{max}, y_{max}).

The aerial photography and detrended lidar found within the moorland line plus the rasterised labels were divided into tiles (also referred to as chips) of 400 x 400 pixels (50 m x 50 m) indexed onto the British National Grid. A total of 5,662,000 50 m aerial photography chips form our uplands dataset.

A training dataset for each feature was created by combining only the label chips with features present with the aerial imagery for the dams and aerial imagery and lidar chips for the grips, gullies and haggs (Figure 7-4Figure 7-4). For the grip, gully and hagg inference dataset all of the uplands aerial imagery and detrended lidar chips were paired. The dam inference uses aerial imagery only.

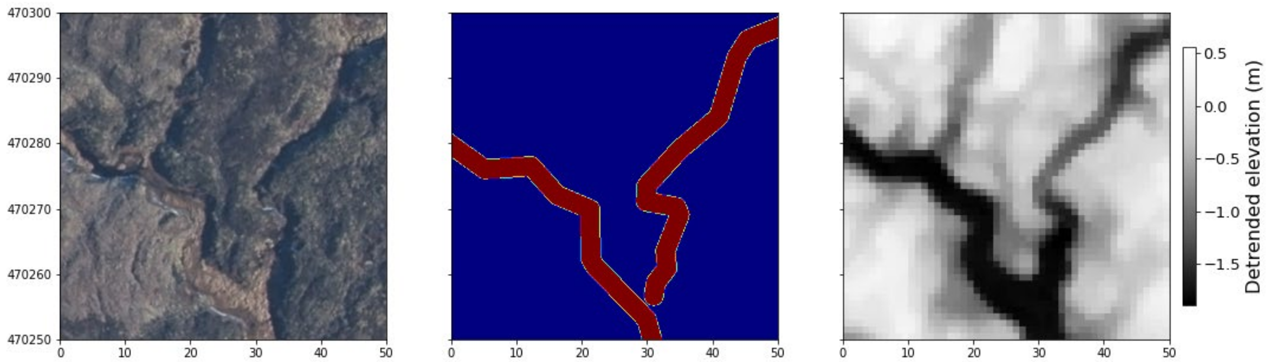


Figure 7-4 Example of matched 50 m chips of aerial imagery (left), gully feature label (middle) and detrended lidar (right). Aerial imagery © 2015 Getmapping plc and Bluesky International Ltd.

Model Selection, Training and Validation

The digitised data was split into a training, validation and test dataset. For detecting grips, gullies and haggs, a semantic segmentation model was used. This type of model was selected as it detects the pixels within an image that belong to the feature of interest, allowing the determination of exact location of each feature as well as its shape and dimensions. Several different semantic segmentation model architectures and hyperparameters were compared by training each on the training dataset and comparing performance on the validation dataset. For all three features, the best performing model was found to be a Feature Pyramid Network (FPN) architecture with ResNet34 backbone.

Table 7-8 Accuracy metrics for drainage and erosion features

Accuracy Metrics	Train	Validation	Test
Grips			
IoU	0.27	0.28	0.28
Recall	0.37	0.39	0.40
Precision	0.47	0.47	0.47
Accuracy	0.95	0.95	0.95
Kappa	0.35	0.36	0.37
F1-Score	0.37	0.38	0.39
Gullies			
IoU	0.27	0.29	0.29
Recall	0.61	0.63	0.63
Precision	0.32	0.33	0.33
Accuracy	0.86	0.86	0.86
Kappa	0.32	0.33	0.33
F1-Score	0.37	0.39	0.39
Haggs			
IoU	0.03	0.03	0.03
Recall	0.08	0.08	0.08
Precision	0.06	0.06	0.06
Accuracy	0.92	0.92	0.93
Kappa	0.02	0.03	0.03
F1-Score	0.05	0.05	0.05
Dams			
MAP50	0.81	0.71	0.64
Recall	0.79	0.62	0.59
Precision	0.85	0.74	0.64

For detecting the dams, knowing their shape and size was not necessary, only their location. We therefore used an object detection model, which draws a box around each object of interest. Several different object detection model architectures and hyperparameters were trained and their performance tested on the validation dataset. The best performing model was found to be a YOLOv8 model.

The performance metrics for the best performing model for each feature are shown in Table 7-8. See Annex 6 for metric definitions. Detailed information on the model selection process and final model parameters are available in Annex 6.

Inference

The trained models were applied to new aerial imagery to generate outputs, a process known as inference. Unlike the training phase, which involves optimising model parameters using labelled data, inference focuses on utilising the trained models to predict the location of the peatland surface features. This process involves feeding the imagery into the model, which then analyses it based on patterns and features learned during training to produce the final mapped outputs.

Inference was carried out on the same size chips (400 x 400 pixels) used for model training. For the whole uplands dataset (approximately 13,000 km^2) the inference step took approximately 1.5 hours using a Databricks compute cluster with 40 workers, each with 112 GB of memory and 16 cores.

Post-processing

Grips, Gullies and Haggs

The model outputs are raster images with each pixel assigned a probability of being a feature. To produce mappable features with attributes, these rasters are converted to vector format. All pixels with a probability score above 0.5 for gullies and haggs and 0.7 for grips are selected and polygons are drawn around the selected pixels. The sensitivity of the outputs to probability thresholds between 0.1 and 0.9 however is low. The model scores display a binomial distribution centred on very low or very high confidences (Figure 7-5Figure 7-5).

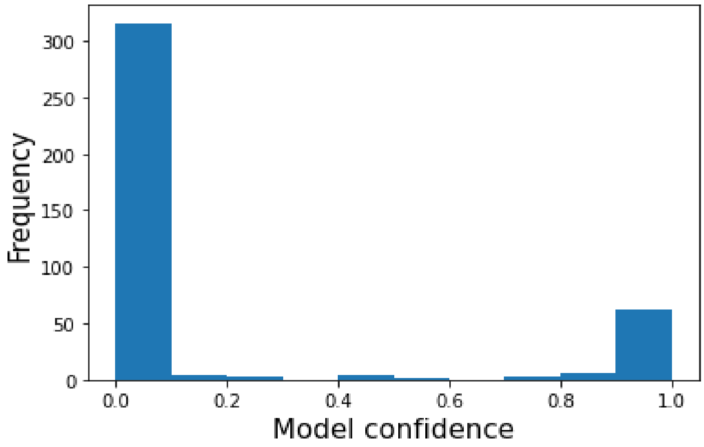


Figure 7-5 Grip model confidence scores for all 12.5cm pixels in a randomly selected 50m x 50m uplands grid.

We carried out further processing on the feature polygons to smooth the very pixelated edges of polygons arising from the high resolution of the imagery used and to join polygons that are cut at the 50 m chip edge. Smoothing polygon edges reduces the number of geometry entries in the final outputs, significantly reducing file sizes. For further details see Annex 6.

Dams

For the predicted dam bounding boxes non-maximum-suppression was applied with an IoU threshold of 0.5. This means that when any bounding boxes overlap by more than 50% only the box with the highest confidence will be used. This avoids having multiple predictions for the same object. For reasons detailed in Annex 6, any dams below a confidence threshold of 0.1 were removed.

Both grips and gullies are often dammed in restoration projects, but for these outputs we have decided to only include dams on grips. There are two reasons for this, firstly gully dams often look different from grip dams and there were very few gully dams included in our training data, making it hard for the model to detect gully dams. Secondly, there are a lot more gullies generally than grips, and the outputs of the gully model tended to be noisier than the grip model. Including dams which intersected with gullies therefore meant introducing significantly more false positives. We have therefore only included dams which intersect with a grip based on the grip model detection. Finally, the remaining predicted bounding boxes were joined based on the 100km index ID and written out as GeoJSON files.

Dimensions

Each grip and gully feature was assigned an average width, a minimum, mean and maximum depth and for linear features (i.e. those where the width divided by length of a fitted minimum rotated rectangle is less than 0.5) with a perimeter greater than 100m a slope value. Detailed methodologies for calculating dimensions are available in Annex 6.

Results

Figure 7-6Figure 7-6 shows selected examples of the mapped features. For all features, the modelled results appear to be generally accurate and initial feedback has shown this will be a valuable dataset for users. There are some situations where the models struggle, leading either to misclassification of features or missing features. Users should therefore be aware of the limitations of the datasets and use in line with guidance. The following quality assurance section details known areas of better/poorer model performance.

