Revised manuscript clean without changes marked (For Production)

A machine learning approach to tungsten prospectivity modelling using knowledge-driven feature extraction and model confidence

Christopher M. Yeomans[a,b,*], Robin K. Shail[a], Stephen Grebby[c], Vesa Nykänen[d], Maarit Middleton[d] and Paul A.J. Lusty[b]

[a]Camborne School of Mines, College of Engineering, Mathematics and Physical Sciences, University of Exeter, Penryn Campus, Penryn, Cornwall, TR10 9FE, UK

[b]British Geological Survey, Environmental Science Centre, Keyworth, Nottinghamshire, NG12 5GG, UK

[c]University of Nottingham, Nottingham Geospatial Institute, Innovation Park, Nottingham, NG7 2TU, UK

[d]Geological Survey of Finland, P.O. Box 77, FI-96101, Rovaniemi, Finland

Keywords

Machine learning; Mineral prospectivity modelling; Mineral exploration; Random Forest™; Tungsten; SW England

Abstract

Novel mineral prospectivity modelling presented here applies knowledge-driven feature extraction to a data-driven machine learning approach for tungsten mineralisation. The method emphasises the importance of appropriate model evaluation and develops a new Confidence Metric to generate spatially refined and robust exploration targets. The data-driven Random Forest™ algorithm is employed to model tungsten mineralisation in SW England using a range of geological, geochemical and geophysical evidence layers which include a depth to granite evidence layer. Two models are presented, one using standardised input variables and a second that implements fuzzy set theory as part of an augmented feature extraction step. The use of fuzzy data transformations mean feature extraction can incorporate some user-knowledge about the mineralisation into the model. The typically subjective approach is guided using the Receiver Operating Characteristics (ROC) curve tool where transformed data are compared to known training samples. The modelling is conducted using 34 known true positive samples with 10 sets of randomly generated true negative samples to test the random effect on the model. The two models have similar accuracy but show different spatial distributions when identifying highly prospective targets. Areal analysis shows that the fuzzy-transformed model is a better discriminator and highlights three areas of high prospectivity that were not previously known. The Confidence Metric, derived from model variance, is employed to further evaluate the models. The new metric is useful for refining exploration targets and highlighting the most robust areas for follow-up investigation. The fuzzy-transformed model is shown to contain larger areas of high model confidence compared to the model using standardised variables. Finally, legacy mining data, from drilling reports and mine descriptions, is used to further validate the fuzzy-transformed model and gauge the depth of potential deposits. Descriptions of mineralisation corroborate that the targets generated in these models could be undercover at depths of less than 300 m. In summary, the modelling workflow presented herein provides a novel integration of knowledge-driven feature extraction with data-driven machine learning modelling, while the newly derived Confidence Metric generates reliable mineral exploration targets.

---

1. Introduction

The use of Machine Learning Algorithms (MLAs) for mineral prospectivity modelling has been driven by the increasing size of individual datasets and the range of data types available for mineral exploration. MLAs are computationally efficient and can deal with large, high-dimensional input datasets, non-Gaussian distributions, and generate robust exploration targets from few training samples (Carranza and Laborte, 2015a, 2015b; Rodriguez-Galiano et al., 2015). The approach requires some *a priori* data to train the model, indicating that it is a data-driven method. However, the number of training samples can be <20 which is a significant improvement compared to other data-driven methods such as Weights-of-Evidence (Carranza and Laborte, 2015b). MLAs are now commonplace in mineral prospectivity modelling. The Random Forest, Support Vector Machine and Artificial Neural Network algorithms are regularly implemented and it is the Random Forest MLA that is proving most effective in comparison studies (Rodriguez-Galiano et al., 2015; Sun et al., 2019).

Prospectivity modelling is often conducted at a large-scale, encompassing national or regional areas to determine new exploration targets. Studies have become increasingly effective due to investment in the acquisition of high-resolution airborne geophysical, satellite and geochemical datasets over large areas (Kreuzer et al., 2010; Bahiru and Woldai, 2016). Furthermore, the commitment from national geological surveys to undertake airborne geophysical surveys and geochemical baseline studies for both mineral exploration and environmental purposes has led to high-quality datasets often being freely available.

