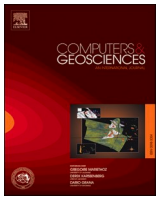
Computers & Geosciences 196 (2025) 105814

Contents lists available at ScienceDirect

Computers and Geosciences

journal homepage: www.elsevier.com/locate/cageo



Capturing exposed bedrock in the upland regions of Great Britain: A geomorphometric focused random forest approach

Chris Williams [a,*], Katie Whitbread [b], Alex Hall [b], Sam Roberson [b], Andrew Finlayson [b], Romesh N. Palamakumbura [b], Andrew Hulbert [a], Matthew Paice [c]

[a] British Geological Survey, Nicker Hill, Keyworth, Nottinghamshire, NG12 5GG, UK
[b] British Geological Survey, The Lyell Centre, Research Avenue South, Edinburgh, EH14 4AP, UK
[c] Centre for Environmental Data Analysis, Science & Technology Facilities Council, Rutherford Appleton Laboratory, Harwell Campus, Didcot, OX11 0QX, UK

ARTICLE INFO

Keywords:
Random forest
Bedrock exposure
Digital mapping
Postglacial landscape

ABSTRACT

Rock exposure distribution maps provide invaluable information for a range of applications from geohazard assessment through to aggregate resource potential assessments. Despite the usefulness of such information, it is only available to a limited extent across Great Britain (GB). Recent developments in the application of machine learning approaches to map exposed rock distribution rely on existing geological and land cover maps as the key input data for model training. We present a catchment-scale approach for delivering high-resolution rock exposure maps for GB mountain terrains. Our application has two objectives: establish a consistent and cross-applicable approach enabling feature identification from elevation datasets; use the results and diagnostics of the application to assist in further environmental process understanding. We utilize manual aerial image interpretation, and a suite of geomorphic terrain variables generated from a 5 m Digital Terrain Model as inputs to a distributed random forest model. Eight separate catchment models were derived from the training datasets using a leave-one-out approach. Aggregated results indicate a model accuracy of 79%, with a relatively high model sensitivity (78%) at the cost of relatively low precision (20%). Variable importance assessment highlighted patterns consistent with expected geomorphic controls on rock exposure related to gravity-driven slope processes in mountain landscapes. These results highlight the potential of multi-variant approaches for high-resolution rock exposure mapping, and lay a foundation for further development, particularly in relation to opportunities for further training data capture to ensure model accuracy. The ability to associate features based on geomorphological variables - indicative of landscape processes including erosion and deposition - presents opportunities that go beyond rock exposure such as for critical mineral and resource assessment. This approach will be applied for initial site characterisation as part of future onshore and offshore geological survey activities where high-resolution terrain and bathymetric data are available.

1. Introduction

Quantifiable distributions of surface bedrock exposures in upland areas provide useful information for a range of applications, from natural resource potential assessments (Scarpone et al., 2017; Hengl et al., 2017), geological site characterization as part of geotechnical surveys (Sarkar et al., 2004), to geohazard assessments (e.g. Yang et al., 2024) and studies of groundwater resource (Wilson et al., 2007). More specifically, rock exposure information can be used to support decision making regarding slope stability assessments, transport developments (e.g. Karlson et al., 2016), energy infrastructure development, forestry and land management, as well as providing inputs to process-based physical and 3D stratigraphic models (Scarpone et al., 2017; Ganerød et al., 2023). Crucially, rock exposure data provide boundary conditions for constraining superficial sediment thicknesses (or depth to bedrock) models, i.e. points where sediment thickness is equal to zero. In Great Britain (GB), existing models of superficial deposit thickness (Lawley and Garcia-Bajo, 2009) are compromised by the paucity of these data in upland areas, yet acquisition of bedrock exposure data through field mapping is limited by the inaccessibility of upland areas throughout GB, and the challenging environmental conditions frequently encountered. Approaches for identifying exposed bedrock in upland terrains using

* Corresponding author.
E-mail address: chrwil@bgs.ac.uk (C. Williams).

https://doi.org/10.1016/j.cageo.2024.105814
Received 12 April 2024; Received in revised form 3 December 2024; Accepted 3 December 2024
Available online 6 December 2024

topographic metrics have been established in previous studies. These metrics, including slope (DiBiase et al., 2012) and roughness (Milodowski et al., 2015), can be related to erosion and depositional processes occurring at the ground surface and therefore have a causal link to rock exposure when analysed at relevant spatial scales. Alternative approaches for mapping rock exposure using remote sensing imagery rely on colour or infra-red signatures that relate to the presence or absence of vegetation through colour, thermal properties and moisture content. Automated image classification methods are now widely used for land cover (i.e. vegetation) and habitat mapping (e.g. Marston et al., 2023).

The use of both topographic metrics and remote sensing imagery for regional mapping of exposed rock have been explored in two recent studies. Scarpone et al. (2017) map exposed rock in British Columbia using a random forest (RF) classifier and 43 variables including remote sensing and topographic parameters sampled at 100 m resolution. They found that variable importance was highest for Landsat derivatives related to biomass and temperature or soil moisture, with a ruggedness metric the third most important predictor of bedrock exposure. Ganerød et al. (2023) trialed a deep learning method in Norway to compare use of 30 m resolution Landsat and 10 m resolution DEM derivatives, finding that the terrain derivatives proved to be better predictors of exposed rock than the Landsat.

Although the different methods used may be a factor affecting the outcomes of these studies, the type and resolution of the training data used and the vegetation type of the study area, are also likely to have influenced the predictive value of Landsat and DEM derivatives. In particular, the differing resolutions of the training datasets, and the vegetation contexts of the study areas are significant factors. Both Milodowski et al. (2015) and Ganerød et al. (2023) highlight the difficulty of identifying exposed bedrock (i.e. bedrock with soil cover less than 10 cm) beneath tree canopies or other vegetation; they note that the use of topographic metrics for mapping rock exposure provides advantages over image classification, which are also affected by variation in illumination and cloud cover.

In GB, upland landscapes are largely vegetated, comprising moorland, acid grassland and forest, with thin, patchy minerogenic soils mantling bedrock (Fig. 1D). Existing land cover maps developed using an RF classifier based on Sentinel-2 datasets and other contextual information (Marston et al., 2023) tend to under-predict exposed rock outcrops in this terrain (Fig. 1A). By contrast geological maps, which are based on a British Geological Survey convention which limit mapping to areas where deposits are estimated to be thicker than 1 m (British Geological Survey, 2012), tend to omit the presence of thin, patchy sediment cover on upland slopes (Fig. 1B).

In this article we present a method for quantifying the spatial extent of exposed bedrock in upland areas of GB using high-resolution elevation data ($\geq$ 5 m). Our specific objectives are to.

1. Establish a consistent and widely-applicable method for predicting bedrock outcropping at surface using nationally available elevation data,
2. Attribute the relative importance of terrain derivatives covariables to further our understanding of environmental process operating within specific catchments.

We use an RF approach for automated mapping (cf. Scarpone et al., 2017) rather than the deep learning method of Ganerød et al. (2023) because it provides information about covariable importance, supporting our second objective. To avoid replicating the limitations of existing GB land cover and geological datasets with respect to exposed rock mapping, we create objective rock exposure datasets for model training through manual mapping of rock outcrop from high-resolution aerial photographs and field observations.

Multi-scale geomorphometric covariables are calculated from a 5 m digital terrain model derived from Lidar (Bluesky International Limited, 2021). The use of terrain data reflects the higher performance of these metrics in the comparable vegetated temperate landscape studied by Ganerød et al. (2023) and reduces problems with illumination and cloud cover inherent in most remote sensing imagery. Training data and geomorphometric terrain variables were collected for 8 catchments within the Montane and Valley (MV) domain (Booth et al., 2015). A total of eight model runs were conducted utilising a leave-out-one approach to verify model predictions and assess covariable importance. We consider the importance of different topographic covariables as predictors of rock exposure and the distributions of false positives and false negatives in the context of geological processes affecting the MV domain area.

2. Study design

The UK has highly varied geology and a complex glacial process history which may give rise to regional variability in the relationship between rock exposure and topographic parameters. We focus on development of a method suitable for UK uplands which typically have sparse distributions of observations. We identify these areas based on the MV domain of the Quaternary domains classification for Great Britain (GB) (Booth et al., 2012). This scheme subdivides GB into eleven semi-contiguous areas, or 'domains', following a land systems approach (Benn and Evans, 1998). These domains are closely related to Pleistocene glacial cycles, firstly subdividing the UK into glaciated versus non-glaciated zones, constrained by the maximum Quaternary glacial limits, followed by a series of regional landscape domains (Booth et al., 2015). The domains are predominantly characterised by topographic features, allowing rule-based predictive models based on geomorphometric parameters to be robustly applied across them. The MV domain is described as mountainous terrain with precipitous slopes, craggy glaciated troughs, deep cirques, fjords and numerous lochs residing in ice scoured basins and hollows, or deeply dissected upland plateaus (Booth et al., 2012).

We focus on eight valley catchments within the MV domain (Fig. 2), selected as representative catchments to enable the development and analysis of a robust predictive model for this landscape type. The catchments are chosen to encompass a range of landscape features, from catchment floors, a variety of slopes, and ridges at watershed boundaries, as well as a variety of bedrock lithologies.

Catchment locations and descriptions are presented in.

Table 1. To ensure high model performance, these catchments share the following characteristics: (i) limited urban areas and infrastructure, (ii) minimal areas of forestry, and (iii) availability of high resolution aerial imagery (25 cm). Fig. 3 provides maps of key geomorphometric variables for a selection of these catchments to give an impression of the landscapes used to train the model.

3. Methodology

A multi-stage method was designed to predict the location of bedrock exposed at surface. This consisted of (1) developing training datasets of observed exposed bedrock via aerial photograph interpretation, (2) the generation of a suite of geomorphometric parameters based on a 5 m spatial resolution DTM, and (3) the training and application of a RF classifier together with various quality testing components.