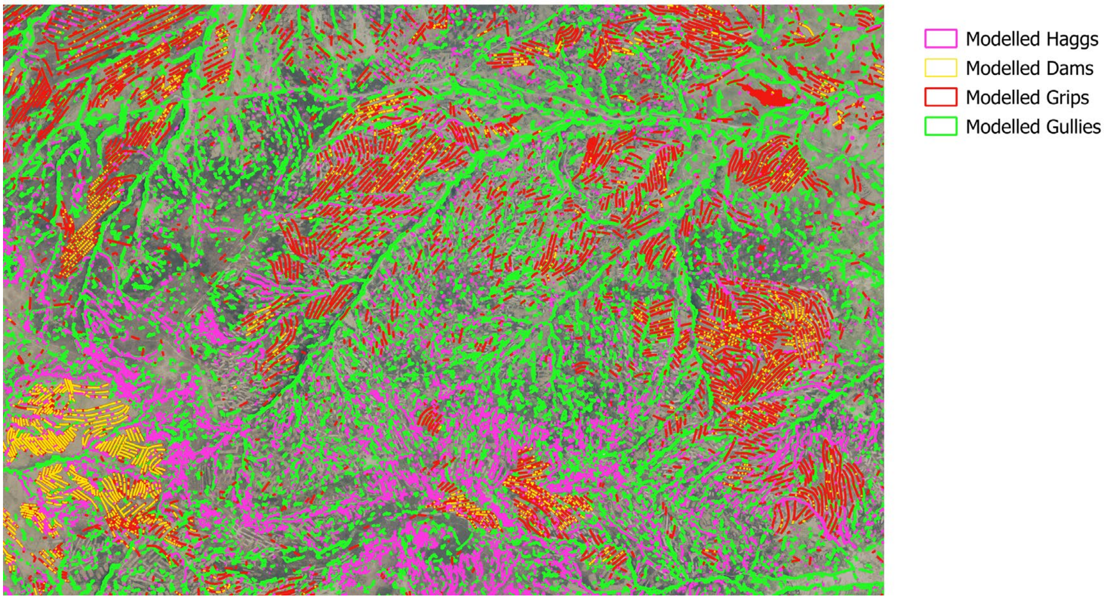


Figure 7-6 A large area of mapped peatland in North Yorkshire.
Aerial imagery © 2015 Getmapping plc and Bluesky International Ltd

Qualitative QA

To assess the quality of the results and understand any limitations, a review was undertaken internally, predominantly using a desk-based review where the modelled results were compared to aerial imagery and LiDAR data. An earlier iteration of the results was also reviewed on the ground in the Peak District. The findings from these reviews for each feature is detailed below.

It is worth noting that due to the extensive area that has been mapped, it is inevitable that some areas and situations will not have been captured in this review. There may be other cases where the models do not perform as expected that are not captured here.

Grips

Grip features generally appear to map well, especially in areas of densely packed grips. Some issues can be seen within areas of grips where small sections are missed by the model, and possible false positives outside areas of grips. The area of grips shown in Figure 7-7 in the Peak District has mapped the dense grip areas quite well, but some small features are detected which might be false positives. Further ground survey work would be needed to better understand these predictions.

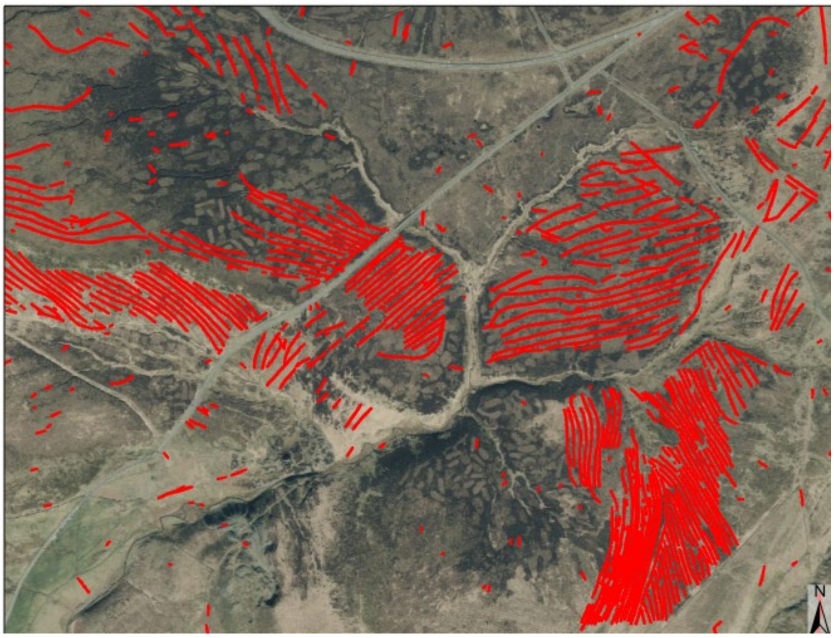


Figure 7-7 Large area of grips in the Peak District Aerial imagery © 2015 Getmapping plc and Bluesky International Ltd.

Gullies

Figure 7-8 shows the gully outputs for an area in the Forest of Bowland. The gullies in general appear to be mapped very successfully, particularly for larger gullies visible in the aerial imagery.

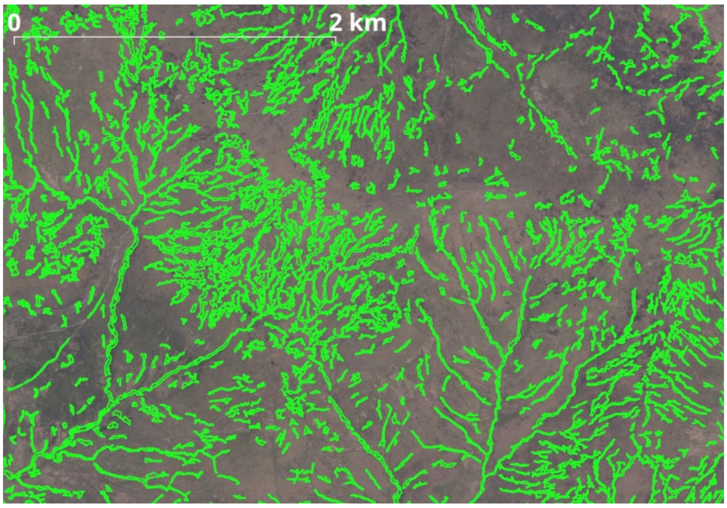


Figure 7-8: A large area of mapped gullies in the Forest of Bowland. Gullies detected by the model are shown in green. Aerial imagery © 2015 Getmapping plc and Bluesky International Ltd.

They also show clear flow pathways with smaller gullies branching from larger ones and often showing complex dendritic areas on the top of hills where severe erosion has occurred.

Larger gullies also often closely follow watercourses labelled in Open Street Map. The agreement with this independent crowd-sourced open data source is an indication that watercourses are correctly being identified by the model. In some cases, the quality of the outputs is hard to determine from a desk-based review, particularly when the model has classified an area as a gully which is not clearly visible in the aerial imagery or in the LIDAR. This seems to happen more often in areas where training data was not present, such as Dartmoor and the Lake District, which may indicate performance is not as reliable in these areas. Verification in the field is needed to understand this further. Gully features occasionally end or begin along the borders of individual image chips.

There are also cases where sections of gullies appear to have been missed. Some of these cases will likely be due to the model underperforming, but in other cases this may be due to the complexity of the gully network on the ground. For example, when we visited an area of dense gullies in the Peak District, we found that in many cases where the model had missed gully sections this was due to the presence of peat pipes where the gully was no longer visible from the air.

Haggs

Figure 7-9 (left) shows an example hagg output. The overhanging nature of hagg peatland surface features means visual identification and assessment of model performance from aerial imagery is challenging. However, haggs are often modelled along the edges of gullies which is largely where they are expected to exist on the ground. The outputs also seem to agree well with the ground truth data, as shown in Figure 7-9 (right). The model predictions often appear to be where there are areas of shadow or sudden colour change in the image. While this often appears to lead to correct identification, there are also cases where the model appears to misidentify other features which appear similar as haggs. The model performance also appears to depend on location and, as with other features, appears to perform less well in areas where training data is limited such as Dartmoor and the Lake District.

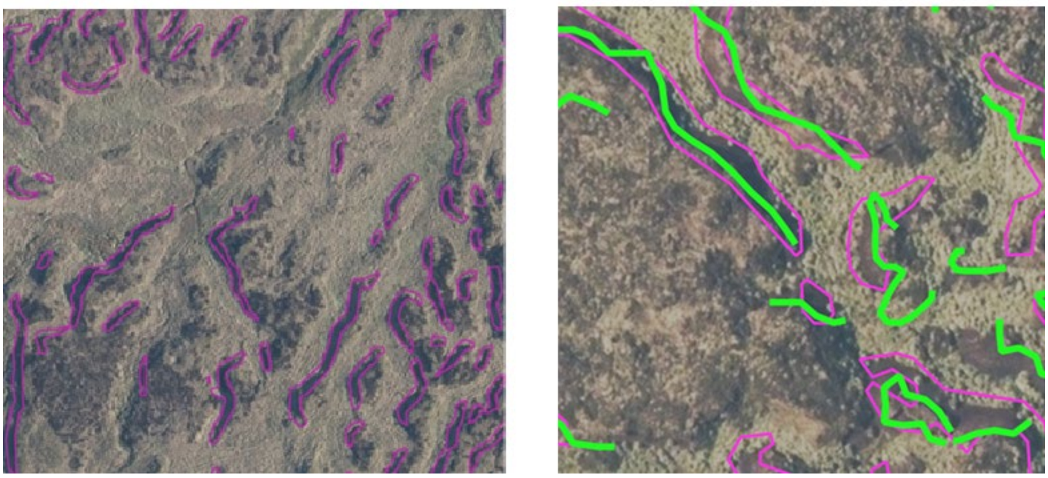


Figure 7-9 **left**: Example of Hagg model output (purple), Yorkshire. **right**: Comparison of modelled haggs (pink) with ground truth data (green) for an area on the North York Moors. Aerial imagery © 2015 Getmapping plc and Bluesky International Ltd.