Classical prospectivity modelling has been dominated by the Weights-of-Evidence and Fuzzy Logic methods. MLAs are a more effective data-driven method compared to Weights-of-Evidence but are dependent on an effective set of training data and their ability to generalise unseen data when defining new deposits. The Fuzzy Logic technique is knowledge-based and founded on fuzzy set theory. The approach allows user-knowledge to be incorporated into the model through various data transformations chosen by the user (Zadeh, 1965; An et al., 1991; Bonham-Carter, 1994). The advantage of this is the ability to weight different data and to introduce some dependencies between variables that may be inferred by the user but not captured in the data within a conceptual deposit model. Until recently, this technique has been considered highly subjective, but work by Nykänen et al. (2015, 2017) provides a means of guiding the data processing by iteratively tuning evidence layers using an evaluation metric. Another method by Burkin et al. (2019) incorporates feature evidence into the initial evidence layer to mitigate interpretative bias of the conceptual model by the user. This approach allows multiple evidence layers to be produced from the same data – mimicking the interpretation of several users – and subsequently combines these through an objective approach (Burkin et al., 2019). The quantitative approach of the former and qualitative approach of the latter are often complementary during feature extraction. In this study we use fuzzy transformations as part of the feature extraction step in MLA modelling. We take the approach of Nykänen et al. (2015, 2017) to ensure the user-knowledge that is introduced to potentially improve a data-driven analysis is quantifiable.

MLAs also offer key post-hoc metrics to evaluate the model beyond the standard accuracy metrics. These include model variance and information entropy, which have been investigated, respectively, by Cracknell and Reading (2013) and Kuhn et al. (2018). Cracknell and Reading (2013) demonstrated the value of assessing model variance for a multi-class problem when mapping lithology to highlight fault zones, whereas Kuhn et al. (2018) used information entropy to guide field sampling campaigns to assist with geological mapping. These metrics are useful for highlighting potentially erroneous aspects of a model, which cannot be found when evaluation is based on a single accuracy metric, but have not been implemented within a mineral prospectivity modelling framework.

Herein, we demonstrate the use of fuzzy set theory for feature extraction, as well as post-hoc metrics, for tungsten mineralisation in SW England using a Random Forest MLA. We explore how incorporating knowledge-driven principles as part of feature extraction within a data-driven modelling workflow can improve the final results and compare this to a model using standardised (zero mean and equal variance) input variables. Furthermore, the models are spatially evaluated using model variance and a newly derived Confidence Metric which are applied to generate robust targets for mineral exploration with a refined area. Finally, legacy mining data are used to further validate new targets and give a depth estimate to mineralisation.

1.1. Geological framework

SW England is a world-class tin-tungsten province and provides an excellent case study location for prospectivity modelling due to the recent acquisition of high-resolution airborne geophysical and geochemical datasets (Beamish et al., 2014; British Geological Survey, 2016). The regional geology (Figure 1) is dominated by low-grade regionally metamorphosed Devonian-Carboniferous successions that were deformed during the Variscan Orogeny; these were subsequently intruded by the Early Permian Cornubian Batholith (Leveridge and Hartley, 2006; Scrivener, 2006; Shail and Leveridge, 2009; Simons et al., 2016). The batholith is closely associated with a tin-tungsten orefield that has also been exploited for copper, zinc, lead, silver, antimony, arsenic, uranium and a number of other subordinate metals (Jackson et al., 1989). Tungsten vein mineralisation was governed by the coeval post-Variscan regional tectonic and structural development and magmatic and magmatic-hydrothermal evolution of the batholith; these are outlined briefly below.

1.1.1. Regional tectonics and structural geology

The regional structural geological evolution records two episodes of deformation (D1 and D2) relating to Variscan convergence and continental collision, e.g. Sanderson and Dearman (1973); Rattey and Sanderson (1984); Alexander and Shail (1996). These were associated with the development of NNW-directed thrust faults and NNW-SSE transfer faults within

Devonian and Carboniferous successions (Dearman, 1963, 1970; Coward and Smallwood, 1984; Shail and Alexander, 1997).

