EGU2020-5017

#shareEGU20

Mapping surface rock exposures to enhance geohazard susceptibility assessment: a random forest approach

Chris Williams, Andrew Finlayson, Romesh Palamakumbura, Tim Kearsey, Severine Cornillon, Sam Roberson and Katie Whitbread

British Geological Survey, UK

Contact: chrwil@bgs.ac.uk

Overview

- ❑ Rationale
- ❑ What we want to achieve
- ❑ Real world applications
- ❑ Current datasets and limitations
- ❑ Methodology
- ❑ Building training datasets
  - ❑ Manual data capture
  - ❑ Alternative resources
- ❑ Building a model
  - ❑ Data inputs
  - ❑ Preliminary results
- ❑ Next steps

Project rationale

- Knowing where bedrock is exposed in an area is useful information when considering for example geohazards, natural resource mapping, understanding the hydrogeological functioning of catchments and site considerations for engineering ([Scarpone et al. 2017](#))
- Exposed rock information is sparse in the UK, particularly in upland areas – this is partly related to a lack of boreholes in these areas which have been the key input to past modelling activities
- Improving our knowledge for rock exposure, we can improve
  - how we model sediment thickness
  - how we understand landscape history and wider environmental processes
  - our assessment of geohazard susceptibility (e.g. landslides)
  - what information we can incorporate into the geoscience products and dataset that we develop

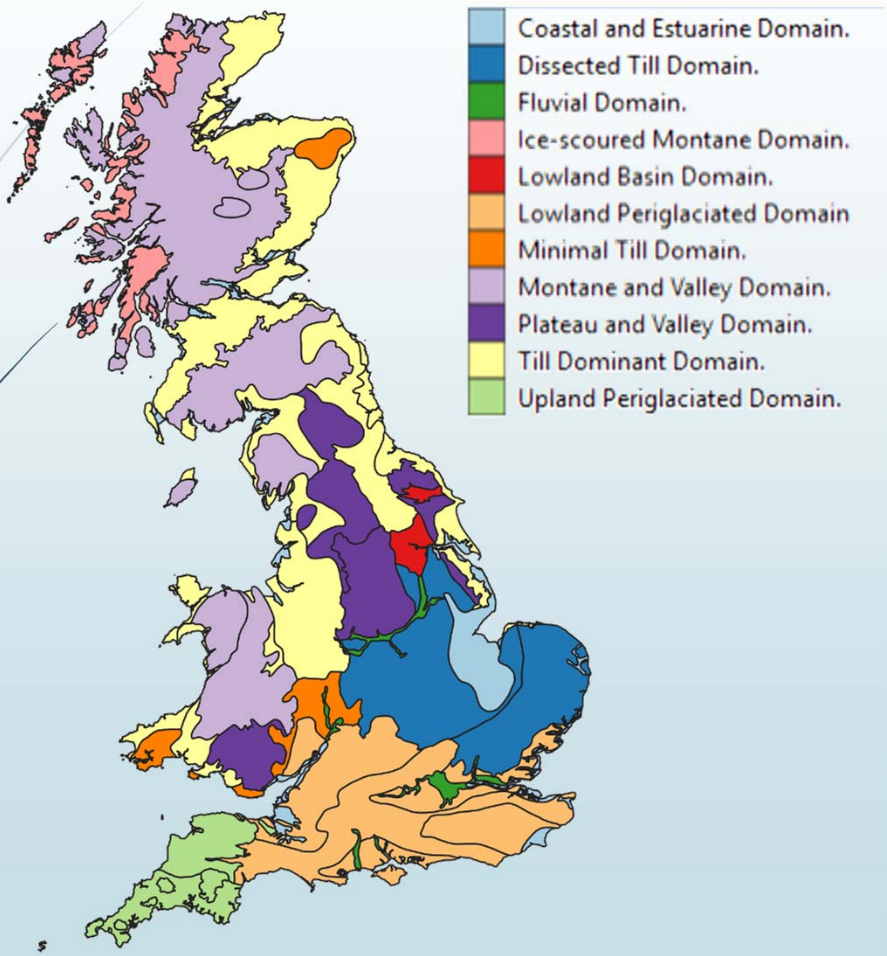
What we want to achieve

- A dataset consisting of rock presence or absence for upland regions of the UK
- A robust and repeatable process specific catchment modelling workflow
- Improved understanding of location specific processes that have resulted in surface rock exposure
- The resultant exposed rock presence/absence will act as an input to modelling efforts focussed on constraining superficial sediment thickness which underpins the work of the BGS in geohazards whilst also providing a useful dataset in its own right to a range of users including:
    - Utility sector
    - Transport sector
    - Engineering sector

Methodology

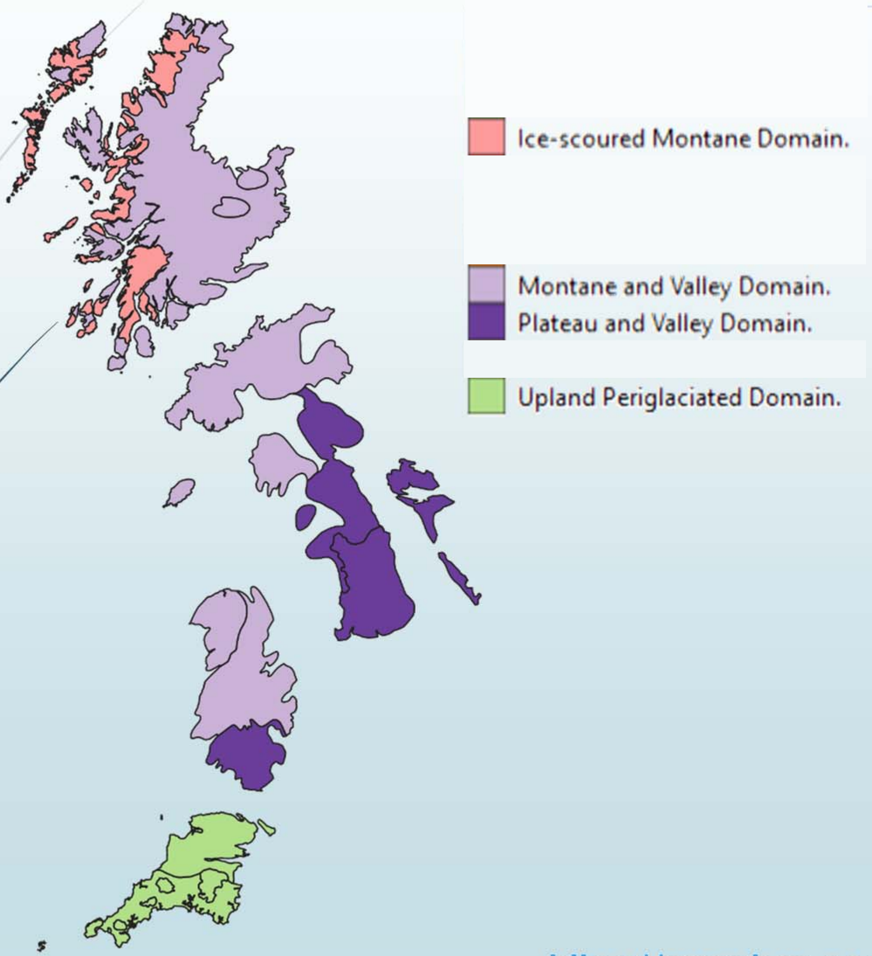
1. Consider the UK as a series of domains that have been subject to common environmental processes
2. For a select number of catchments within each domain, build up a catalogue of testing and training data:
   - ❑ rock exposure presence/absence datasets
   - ❑ geomorphometric parameters (terrain derivatives)
3. Qualify remote observations with field surveys
4. Build a random forest classifier for test sites
   - ❑ assess model sensitivity to input variables
   - ❑ test model application on inter- and intra-domain sites
5. Apply the resultant model(s) across UK upland areas to create a national upland rock exposure dataset