An important aspect to note is with the ground truth data itself. This data was often collected for restoration purposes, not for model training. This means data collection may be geographically imprecise and not comprehensive. This affects both the performance of the model, but also the ability of the performance metrics to accurately reflect that performance. This issue applies to the gullies, grips and haggs, though is perhaps most pronounced in the hagg modelling. Figure 7-10Figure 7-10 shows an example image from training, highlighting the generally positive model performance against the poor ground truth, resulting in poor performance metrics.

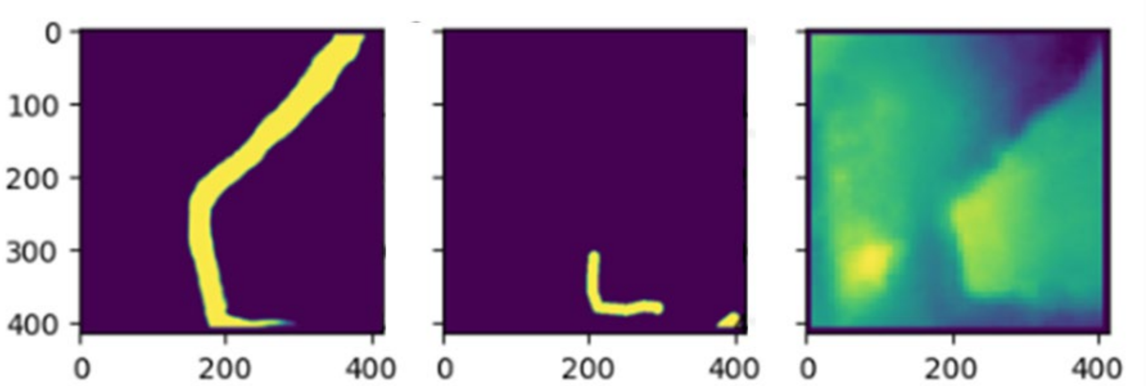


Figure 7-10 Example image of a hagg from model training. The **left** image shows the results of the model, with the **right** image showing the feature in LiDAR. Despite the ability of the model to successfully capture the feature from the LiDAR, the **middle** image shows the ground truth. Due to the poor ground truth, the metrics for this image are as follows: IOU: 0.0, f1: 0.0, recall: 0.0, precision: 0.0, accuracy: 0.88.

Dams

For the dams the results appear to be broadly accurate, with the model performing very well in many areas, clearly capturing the dams visible in aerial imagery (Figure 7-11Figure 7-11Figure 7-11). However, in some areas the model did not perform as well, either detecting high numbers of false positives or not detecting some dams clearly visible in the aerial imagery, so some caution is needed when using the outputs.

There are several possible reasons for this underperformance, the main one being the limited amount of training data available to train the model. This means that the model will struggle to identify dams which differ significantly from those included in the training data. For example, the model appears to often miss stone dams and dams which are most obvious due to their bow-shaped pools because these dams were not common in the training data.

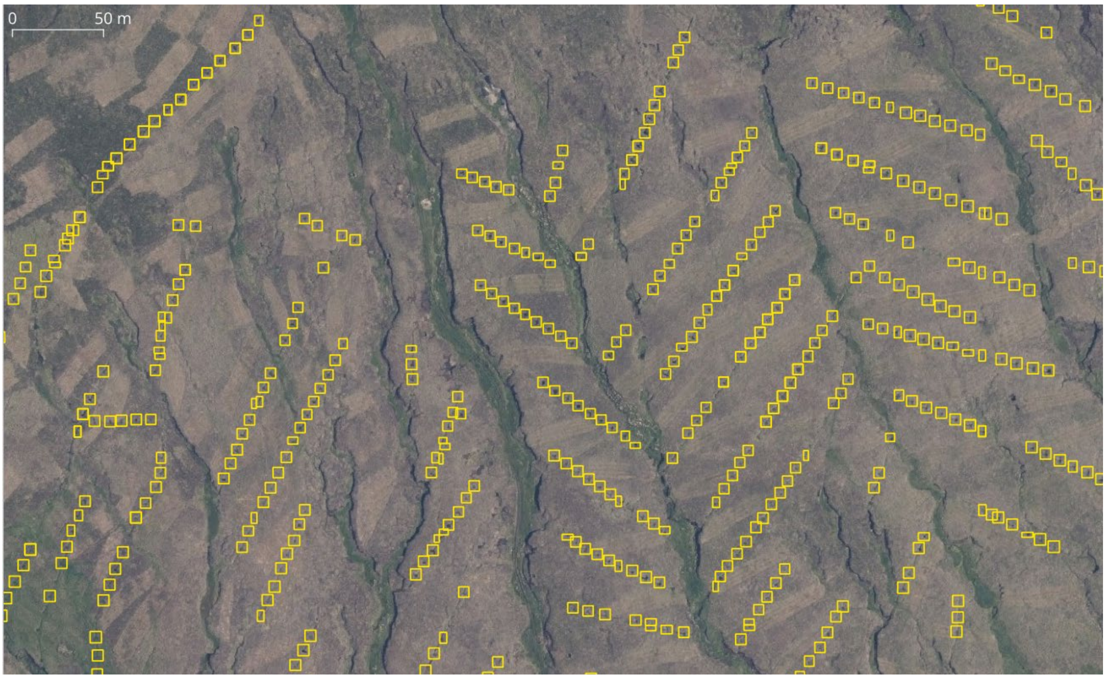


Figure 7-11: Example of dam model output in the North Pennines. Dams detected by the model are shown in yellow. Aerial imagery © 2015 Getmapping plc and Bluesky International Ltd.

Similarly, the model underperformed in areas where the vegetation or other characteristics of the landscape varied significantly from those in the training data such as in the Dartmoor area where the vegetation is very different, or areas with lots of bushes or large rocks that can be mistaken for dams. This is due to their similar appearance, with the rock or bush looking like the raised area of a dam and the shadow resembling the pool.

There are a few options which may improve performance.

1. Increasing the amount and variety of training data to ensure a wide range of dam type and vegetation types are included in the model training. Also having a multiclass model with different classes for different dam types such as peat dams and stone dams would likely allow the model to perform better for each dam type.
2. Including areas which do not contain any dams in the training data to help the model to learn to recognise objects which are not dams but which resemble dams, e.g. bushes and boulders along grips and gullies.

User Testing

In September 2024, peatland stakeholders provided feedback on an early iteration of AI4Peat outputs. A questionnaire was used to assess the usefulness of datasets from the England Peat Map portal, and a supplementary workshop allowed for further discussion and evaluation.

Stakeholders generally gave a positive review of the data outputs, finding them useful while also identifying limitations aligned with those outlined in the report.

The model shows strong potential in identifying features such as gullies, grips, and hagged peat, but challenges remain with smaller features, shadows, and misclassifications. For example, animal tracks are sometimes mistaken for grips/gullies, and gorse is misclassified as haggs. While larger drainage features are often well-identified, intricate details like dendritic erosion networks and grip dams are occasionally overlooked.

Specific feedback revealed that gullies, although often accurate, can appear fragmented rather than continuous. Haggs and gullies were sometimes misclassified, particularly in areas with mining spoil. Grips are generally accurate but occasionally misidentify boundary walls, tracks, or least. Grip dams are inconsistently detected, even in areas with blocked ditches.

A consistent request was for some kind of confidence scores for each feature identified. Such metrics could guide users in understanding the reliability of the outputs. Confidence metrics have been provided for the dams, but providing a confidence score for the other features is a more complex task and has therefore not been included in this publication.

Quantitative QA

Calculating the accuracy of the post-processed mapped features is challenging when the ground-truth data that is used to evaluate the outputs is not always accurate in itself (see Figure 7-9 and Figure 7-10). The main difficulties in producing accuracy metrics arise from i) ground-truth polygons not always following the feature exactly, meaning the modelled outputs are sometimes more accurate than the digitised labels; ii) not all features in an area have been digitised so false positives will be overestimated if the model identifies these and iii) modelled features may not overlap directly with features on the ground but are within a proximity tolerated by end-users and thus considered a successful identification.

To overcome these issues, we used a grid-based method of providing success scores to the outputs. Figure 7-12 illustrates the resulting metrics for an example region. For full details on this methodology please see Annex 6. Using this grid based method, we can again calculate accuracy, precision, recall and the F1-score for the final outputs.

Table 7-9 details these grid-based performance metrics for a region in the West Pennines. 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.

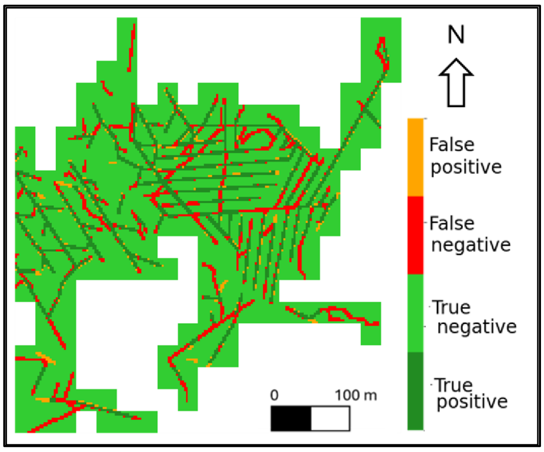


Figure 7-12 Example of the pixel-based method of calculating accuracy metrics for grips at a 5 m resolution for a 1 km^2 (SD6418) region of the West Pennine ground-truth dataset.