Post-convergence NNW-SSE extension (D3) commenced in the latest Carboniferous and brought about reactivation of Variscan thrust faults. Continued extension generated new higher angle extensional faults through much of the Early Permian (Figure 2; Shail and Wilkinson, 1994; Alexander and Shail, 1995, 1996). Subsequent minor, Permian, ENE-WSW (D4) and NNW-SSE (D5) intraplate shortening events are also recognised (Hobson and Sanderson, 1983; Rattey and Sanderson, 1984; Shail and Alexander, 1997). The D3-D5 events spanned batholith construction and mineralisation and their brittle expression, as faults and tensile fractures, were essential for the migration of magmatic-hydrothermal fluids and the development of lodes and sheeted veins (Shail and Wilkinson, 1994; Shail and Alexander, 1997). Tungsten deposits formed in cuspate bodies of granite and their immediately adjacent host rock (Hosking and Trounson, 1959; Jackson et al., 1989; Ball et al., 1998). These deposits are commonly proximal to major NW-SE faults, e.g. Hemerdon, Redmoor, Cligga Head (Figure 3), that have acted as strike-slip transfer faults during Early Permian NW-SE extension, and appear to have influenced both magmatism and mineralisation (Shail and Wilkinson, 1994; Shail et al., 2017).

1.1.2. Permian granite batholith

The Cornubian Batholith comprises five principal granite types: G1, two-mica granite; G2, muscovite granite; G3, biotite granite; G4, tourmaline granite; G5, topaz granite (Simons et al., 2016). The association between granite type and mineral prospectivity is not well-constrained; granite types close to surface are sometimes older than, and unrelated to, the mineralisation they host, e.g. Carnmenellis Granite (Moscati and Neymark, 2020). Nevertheless, there is a strong association between W mineralisation and muscovite granites (G2); these typically form small stocks and have been interpreted as a differentiation product of two-mica (G1) granites, which also have an association with W mineralisation (Simons et al., 2016, 2017). Tourmaline granites (G4) are common in areas of significant tin mineralisation and have been interpreted as the precursor differentiated magmas that released Sn-bearing magmatic-hydrothermal fluids (e.g. Müller et al., 2006). Topaz granites (G5) host very low-grade disseminated Sn-W-Ta-Nb mineralisation and have been inferred to be the source of substantial tourmalinisation haloes and associated Sn-W mineralisation in the surrounding host rocks (Manning and Hill, 1990).

1.1.3. Tungsten mineralisation and exploration

Tungsten mineralisation in SW England, as reported in the British Geological Survey (BGS) GeoIndex (2018), is shown in Figure 3. Additional tungsten occurrences are known, and described in Dines (1956), but are not readily available in digital form and so were used solely for qualitative evaluation.

Tungsten mineralisation is overwhelmingly hosted by sheeted veins and lodes. Wolframite is the dominant ore mineral; scheelite is often present but usually minor (Jackson et al. 1989). Sheeted veins typically comprise quartz ± tourmaline ± K-feldspar ± tourmaline-wolframite ± cassiterite ± arsenopyrite and commonly display greisened margins. They occur in well-exposed stocks or dykes of muscovite (G2) granite, and their immediately adjacent host rocks, and have been described in detail, e.g. Cligga Head (Hall, 1971; Moore and Jackson, 1977), St Michael's Mount (Dominy et al., 1995) and Hemerdon (Cameron, 1951; Dines, 1956; Shail et al., 2017). The Hemerdon deposit was recently operated by Wolf Minerals Limited and produced tungsten and tin concentrates during 2015—2018. Lode mineralisation usually occurs in two mica (G1) granites, e.g. Carnmenellis and Bodmin Moor, and muscovite (G2) granites, and their immediately adjacent host rocks; assemblages can be similar to those in sheeted veins, e.g. East Pool and Agar Mine and Castle-an-Dinas Mine (Dines, 1956). However, wolframite also occurs in complex polymetallic lodes comprising quartz ± tourmaline ± chlorite ± fluorite ± cassiterite ± arsenopyrite ± chalcopyrite ± sphalerite, e.g. Roskear Complex Lode (Dines, 1956).

These magmatic-hydrothermal systems are Early Permian in age and synchronous with batholith construction, based on Ar-Ar dating of muscovite wallrock alteration and U-Pb dating of cassiterite (Chen et al. 1993; Chesley et al., 1993; Moscati and Neymark, 2020; Tapster and Bright, 2020). Fluid inclusion studies, on vein quartz and cogenetic wolframite-cassiterite, indicate typical magmatic-hydrothermal fluids temperatures in the range 300-400°C (Jackson et al., 1977, 1989; Campbell and Panter, 1990; Smith et al., 1996). The majority of vein and lode systems formed in response to Early Permian N-S regional extension (Moore, 1975; Shail and Wilkinson, 1994) but coeval NW-SE transfer faults also appear to have influenced magmatism and mineralisation (e.g. Shail and Wilkinson, 1994; Shail et al., 2017).