Common process domains



- We consider the UK in terms of separate quaternary domains where each domain accounts for a landscape that has been exposed to common processes (see Booth et al., 2015)
- The designation of these areas is based on the *land systems approach* developed (Eyles, 1983; Benn and Evans, 1998)

Common process domains



- In this study we are focusing on those domains in upland areas
- For each of these domains, a number of test catchments are selected for which we are developing training/testing datasets for model development

https://www.bgs.ac.uk/research/climatechange/quaternaryDomains/home.html

Test site:
Glengyle, Loch Lomond and Trossachs National Park, Scotland

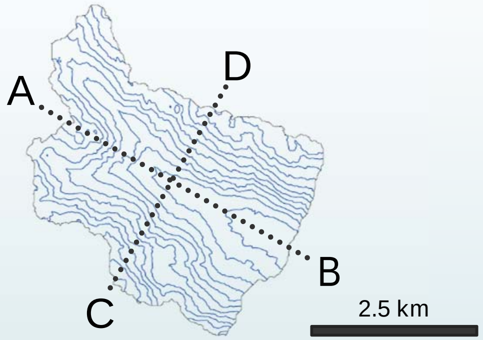
Quaternary domain: Montane and Valley

Location: 56.292125 °N, -4.636513 °E

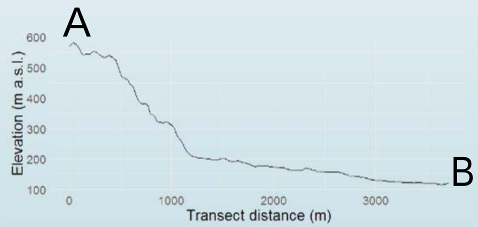
Area: 11.6 km^2

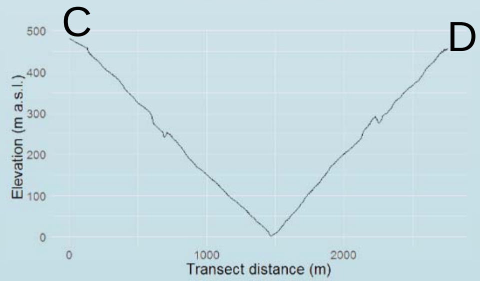
Google Earth link

glengyle_catchment.kmz

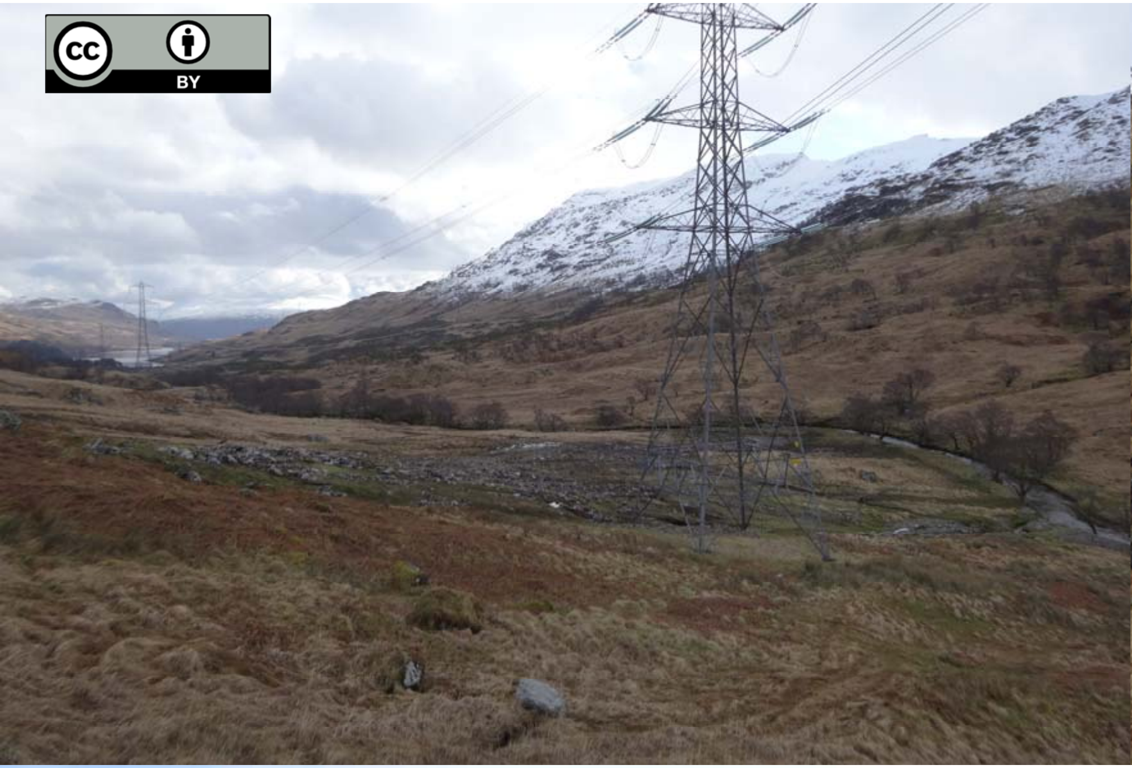