3.1. Manual data capture and training dataset development

A team of two to four expert geologists generated binary point data on a pixel-by-pixel basis for exposed rock outcrops in each catchment using high resolution aerial photographs (25 cm) within a GIS. These point data were converted into binary raster images (10 m resolution), scoring one where observations of rock at surface by all operators were coincident, and a value of zero where rock was not unanimously identified. The expert consideration of environmental process and feature

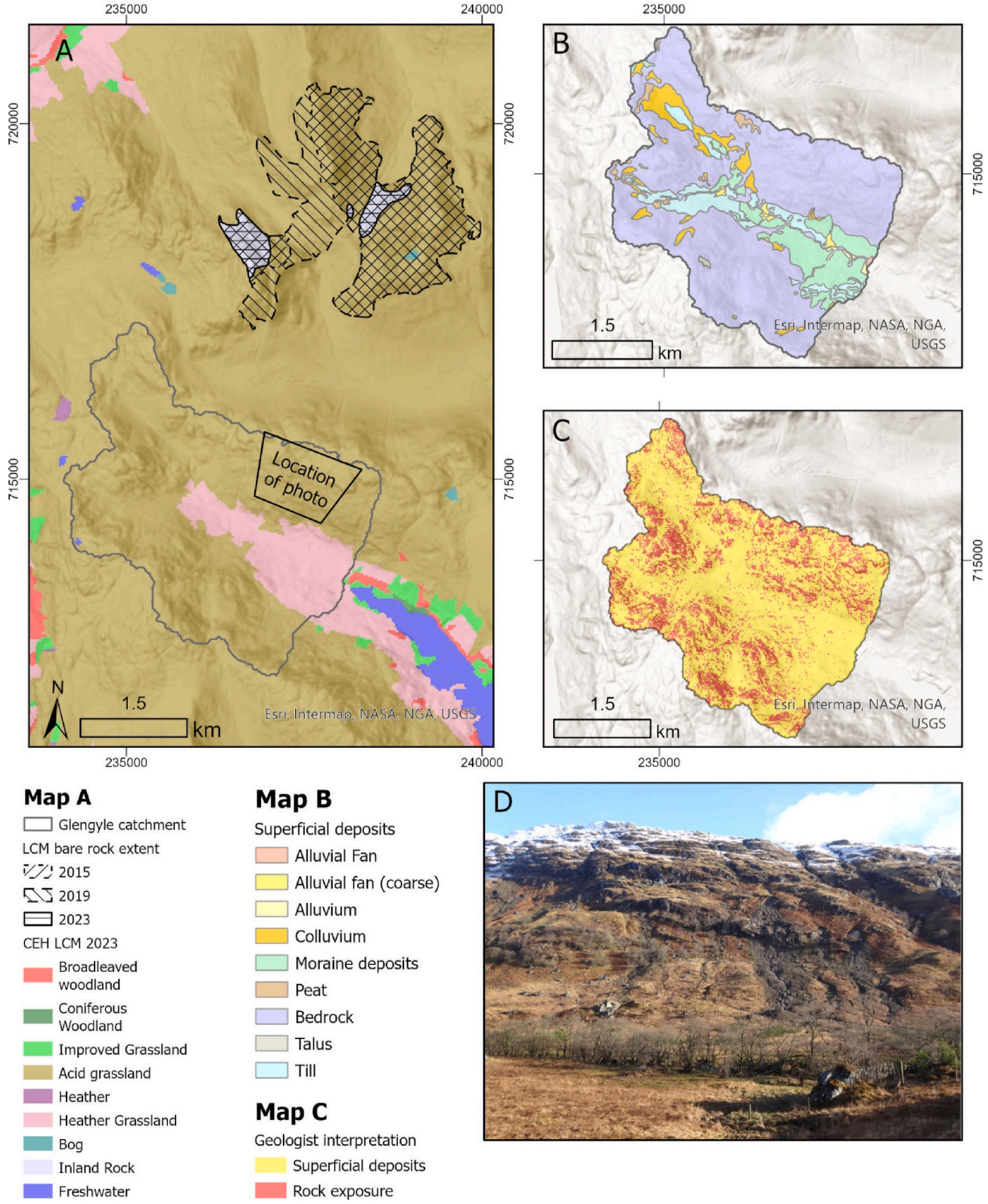


Fig. 1. Comparison of existing landcover and geological datasets for the Glengyle catchment, near Loch Katrine in the southern Scottish Highlands. A) the CEH 25 m Landcover map for 2023, no "Inland Rock" areas are shown within the Glengyle catchment area. Patches of Inland Rock located over the mountains to the north vary considerably in extent between years (4.3 km^2 in 2015, 5.8 km^2 in 2019 and 0.4 km^2 in 2023). Data owned by UK Centre for Ecology & Hydrology © Database Right/ Copyright UKCEH. B) 1:10,000 scale geological mapping of superficial deposits (from 2015). C) a geologist's interpretation of rock exposure in the Glengyle catchment (10m grid) based on manual inspection of aerial photograph imagery from 2016 (0.25m resolution). D) Image from the Glengyle catchment showing slopes with steep rock outcrops interspersed with areas of colluvium. Note the debris flows sourced from colluvial patches. The photo was taken facing north, location shown in map A.

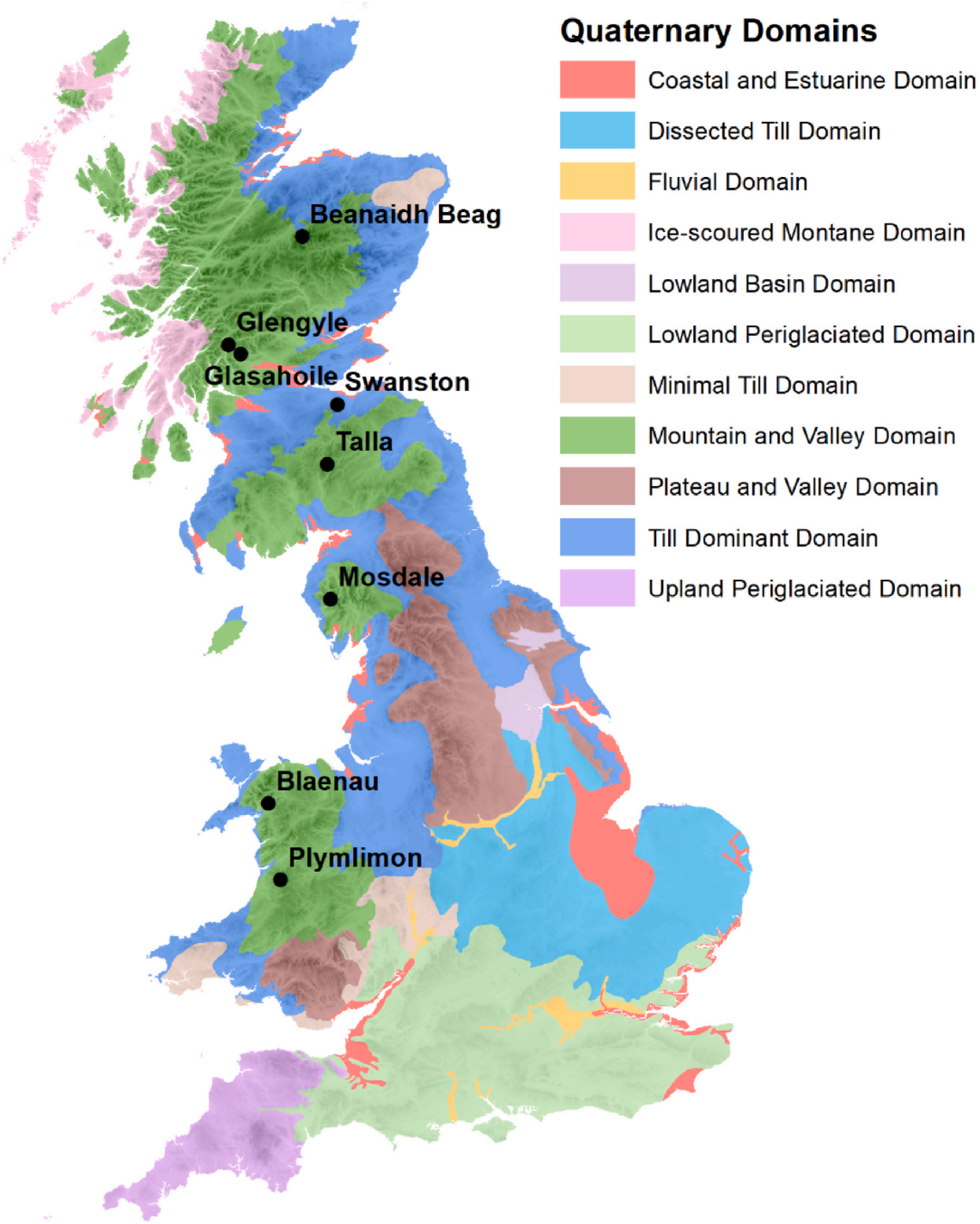


Fig. 2. Map of GB Quaternary domains after Booth et al. (2012) indicating the location of training catchments used in this study.

interpretation based on a range of environmental constraints (i.e. landform morphology, position within the catchment, and adjacent pixel classifications) is what differentiates this approach from more traditional image classification solutions, where feature identification is based on colour balance alone.

3.2. Geomorphometric derivatives

A suite of terrain derivative rasterized surfaces were generated as model covariates using a 5 m DTM, with a mean estimated uncertainty of 0.5 m (Bluesky International Limited, 2021). Terrain derivatives were calculated using a Python framework that integrated a range of geospatial software packages including Numpy, SciPy, GDAL and GRASS GIS (Table 2). Roughness-related metrics were calculated using a multi-scale approach to account for roughness values over different spatial scales following Amatulli et al. (2020), implemented by adapting processing window size.

Most terrain derivatives (Table 2) are widely used, but for completeness the less common are described below.

- Multi-resolution Valley Bottom Flatness (MrVBF) and Multi-resolution Index of Ridge Top Flatness (MrRTF) (Gallant and Dowling, 2003) provides an index that can be used to respectively

Table 1
Summary information relating to the catchments considered in this study, ordered by area Generalised domain area descriptions are taken from Booth et al. (2015).

Catchment	Locale	Description	Area (km^2)	Elevation range (m a. s.l.)
Swanston	Southern Uplands (Scotland)	Dissected rolling plateau with local steep-sided glaciated troughs, especially in the Galloway and Moffat hills. Extensive paraglacial deposits on slopes. Restricted glaciofluvial deposits. Devonian extrusive volcanics (andesite, basalt and rhyolite).	1.2	193 to 492
Plymlimon	Cambrian mountains (Wales)	Extensive dissected upland erosional plateau with areas of higher relief. Extensive till on upland surfaces, commonly solifluted. Valley slopes dominated by periglacial deposits. Glacial and glaciofluvial deposits in valley bottoms. Ordovician continental slope sandstones, siltstones and mudstones.	3.0	347 to 745
Beanaidh Bheag	Cairngorm & Gaick plateaux (Scotland)	Glacial erosion and deposition mainly restricted to corries. Widespread survival of deeply weathered Devonian granitic rock.	3.8	594 to 1294
Blaenau	Snowdonian mountains (Wales)	Steep slopes, glaciated troughs, extensive paraglacial deposits on slopes with glacial and glaciofluvial deposits in valley bottoms. Upland plateau area at catchment head. Ordovician mudstone.	4.3	246 to 688
Talla	Southern Uplands (Scotland)	Dissected rolling plateau with local steep-sided glaciated troughs, especially in the Galloway and Moffat hills. Extensive paraglacial deposits on slopes. Restricted glaciofluvial deposits. Gala Group sandstones, mudstones and conglomerate bedrock.	5.5	405 to 801
Mosdale	Shap and Howgill Fells (England)	Steep slopes, glaciated troughs, extensive paraglacial deposits on slopes with glacial and glaciofluvial deposits in valley bottoms. Borrowdale Volcanic Group (andesitic sills).	6.7	85 to 892
Glasahoile	Grampian Highlands (Scotland)	Steep-sided glaciated troughs and extensive paraglacial deposits on slopes. Severity of glacial erosion declines eastwards. Cambrian psammites and semipelite.	6.8	118 to 717
Glen Gyle	Grampian Highlands (Scotland)	Steep-sided glaciated troughs and extensive paraglacial deposits on slopes. Severity of glacial erosion declines eastwards. Cambrian psammites and semipelite.	11.6	116 to 765

characterise valley and ridge locations – larger values of both MrVBF and MrRTF indicate flatter valleys and more prominent ridges.