Exploration has been selective and focused around known tungsten deposits. Andrews et al. (1987) conducted soil geochemical studies around the Hemerdon deposit, which involved three transects and identified geochemical anomalies, although no follow up trenching is known. Geochemical exploration at Redmoor, which made use of an extensive diamond and percussive drilling campaign as well as samples of float (rock fragments in soil), attempted to define an alteration halo (Newall and Newall, 1989; Newall, 1994). The work used factor analysis to identify a "mineralisation factor" for the elements As, Cu, W, Sn, Na* and Zr (where * indicates a negative correlation). Beer et al. (1986) identified clear geochemical anomalies for tungsten, based on percussive drilling along traverses, near to the Castle-an-Dinas tungsten lode. The Mulberry and Wheal Prosper area was investigated by Bennett et al. (1981) who found both tungsten and tin soil geochemical anomalies, in proximity to Meadfoot Group calc-silicate units. Regional investigations were undertaken by Moore and Camm (1982) and James and Moore (1985) using space-borne Landsat MSS and Seasat data to map regional structures associated with tungsten mineralisation.

2. Data and Methods

The workflow illustrated in Figure 4 shows the steps required to incorporate knowledge-based feature extraction into a data-driven modelling workflow to generate spatially refined robust targets for mineral exploration. These include defining the conceptual deposit model, initial data preparation (see Supplementary Information), feature extraction using fuzzy transformations and machine learning modelling. It should be noted that, herein, the terms *evidence layer* and *input variable* are used interchangeably.

2.1. Conceptual tungsten deposit model

The conceptual mineral deposit model enables the user to identify key exploration criteria. These are represented by evidence layers, generated from available datasets. Regional geological, geochemical and geophysical datasets have been incorporated in this work to identify tungsten mineralisation in SW England. The contribution of these evidence layers to the conceptual deposit model is described below.

Prior mineral exploration and geological investigations provide a substantial body of research on which to build a regional conceptual tungsten deposit model for SW England (Hosking and Trounson, 1959; Hall, 1971; Moore and Camm, 1982; Andrews et al., 1987; Moore and Jackson, 1977; Jackson et al., 1989; Newall and Newall, 1989; Newall, 1994; Ball et al., 1998, 2002; Shail et al., 2017). Based on these observations, a conceptual deposit model has been developed to capture the common characteristics of known tungsten deposits (Figure 5). The model is based on a range of readily available geological, geochemical and geophysical datasets. Geological data comprises: (1) the mapped extent of granite plutons based on British Geological Survey 1:50 000 data, and (2) a depth to granite layer determined from the LiDAR Digital Terrain Model (DTM) and the granite surface model, based on regional gravity data, created by Willis-Richards and Jackson (1989). Geochemical datasets include soil and stream-sediment data from the G-BASE survey (British Geological Survey, 2016), Tellus South West airborne geophysical surveys (Beamish et al., 2014; Ferraccioli et al., 2014) and lineament data (Yeomans et al., 2019).

The evidence layers generated from these datasets have been prepared within the ESRI ArcGIS Desktop software package. These data were resampled to a common extent and resolution based on the airborne geophysical data (40 m pixels), and standardised to zero mean and equal variance; as is usual in many machine learning approaches (Camps-Valls et al., 2007; Hastie et al., 2009; Cracknell and Reading, 2015, 2014). The data preparation steps for each layer are presented in the Supplementary Information (S1).

2.1.1. Geological evidence layers

The geological exploration criteria defined here are based on the observation that tungsten mineralisation generally occurs, in granites or their host rocks, close to the margins of “cuspate” granite bodies or cupolas, at the roof of the batholith (Hosking and Trounson, 1959; Beer et al., 1975; Dominy et al., 1995; Ball et al., 1998). An evidence layer for proximity-to granite was prepared using the British Geological Survey 1:50 000 data to capture the XY locations of granite contacts. A proximity-to granite layer was also prepared to capture the depth to the granite contact in areas that may have blind mineralisation. The granite surface from the 3D model created by Willis-Richards and Jackson (1989) is subtracted from the LiDAR DTM and included as a proximity-to layer that captures the proximity-to granite in Z (depth) to identify shallow granite bodies. Due to some areas of the model protruding above surface, the evidence layer was classified into seven groups to allow down-weighting of the protruding areas.