50 metre contour profile





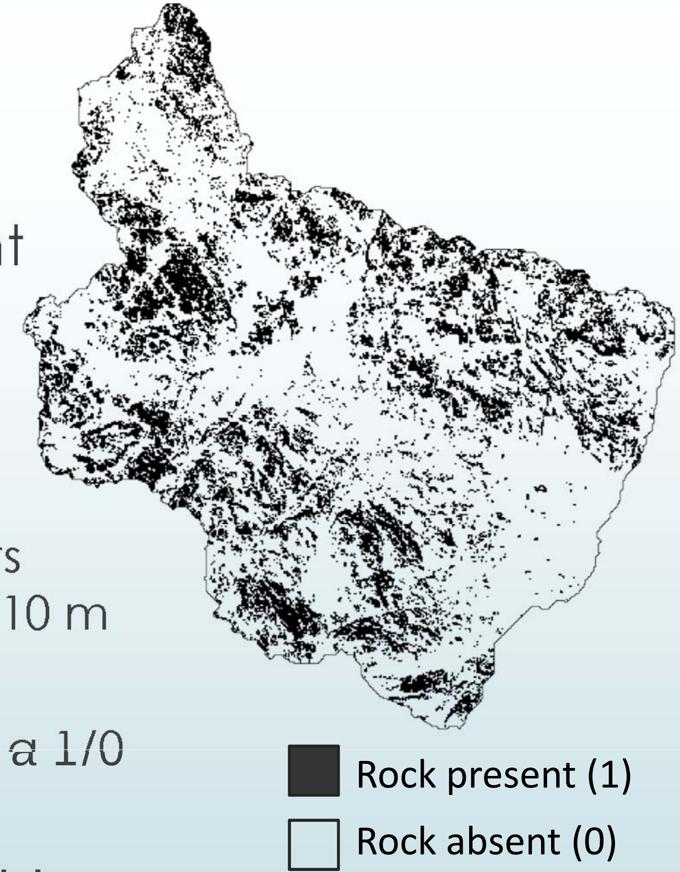
- ❑ The catchment is on the edge of Loch Katrine which is a major reservoir for the city of Glasgow
- ❑ Within the catchment, there is a large amount of power infrastructure with an associated access road
- ❑ This catchment and the surrounding area has been subject to extensive debris flow activity in recent years
- ❑ No mapped rock exposure or sediment thickness data were available here prior to this study



Building training datasets

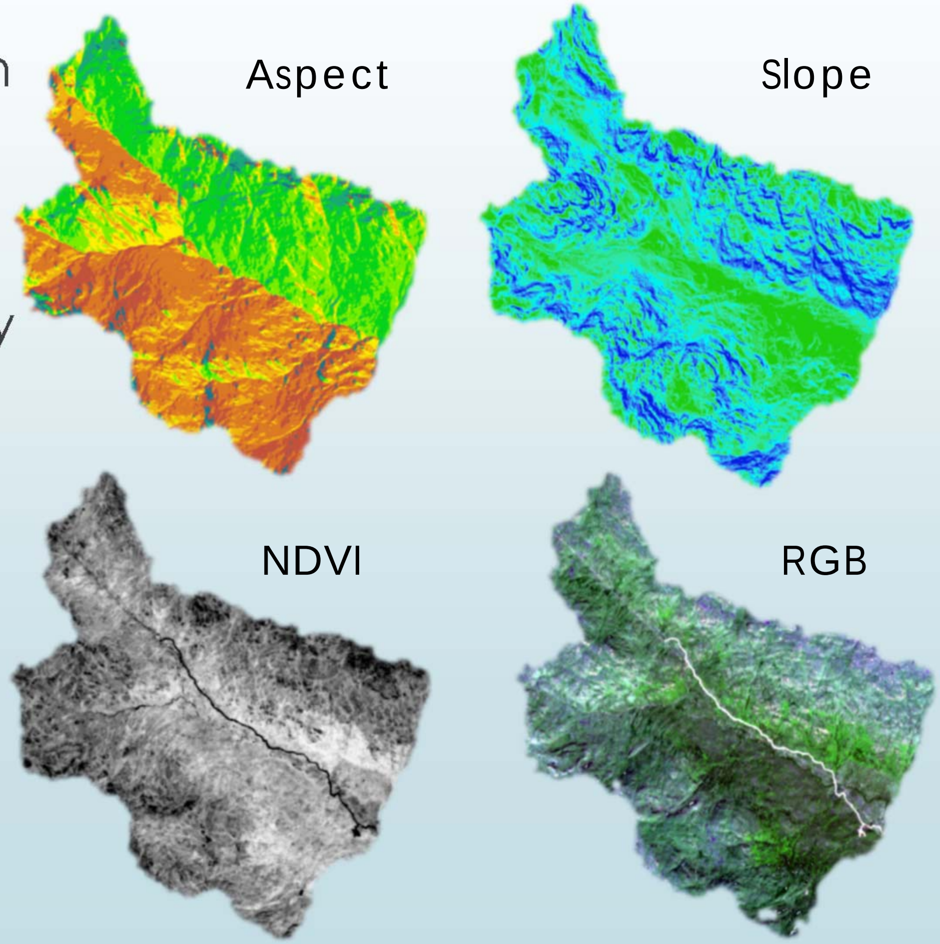
For each domain, train/test catchment data include:

- Rock presence and absence from aerial photography
  - these data are created by 4-5 geologists who manual digitize points relative to a 10 m grid where rock through to be present
  - Point data are then rasterised to create a 1/0 presence/absence raster
- For some of these catchments, a field survey is then carried out to validate the digitising approach