Table 7-9 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

Evaluation

The AI4Peat project has successfully developed a detailed spatial dataset of peatland surface features, including grips, gullies, haggs, and dams, leveraging aerial imagery and machine learning. By enabling rapid and large-scale mapping, this approach offers significant benefits for peatland monitoring, restoration planning, and environmental research. Furthermore, the project highlights the potential of AI and machine learning in environmental monitoring to automate detection and tracking on a national scale.

Despite its successes, the project also faced several limitations. Model performance was affected by the availability and quality of training data, particularly in regions like Dartmoor and the Lake District, where sparse or misaligned data reduced accuracy. Additionally, the calculated dimensions of features, such as slope and depth, require further validation, as they are limited by the resolution of LIDAR data and the high computer costs of the calculation methods. Environmental factors, such as vegetation shadows and human-made structures, also posed challenges for accurate classification. For this reason, users should make sure they are aware of the limitations of the data, follow user guidelines and where possible undertake their own quality checks.

Future work should focus on enhancing accuracy by incorporating higher-quality and more diverse training data, exploring additional remote sensing datasets, and expanding the spatial and temporal scope of the analysis. Applying models to historical aerial imagery and extending coverage to lowland peatlands are promising avenues for monitoring change and facilitating restoration efforts.

8. Discussion

The England Peat Map project has created an up-to-date inventory of England's peatlands, mapping peaty soil extent and depth, their vegetation cover over the nation, and areas of bare peat and drainage & erosion features across upland peatlands. It has also collated a substantial repository of data, including new survey data, that is free for anyone to use and has developed survey methods and data sharing standards.

8.1. Modelling

Extent and depth

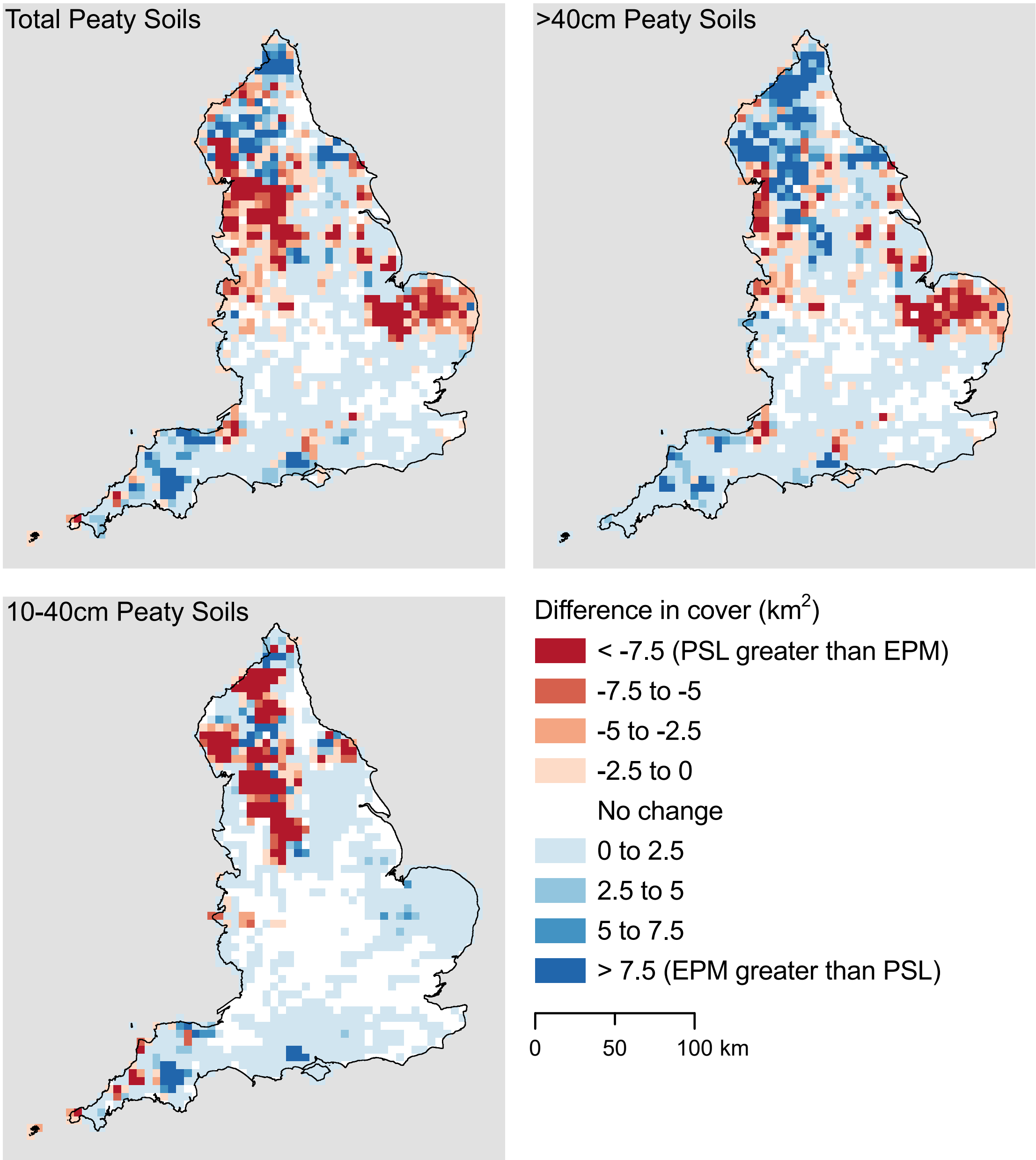
We compared the EPM peaty soil extent and depth model with other sources of mapped peat (see Table 8-1). The new estimate of the total area of peaty soils in England from the EPM peaty soil extent model is 11,047km^2. This compares with 12,071 km^2 of peaty soils in the Peaty Soils Location (PSL) (Natural England, 2010) dataset. It is important to recognise that the majority of difference between these two datasets is not due to changes in peaty soil extent over the last 15 years. They were created using very different methods and using different source data. The apparent difference in overall area is in line with our expectations.

Table 8-1 Comparison between areas mapped as peaty soil: Peaty Soil Location (Natural England, 2010) and England Peat Map

Peaty soils	PSL Area (km2)	EPM Area (km2)	Difference Area (km2)
Shallow	5,272	3,971	-1,301
Deep	6,799	7,076	+276
Shallow + Deep	12,071	11,047	-1,024
Peaty pockets	2,114	n/a	n/a

The biggest difference is in shallow peaty soils, which cover an area more than 25% smaller in the EPM estimate whereas deep peaty soils cover a slightly larger area than previously estimated. The extent of “Soil with Peaty Pockets” mapped in the PSL could not be compared as the location of the ‘pockets’ within those areas was not mapped, and the majority of the area included in that class is recorded as “mostly non-peaty”. Some of the difference between the two maps is due to an improvement in the mapping of these peaty pockets and may explain the increase in the mapped area of deep peaty soils.

Differences between the two products is regionally variable. We calculated the area covered by peaty soil in each of the maps for each 10 x 10 km Ordnance Survey grid square (100 km^2) and then subtracted the EPM figure from the PSL figure (see Figure 8-1). Although the total area of deep peaty soils has changed relatively little between the two maps, there is a strikingly lower prediction for the area covered by deep peaty soils (Figure 8-1 B) in lowland areas, particularly in East Anglia.



Map produced by the England Peat Map Team, Date: 31/03/2025.

Figure 8-1 Comparison between maps of peaty soil extent by depth class. These three maps show differences in peat cover between the EPM output and the Peaty Soils Location (PSL) (Natural England, 2010) map, for: (top left) all peaty soil, (top right) deep peaty soils and (bottom left) shallow peaty soils. The maps show the difference in area of peaty soils per 100 km^2. Red colours indicate that EPM predicts less peaty soil than PSL, blue colours indicate that EPM predicts more peaty soil. The PSL "Soil with Peaty Pockets" class is excluded because it shows areas of "mostly non-peat soils" (Natural England, 2010). Note that these maps show differences in mapping, not change in peaty soils.

While our confidence in the predicted depth of lowland peaty soils is lower than for upland peaty soils, due to the relative sparseness of data, the difference can be explained by the large time span between the fieldwork supporting the two estimates. The lowland component of the PSL is underpinned by source data dating back to the Lowland Peat Survey of the 1980s (Avery, 1980, Burton and Hodgson, 1987), with decades of peatland degradation and erosion having already taken place when the PSL was published in 2010, and even longer until the field surveys underpinning the EPM model were carried out between 2022 and 2024. We are therefore confident that the general direction of our prediction reasonable in these areas, at least for deep peaty soils. Our estimate of shallow peaty soils has increased, which fits equally with this understanding. Upland areas by contrast show a greater area of deep peaty soils compared to the PSL, and it is likely that this is due to existing pockets of peat in areas such as the Border Mires in Northumbria, being under-recorded in the PSL.

Vegetation

The vegetation and land cover map generally performs well in differentiating between bog and fen vegetation, providing a clear distinction between these two broader habitat groups in most cases. It is also capable of identifying large areas of bare peat reliably. It faces challenges in differentiating between some habitats within these groups, for instance in transition zones tall and short fen vegetation can sometimes be misclassified, especially around the 50 cm height threshold between categories. It also struggles to accurately represent mosaic habitats, particularly in areas dominated by Calluna vulgaris and Eriophorum spp. The mixed composition of these habitats makes them difficult to distinguish, particularly if the mosaic is within the 10m x 10m pixel resolution of our output. It also struggles in some instances to distinguish features that are visually similar to bare peat, such as shadows and darker patches of Calluna vulgaris, which are sometimes misclassified. Careful consideration is needed when interpreting the maps in these areas.