Structural information was also included, based on observations by Shail et al. (2017), using regional lineament data derived from the airborne geophysics by Yeomans et al. (2019). A proximity-to structures layer using a Euclidean distance algorithm was prepared based on NW-SE lineaments with lengths > 1200 m; these lineaments are interpreted to be primarily fault-controlled. Furthermore, a density map of all NW-SE lineaments was created to capture areas of high fracturing that may favour mineralisation.

2.1.2. Geochemical evidence layers

Regional soil and stream-sediment geochemical data from the G-BASE survey (British Geological Survey, 2016) were used to derive geochemical evidence layers. The soil samples were collected at a depth of 0-20 cm and sieved to 2 mm. Stream-sediment samples were analysed using X-ray Fluorescence Spectroscopy with no digestive reagent. Strict Quality Assessment and Quality Control was conducted by the British Geological Survey prior to release through the G-BASE survey; detailed by Wragg et al. (2018).

Geochemical evidence layers have been created through an Inverse-Distance Weighting (IDW) algorithm based on preparation steps by Carranza (2010) and are summarised in Table 1. Geochemical evidence layers are duplicated for both soil and stream-sediment datasets discussed below, excluding the K/(Zr/Eu) layer. This ratio is exclusive to the stream-sediment data due the absence of rare earth element analyses for soil samples. These data are considered in three groups representing mineralisation, aureole and granite geochemistry.

For mineralisation geochemistry, data on W as well as Sn, due to their common association, is included (Cameron, 1951; Dines, 1956; Hall, 1971; Moore and Jackson, 1977; Jackson et al., 1989). The use of As, Bi, Sb, Na*, Rb and Cs (where * indicates a negative correlation) is based on the previous exploration campaigns.

As, Bi and Sb are used as indicators for mineralisation where tungsten and tin may be unobserved. They occur at distance from the deposit (Andrews et al., 1987), therefore, these elements may be a vector element in soil geochemistry for mineralisation at depth (or laterally) where the main tungsten mineralisation is undercover and assuming there has been minimal soil transport. Sb was considered to be an unreliable indicator element by Ball et al. (2002) but is included in this study to determine its importance.

The inclusion of Na*, Rb and Cs and ratios such as K/Rb* and K/Cs* is based on aureole geochemistry and alteration in mineralised country rocks surrounding granite cupolas (Newall and Newall, 1989; Ball et al., 1998). Other elements that are enriched include Li and F (Andrews et al., 1987; Newall and Newall, 1989; Newall, 1994; Ball et al., 1998), but there are insufficient analyses for these elements across the region and they have therefore not been included.

Lithogeochemical evidence layers are focused on granite types and these are defined using two ratios. Ti/Sn* is useful for determining a general granite signature (Ball et al., 1984, 1998) but fails to separate granite types. By interrogating geochemical data from Simons et al. (2016), an indicator ratio has been determined, K/(Zr/Eu), that separates the G2 granite from other granite types (Figure 6), albeit with some close associations with the G1a type. Other useful ratios have been identified, such as Zr/Fe_2O_3, Nb/Zr and Ba/Rb, but they are not effective discriminators of G2 granites (Simons et al., 2016). Potential indicator elements for G2 granite types include Be and Li (Simons et al., 2017); however, these are not included in the available soil and stream-sediment geochemical datasets for the region.

2.1.3. Geophysical evidence layers

The geophysical evidence layers defined in the conceptual deposit model incorporate airborne radiometric data from the Tellus South West project. The magmatic-hydrothermal aureole around granite plutons in SW England is highlighted by tan^{-1}(K/eU*). It is included to capture hydrothermal alteration where elevated uranium concentrations indicate that mineralising fluids may have circulated; as with geochemical ratios the evidence layer is an inverse relationship. The inverse tangent function is applied to the ratio and results in a non-linear normalisation with the data scaled from -1.57 to +1.57, which limits the effects of outliers and potentially infinite values (Schetselaar, 2002; IAEA, 2003).

2.1.4. Training and validation data

A set of 34 known regional tungsten occurrences was compiled from the Mineral Occurrence Database, maintained by the BGS GeoIndex (2018), and were used as true positive samples. True negative samples are also necessary to accurately model and validate unfavourable areas in the prospecitivity models. An equal number of true negative samples were randomly generated to ensure balanced training classes and minimise error rates (Mellor et al., 2015). A minimum buffer of 400 m was applied to minimise spatial correlation with true positive samples and other true negative samples. Furthermore, instead of one set of true negative samples, 10 sets of 34 true negative samples generated as suggested by Nykänen et al. (2017).