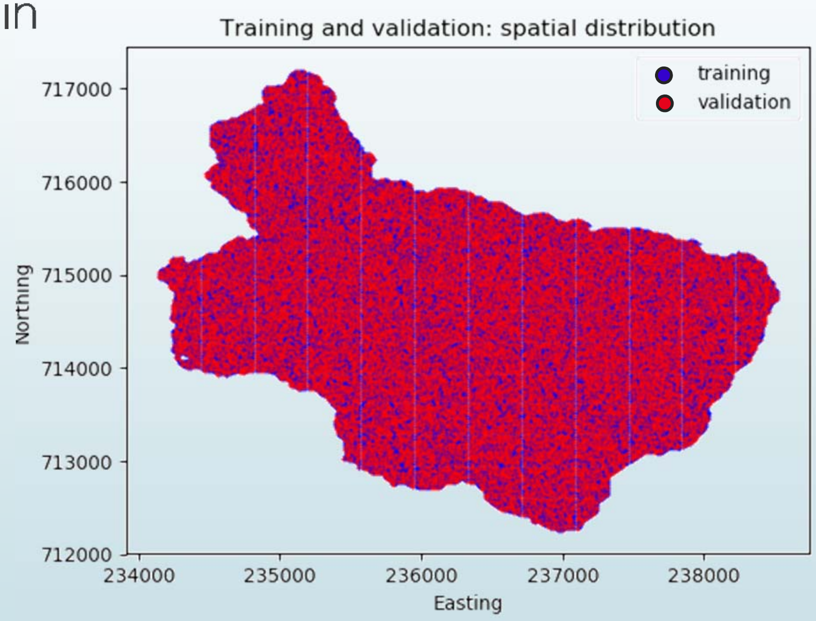
Building training datasets

- ❑ Terrain derivatives based on a 5 m resolution digital terrain model:
  - ❑ Slope, aspect, curvature, topographic position index, multiresolution index of valley bottom and ridge top flatness....
- ❑ Earth observation data
  - ❑ RGB aerial photographs
  - ❑ Sentinel 2 NDVI
- ❑ Mapped geology datasets
  - ❑ Superficial and bedrock



Building the model

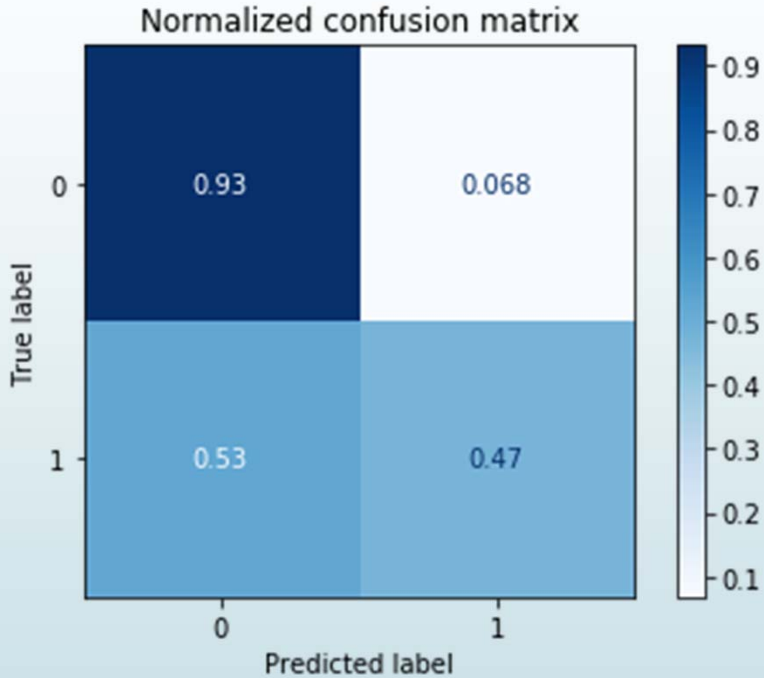
- We build a random forest classifier to predict rock exposure presence and absence based on 29 variables (terrain derivatives, EO, bedrock geology...)
- These gridded variables are first mapped to coordinates across the area of interest with a 10 m spacing
- Categorical data are one-hot encoded
- All data are scaled between 1 and 0
- A **train:test** split of **60:40** is used and then passed to a random forest classifier using the python package [scikit learn]()



- Training data rows: **68460**
- Test data rows: **45640**

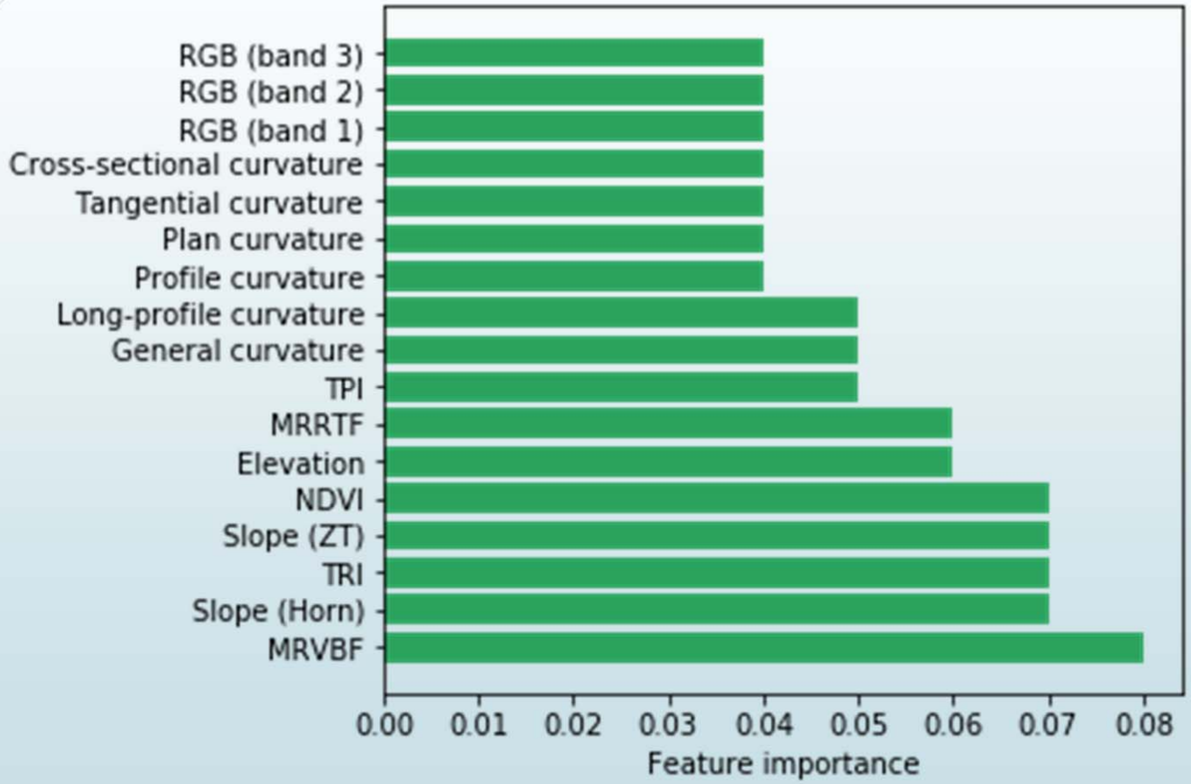
Results: confusion matrix

- ❑ Initial model based on test:train
- ❑ Overall accuracy (predicted vs test): 81.4 %
- ❑ Tendency to under-predict rock presence given the training data*
- ❑ Scaling of training variables appears to have little impact on the modelling
  - ❑ With scaling (0-1): 81.6 %
  - ❑ Without scaling : 81.5 %
- ❑ No hyperparameter tuning or model optimisation has been undertaken at this point