- Easterness and northerness are the cosine of slope and sine of slope respectively, enabling aspect to be treated as a continuous variable (Olaya, 2009).
- Vector strength and Fisher's k provide roughness information based on unit vectors relating to regularly spaced elevation values (Grohmann et al., 2011). In smooth locations, vector strength values will be high, while rough areas would be low. Fisher's k (based on Fisher, 1953) provides a measure of the uncertainty linked with the potential dispersion associated with a surface and contrasts with vector strength. Smaller values of k are expected in smooth areas, whereas relatively high values of k are expected for rougher areas where the potential for dispersion and therefore uncertainty associated with dispersion direction, is greater (Grohmann et al., 2011).

3.3. Training data selection

The full training dataset across all eight catchments consisted of 406,835 entries in total: 31 variables, 29 of which were continuous and three of which were categorical. The dependant variable, rock at surface, was coded as a binary category, i.e. rock present/rock absent at surface. The data were unbalanced with respect to the dependant variable: 6.3% of entries were positive for rock at surface, the remaining 93.7% negative. Tree-based models are sensitive to class imbalance, i.e. bias (Liu et al., 2009), therefore the minority class was upscaled during the training runs following test/train data splitting, creating a model trained on an equal number of positive and negative entries.

3.4. Classification model design

A Distributed Random Forest was selected as our Machine Learning (ML) algorithm (Chen et al., 2017), as they are less prone to over-fitting, tend to minimise the effects of co-linearity (Cutler et al., 2007) and are therefore more computationally efficient. RF models allow for straightforward assessment of variable importance, a key requirement given that we wished to use terrain derivatives as proxies for understanding environmental processes operating within a catchment. Importantly, RF models can characterise non-linear interactions between co-variables; a realistic portrayal of process interactions in a natural setting. The H2O framework (LeDell and Poirier, 2020) was used to train and create a model as it can accept mixed data (categorical and continuous) with minimal pre-processing. A grid search was used to optimise model hyperparameters, using a training set of 10% of the overall training set. Hyperparameters for the final models were defined as: ntrees = 500, max. tree depth = 50, min. rows = 20.

3.5. Classification model cross-validation

Tests of model performance were conducted via cross-validation, splitting the training and testing datasets using a leave-one-out

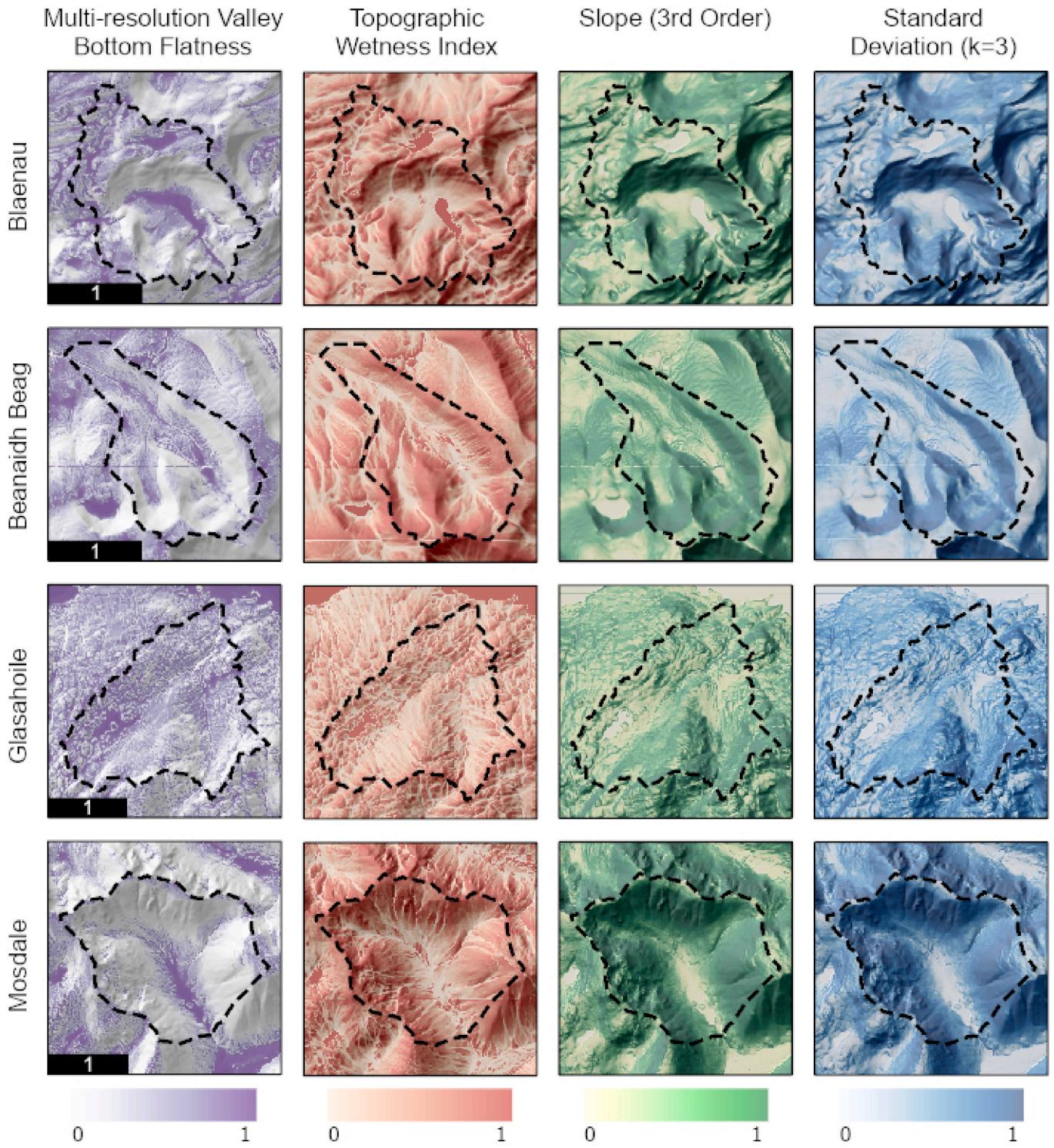


Fig. 3. Maps of key derivatives for selected catchments. Colour ramps have been normalised for comparison. 1 km scale is shown for each catchment and north is towards the top of the page in all cases. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

approach. Data from a single study area (catchment) were held out for testing, and a model trained using the seven remaining catchments. This was repeated for all eight catchments, generating eight separate models and eight separate test sets. For details on the specific programmatic implementation of this approach, please refer to the model code and associated documentation.

Predictions are given as confidence of rock at surface (p1), where 0 indicates the model has 0% confidence of rock at surface and 1 indicates 100% confidence per pixel. Model performance was assed using accuracy, precision and sensitivity statistics. Accuracy is given as:

$$\frac{T_p + T_n}{T_p + T_n + F_p + F_n} \tag{1}$$

where T_p is the number of true positives, T_n is the number of true negatives, F_p is the number of false positives and F_n is the number of false negatives. Precision is given as:

Table 2
Calculated derivatives together with associated software and processing scales with which they were generated. Elevation is included here to illustrate all continuous data available for each catchment.

Derivative	Processing window (pixels | metres)	Algorithm	Reference/Software
Elevation	N/A	Values as provided	Bluesky International Limited, 2021
Aspect (2nd order)	3 | 15	9 parameter 2nd order polynomial	Zevenbergen and Thorne (1987)
Slope (2nd order)	3 | 15		
Convergence	3 | 15	GRASS: r. convergence	GRASS
Multiresolution index of the ridge top flatness (MrRTF)	3 | 15	GRASS: r. valley.bottom *t* = 42[a]	GRASS; Gallant and Dowling, 2003
Multiresolution index of valley bottom flatness (MrVBF)	3 | 15		
Easterness	3 | 15	GRASS: r. northerness. easterness	GRASS; Olaya (2009)
Northerness	3 | 15		
Northerness slope	3 | 15		
Vector strength	3 | 15, 5 | 25, 7 | 35, 9 | 45	GRASS: r. roughness. vector	GRASS; Grohmann et al., 2011
Fisher's *k*	3 | 15, 5 | 25, 7 | 35, 9 | 45		
Geomorphons	3 | 15	GRASS: r. geomorphon	GRASS; Stepinski and Jasiewicz, 2011; Jasiewicz and Stepinski (2013)
Aspect (3rd order)	3 | 15	9 parameter 3rd order polynomial:	Horn (1981); Mitasova (1985); Hofierka et al. (2009)
Plan curvature (3rd order)	3 | 15		
Slope (3rd order)	3 | 15	GRASS: r.slope. aspect	
Tangential curvature	3 | 15		
Standard Deviation	3 | 15, 5 | 25, 7 | 35, 9 | 45	Standard deviation	–
Topographic Roughness Index	3 | 15, 5 | 25, 7 | 35, 9 | 45	9 parameter variance	Adapted from Riley et al. (1999)
Topographic Wetness Index	3 | 15	GRASS: r. topidx	GRASS; Cho (2000)

[a] for the MRVBF and MRRTF variables, a value for the required t_slope parameter in the GRASS implementation of the function of 42 was used based on the input DTM pixel size of 5 m while considering the power function provided in Gallant and Dowling (2003).

$$\frac{T_p}{T_p + F_p} \tag{2}$$

Sensitivity is defined as:

$$\frac{T_p}{T_p + F_n} \tag{3}$$

For each model, the following were calculated.

- The Receiver Operator Characteristic (ROC) curve, which characterises the balance between sensitivity and specificity when different thresholds are used to define a positive prediction from model confidence.
- The average Youden's J statistic for all eight models derived from the ROC curve. This statistic is used to select a decision threshold value that is a good compromise between model sensitivity and specificity (Witten et al., 2017). This is referred to as the 'global decision threshold'.
- A confusion matrix which represents the accuracy of the eight models *en masse*, when tested against each of their respective catchment test sets.

The purpose of calculating metrics across all catchments was to obtain a characterisation of model performance across a series of disconnected datasets, rather than testing on a single catchment. The global decision threshold value was applied to obtain results for all individual catchments, necessary because in a real-world application the actual values for a new catchment may not be known. In such scenarios a decision threshold can only be determined from existing data, so the global decision threshold is taken as a robust measure of overall model performance potential.

4. Results

4.1. Rock at surface prediction: multi-catchment results

Aggregating predictions for all eight models, across all eight respective test sets resulted in an overall model accuracy of 78.98%. The model sensitivity was 77.71%, which is the proportion the model correctly identifies instances of rock at surface. This was at the cost of a relatively low precision of 19.98%, meaning for every correctly identified instance of rock at surface, the models identified four false positives.

The low precision could be traded for lower accuracy by raising the prediction threshold value. However, the optimum value used for the final confusion matrix (Fig. 4), as defined by the average Youden's J statistic, was 0.24. This adjusted threshold was used to account for class imbalances, as opposed to the more typical threshold of 0.5 which is more appropriate for balanced datasets. This was then applied to the predictions for each individual catchment. The ROC curve for the eight aggregated models displays the tendency for false positive rate to initially grow rapidly as the predictive threshold is increased (Fig. 5). This is indicative of models with a low precision relative to sensitivity where a low decision threshold is required to correctly categorise most positive instances.