These sample sets were randomly subset 70:30 into 23 training and 11 validation data for use in the fuzzy feature extraction methods discussed below. Multiple random sets of true negative samples allow for testing of the random effect of point selection using the Receiver Operating Characteristics (ROC) curve tool and the Area Under Curve (AUC) value (Nykänen et al. 2017). By repeating the ROC curve analysis 10 times using randomly generated true negative samples, Nykänen et al. (2017) demonstrated that a more robust metric is obtained that highlights the potential for random variability in the AUC statistic.

For feature extraction, the training sample subsets are used to generate 10 ROC curve analyses and determine the relevance and sensitivity of the evidence layer and tune the parameters of the fuzzy transformation or combination.

For modelling, the 10 sets of 34 true negative samples were combined into a single dataset and reselected randomly into new training and validation subsets using the same 70:30 split. The reselection of random points is aimed at reducing the likelihood of overfitting due to feature extraction being honed by the same training data used for modelling. Model training data used the true positive training subset and the first random true negative training subset. The model testing (and final AUC values) used validation samples from all 10 reselected true negative validation subsets as part of the ROC curve analysis for model evaluation.

2.2. Fuzzy feature extraction

The advent of high-resolution datasets of various types has meant that mineral prospectivity models often include high numbers of input variables which increase the dimensionality. Minimising the number of variables reduces data redundancy, which can improve classification accuracy and reduce computation times (Witten et al., 2017). This process also mitigates the "curse-of-dimensionality", also known as the "Hughes effect" (Hughes, 1968), whereby the number of training samples required to capture data variance increases disproportionately with the number of variables. This is an important consideration when only a small number of training samples is available. For these reasons, the extraction of the most relevant features or characteristics within the evidence layers used in the prospectivity modelling is of paramount importance.

A common and simple means of feature extraction is to use operators, such as multiplication or division, to amplify the interactions between different variables (Henery, 1994a, 1994b). Some of these may also have the benefit of mitigating noise and removing correlated data (Hastie et al., 2009); e.g. radioelement ratios (IAEA, 2003). Another option is to highlight features using data transformations or image enhancements. There is a broad range of task-specific transformations and enhancements that, when used with an appropriate MLA, result in a high degree of accuracy (Sukumar et al., 2014).

In mineral prospectivity modelling, it is common to include 'proximity-to' evidence layers which is an example of feature extraction, e.g. proximity-to structures. Many prospectivity models attempt to refine the number of evidence layers using factor analysis, principal component analysis or the singularity method to extract new features (Abedi et al., 2013; Zhao et al., 2015; Wang et al., 2017a; Wang et al., 2017b; Wang et al., 2018). The Fuzzy Logic approach incorporates the transformation and weighting of data and is also an example of the feature extraction process where the fuzzy transformations and operators enhance and accentuate particular characteristics.

The feature extraction methods discussed in this section concerns the reduction and enhancement of the standardized variables generated during data preparation (see Supplementary Information). This was conducted in ESRI ArcGIS software and the ArcSDM 5 package, maintained by the Geological Survey of Finland (GTK, 2019), which compiles various tools for mineral prospectivity modelling. It includes the ROC curve tool that is used to guide the subjective fuzzy data transformations.

2.2.1. The Receiver Operating Characteristics (ROC) curve tool

The output for mineral prospectivity modelling using MLAs is often a binary classification. However, it is the class probabilities, the likelihood that a pixel is classified correctly, that are of value when considering prospectivity (Harris et al., 2015). It is good practice to evaluate the accuracy of the prospectivity models, most commonly through the ROC curve tool (Agterberg and Bonham-Carter, 2005; Fawcett, 2006; Robinson and Larkins, 2007; Nykänen, 2008). This uses *True Positives* (TP), *True Negatives* (TN), *False Positives* (FP) and *False Negatives* (FN) to determine a range of metrics including *Sensitivity* (Equation 1) and *Specificity* (Equation 2).

$$Sensitivity = \frac{TP}{TP+FN} \quad (1)$$

$$Specificity = \frac{TN}{TN+FP} \quad (2)$$

The ROC curve tool plots *Sensitivity* against *1 - Specificity* and this can be used to calculate the AUC. From a modelling perspective, the AUC values provide an accuracy measure with a range between 0 and 1 where 0.5 represents a random result. During feature extraction, more reliable features that capture the traits of true positive samples, are achieved by maximising the AUC value by tuning the enhancement parameters. A minimised AUC value is still useful in this instant as it represents a correlation with true negative samples and thus has an inverse relationship to the model.