The vegetation and land cover map integrates existing data sources to supplement its own mapped data, ensuring interoperability with key Defra group datasets, including the Living England Habitat Probability Map and the National Forest Inventory. This represents new evidence that did not exist before this map and complements other land cover products like Living England, as similar methodologies are used and much of the same underlying data is incorporated. Using a consistent evidence base enhances its value for supporting nature recovery through Defra policy and wider strategies. Importantly, it is a free and open product and ensures accessibility for researchers, policymakers and land managers.

A key advantage of modelled vegetation and land cover products is their ability to incorporate new data and satellite/aerial imagery over time. This adaptability allows for continuous monitoring of peatland condition, tracking the progress of restoration efforts and improving our understanding of environmental changes which are crucial for greenhouse gas emissions accounting. In the future, change detection techniques could further enhance this capability by identifying and analysing trajectories of change, providing valuable insights into long-term habitat dynamics and the effectiveness of restoration interventions.

Upland peat drainage and erosion

The AI4Peat project team provided a detailed spatial dataset of peatland surface features including grips, gullies, haggs and dams by leveraging aerial imagery and machine learning models. This modelling approach allows mapping to be done on a much larger scale and much shorter timescales than is possible through manual digitisation or field work. It also provides opportunities for future updates, allowing change to be tracked over time. More broadly, this project demonstrates the potential of using AI and Machine Learning in environmental monitoring to provide faster, automated detection and tracking of the environment at a national level.

There were some challenges, arising from the available training data, the complexity of the peatland landscape, and the evolving nature of AI and machine learning in this context. The accuracy of the model depends heavily on the amount and variety of training data. In some areas, such as Dartmoor, Exmoor, and the Lake District, the lack of training data has resulted in reduced accuracy. For instance, the model may misclassify large rocks as dams or fail to recognize features in regions with unique vegetation types not included in the training data.

The quality of the training data affects the model's ability to learn effectively. A lot of the training data was not created for the purpose of training models and is therefore not always accurate (e.g. grip training data was often offset slightly from the actual grip location). This will likely impact the model’s performance. Errors in the training dataset can also lead to inaccurate performance metrics, and this may have influenced the selection of the final model. The metrics used to assess models might also not fully reflect real-world performance, which underscores the importance of cautious interpretation.

The calculated dimensions of features have not been as thoroughly reviewed as their locations and should therefore be used with caution. While these measurements aim to balance accuracy and computational efficiency, different calculation methods could produce slightly varying results and, particularly for slope and depth, are limited by the resolution of the LIDAR data. Further user research is needed to understand how these dimensions will be applied in practice to ensure that they are fit for purpose and to direct future analysis.

Peatland features can vary significantly depending on local conditions. Shadows from vegetation, seasonal changes, or human-made structures like boundary walls and footpaths can all lead to misclassifications. For example, shadows are sometimes identified as haggs, while linear features such as walls may be mistaken for grips.

8.2. Field Survey

The EPM field survey was primarily conducted to fill gaps in data available from other sources. Existing survey data was heavily biased towards the uplands and towards sites being explored for restoration. EPM field survey therefore had to target areas that were less straightforward to obtain permission to access and less concentrated in space. Contacting owners and occupiers and obtaining permission was a substantial undertaking, and we found relatively few potential contractors in the market with the necessary combination of understanding, skills, and tools (especially GIS) to deliver this efficiently. We also found that there were relatively few soil surveyors with the necessary skills in the market, and that general ecological and botanical surveyors often had little experience of soil survey. When conducting a very dispersed national survey, this presented logistical challenges. We were also continually collating existing survey data, so our understanding of the data gaps evolved in parallel with survey delivery, further complicating the survey logistics.

Almost the full range of peatland types were surveyed, from the very wet to the very dry. Optimally these would be surveyed with different tools, which would have entailed excessive costs for what is a national survey, and after testing peat corers against peat probes, the latter was chosen. It was found that laboratory testing of organic matter content, through loss on ignition, often disagreed with the field determination of surveyors (see Annex 6), so the majority of field identifications of peaty soil were verified with samples. Additional complication arose from providing data for different model types (soil and vegetation), resulting in the development of two separate protocols.

8.3. Data

Peatland restoration projects drive the production of large amounts of data. Funds for peatland restoration commonly require initial surveys followed by subsequent monitoring whilst research into the effects of restoration and surveys informing site management also require periodic measurement of peatland condition. Other peatland data is gathered as part of scientific surveys, for regulatory purposes, infrastructure projects, and for habitat and species conservation. There is no central coordination of this data gathering, so there are large variations in methods, standards, availability and licensing. However, opportunities did emerge for more consistent data gathering.

One of the major funders for peatland restoration in recent years, the Nature for Capital Peatland Grants Scheme (NCPGS), is administered by Natural England. The EPM and NCPGS teams collaborated to develop standard data formats and ensure the data would be useful for our purposes but also licensed for publishing so that it can be used by others. A different model was available in the north of England, where the Great North Bog partnership was already engaged in regional coordination of activities and data. The pooled data of the GNB partners is substantial and a major effort was made to mobilise it for EPM and beyond. Smaller peat partnerships and other organisations throughout the country were also immensely helpful.

All EPM outputs rely to a large extent on data collected by others, most significantly the surface features models, which used only pre-existing survey data. We recognise that these data sources were fundamental to the success of the project and are very grateful to the organisations and individuals who provided the data for us to use.

The most resource demanding aspect of mobilising such a large amount of disparate data was ensuring intellectual property rights were respected. Most data arrived without an explicit data licence so had to be negotiated. Additional staff resources from several teams were required to identify and engage with data owners. While the vast majority were willing to allow use of the data, securing explicit permission to publish it was less straightforward. Organisational data owners had perfectly legitimate procedures that made it more time consuming to agree data licenses.

The creation, storage, exchange, and use of peat survey data would be made simpler and easier by the development and adoption of a set of standards. EPM Project have developed and published an initial draft data exchange standard for Peat Depth data < Working Towards a Peat Data Standard | IUCN UK Peatland Programme>. This has helped the project to manage peat presence and depth measurement data, however much work remains necessary before peat data standards can be adopted and benefits realised. Future priorities include standards for peatland surface features data (grips, gullies, hagging, bare peat), and peatland restoration data, which could be based on existing formats in use e.g. within the National Trust or peat partnerships. It is also important to recognise that mobilising data can be complex, costly and should not be underestimated. Significant staff resources were required for this project to engage with data owners to agree and manage the data licences and licencing process, before any data could be shared.

8.4. Using the England Peat Map

The significance of the EPM products represents a major upgrade to the available evidence base about England’s peat resources, because:

- They are detailed, based on recent survey data and state-of-the-art modelling techniques, are of known quality, and are freely available to everyone
- They include the first published England-wide map of predicted peat depths, and upland peatland surface features as an indicator of condition
- They are updateable with new data, to achieve further accuracy improvements

- They can be built on to develop new condition monitoring methods
- Their importance is demonstrated by the range of expected future uses and applications.

Potential applications

Our expectation is that the EPM, as a new product which significantly improves the evidence base available about England's peatlands, will find a wide range of uses across Defra group and amongst the wider peatland community. We anticipate it will broadly support policy development in Defra and be used to improve the Greenhouse Gas Inventory. During project delivery, a diverse range of preliminary discussions were held with Defra and Natural England colleagues to try and understand potential uses across areas including protecting peatlands from undesirable tree planting, protected sites strategies, targeting peatland restoration, nature recovery, Local Nature Recovery Strategies, and more. It is only really after the publication of the final EPM outputs, accompanied by the information we have provided on constraints and limitations, that these discussions can develop to ensure the full benefits of EPM can be realised.

Constraints and limitations

Model performance

The accuracies we were able to achieve were constrained by data availability and data quality. Despite collating more than 1,700 peat survey datasets across England, survey data availability from lowland peatland areas was limited, reflecting the historical lack of restoration in the lowlands beyond a few well-known sites (and reflected in the lower data density described in section 6 above). Furthermore, EPM field survey did not yield as much lowland data as we aimed for. Survey data gaps will remain even though future data will be available from several sources (such as Landscape Recovery projects, Nature for Climate Peatland Grant Scheme monitoring data, Environment Agency Lowland Agricultural Peat Water Discovery Grant Pilots, the NCEA England Ecosystem Survey, and other monitoring programmes). Targeted additional surveys will be required to fix this issue and improve lowland model accuracies, and other weaknesses that have been identified such as isolated small mires such as valley fens which were also poorly described by the available survey data. We hope that by sharing as much of our field survey data as possible, others can also identify areas of poor coverage and contribute new surveys and new data to improve the models in future.

Predictor variables were adequate for model development in upland areas, however again in the lowlands the lack of detailed maps of historical land use, which would have been useful additional predictors of the presence and/or depth of lowland peaty soils, and the lack of topographical variation, contributed to the lower accuracy of lowland extent and depth models generally. In addition, some potentially useful commercially available predictor data were not used due to data licence constraints – particularly British Geological Survey 1:50,000 scale mapping and Cranfield University's National Soil Inventory data were available to us, but using them would have hindered publication of our outputs under an Open Government Licence.