Variable importance was reasonably consistent across all models. Slope (2nd order), slope (3rd order), MrVBF, topographic wetness index and elevation were the five most important variables for all models, although not always in the same order. The top ten most important variables, averaged across all eight models are presented in Table 3.

4.2. Rock at surface prediction: catchment specific results

Model performance is compared for individual catchments using key metrics (Table 4). Catchment models have a relatively high accuracy (>70%) and sensitivity (>79%), except for Blaenau) at the expense of lower precision (<34%).

Precision is very low for Plymlimon, Glasahoile, Swanston and Talla, with 5% or less of the total predictions of rock as true positives. All these catchments have areas of observed rock that cover less than 1% of the catchment area (Table 4; e.g. Talla, Fig. 6). Precision is higher for Mosdale, Glen Gyle, Blaenau and Beanaidh Beag, where 16–34% of the total predictions of rock are positives. In these catchments, observed rock covers 2–11 % of the catchment area (e.g. Glen Gyle, Fig. 7).

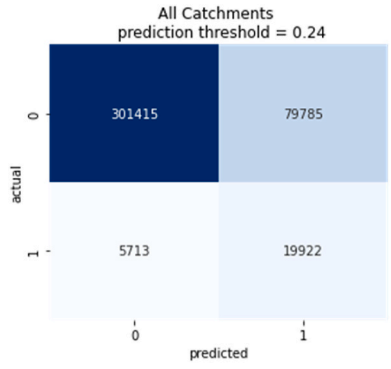


Fig. 4. Confusion matrix based on combined test data from the separate catchment cross-validation models.

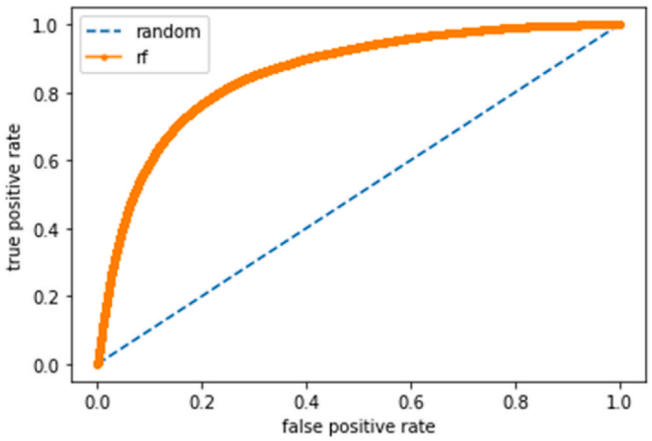


Fig. 5. ROC curve for cross validated predictions on the 8 catchments test holdout data split. Note that the threshold is low to the bottom left, increasing to the top right.

Table 3
Mean relative importance of variables from the combined results of the catchment cross-validation models. Only the top ten variables are shown.

Variable	Scaled importance
Topographic wetness index	0.94
MrVBF	0.81
Slope (3rd order)	0.68
Slope (2nd order)	0.56
Elevation	0.41
Standard deviation (15 m^2 kernel)	0.35
Standard deviation (25 m^2 kernel)	0.34
Convergence (15 m^2 kernel)	0.30
Topographic roughness index (25 m^2 kernel)	0.28
Topographic roughness index (15 m^2 kernel)	0.26

Blaenau has the lowest sensitivity (0.52) and accuracy (0.71) despite moderate precision (0.19) and the highest proportion of observed rock for the catchment area at 11.2% (Fig. 7). However, Blaenau has a lower proportion of predicted rock (29.6%) than both Glen Gyle and Mosdale (>30%) which have only 4% and 5% of the catchment mapped as rock respectively. Comparison of mapped and predicted rock for Blaenau (Fig. 8) highlights a large area in the south of the catchment which is mapped as rock that does not show up in the model prediction and may be responsible for the low sensitivity and accuracy for this catchment.

4.3. Geomorphometry covariance and predicted rock presence/absence

4.3.1. Spatial autocorrelation of derivatives

To provide an understanding of how the variables are spatially distributed within catchments, spatial autocorrelation was calculated for key variables and rock presence using Moran's I statistic (Fig. 9). Beanaidh Beag, Blaenau and Mosdale exhibit the highest Moran's I, indicating localised clustering in larger contiguous areas of rock (see the results for Blaenau in Fig. 8). By contrast, Plymlimon, Swanston, Glasahoile and Glen Gyle have Moran's I scores of less than 0.5 and show more dispersed distributions of rock presence (see the results for Glen Gyle in Fig. 7).

In general, stronger spatial autocorrelation signatures are associated with key derivatives for the catchments where rock presence is also more autocorrelated (Fig. 9). The effect of data resolution and sampling window on the autocorrelation of derivatives is illustrated by the systematic increases in Moran's I values for the vector strength, Fisher's *k* and standard deviation parameters with increasing kernel size.

4.4. Derivative covariance and predicted rock presence/absence

The distribution of rock presence and absence relative to key derivative distributions highlights that rock presence is generally associated with higher slope angles, and a lower TWI, but there is considerable variation (Fig. 10). Except for elevation, rock presence is reflected by a relatively narrow range of values for each derivative.

4.4.1. Derivative covariance and catchment variability

To assess for catchment-specific effects, the association of derivatives and rock presence is shown for the different catchments in Fig. 11. This highlights that rock presence plots relatively consistently with key variables. The only exceptions to this include Beanaidh Beag, which is characterised by an overall higher elevation profile and Glen Gyle that has a small proportion of rock presence associated with higher MrVBF values. Mosdale, Glen Gyle and Blaenau have rock presence across a large range of TWI, slope and elevation values.

5. Discussion

5.1. Model design and performance

Our approach of predicting exposed bedrock from a range of terrain derivatives using a random forest method resulted in models with relatively high sensitivity but low precision. All eight test catchments had high levels of over-prediction of rock, particularly in catchments with very limited mapped rock exposures. This low precision means the model will be best suited to real world applications where false positives can be tolerated, but false negatives are less acceptable.

The training set for each model contained data from seven distinct catchments, with marked variability in the amount of rock present in the selected catchments. Testing on a large catchment required the catchment be omitted from the training set to create its respective model, which removed a proportionally larger amount of data than for a smaller catchment. The use of a greater number of appropriately varied catchments for training would mitigate this problem.

Table 4
Catchment specific model performance - based on rounded and threshold adjusted positive prediction values. Moran's I is presented here for each catchment to illustrate the varying spatial autocorrelation of rock exposure within each catchment.

Catchment	Observed rock as a proportion of the catchment	Predicted rock as a proportion of the catchment	Accuracy	Sensitivity	Precision	Area under ROC curve	Moran's I (rock presence)
Beanaidh Bheag	0.02	0.14	0.87	0.98	0.16	0.96	0.76
Blaenau	0.11	0.30	0.71	0.52	0.19	0.68	0.76
Glasahoile	0.01	0.21	0.78	0.85	0.04	0.90	0.48
Glen Gyle	0.04	0.31	0.72	0.79	0.23	0.83	0.49
Mosdale	0.05	0.32	0.78	0.88	0.34	0.89	0.68
Plymlimon	0.004	0.16	0.83	0.94	0.02	0.96	0.32
Swanston	0.005	0.16	0.84	0.82	0.05	0.91	0.39
Talla	0.0002	0.09	0.91	0.90	0.05	0.96	0.58

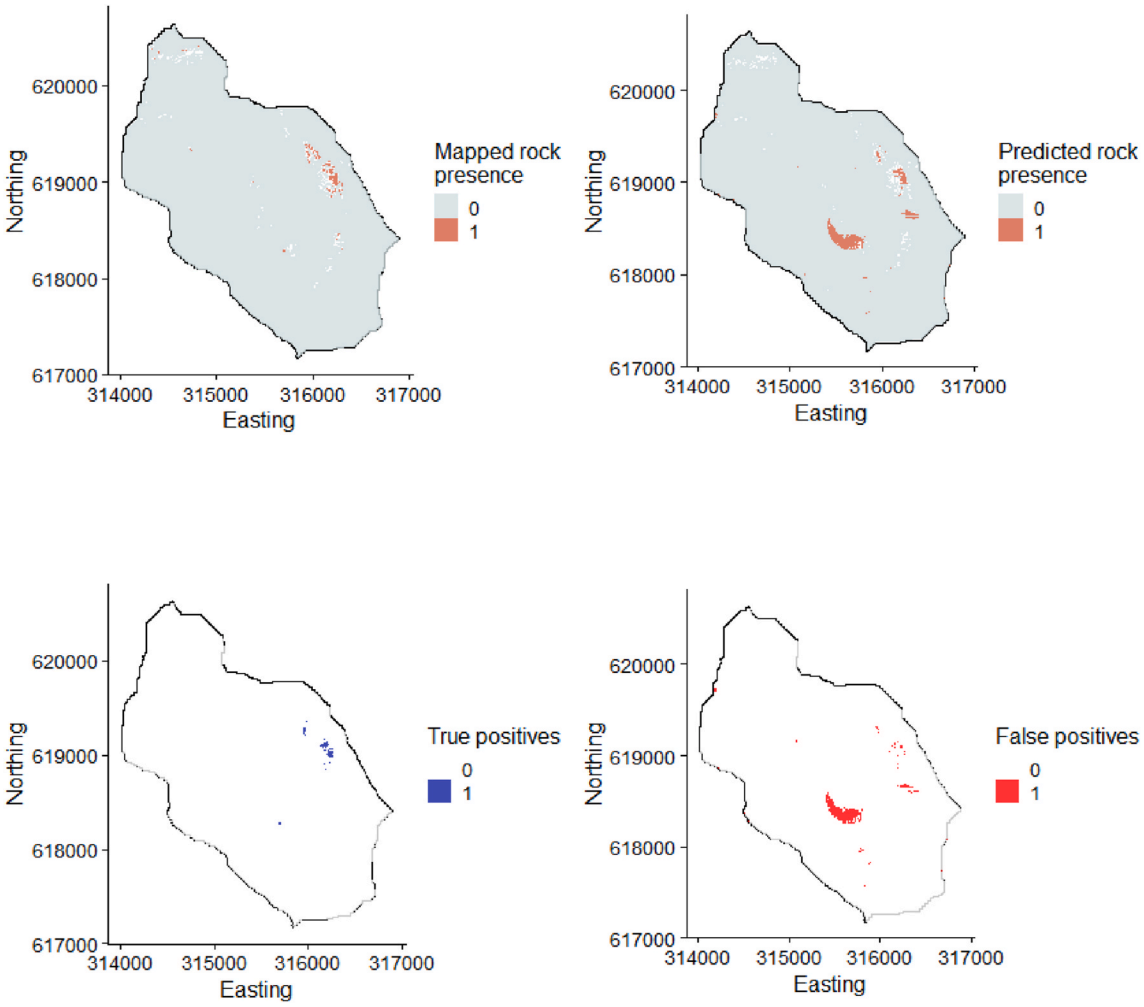


Fig. 6. Results of model prediction for the Talla catchment, representing observed rock presence, predicted rock presence, true positives and false positives. Axis coordinates are in British National Grid.