2.2.2. Fuzzy membership transformation

The subjective nature of fuzzy set theory and the Fuzzy Logic method can be circumvented by refining input variables using the ROC curve tool developed by Nykänen et al. (2015, 2017). The approach provides a quantitative metric for assessing subjective aspects of the Fuzzy Logic technique, namely the application of the fuzzy membership function and fuzzy operators such as *FuzzyOR* (An et al., 1991; Bonham-Carter, 1994). The tool optimises the output of these functions and operators and allows tuning of the features to reflect the characteristics of known deposits. In turn, the correlation of an input layer can be used to indicate whether it is correctly included as part of the conceptual deposit model.

The method applied here used an iterative approach to assess the fuzzy membership function where initial evidence layers are transformed by determining a *spread* and *midpoint*. Once a variable was determined to be ascending or descending, e.g. the target values are small or large, respectively, the *spread* and *midpoint* were tuned to create a layer with the best AUC value with associated mean, median and standard deviation. This approach provides information on the variability caused by random points and of feature sensitivity, whilst minimising the chance of a biased true negative sample set affecting the transformation. Note that the Proximity-to Granite in Z layer was generated using the Table of Contents (TOC) function from the ArcSDM 5 package.

A list of the final input variables and the optimised parameters used for the fuzzy membership functions is provided in Table 2; full results for all tested parameters are presented in the Supplementary Information (S1). It is clear that some input variables have a much higher AUC than others. Nykänen et al. (2017) suggest there is value in the inclusion of a variable even where AUC values are close to 0.5 (random correlation) because it may provide mutually beneficial information to a subsequent combination of variables later in the analysis.

2.2.3. Fuzzy operator combinations

Following fuzzy membership transformation, some associated input variables were combined into single layers to not only enhance the feature, but to also assist with dimensionality reduction. Elements with geochemical analyses in the form of both soil and stream-sediment data were integrated into single variables to represent the overall anomalies for that element (Figure 7). The same approach was also applied to geochemical ratios, with the exception of K/(Zr/Eu), as this was only created for stream-sediment data due to the absence of soil REE analyses for the soil data. A visual inspection of the data was conducted prior to integration to ensure that the values for each variable were comparable.

The *fuzzyOR* operator is considered to be the best tool to combine two elements or ratios into a single input variable to maximise potential anomalies (Bonham-Carter, 1994), as well as reduce dimensionality in the model, and it is used here to maximise indications of geochemical anomalies from both datasets. These were subsequently reassessed using the ROC curve tool and new AUC values were calculated (Table 3). For W, Sn, As and Na, this results in a synergistic effect where the AUC is greater than both AUC values for the individual datasets. For Bi, Sb, Rb, Cs, K/Cs, K/Rb and Ti/Sn, the AUC values fall between the lower and upper values derived for the original datasets.

2.3. Machine learning for prospectivity modelling

Various MLAs are available for prospectivity modelling, however, it is the Random Forest algorithm that has consistently proven to be highly effective in comparison to Support Vector Machines and Artificial Neural Networks (Carranza and Laborte, 2015a, 2015b; Rodriguez-Galiano et al., 2015; Carranza and Laborte, 2016; Sun et al., 2019). For this reason, two Random Forest models are presented for prospectivity modelling, using: (i) standardized input variables with no transformation; (ii) variables transformed using the guided fuzzy set theory approach of Nykänen et al. (2015, 2017).

An advantage of using MLAs for mineral prospectivity modelling is the evaluation metrics available for each algorithm. Many classification methods allow the probability of a pixel being correctly classified, the class probabilities, to be interrogated. For mineral prospectivity modelling, class probabilities are often presented as the final result, but these can be further manipulated through model variance (Kohavi and Wolpert, 1996; Cracknell and Reading, 2013). Model variance was implemented as part of lithological mapping by Cracknell and Reading (2013) in the Broken Hill area of New South Wales, Australia where higher variance was an indicator for the presence of fault zones and was termed “the upside of uncertainty”. This was further investigated using information entropy (Kuhn et al., 2018).