It is also important to remember that there are additional sources of error in any field survey. The gold standard for peat depth or thickness measurement is to use a peat corer to extract the soil, but it is very time consuming. Peat probing (see Annex 3) using semi-rigid threaded poles is widely accepted in the uplands as a reliable technique for measuring peat depth or thickness - however results will vary depending on the precise method, equipment used, observer differences, substrate differences, water content of the peat, and other factors such as how degraded or compacted the peat is. These errors are rarely, if ever, quantified in peat depth measurements (see section 8.5 below). We believe, however, from QA checks of EPM field survey depth data, that the errors associated with EPM depth models are likely to be comparable to the errors associated with manual peat probing in many circumstances. This is important to bear in mind when using the depth model and reviewing the confidence rating associated with an area.

Scale of appropriate use

EPM outputs for peaty soil extent, depth and vegetation are mapped at 10m pixel resolution as this reflects the high-resolution predictor data we have available. Surface features model outputs are based on even higher resolution (12.5cm) air photography imagery. From the various assessments we made of accuracy (standard model metrics, and expert user opinion collected during feedback exercises and semi-structured feedback interviews), we developed some clear and simple guidance on the use of our products.

At **site scale**, use EPM model outputs with caution. Consider collecting additional ground truth data to gain a better understanding of the accuracy of EPM outputs on your site. EPM output predictions are to be understood as the predicted average measurement for the output pixel. Results should be seen as indicative, and whilst EPM can provide a guide as to what may be found on a site on average, in some cases this will not be reliable. There will be some instances where areas of known peaty soils have not been predicted by the extent model, conversely areas where the extent model has confidently, but wrongly, predicted peaty soil occurrence. Where known peaty soil sites are not predicted this will normally reflect a lack of survey data for these areas, which is something we aim to address in a future update. EPM is therefore unlikely to be a suitable substitute for field survey at site scale. As geographic scale increases (i.e. the map becomes more 'zoomed out') results are likely to become more reliable. Consider limiting the display of EPM models in your map systems to a suitable minimum scale threshold (e.g. 1:50,000), to avoid giving a misleading impression of model accuracy.

At **landscape** scale, model outputs are more reliable. Summaries of EPM models (e.g. total peat areas or volumes, areas of each condition class) within e.g. Protected Landscapes or large administrative areas are likely to be reliable.

At **national** (England) scale, EPM model outputs will be the best available evidence, e.g. for summary statistics, and for national reporting purposes.

Buried, wasted, and shallow peat

Buried, wasted, and shallow peats were represented in our models in the following ways:

Buried peat refers to peaty soil horizons that occur buried beneath mineral soil horizons in the soil profile. In some lowland areas of England, multiple layers of buried peat can occur, making prediction of the total thickness of peaty soil horizons buried within the soil profile very difficult. We trained the EPM extent and depth models only on peaty soil horizons which start within the top 1m of the soil profile. Consequently, buried peat deposits are not separately identified from surface peat deposits in our models. Predicted peat depth values may therefore refer to surface and/or buried peat deposits (where those buried deposits start in the first metre) or both combined.

Wasted peat refers to peaty soil that has, through years of agricultural use, become degraded and mixed (ploughed in) with underlying mineral soil. We have not identified wasted peat specifically as a 'type' but have recorded peaty soil horizons and depths within the top 1m of the soil profile in areas where wasted peat is a common feature of the landscape. At many of these survey sites we have also taken soil samples to confirm organic matter content. The extent and depth models therefore treat wasted peat in the same way as any other peaty soil - i.e. we predict the probability of peaty soil occurrence in each pixel of the output, and predict the total thickness of all peaty soil horizons found within the top 1m of the soil. An approximation of the extent of wasted peaty soil could be extracted from the EPM data by locating lowland peaty soil pixels with a depth less than 30cm located in areas with intensive agricultural land cover types.

Shallow peat thicknesses of less than 10cm are unlikely to be recorded accurately in the field. They tend to include a variable element of surface organic matter (humous layer) and the exact beginning and end of the peaty soil horizon is very difficult to determine. EPM have therefore used a minimum thickness of 10cm as the shallowest depth of the peaty soil depth model. This means that where the depth model predicts peaty soil presence but it is less than 10cm thick, these pixels are assigned a 10cm value in a post-processing step.

8.5. Potential Next Steps

The following identifies ways of addressing known constraints and limitations of the England Peat Map.

Predictions could be improved by **addressing known accuracy issues** (particularly the accuracy of peaty soil extent in lowland agricultural areas, and peaty soil depth in areas of deep peat), through a combination of additional targeted field survey data to train and validate models, and updated and improved predictors, including from other as-yet unused sources. Further improvements could be gained by developing **'hybrid' products** where modelled outputs are supplemented with more accurate and reliable additional data sources.

This project has demonstrated the high value of using **field data from existing surveys**. Future improvement of EPM models should prioritise this, and use new field survey primarily to fill gaps in coverage. Survey data from Defra-funded activities (such as England Ecosystem Survey, ELMS, NCPGS, NE's Long-Term Monitoring Network, Trees Action Plan Delivery surveys) has already been identified but in some cases could be made more readily available and more standardised. Private-funded data sources particularly Peatland Code applications and monitoring surveys should also be considered. Ideally all data produced as part of publicly funded projects would be published openly and in accordance with the FAIR data principles.

To facilitate re-use of field data, **peat data standards** should be developed and promoted, ideally at the UK level. This should include not just peat survey activities, but also restoration and other important peat-related data types. Natural England has been working with **stakeholders** (particularly National Trust and IUCN UK Peatlands Programme) on standards for peat data, but an organisational owner of standards at a UK level would help with embedding and maintaining standards for the whole UK peatland community. Future standards should include surface features, and bare peat, restoration and other important peat-related data types.

EPM model outputs could be combined with **data from restoration** planning and activity data to provide improved capabilities around monitoring and reporting for peatland restoration.

Continuation of England-scale peat mapping work provides opportunities to incorporate additional priorities such as:

- the incorporation of new products into Greenhouse Gas Inventory reporting processes,
- the development of new peatland condition monitoring tools including change detection of vegetation classes and promising elements of surface motion pilot work,
- ensuring summary statistics continue to be updated and available,
- an investigation of the accuracies involved in peat probing and other survey methods,
- the development of a combined condition index map,

- the development of new models of peaty soil total carbon content and mire types based on newly published eco-hydrological guidelines (Wheeler and others, 2023), and
- support for the uptake of project field survey protocols by citizen scientists.

Peaty soil is currently mapped differently across the UK leading to very different outputs. This would be addressed by the development of a **UK peat map**, using combined training data and the best models from the four countries.

9. References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Irving, S., Geoffrey, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D.G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y. and Zheng, X. 2016. TensorFlow: A System for Large-Scale Machine Learning, in *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation*. Savannah, GA, USA, pp. 265–283.

Artz, R., Evans, C., Crosher, I., Hancock, M., Scott-Campbell, M., Pilkington, M., Jones, P., Chandler, D., McBride, A. (SNH), Ross, K. (Falklands C. and Weyl, R. (NIEA) 2019. Commission of Inquiry on Peatlands : The State of UK Peatlands- an update. IUCN. Available at: https://www.iucn-uk-peatlandprogramme.org/resources/commission-inquiry.

Bain, C.G., Bonn, A., Stoneman, R., Chapman, S., Coupar, A., Evans, M., Gearey, B., Howat, M., Joosten, H., Keenleyside, C., Labadz, J., Lindsay, R., Littlewood, N.A., Lunt, P., Miller, C.J., Moxey, A., Orr, H., Reed, M., Smith, P., Swales, V., Thompson, D.B.A., Thompson, P.S., Van de Noort, R., Wilson, J.D. and Worrall, F. 2011. IUCN UK Commission of Inquiry on Peatlands. IUCN, Edinburgh. Available at: https://www.iucn-uk-peatlandprogramme.org/resources/commission-inquiry.

Bord na Mona 1985. Fuel peat in developing countries, *World Bank Technical Paper*. World Bank, Washington DC. Available at: http://documents.worldbank.org/curated/en/1985/06/439714/fuel-peat-developing-countries.

Campbell, H.A., Urbano, F., Davidson, S., Dettki, H. and Cagnacci, F. 2016. A plea for standards in reporting data collected by animal-borne electronic devices, *Animal Biotelemetry*, 4(1), pp. 1–4. Available at: https://doi.org/10.1186/s40317-015-0096-x.

Chen, T. and Guestrin, C. 2016. XGBoost: A scalable tree boosting system, in *Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining*. San Francisco, CA, USA, pp. 785–794. Available at: https://doi.org/10.1145/2939672.2939785.

Defra 2021. England Peat Action Plan, (May), pp. 1–40. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1010786/england-peat-action-plan.pdf.

Evans, C., Artz, R., Moxley, J., Smyth, M.-A., Taylor, E., Archer, N., Burden, A., Williamson, J., Donnelly, D., Thomson, A., Buys, G., Malcolm, H., Wilson, D., Renou-Wilson, F. and Potts, J. 2017. Implementation of an Emissions Inventory for UK Peatlands. Report to the Department for Business, Energy & Industrial Strategy. Centre for Ecology and Hydrology, Bangor.