** the training data is based on the interpretation of a human mapper – the "quality" of the training data will be considered in future by comparing multiple training sets from differ mappers*

Results

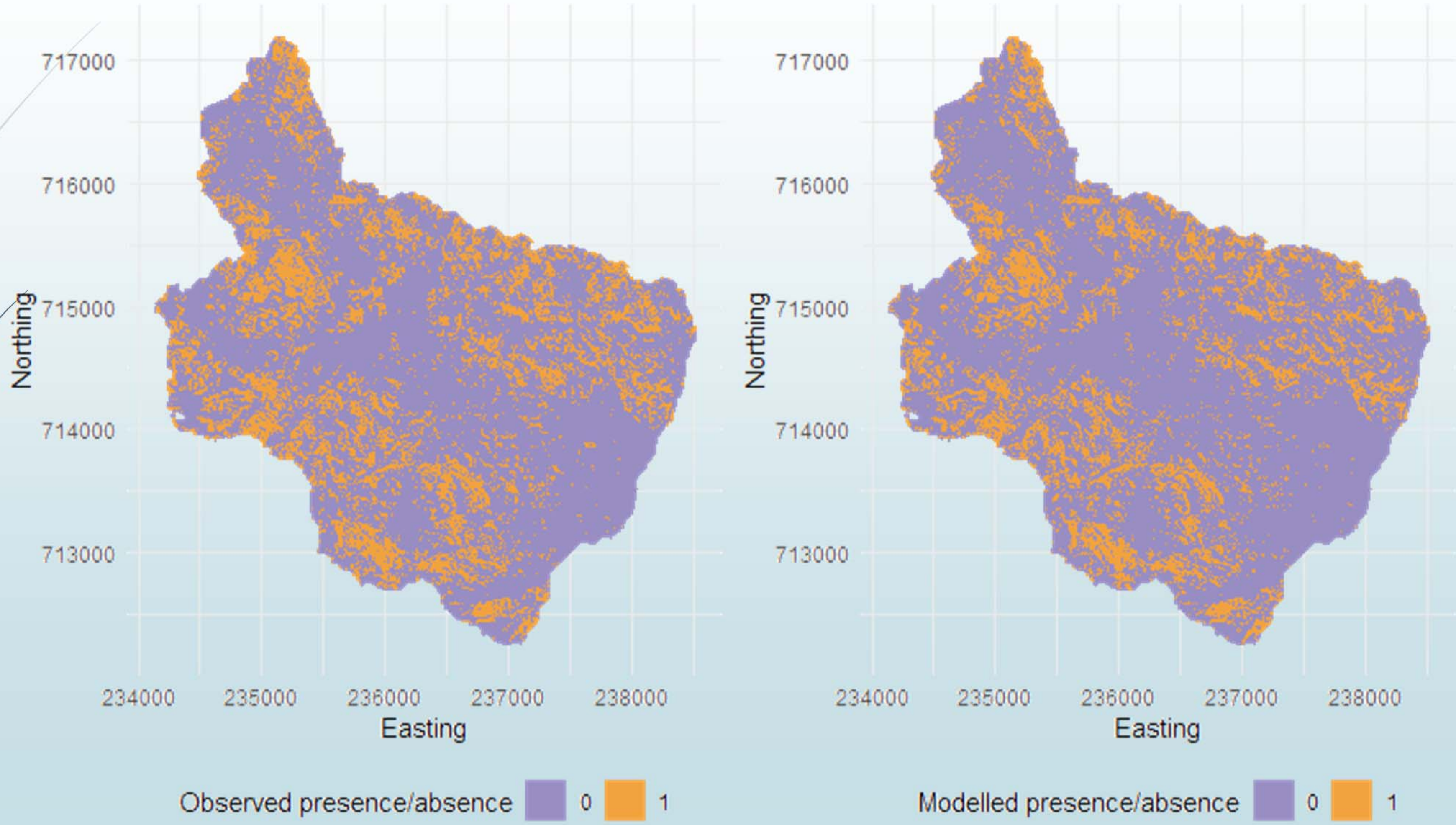


- ❑ Geology type variables were found to be of minimal importance within the model
- ❑ The spatial scale underpinning the calculation of variables including MRVBF, MRRTF and TPI may be important and requires further testing