5.2. Model training data

Observational training data for our model was captured by multiple expert geological interpreters, then filtered to isolate points of unanimous agreement. Uncertainties arise from the limitations of interpreting from aerial images, including illumination angles, vegetation cover (including tree cover along narrow river corridors), shadow effects, clouds, and the inability to capture and steep slopes. This last point is mitigated, to a certain extent, by using terrain derivatives in the RF model (cf. Milodowski et al., 2015).

Some discrepancies between aerial images and DTM data arise due to temporal mismatch, i.e. the action of geomorphic processes such as debris flows occurring between the capture of datasets. However, the comparatively small scale of geomorphic processes in UK uplands (i.e. tens to hundreds of metres) means that any impact on model results is likely to be small. Utilising recent imagery and ensuring minimal gap between capture of terrain and aerial imagery will minimise these effects.

The training data effectively defines the nature of 'bare rock' being modelled and machine learning outcomes will reflect the key associations and limitations of the training data. Scarpone et al. (2017) trained their model using bare rock classifications derived from habitat-based landcover datasets, which may be a significant factor in the relative importance of Landsat imagery in the model outcome. By contrast Ganerød et al. (2023) utilize geological maps as training data, and the finding that terrain derivatives provide better prediction may reflect the fact that areas mapped as "bare rock" and "thin organic soils on bedrock" (i.e. soils 10–30 cm thick) were merged within the training dataset. The latter includes vegetated areas thus limiting the value of 'colour' and moisture properties as predictors of rock exposure.

In this study we used an independent dataset of observations of rock exposure to avoid replicating the limitations of existing landcover and geological datasets. Alternative datasets and methods for the capture of training data could be explored in future, including the use of outcrop sites marked on geological field maps, ground walkover surveys, oblique photography (e.g. DiBiase et al., 2012), and terrestrial laser surveys.

5.3. Scaling effects

Rock outcrops in GB vary considerably in size from metres to kilometres. The choice of grid resolution for rock exposure mapping/modelling needs to reflect the scale of rock outcrops that are relevant to (a) the required mapping outcome, and (b) the scale of features and the process-scales associated with their formation (Hengl, 2006; Tarolli and Tarboton, 2006; Tarolli, 2014).

For example, a coarse grid resolution (such as 100 m) will smooth terrain metrics such as slope and roughness (Molnar and Julien, 1998), potentially reducing the likelihood of identifying rock exposures on crags and decimetre-scale rock ridges that occur on otherwise sediment mantled slopes. Conversely, using too fine a grid resolution may lead to spurious associations such as with metre-scale boulders present in rockfall, landslide or moraine deposits (c.f. Tarolli and Tarboton, 2006).

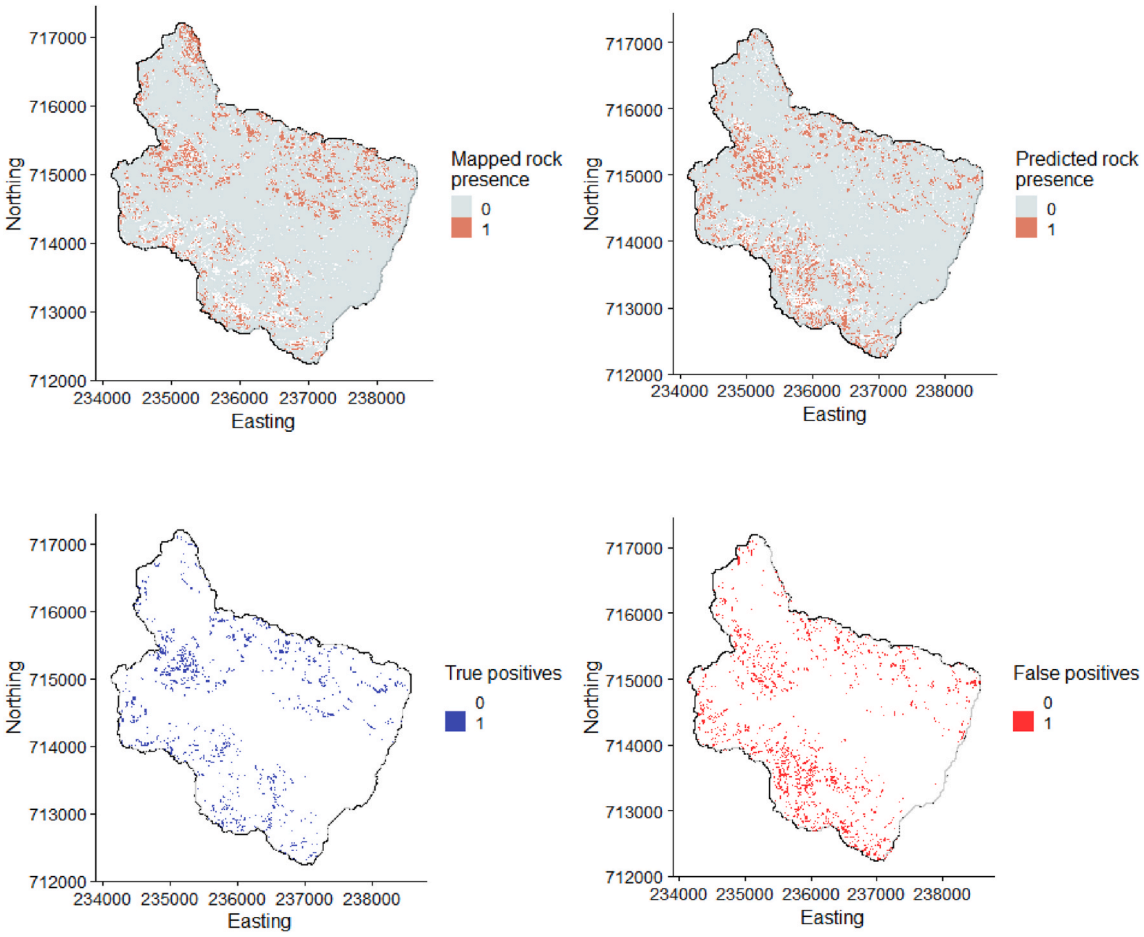


Fig. 7. Results of model prediction for the Glen Gyle catchment, representing observed rock presence, predicted rock presence, true positives and false positives. Axis coordinates are in British National Grid.

The use of 100 m grided training data by Scarpone et al. (2017) may thus be an additional factor limiting the sensitivity of their modelling to terrain predictors. We use a 10 m grid to derive a range of multi-resolution datasets, finding that the importance of slope and roughness variables decreases with kernel size, with the 3x3 window yielding the highest variable importance. Although limited by the base resolution of our training data, this result is consistent with the relevant scales of rock outcrops within the UK MV terrain (typically 10–100 m). There is potential for further work to compare the relative importance of scale derivatives in different UK landscape domains where rock exposure may have different geomorphic process associations.

5.4. Modelling and geological context

5.4.1. Geomorphic significance of key variables

Of the 31 terrain derivatives, all models had the same top ten variables (Table 3). The high contribution of slope, elevation, TWI and MrVBF across the full range of catchment morphologies sampled is consistent with gravity-driven slope processes controlling the distribution of soil and other sediment deposits on hillslopes (Patton et al., 2018). The top contributing co-variable, TWI, reflects the accumulation of flow at a given point and therefore scales inversely with the occurrence of rock at surface (Fig. 9), identifying areas where limited sediment can accumulate due to either high slopes or a lack of material inputs from upslope (i.e. where there is rock at surface). Similarly, areas with lower slope values are more likely to be sediment covered due to the diffusion-like nature of slope processes.

High values of standard deviation and roughness at length-scales less than 30 m are thought to be indicative of areas where bedrock is present at surface regardless of elevation or position on slopes (e.g. Beven and Kirkby, 1979). This potentially reflects the influence of geological heterogeneities, such as bed thickness and fracture density, which locally influence weathering, erosion rates and slope stability (e.g. Coblentz et al., 2014; Milodowski et al., 2015; Falcini et al., 2022). By contrast in areas of greater sediment availability, depositional processes typically lead to the formation of relatively smooth, convex slope surfaces (Carson and Kirkby, 1972).

5.4.2. Model performance within catchments – the role of landforms

Differences in spatial autocorrelation between catchments reflect the influence of different landscape morphologies within the MV domain on the model performance. The spatial autocorrelation of rock exposure is highest (i.e. most clustered) for Beanaidh Beag and Blaenau, and lowest (i.e. most dispersed) for Plymlimon, Glasahoile and Glen Gyle.

Plymlimon, Glasahoile and Glen Gyle catchments lack well-developed corries and are characterised by rugged slopes and narrower valley floors with patchy superficial deposits (Fig. 3 - Glasahoile). Rock exposures are dispersed across the catchment in small clusters occurring across a range of elevations and are associated with high slopes, low TWI, and low to moderate MrVBF. These catchments with lower autocorrelation show a more even distribution of false positives, which tend to be located at the margins of the areas of rock exposure and reflect higher model uncertainty in areas that are marginal to small rock outcrops (e.g. the Glen Gyle catchment, Fig. 7).

The stronger autocorrelation seen in the catchments of Beanaidh Beag and Bleanau is associated with the presence of well-developed