Evans, C., Morrison, R., Burden, A., Williamson, J., Baird, A., Brown, E., Callaghan, N., Chapman, P., Cumming, A., Dean, H., Dixon, S., Dooling, G., Evans, J., Gauci, V., Grayson, R., Haddaway, N., He, Y., Heppel, K., Holden, J., Hughes, S., Kaduk, J., Jones, D., Matthews, R., Menichino, N., Misselbrook, T., Page, S., Pan, G., Peakock, M., Rayment, M., Ridley, L., Robinson, I., Rylett, D., Scowen, M., Stanley, K. and Worrall, F. 2016. Lowland peatland systems in England and Wales – evaluating greenhouse gas fluxes and carbon balances. Final report to Defra on Project SP1210. Centre for Ecology and Hydrology, Bangor. Available at: http://randd.defra.gov.uk/Default.aspx?Module=More&Location=None&ProjectID=17584.

Finlayson, A., Marchant, B.P., Whitbread, K., Hughes, L. and Barron, H.F. 2020. Estimating organic surface horizon depth for peat and peaty soils across a Scottish upland catchment using linear mixed models with topographic and geological covariates, *Soil Use and Management* [Preprint]. Available at: https://doi.org/10.1111/sum.12596.

Food and Agriculture Organization 1988. World Reference Base for Soil Resources. World Soil Resources Report No. 84. FAO, Rome, Italy.

Hamer, A.M. 2021. Mapping agricultural land in support of opium monitoring in Afghanistan with Convolutional Neural Networks. Cranfield University. Available at: https://dspace.lib.cranfield.ac.uk/handle/1826/20991.

Higgins, A. 2011. Towards an assessment of the state of UK Peatlands - JNCC Report No 445, *Inquiry*. Joint Nature Conservation Committee, Peterborough. Available at: https://doi.org/ISSN 0963-8091.

Hijmans, R.J. 2024. terra: Spatial Data Analysis. Available at: https://cran.r-project.org/package=terra.

HM Government 2021. Net Zero Strategy: Build Back Greener. HM Government, London, UK. Available at: https://www.gov.uk/government/publications/net-zero-strategy.

Horsburgh, J.S., Tarboton, D.G., Piasecki, M., Maidment, D.R., Zaslavsky, I., Valentine, D. and Whitenack, T. 2009. An integrated system for publishing environmental observations data, *Environmental Modelling and Software*, 24(8), pp. 879–888. Available at: https://doi.org/10.1016/j.envsoft.2009.01.002.

IGGI 2005. The Principles of Good Data Management, *Data Management*, pp. 1–35. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/14867/Good_dataMan.pdf.

IPCC 2014. 2013 Supplement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories: Wetlands, Hiraishi, T., Krug, T., Tanabe, K., Srivastava, N., Baasansuren, J., Fukuda, M. and Troxler, T.G. (eds). IPCC, Switzerland. Available at: https://www.ipcc-nggip.iges.or.jp/public/wetlands/pdf/Wetlands_Supplement_Entire_Report.pdf.

IUCN 2023. Peatland Code Field Protocol. IUCN. Available at: www.iucn-uk-peatlandprogramme.org/peatland-code/introduction-peatland-code/projects.

IUCN 2025. Peatland Benefits. IUCN. Available at: https://www.iucn-uk-peatlandprogramme.org/about-peatlands/peatland-benefits (Accessed: 6 March 2025).

Jasiewicz, J. and Stepinski, T.F. 2013. Geomorphons-a pattern recognition approach to classification and mapping of landforms, *Geomorphology*, 182, pp. 147–156. Available at: https://doi.org/10.1016/j.geomorph.2012.11.005.

Joosten, H. and Clarke, D. 2002. Wise Use of Mires and Peatlands, *International Mire Conservation Group / International Peat Society*, p. 304.

Kuhn, M. and Wickham, H. 2020. Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles. Available at: https://www.tidymodels.org.

Li, J., Ma, R., Cao, Z., Xue, K., Xiong, J., Hu, M. and Feng, X. 2022. Satellite Detection of Surface Water Extent: A Review of Methodology, *Water*, 14(7), p. 1148. Available at: https://doi.org/10.3390/w14071148.

Lindsay, R. 2010. Peatbogs and carbon: a critical synthesis, *Unpublished report to the RSPB*, p. 315. Available at: https://birdwatch.rspb.org.uk/Images/Peatbogs_and_carbon_tcm9-255200.pdf.

Lourenco, M., Fitchett, J.M. and Woodborne, S. 2023. Peat definitions: A critical review, *Progress in Physical Geography*, 47(4), pp. 506–520. Available at: https://doi.org/10.1177/03091333221118353.

Minasny, B., Berglund, Ö., Connolly, J., Hedley, C., de Vries, F., Gimona, A., Kempen, B., Kidd, D., Lilja, H., Malone, B., McBratney, A., Roudier, P., O'Rourke, S., Rudiyanto, Padarian, J., Poggio, L., ten Caten, A., Thompson, D., Tuve, C. and Widyatmanti, W. 2019. Digital mapping of peatlands – A critical review, *Earth-Science Reviews*, 196. Available at: https://doi.org/10.1016/j.earscirev.2019.05.014.

Natural England 2008. Technical Information Note TIN037: Soil Texture. Available at: https://publications.naturalengland.org.uk/publication/32016.

Natural England 2010. England's peatlands: carbon storage and greenhouse gases (NE257). Natural England, Peterborough. Available at: http://publications.naturalengland.org.uk/publication/30021.

Natural England 2021. National Character Areas (England). Natural England. Available at: https://environment.data.gov.uk/dataset/f955fff5-a132-4e25-8c7f-0d43664ff42e (Accessed: 1 October 2024).

Natural England 2023. Data Exchange Standard for Peat Surveys. Available at: https://www.iucn-uk-peatlandprogramme.org/sites/default/files/2023-06/Data Exchange Standard for Peat Surveys v1.0.pdf (Accessed: 23 January 2025).

Office for National Statistics 2019. UK natural capital: peatlands. Available at: https://www.ons.gov.uk/economy/environmentalaccounts/bulletins/uknaturalcapitalforpeatlands/naturalcapitalaccounts (Accessed: 24 January 2025).

Parry, L.E., West, L.J., Holden, J. and Chapman, P.J. 2014. Evaluating approaches for estimating peat depth, *Journal of Geophysical Research: Biogeosciences*, 119(4), pp. 567–576. Available at: https://doi.org/10.1002/2013JG002411.

Penny Anderson Associates 2012. Natural England Commissioned Report NECR089: Mapping the status of upland peat using aerial photographs. Natural England. Available at: https://publications.naturalengland.org.uk/publication/369581.

Poggio, L., Lassauce, A. and Gimona, A. 2019. Modelling the extent of northern peat soil and its uncertainty with Sentinel: Scotland as example of highly cloudy region, *Geoderma*, 346(February), pp. 63–74. Available at: https://doi.org/10.1016/j.geoderma.2019.03.017.

R Core Team 2024. R: A Language and Environment for Statistical Computing. Vienna, Austria. Available at: https://www.r-project.org/.

Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J.J., Schröder, B., Thuiller, W., Warton, D.I., Wintle, B.A., Hartig, F. and Dormann, C.F. 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure, *Ecography*, 40(8), pp. 913–929. Available at: https://doi.org/10.1111/ecog.02881.

Shelhamer, E., Long, J. and Darrell, T. 2015. Fully convolutional networks for semantic segmentation, in *IEEE Conference on Computer Vision and Pattern Recognition*. Boston, USA, pp. 3431–3440. Available at: https://doi.org/10.1109/CVPR.2015.7298965.

Shorten, C. and Khoshgoftaar, T.M. 2019. A Survey on Image Data Augmentation for Deep Learning, *Journal of Big Data*, 6(1). Available at: https://doi.org/10.1186/s40537-019-0197-0.

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.

Wheeler, B., Eades, P., Tratt, R., Shaw, S. and King, L. 2023. Working Towards the Development of Ecohydrological Guidelines for Blanket Bogs and Allied Minerotrophic Habitats (Phase 2). Sheffield, UK. Available at: https://www.iucn-uk-peatlandprogramme.org/sites/default/files/2023-12/BlBogGuidelines_Phase 2 Report.pdf.

Wilkinson, M.D., Dumontier, M., Aalbersberg, Ij.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., … Mons, B. 2016. Comment: The FAIR Guiding Principles for scientific data management and stewardship, *Scientific Data*, 3, pp. 1–9. Available at: https://doi.org/10.1038/sdata.2016.18.

Williamson, J., Morton, D., Artz, R., Burden, A., Rowland, C., Tornero, L.C., O'Neill, A., Poggio, L., Khomik, M., Donnelly, D. and Evans, C. 2017. The role of earth observation in an integrated framework for assessing peatland habitat condition and its impact on greenhouse gas accounting. Final report to Defra, Project number MI07. Centre for Ecology and Hydrology, Bangor.

Young, D.M., Parry, L.E., Lee, D. and Ray, S. 2018. Spatial models with covariates improve estimates of peat depth in blanket peatlands, *PLOS ONE*.

Edited by B. Minasny, 13(9), p. e0202691. Available at: https://doi.org/10.1371/journal.pone.0202691.

10. List of Annexes

Annex 1: Field survey protocol – soils

Annex 2: Field survey protocol – vegetation

Annex 3: Peat probing guidance

Annex 4: Detailed hand texturing method and use of soil maps

Annex 5: Data Supplement

Annex 6: Technical Supplement

11. Glossary

This glossary defines how key terms are used in this report. It is grouped into themes of related concepts: Soil, Vegetation and Land Cover, Upland Peat Drainage and Erosion Features and Modelling.