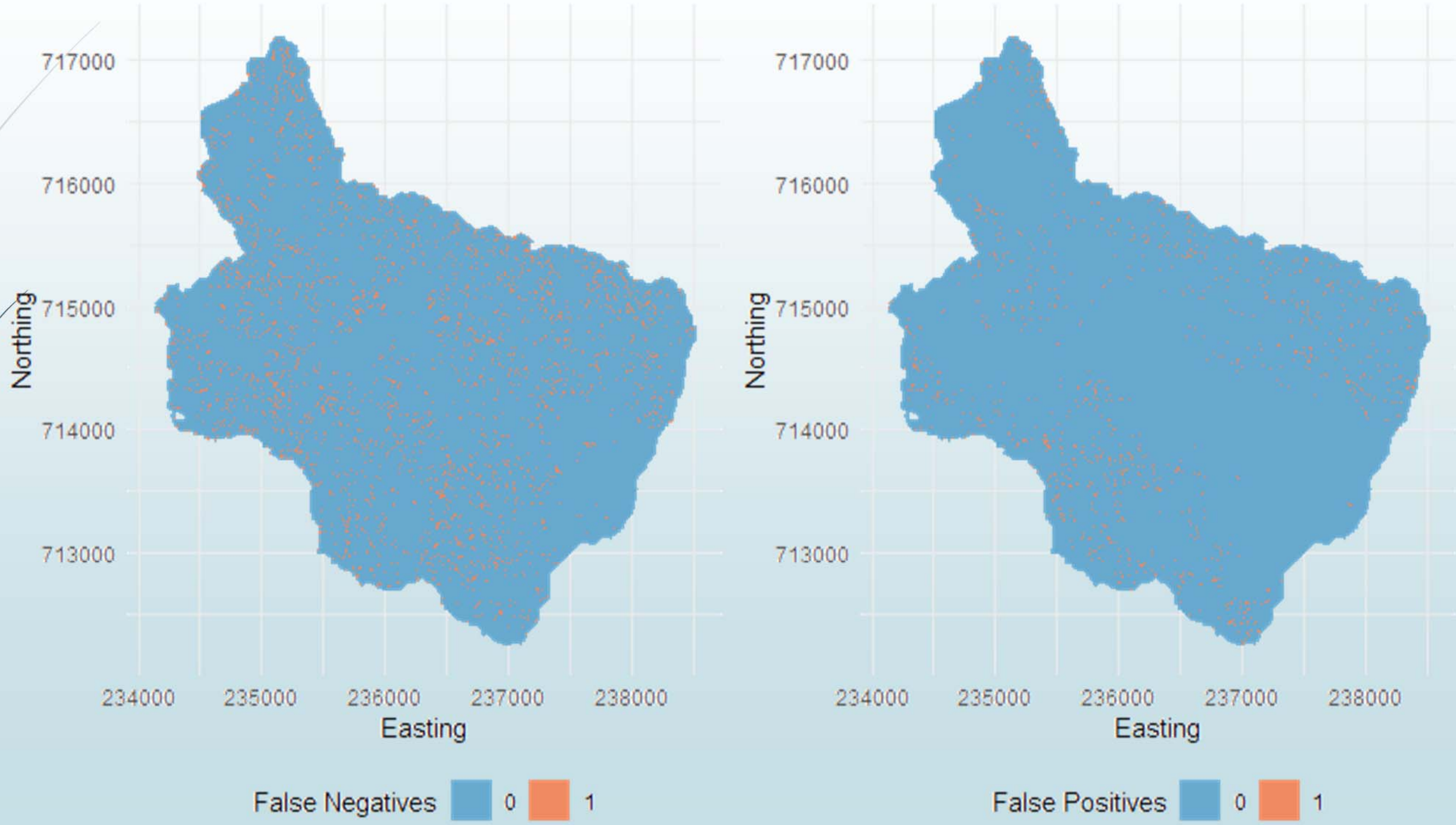
NB/ This is only a selection of the most important variables

MRVBF: Multiresolution index of valley bottom and ridge top flatness
MRRTF: Multiresolution index of ridge top flatness
TRI: Topographic Ridge Index
TPI: Topographic Position Index
Slope (Horn): Slope calculated using the Horn method
Slope (ZT): Slope calculated using the Zevenbergen and Thorne method