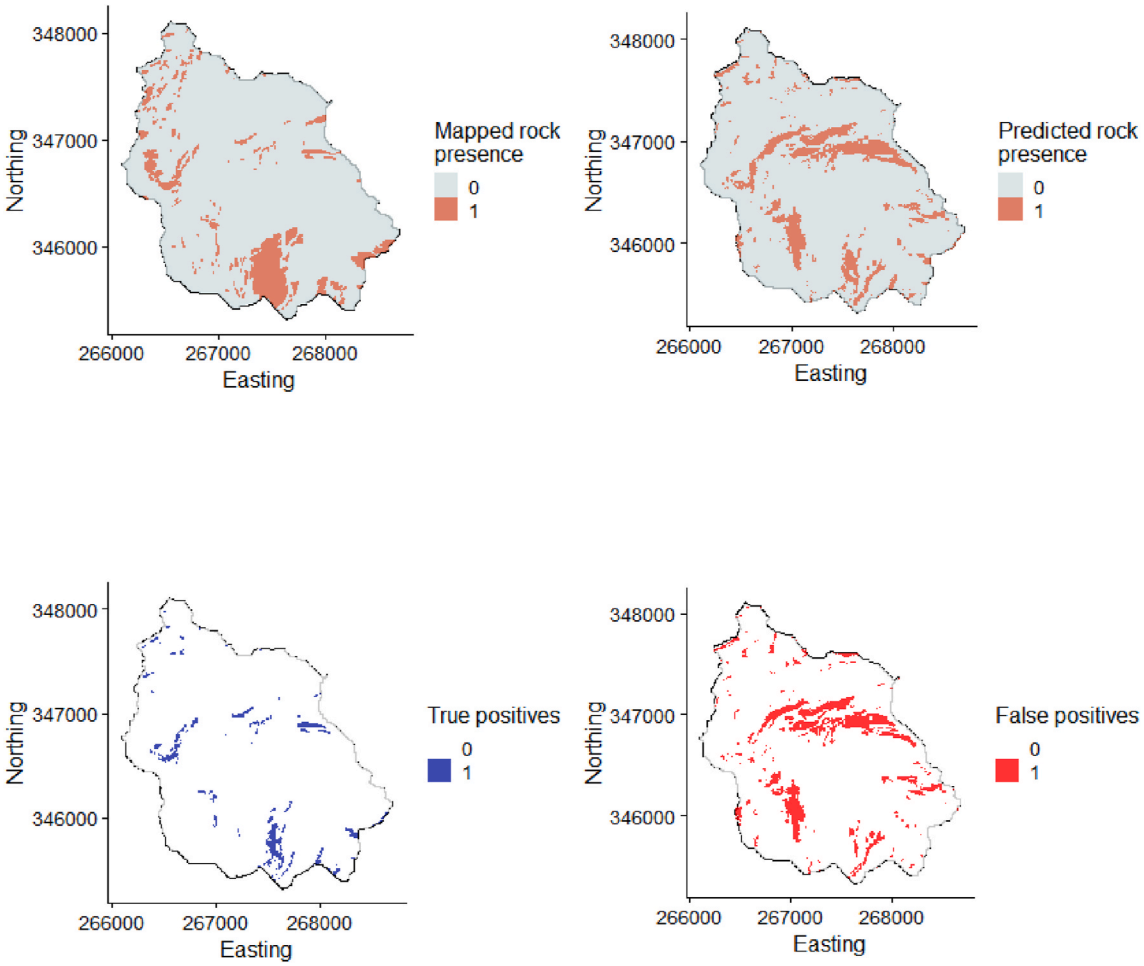


Fig. 8. Results of model prediction for the Blaenau catchment, representing observed rock presence, predicted rock presence, true positives and false positives. Axis coordinates are in British National Grid.

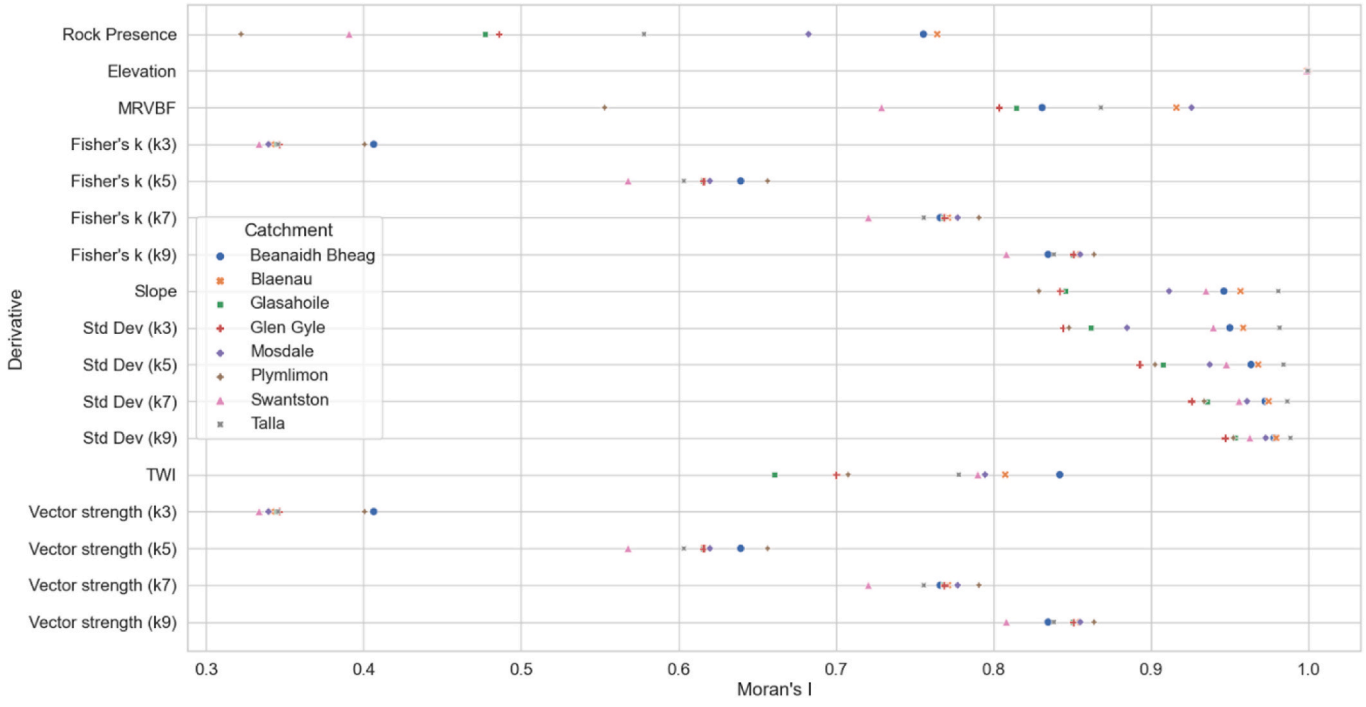


Fig. 9. Moran's I calculated for each derivative for each catchment to demonstrate variations in spatial auto-correlation.

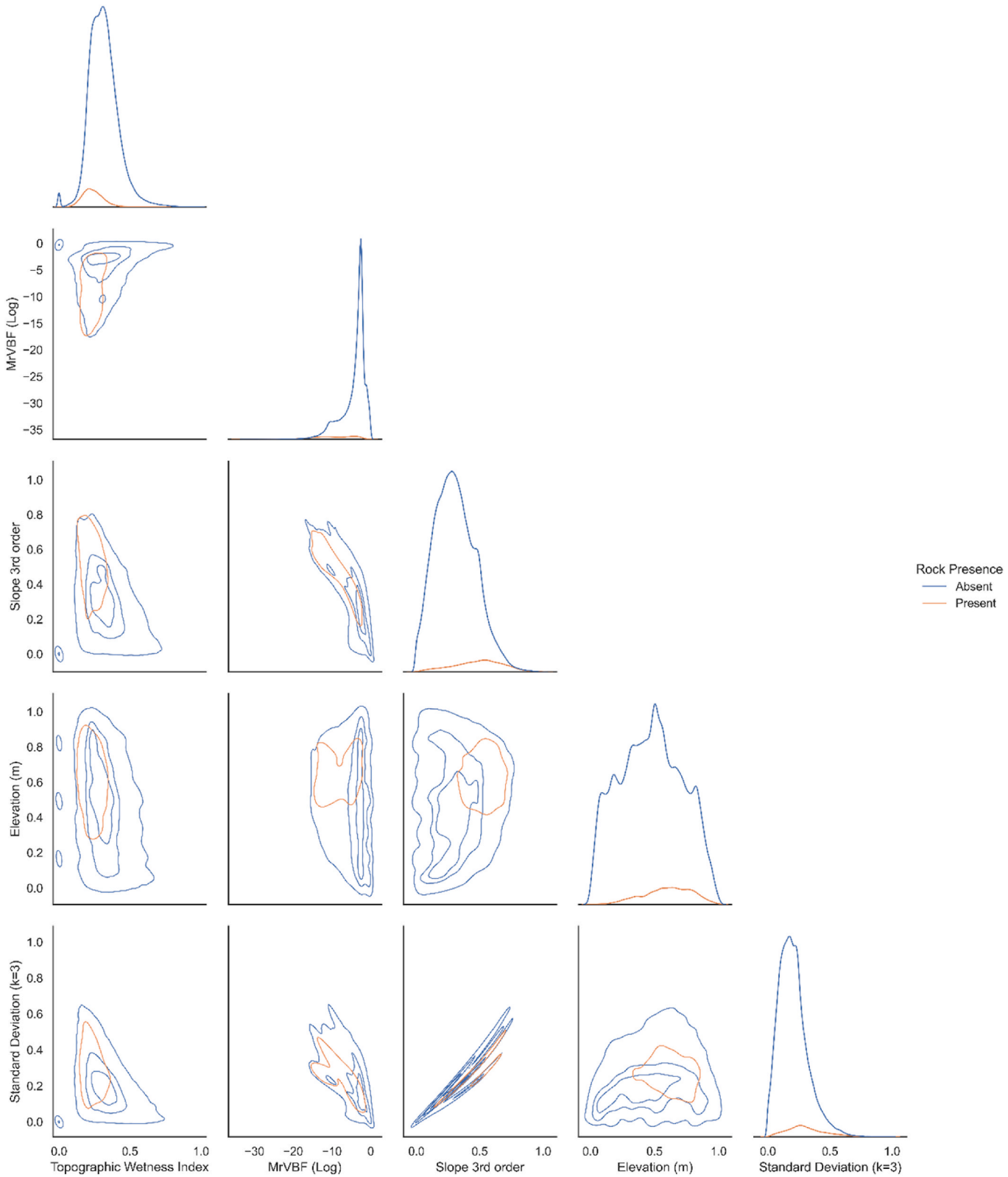


Fig. 10. Predicted rock presence and absence normalised derivative covariance associations.

corries, and wide valley floors that are typically mantled in till and moraine deposits (Fig. 3). In these catchments, rock exposures are concentrated in large contiguous areas on the corrie head and side walls, giving rise to larger areas of exposure linked to high slopes, mid to high elevations and low TWI and MrVBF values. These catchments correspondingly have relatively large contiguous areas of false positives and false negatives. False positives are associated with the presence of shadows in the aerial images and with vertical rock faces separated by vegetated ledges (e.g. Fig. 12). These false positives reflect limitations in the training data, with some degree of compensation arising from the use of terrain metrics – effectively the model is performing better than the training data in these areas (cf. Milodowski et al., 2015).