Soil

Loss on ignition: A method for estimating the amount of organic matter in a sample of soil by:

1) drying and weighing the soil sample;
2) heating it to ignite (i.e. burn) organic matter in the soil; and finally
3) weighing it again.

The percentage difference between the weight after ignition and the dried weight is taken to represent the organic matter that is burned away i.e. the loss on ignition. In our protocol the soil sample is dried at 105 ºC for sixteen hours when it arrives at the laboratory, then a portion of it is weighed, heated at 375 ºC for sixteen hours and weighed again.

Peat is a type of soil "formed from carbon rich dead and decaying plant material under waterlogged conditions" (Bain *et al.* 2011). Definitions of peat soil vary between countries in relation to the minimum percentage of organic matter it is required to contain, and the minimum thickness it is required to have (Bord na Mona, 1985; Food and Agriculture Organization, 1988; Joosten & Clarke, 2002; IPCC, 2014; Lourenco and others, 2023). In the EPM Survey Field Protocol (Annex 1) the term "peat" is reserved for soils with an organic matter content greater than 50%. No minimum thickness criteria is used for the identification of peat. For the use of related terms in the EPM context, including loamy peat, sandy peat, peaty loam and peaty sand, see the definition of peaty soil below.

Peaty soil: In this report, the term peaty soil is used to describe a group of soil texture classes comprising peat, loamy or sandy peat, and peaty loam or sand. Peaty soil contains at least 20% organic matter (25% where there is more than 50% clay) (Natural England, 2008). A threshold thickness of 10 cm of peaty soil has been used for mapping peaty soils. Peaty soil is contrasted with two soil texture classes that have less than 20% organic matter: organo-mineral soil and mineral soil. See Table 11-1 and Annex 1.

Table 11-1 Soil texture classes

Soil texture class	Limiting percentage of organic matter and clay
peaty soil (group)	
Peat	greater than 50% organic matter
Loamy peat or **Sandy peat**	35% to 50% organic matter
Peaty loam or **Peaty sand**	20% to <35% organic matter (25-35% for over 50% clay)
not peaty soil (not grouped)	
Organo-mineral	6% to <20% organic matter (10-25% for over 50% clay)
Mineral	<6% organic matter (<10% for over 50% clay)

Soil texture: Soil texture describes the mixture of different particle sizes in soils along with organic matter content. The texture **class** of a soil is defined by the proportions of sand, silt, clay and organic material in the soil. Several texture classes can be combined into a soil texture **group** (Natural England, 2008). Texture is a fundamental soil property influencing key characteristics such as drainage, water storage, workability, susceptibility to soil erosion and suitability for different uses, and it plays a major part in defining soil 'structure'.

Wasted peat – Peaty soil "that has lost both its peat-forming vegetation and a significant depth of soil" (Higgins, 2011). It is characterised by mixing of soil mineral material with the peat organic matter (Higgins, 2011). The England Peat Map does not use wasted peat as a separate peat class.

Vegetation and land cover

Bare peat: "Peat that has had all its vegetation removed (e.g. by erosion) but has not been affected by a significant change of land use." (Higgins, 2011)

Domin scale: A common way of recording the proportion of an area covered by the living parts of a species of plant in a quadrat. The Domin scale ranges from 1 (less than 4% cover and 'few individuals') to 10 (91% to 100% cover). Because species overlap with other species (e.g. a shrub may overlap with grasses growing under it), the sum of all Domin values in a quadrat may, and often does, exceed 100%. See Annex 6, section 2.2 for discussion as to why this scale was not used in the final EPM Field Survey Protocol - Vegetation (Annex 2).

Upland peat drainage and erosion features

Grips: Artificial ditches that have been cut into the ground across large areas of upland peatlands in England to drain water from a peatland. They are often shallow. (Bruneau, 2014).

Grip blocks / dams: Structures created to prevent or slow the flow of water through grips, often as part of peatland restoration.

Gullies: Fluvial erosion channels which cut into a peat mass, resulting in loss of peat and significant dehydration of adjacent in situ peat. They can be naturally occurring features of peatlands and occur where blanket peats spread to the heads of valleys. However, they also occur where artificial drainage features become eroded, and where other pressures such as wildfire, overgrazing or pollution reduce vegetation cover and exacerbate erosion (Bruneau, 2014).

Haggs: Peat haggs are more varied than grips and gullies and as such are harder to define. As gullies erode and branch, especially on level or gently sloping areas, adjoining gullies can meet. This results in isolated 'islands' of peat. Haggs also form on the sides of gullies, typically on steeper slopes. Both types of haggs can either be entirely cut within the peat mass or extend downward into the mineral substrate (Bruneau, 2014).

Modelling

AI, **Machine learning**, **Deep learning**: A suite of modelling techniques (algorithms) that learn patterns from data to make decisions and predictions.

Biogeographic Zones (**BGZ**): A way of dividing England into 14 zones to improve models that use **Sentinel 2** data. It comprises groups of adjoining **National Character Areas** which have broadly the same Sentinel 2 orbital paths. Using BGZs in models reduces undesired variability of Sentinel 2 imagery in each region because each zone's images are taken during the same pass of a Sentinel 2 satellite. (Trippier and others, 2024)

Broad Landscape Type: see **National Character Area**

Convolutional Neural Network: A type of deep learning algorithm designed to analyse images by detecting patterns.

Digital Soil Mapping: Process of creating digital maps representing soil types and/or properties using computer-assisted techniques (e.g. remote sensing, GIS and computational methods). Typically involving the use of field and/or laboratory observations, environmental data, and quantitative relationships to generate geographically referenced soil databases.

Geomorphon: Categories of common land forms (Jasiewicz & Stepinski, 2013). The 10 most common geomorphons are flat, peak, ridge, shoulder, spur, slope, hollow, footslope, valley and pit. EPM only uses these 10 classes (see Annex 6). They can be calculated from a digital terrain model at different resolutions and threshold angles. EPM uses two resolutions and a threshold angle of 1 degree (see Annex 6). The higher resolution 50m geomorphons classify the shape of the land in a grid of 50m by 50m cells. The lower resolution 1500m geomorphons classify the shape of the land in a grid of 1,500 x 1,500m cells. The higher resolution geomorphons show small variations in local land form (e.g. a dip in a peatland would be recorded as 'valley', 'hollow' or 'pit'), but do not show large features (such as that the dip is in a large flat plain), whereas the lower resolution geomorphons pick up these larger features but not the local variation.

Image chip: A small subset of a larger image used for analysis.

InSAR coherence: A measure of how consistent satellite radar signals are over time. Used to track small changes in elevation on the Earth's surface.

Interpolation: A statistical method to estimate unknown values between known data points.

Kriging: An advanced **interpolation** method that considers both distance and spatial relationships between data points to estimate unknown values. See also **Variogram**.

LiDAR: An instrument that uses laser pulses to create 3D maps of the landscape.

Model training, **model fitting**: Providing an algorithm with data from which it can learn relationships. Here we use it in the context of **AI**, **machine learning** and **deep learning algorithms**. There are usually two kinds of data. 'Training data' is real-world observations of the attribute being modelled (e.g. the presence of peaty soil). 'Predictor data' (also known as 'covariate data') is other information about the area being modelled, such as elevation, geology, satellite imagery. During model training / fitting, the algorithm looks for relationships between training data and predictor data and then attempts to make predictions using these relationships. **Performance metrics** are then used to evaluate the predictions and changes are made to model parameters to attempt to improve ('optimise') performance. This process is repeated many times until the performance metrics stop improving. At this stage the model is said to be **trained**.

Mosaic, **Image Mosaic**: A composite image made by combining multiple satellite photos e.g. cloud-free imagery.

National Character Area (**NCA**): Natural England has defined 159 NCAs to represent areas of distinct and recognisable character at the national scale (Natural England, 2021). Their boundaries follow natural lines in the landscape, not county or district boundaries. Each NCA is also assigned to one of 19 **Broad Landscape Types** (**BLT**) which classifies their physical, ecological and land-use features.

Performance Metrics. Measurements to quantify how well a model makes correct predictions. A full list of these metrics can be found in Appendix 6.

Recursive Feature Elimination: An algorithm for selecting the most important predictors in a dataset by repeatedly fitting a model, eliminating the least important predictor, and re-fitting the model.

Sentinel-1: An active Synthetic Aperture Radar (SAR) satellite constellation operated by the European Space Agency collecting radar imagery of the Earth. ‘Active’ means that it sends radar pulses to the Earth and measures how long the pulse takes to return and how the phase of the signal has shifted.

Sentinel-2: A passive satellite constellation operated by the European Space Agency collecting optical imagery of the Earth. ‘Passive’ means that it only records light emitted or reflected from Earth and does not send its own light pulses.

Variogram: A statistical tool used to understand how observations are related across distances. In a typical variogram, observations that are close to each other have less variation on average than data points that are far away from each other. Observations that are (almost) in the same location also have some variation on average, arising from measurement error (the ‘nugget’). After some distance (the ‘range’), average variation between observations stops increasing (the ‘sill’). Observations within the range can be **interpolated** using **kriging**.

XGBoost: A type of **machine learning algorithm** that improves prediction accuracy by combining multiple decision-making algorithms.

www.gov.uk/natural-england