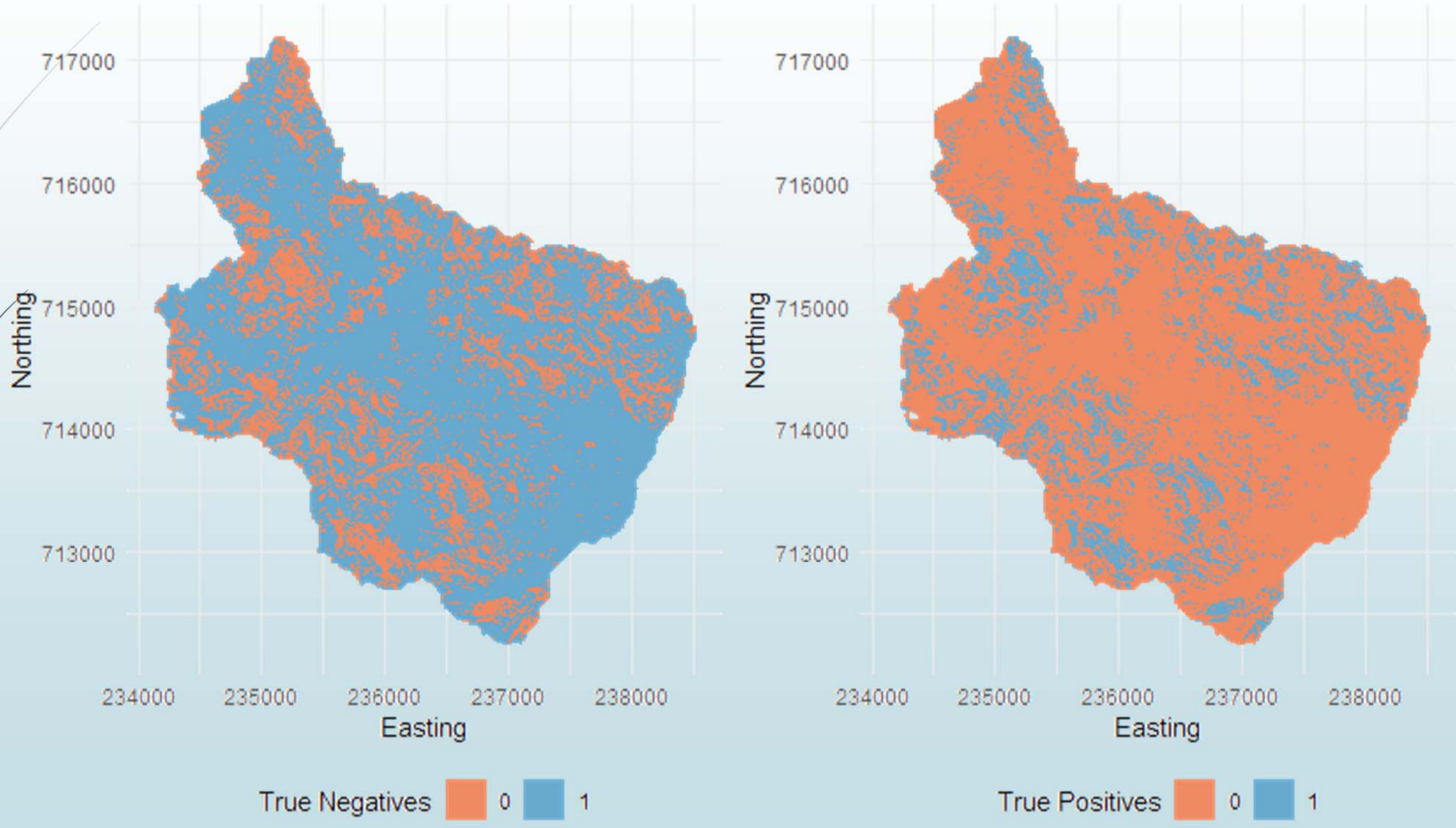
Results: observed vs modelled



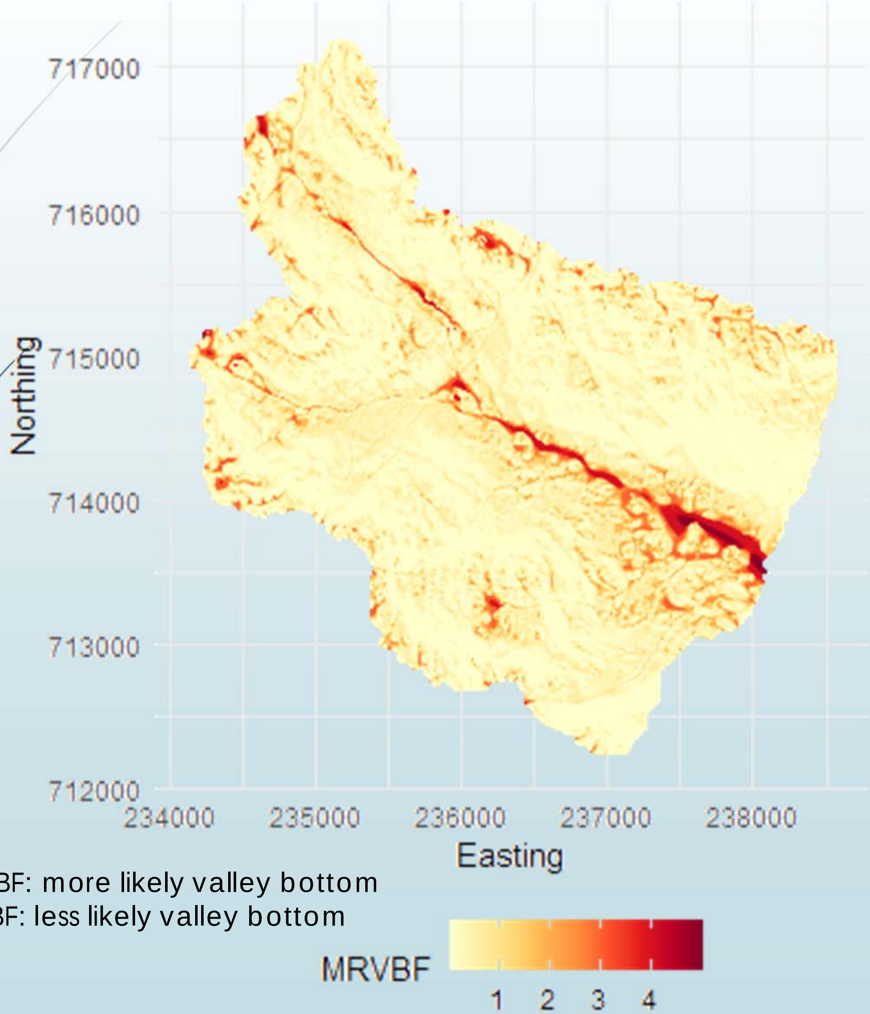
Results: false positives/negatives



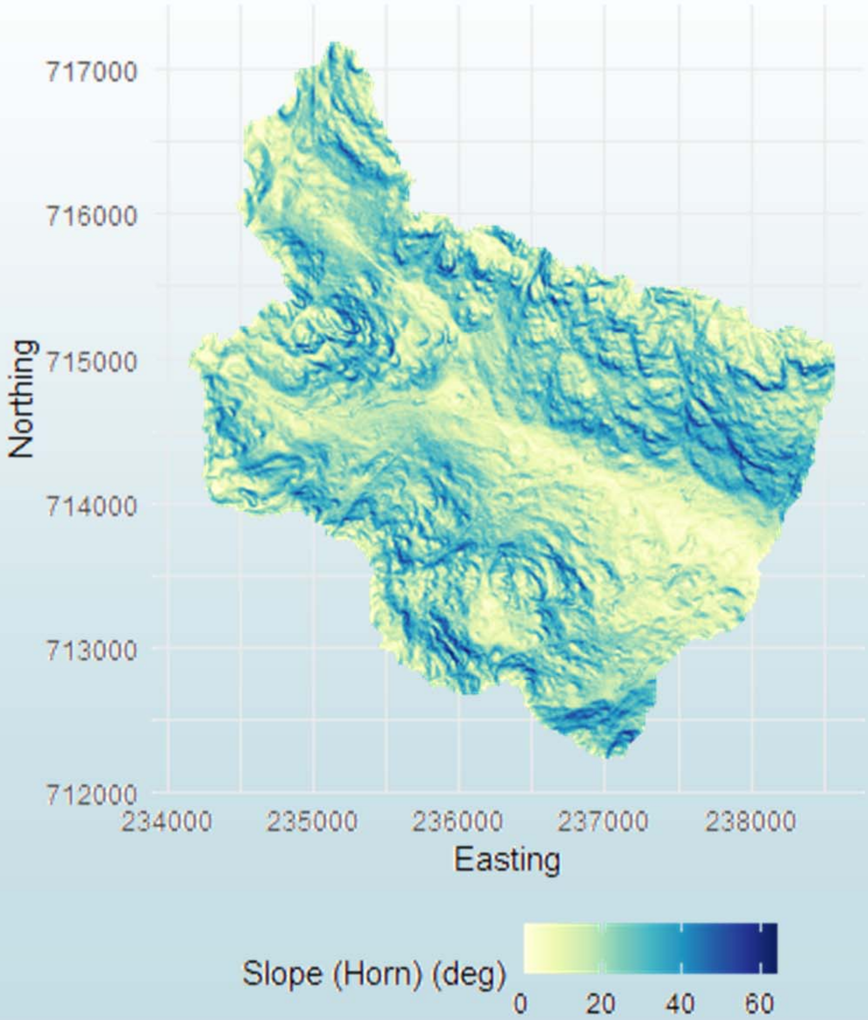
Results: true positives/negatives



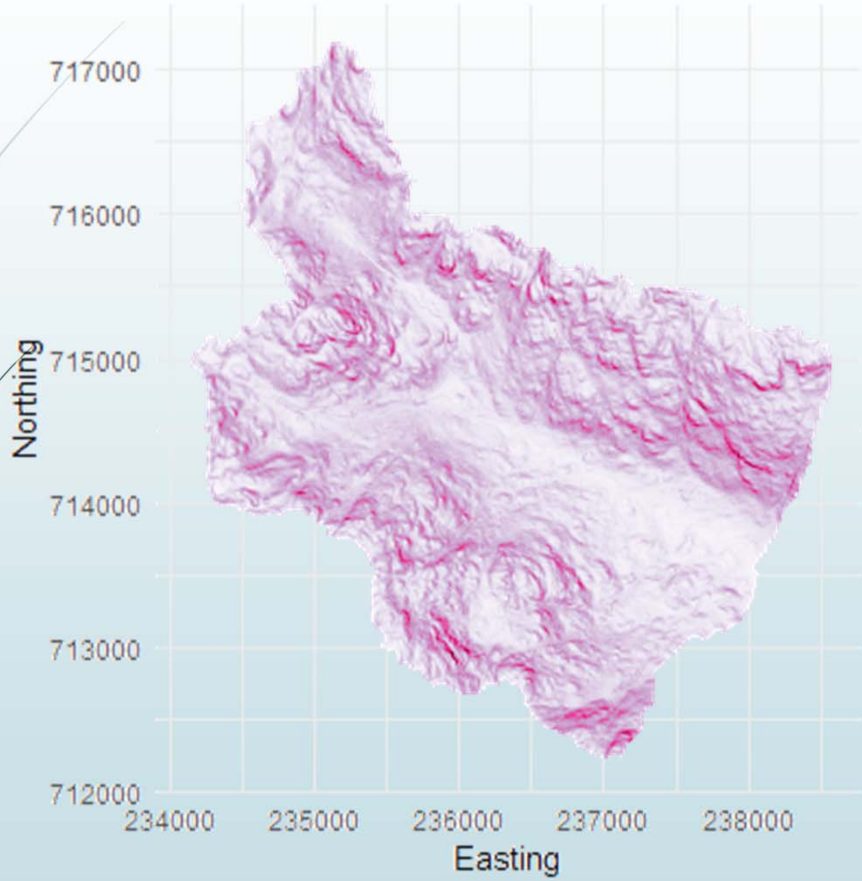
Results: most important variables (1)

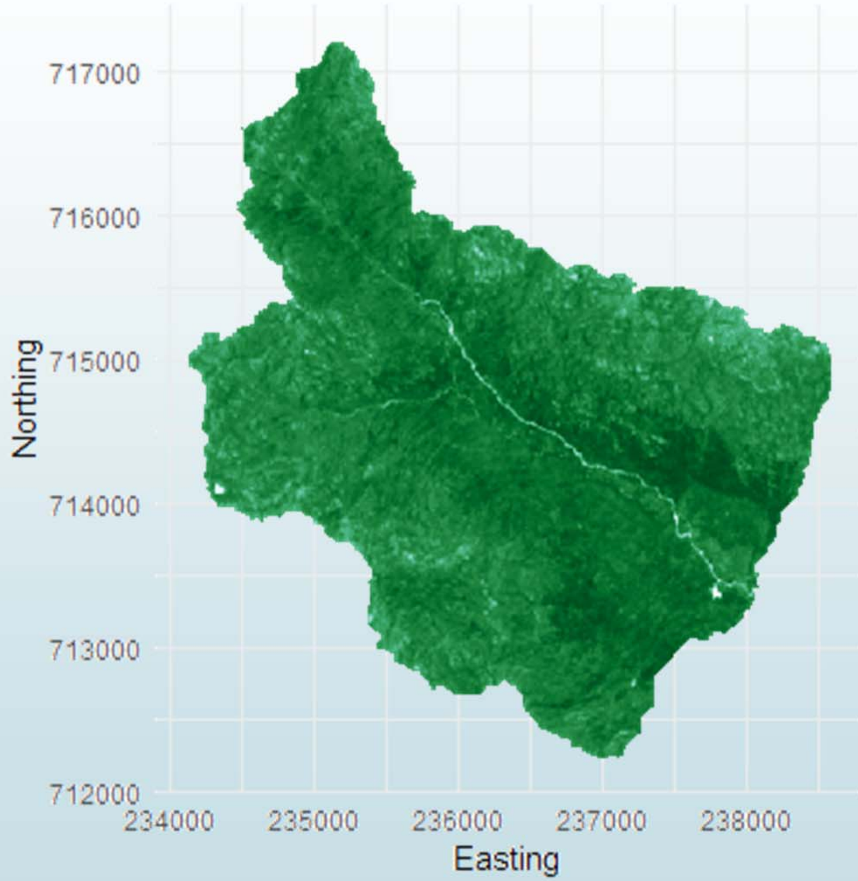


High MRVBF: more likely valley bottom
Low MRVBF: less likely valley bottom

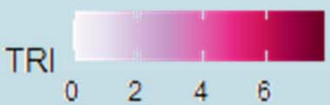


Results: most important variables (2)

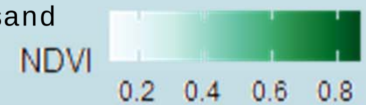




High TRI: more rugged
Low TRI: less rugged
See [Riley et al., 1999]



NDVI 0.2-0.4: shrub and grassland
NDVI close to zero: bare rock/sand
See [Sentinel Hub docs]



Next steps

For the montane and valley domain…

- ❑ Application of the model to different areas within the same domain
- ❑ Hyperparameter tuning of the model
- ❑ Consideration of different scales of terrain derivatives included
- ❑ Analysis of true/false positives/negatives relative to variable spatial distributions
- ❑ Consideration of the variability in the rock mapping training data development considering different mappers

For other domains…

- ❑ Collection of training data for catchments in different domains
- ❑ Development of new models for each of these different domains

References

- Benn, D.I. and Evans, D.J.A. 1998. Glaciers and Glaciation. Arnold, London.
- Booth, S., Merritt, J and Rose, J. 2015. Quaternary Provinces and Domains – a quantitative and qualitative description of British landscape types. Proceedings of the Geologists' Association, 126 (2), pp163-187. https://doi.org/10.1016/j.pgeola.2014.11.002.
- Eyles, N. 1983. Glacial Geology: An Introduction for Engineers and Earth Scientists. Pergamon Press, Oxford.
- Riley S.J., DeGloria, S.D. and Elliot, R. 1999. A Terrain Ruggedness Index That Quantifies Topographic Heterogeneity. International Journal of Sciences, 5(1-4), pp23-27.
- Scarpone, C., Schmidt, M. G., Bulmer, C. E. and Knudby, A. 2017. Semi-automated classification of exposed bedrock cover in British Columbia's Southern Mountains using a Random Forest approach. Geomorphology, 285. pp214-224. http://dx.doi.org/10.1016/j.geomorph.2017.02.013.