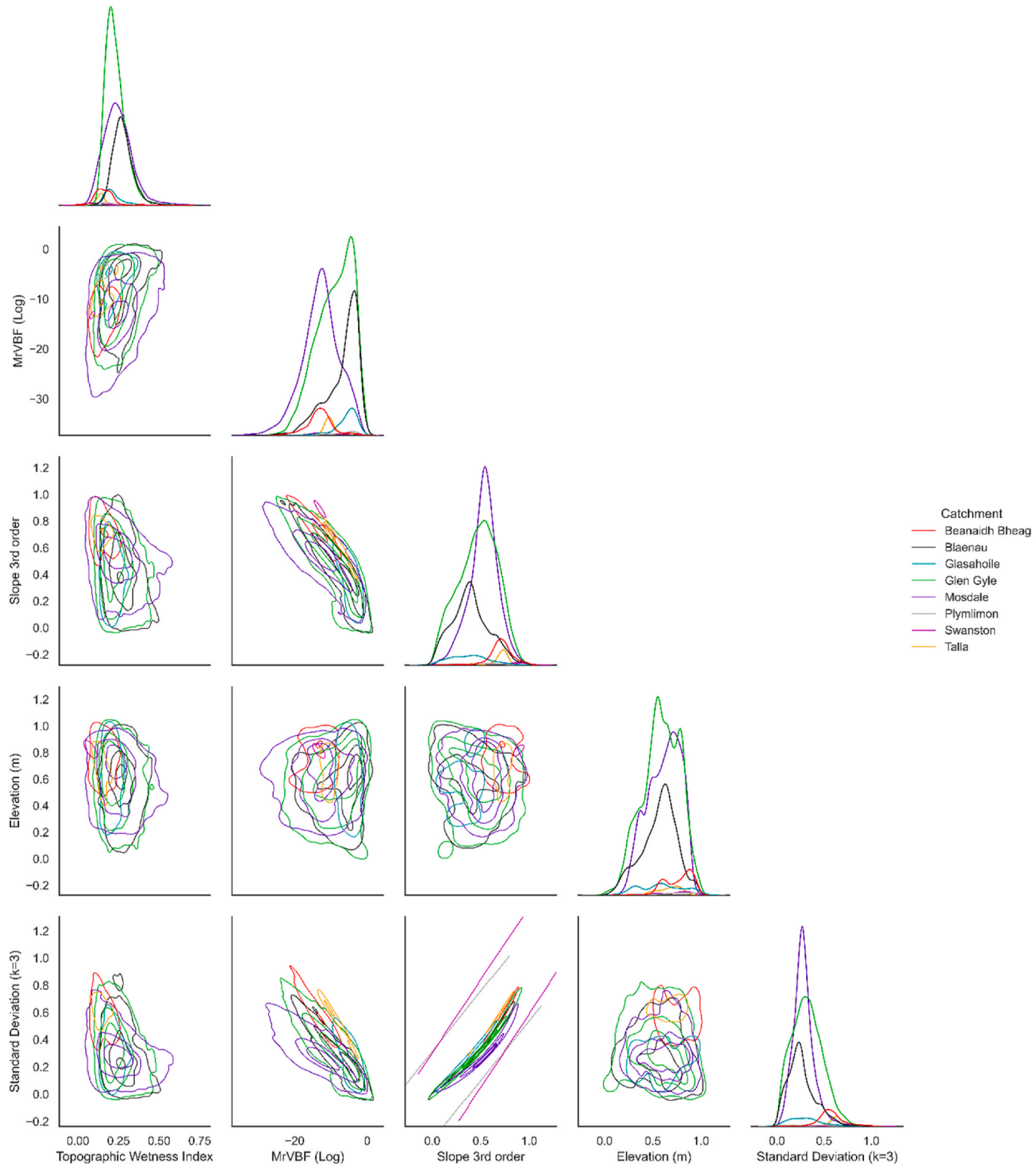


Fig. 11. Predicted rock presence derivative covariance associations by catchment.

Underprediction of rock exposure (false negatives) occurs in the southern part of the Blaenau catchment where a large area of mapped exposure associated with rock slabs are underrepresented in the model prediction (Fig. 12E). The rock slabs occur over an interfluve at moderate to high elevations, and with moderate slope, on the southern side of the corrie. The lack of sediment on these surfaces is likely to reflect localised ice-scouring at the intersection of the corrie glacier with a larger glacier occupying the adjacent valley.

The terrain variables used in this analysis may not sufficiently account for the influence of historical ice-scouring on sediment distributions at low to moderate slopes and elevations. This factor would need to be accounted for in upscaling of the model, and particularly in further applications to the ice-scoured domain of Northern Scotland (Fig. 2; Booth et al., 2015). Additional variables such as higher-resolution

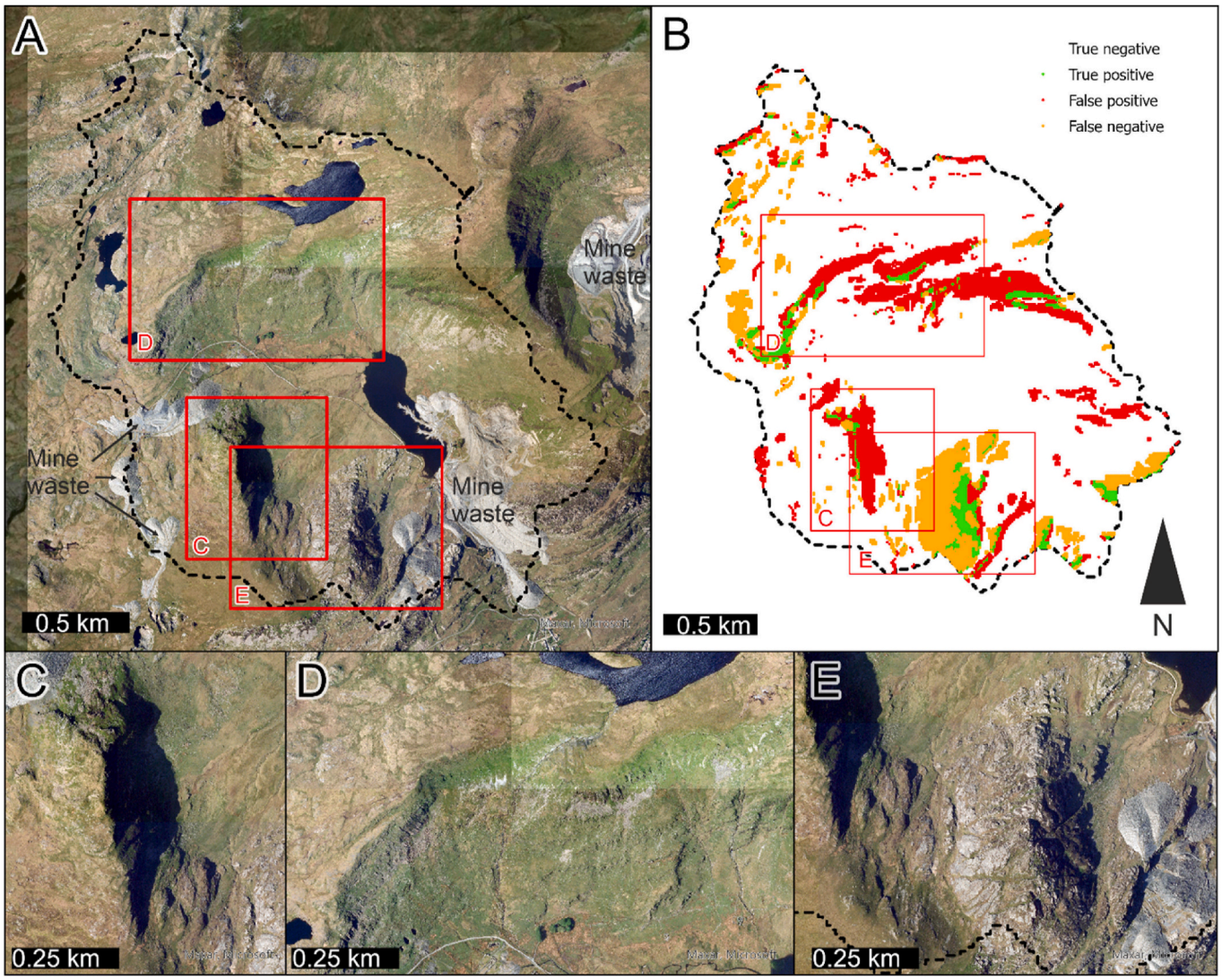


Fig. 12. Illustration of false positives and negatives in the Blaenau catchment: A) aerial image of the whole catchment, B) distribution of predicted rock shown a true positives, false positives and false negatives, C) area showing shadow effect in areal image where the model predicts false positives, D) corrie backwall of stepped crag lines associated with false positives, E) interfluve with exposed rock slabs associated with false negatives. Aerial photography © UKP/Getmapping Licence No. UKP2006/01.

roughness and directional roughness parameters related to specific wavelengths (Hubbard and Hubbard, 1998) could be included in future model iterations for application to ice-scoured topography.

5.5. Development opportunities

A range of the current model limitations might be addressed by using additional covariates and additional, or alternative, sources of independent observational data. The model has been trained in non-urban environments and is therefore suitable for applications in mountainous upland settings that are represented by the training datasets. However, additional training data would be required for wider applications or upscaling, along with approaches that account for the influence of urban areas and infrastructure.

Few of the terrain variables used as inputs had a significant control on the model prediction. It is likely that lower value metrics could be removed from the training data without significantly impacting model performance. This would reduce the training time, allowing for the training of more models and a more comprehensive approach to hyperparameter tuning. A recursive-based approach to tuning could be used (Wicaksono and Supianto, 2018), building on the simple grid-based approach used for this study. However, it should be noted that RFs do tend to be less sensitive to poorly optimised hyperparameters than most other ML approaches (Bernard et al., 2009).

The ability of the models to produce predictions that correlate well with the spatial distribution of rock at surface, but not accurately predict rock at a given point suggests a more complex model may be required if point-prediction is necessary. The RF models only consider the independent variables at a given point to make a prediction. As the data is presented in the form of a 2D array, the values of these variables at nearby points are likely to provide useful information for determining the value at a given point. The use of a Convolutional Neural Network (CNN) (Minetto et al., 2019) may be an alternative approach for improving predictive accuracy. However, use of a CNN would remove some of the benefits of a RF based model; notably easy assessments of variable importance, which were a key output of this study.

6. Conclusions

Modelling rock exposure across GB uplands using a RF classifier and a range of terrain derivatives offers a time-efficient approach for mapping exposed rock compared to traditional field approaches. The use of

catchment-scale cross-validation highlights the high sensitivity but low precision of the model, which means the potential implications of false positives must factored into downstream applications. However, examination of the model highlights that it does a good job of isolating key variables that control the nature of upland sediment and soil processes, and it may out-perform the training data input leading to overestimation of false positive cases. The consistency of identified key variables across models indicates the model is robust, therefore supporting upscaling in the MV domain.

The model may be improved by inclusion of additional derivative datasets such as NDVI. These may be particularly useful in regions where the distribution of deposits related to historical geological contexts cannot be fully accounted for by purely gravity and flow processes, and therefore may be important when expanding the model coverage to other Quaternary domains, particularly the ice-scoured domain.

The preliminary results of this study highlight that field validation will be important for assessing the model outcome and should be used to provide quantitative assessment of model error, as well as evaluating the model performance metrics. Further data input capture and development work may be needed to justify model application in areas with significant anthropogenic features, or along streams.

Although focussed on rock exposure mapping in this work, the machine learning model developed has broad applicability for land system modelling, with opportunities for a range of input and derivative data to be used. Applications may include geochemical mapping, habitat modelling, and superficial deposit mapping amongst many other opportunities.

CRediT authorship contribution statement

Chris Williams: Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Katie Whitbread:** Writing – original draft, Visualization, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Alex Hall:** Writing – original draft, Software, Methodology. **Sam Roberson:** Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Conceptualization. **Andrew Finlayson:** Validation, Methodology, Conceptualization. **Romesh N. Palamakumbura:** Writing – review & editing, Validation, Methodology, Conceptualization. **Andrew Hulbert:** Software. **Matthew Paice:** Software.

Code availability

- Name of the code: geospatial-rf
- Contact: chrwil@bgs.ac.uk
- Hardware requirements: Minimal – desktop or equivalent suitable for running Python
- Program language: Python
- Software required: Python v3.5 (minimum)
- Program size: 16.6 MB
- Source code: https://github.com/BritishGeologicalSurvey/geospatial-rf/releases/tag/v1.0.0-external

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Chris Williams reports financial support was provided by British Geological Survey. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This project has been funded as part of NERC National Capability activities. All terrain derivatives have been derived in part from DTM of Great Britain at 5 m resolution © Bluesky International Limited. The authors would also like to thank Alessandro Novellino, Maarten Krabbendam, Severine Cornillon, Tim Kearsey and Daniela Cuba for their input and assistance during this project.

Data availability

The authors do not have permission to share data.

References

Amatulli, G., McInerney, D., Sethi, T., Stroble, P., Domisch, S., 2020. Geomorpho90m, empirical evaluation and accuracy assessment of global high-resolution geomorphometric layers. Sci. Data 7, 162. https://doi.org/10.1038/s41597-020-0479-6.

British Geological Survey, 2012. Specifications for the preparation of 1:10 000 and 1:25 000 scale geological maps and DiGMapGB-10 data. British Geological Survey Internal Report. IR/12/019.

Benn, D.I., Evans, D.J.A., 1998. Glaciers and Glaciation. Edward Arnold, London, pp. 734–pp.

Bernard, S., Heutte, L., Adam, S., 2009. Influence of hyperparameters on random forest accuracy. In: Benediktsson, J.A., Kittler, J., Roli, F. (Eds.), Multiple Classifier Systems. Springer Berlin Heidelberg, pp. 171–180. ISBN: 978-3-642-02326-2.

Beven, K.J., Kirkby, M.J., 1979. A Physically Based, Variable Contributing Area Model of Basin Hydrology. Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. Hydrol. Sci. Bull. 24, 43–69. https://doi.org/10.1080/02626667909491834.

Bluesky International Limited, 2021. DTM of Great Britain (5 m resolution). https://bluesky-world.com/standard-height-data/.

Booth, K.A., Booth, S.J., Slater, C., 2012. BGS geological cross sections & quaternary domains: user guidance notes. British Geological Survey Internal Report. OR/10/030.

Booth, S., Merritt, J., Rose, J., 2015. Quaternary provinces and domains–a quantitative and qualitative description of British landscape types. Proc. Geologists' Assoc. 126, 163–187.

Carson, M.A., Kirkby, M.J., 1972. Hillslope Form and Process. Cambridge University Press, Cambridge, p. 475.

Chen, J., Li, K., Tang, Z., Bilal, K., Yu, S., Wend, C., Li, K., 2017. A parallel random forest algorithm for big data in a spark cloud computing environment. IEEE Trans. Parallel Distr. Syst. 28 (4), 919–933. https://doi.org/10.1109/TPDS.2016.2603511.

Cho, H., 2000. GIS Hydrological Modeling System by Using Programming Interface of GRASS. Master's Thesis, Department of Civil Engineering. Kyungpook National University, Korea.

Coblentz, D., Pabian, F., Prasad, L., 2014. Quantitative geomorphometrics for terrain characterization. Int. J. Geosci. 5, 247–266. https://doi.org/10.4236/ijg.2014.53026.

Cutler, D.R., Edwards Jr., T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J., 2007. Random forests for classification in ecology. Ecology 88, 2783–2792. https://doi.org/10.1890/07-0539.1.

DiBiase, R.A., Heimsath, A.M., Whipple, K.X., 2012. Hillslope response to tectonic forcing in threshold landscapes. Earth Surf. Process. Landforms 37, 855–865. https://doi.org/10.1002/esp.3205.

Falcini, F.A.M., Krabbendam, M., Selby, K.A., Rippin, D.M., 2022. Using bed-roughness signatures to characterise glacial landform assemblages beneath palaeo-ice sheets. J. Glaciol. 68 (269), 518–532. https://doi.org/10.1017/jog.2021.122.

Fisher, R., 1953. Dispersion on a sphere. Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences 217 (1130), 295–305. http://www.jstor.org/stable/99186.

Gallant, J.C., Dowling, T.I., 2003. A multiresolution index of valley bottom flatness for mapping depositional areas. Water Resour. Res. 39, 1347. https://doi.org/10.1029/2002WR001426, 12.

Ganerød, A.J., Bakkestuen, V., Calovi, M., Fredin, O., Ketil Rød, J., 2023. Where are the outcrops? Automatic delineation of bedrock from sediments using Deep-Learning techniques. Applied Computing and Geosciences 18, 100119.

Grohmann, C.H., Smith, M.J., Riccomini, C., 2011. Multiscale analysis of topographic surface roughness in the midland valley, Scotland. IEEE Trans. Geosci. Rem. Sens. 49 (4), 1200–1213. https://doi.org/10.1109/TGRS.2010.2053546.

Hengl, T., 2006. Finding the right pixel size. Comput. Geosci. 32 (9), 1283–1298. https://doi.org/10.1016/j.cageo.2005.11.008. ISSN 0098-3004.

Hengl, T., de Jesus, J.M., Heuvelink, G.B.M., Gonzalez, M.R., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G.B., Ribeiro, E., Wheeler, I., Mantel, S., Kempen, B., 2017. SoilGrids250m: global gridded soil information based on machine learning. PLoS One 12 (2), e0169748. https://doi.org/10.1371/journal.pone.0169748.

Hofierka, J., Mitasova, H., Neteler, M., 2009. Geomorphometry in GRASS GIS. In: Hengl, T., Reuter, H.I. (Eds.), Geomorphometry: Concepts, Software, Applications, Developments in Soil Science, vol. 33, pp. 387–410. http://www.geomorphometry.org.

Horn, B.K.P., 1981. Hill shading and the reflectance map. Proc. IEEE 69 (1), 14–47.

Hubbard, B., Hubbard, A., 1998. Bedrock surface roughness and the distribution of subglacially precipitated carbonate deposits: implications for formation at Glacier de Tsanfleuron, Switzerland. Earth Surf. Process. Landforms 13 (1), 261–270. https://doi.org/10.1002/(SICI)1096-9837(199803)23:3<261::AID-ESP848>3.0.CO;2-5.

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

Karlson, M., Karlsson, C., Mörtberg, U., Olofsson, B., Balfors, B., 2016. Design and evaluation of railway corridors based on spatial ecological and geological criteria. Transport. Res. Part D 46, 207–228. https://doi.org/10.1016/j.trd.2016.03.012.

Lawley, R., Garcia-Bajo, M., 2009. The national superficial deposit thickness model (version 5). British Geological Survey Internal Report. OR/09/049.

LeDell, E., Poirier, S., 2020. H2O AutoML: scalable automatic machine learning. In: 7th ICML Workshop on Automated Machine Learning (AutoML) url: https://www.automl.org/wpcontent/uploads/2020/07/AutoML_2020_paper_61.pdf.

Liu, X., Wu, J., Zhou, Z., 2009. Exploratory undersampling for class-imbalance learning. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 39 (2), 539–550. https://doi.org/10.1109/TSMCB.2008.2007853.

Marston, C.G., O'Neil, A.W., Morton, R.D., Wood, C.M., Rowland, C.S., 2023. LCM2021 – the UK land cover map 2021. Earth Syst. Sci. Data 15, 4631–4649. https://doi.org/10.5194/essd-15-4631-2023.

Milodowski, D.T., Mudd, S.M., Mitchard, E.T.A., 2015. Topographic roughness as a signature of the emergence of bedrock in eroding landscapes. Earth Surf. Dynam. 3, 483–499. https://doi.org/10.5194/esurf-3-483-2015.

Minetto, R., Pamplona Segundo, M., Sarkar, S., 2019. Hydra: an ensemble of convolutional neural networks for geospatial land classification. IEEE Trans. Geosci. Rem. Sens. 57 (9), 6530–6541. https://doi.org/10.1109/TGRS.2019.2906883.

Mitasova, H., 1985. Cartographic Aspects of Computer Surface Modeling. PhD thesis. Slovak Technical University, Bratislava.

Molnar, D.K., Julien, P.Y., 1998. Estimation of upland erosion using GIS. Comput. Geosci. 24 (2), 183–192. https://doi.org/10.1016/S0098-3004(97)00100-3. ISSN 0098-3004.

Olaya, V., 2009. Basic land-surface parameters. In: Hengl, T., Reuter, H.I. (Eds.), Geomorphometry. Concepts, Software, Applications, Developments in Soil Science, vol. 33. Elsevier. http://cours.st.free.fr/0123743451.pdf.

Patton, N.R., Lohse, K.A., Godsey, S.E., Crosby, B.T., Seyfried, M.S., 2018. Predicting soil thickness on soil mantled hillslopes. Nat. Commun. 9, 3329. https://doi.org/10.1038/s41467-018-05743-y.

Riley, S.J., DeGloria, S.D., Elliot, R., 1999. A terrain ruggedness index that quantifies topographic heterogeneity. Intermt. J. Sci. 5 (1–4), 23–27.

Sarkar, S., Mukherjee, A., Benslimane, A., Stewart, C., 2004. Geotechnical investigation and rock characterization for the east side access project's manhattan segment. International Conference on Case Histories in Geotechnical Engineering 3. https://scholarsmine.mst.edu/icchge/5icchge/session11/3.

Scarpone, C., Schmidt, M.G., Bulmer, C.E., Knudby, A., 2017. Semi-automated classification of exposed bedrock cover in British Columbia's Southern Mountains using a Random Forest approach. Geomorphology 285, 214–224. https://doi.org/10.1016/j.geomorph.2017.02.013.

Stepinski, T., Jasiewicz, J., 2011. Geomorphons - a new approach to classification of landforms. In: Hengl, T., Evans, I.S., Wilson, J.P., Gould, M. Redlands (Eds.), Proceedings of Geomorphometry, pp. 109–112 (PDF).

Tarolli, P., 2014. High-resolution topography for understanding Earth surface processes: opportunities and challenges. Geomorphology 216, 295–312. https://doi.org/10.1016/j.geomorph.2014.03.008. ISSN 0169-555X.

Tarolli, P., Tarboton, D.G., 2006. A new method for determination of most likely landslide initiation points and the evaluation of digital terrain model scale in terrain stability mapping. Hydrol. Earth Syst. Sci. 10 (5), 663–677.

Wicaksono, A.S., Supianto, A.A., 2018. Hyper parameter optimization using genetic algorithm on machine learning methods for online news popularity prediction. Int. J. Adv. Comput. Sci. Appl. 9 (12), 263–267.

Wilson, M.F.J., O'Connell, B., Brown, C., Guinan, J.C., Grehan, A.J., 2007. Multiscale terrain analysis of multibeam bathymetry data for habitat mapping on the continental slope. Mar. Geodesy 30 (1–2), 3–35. https://doi.org/10.1080/01490410701295962.

Witten, I.H., Frank, E., Hall, M.A., Pal, C.J., 2017. Chapter 6 - trees and rules. In: Witten, I.H., Frank, E., Hall, M.A., Pal, C.J. (Eds.), Data Mining - Practical Machine Learning Tools and Techniques, fourth ed. Morgan Kaufmann, pp. 209–242. https://doi.org/10.1016/B978-0-12-804291-5.00006-4. ISBN: 978-0-12-804291-5.

Yang, W., Niu, R., Si, R., Li, J., 2024. Geological hazard susceptibility analysis and developmental characteristics based on slope unit, using the xinxian county, henan province as an example. Sensors 24 (8), 2457. https://doi.org/10.3390/s24082457.

Zevenbergen, L.W., Thorne, C.R., 1987. Quantitative analysis of land surface topography. Earth Surf. Process. Landforms 12, 47–56. https://doi.org/10.1002/esp.3290120107.