There is often a predilection for distilling model performance to a single accuracy metric. However, this is not ideal, especially with spatial data where some aspects of the model may be well-constrained and other components highly suspect. By incorporating a spatial assessment of model reliability into the evaluation process, the user can enhance the analysis and mitigate the potential limitations of a single accuracy metric. To this end, we develop a new Confidence Metric, founded on model variance, to evaluate the model and further investigate the extent of prospective areas before giving some quantification of the depth to potential targets.

2.3.1. Random Forest modelling

Prospectivity modelling was performed using the *R* statistical computing language (R Core Team, 2019). A binary MLA classification model was created where two classes were used (unfavourable and favourable) to determine a simple class probability model. The Random Forest models were implemented using the *caret* (Kuhn et al., 2019), *raster* (Hijmans, 2019) and *rgdal* (Bivand et al., 2019) packages. A full description of the *R* workflow is presented in the Supplementary Information (S2).

The Random Forest method is an ensemble decision tree machine learning algorithm first described by Breiman (2001). The method has become increasingly popular in geoscience and has been used in prospectivity modelling for a range of ore deposit types (e.g. O’Brien et al., 2014; Harris et al., 2015; Carranza & Laborte 2015a, 2015b, 2016; Gao et al., 2017; Hariharan et al., 2017; Li et al., 2019; Sun et al., 2019). The approach combines multiple binary-split trees which limits overfitting that can occur through multi-split trees (Hastie et al., 2009). The Random Forest algorithm, illustrated in Figure 8, utilises multiple decision trees (the forest) which attempt to split a random selection of input variables. The number of random variables is controlled by the user-defined *mtry* value that can be determined using a random or grid search to find the best value, or, as in this study, by calculating the square root of the number of input variables (Breiman, 2001; Gislason et al., 2006; Belgiu and Drăguţ, 2016). A further parameter must be set, *ntree*, which dictates the number of binary trees in the forest and controls the reproducibility of the results. Based on a review by Belgiu and Drăguţ (2016), *ntree* is commonly set to 500 for most classification problems using remote sensing data. Carranza and Laborte (2015b) increased *ntree* to 20 000 in order to achieve stable predictions and lower the prediction error for a training set of 12 samples. Given the comparably small training sample size in this study (23 training samples and 11 validation samples), the *ntree* value of 20 000 was also adopted here.

A total of 28 input variables are included in the standardised model (see Table 2), while 17 variables are included in the fuzzy-transformed model following combination of duplicate geochemical elements using the *fuzzyOR* operator (see Table 3). All fuzzy-transformed and combined data were included in the modelling process despite the potentially low relevance of Sb. The inclusion of Sb is due to its minor positive correlation with known deposits that may still contribute some relevant information.

The models were evaluated using the ROC curve tool to derive the mean and median AUC values and associated standard deviation for each model using the true positive validation subset and the 10 randomly reselected true negative validation subsets (described in Section 2.1.4).

2.3.2. The Confidence Metric

Spatial evaluation of the model can be undertaken by calculating the model variance (Equation 3) of the class probabilities to derive an uncertainty value (Kohavi and Wolpert, 1996). This approach was implemented by Cracknell and Reading (2013) to show areas where the classification is less reliable. In this study, model variance is exploited to determine whether favourable targets are truly robust in the mineral prospectivity model. By combining model variance and the class probabilities into the new Confidence Metric using Equation 4, exploration targets can be refined to highlight the areas of highest confidence in the model.

$$model\ variance\ (v) = \frac{1-\sum p_c^2}{1-\sum\left(\frac{1}{c}\right)} \qquad (3)$$

Where p_c is the class probability for each class per pixel and c is the total number of classes.

$$confidence\ (p_{conf}) = \frac{(p_c-v)_i - min(p_c-v)}{max(p_c-v)-min(p_c-v)} \qquad (4)$$

Where i indicates a per pixel subtraction.

By subtracting the model variance, the values of pixels with high uncertainty are reduced accordingly, leaving only the most reliable areas with high class probabilities. In some cases, this can reduce the value to less than zero and, for the purposes of comparison, Equation 4 normalises the output to a range of 0 to 1.

2.3.3. Areal evaluation

The spatial distribution of the prospectivity is quantitatively evaluated using areal analysis. Total areal extents are calculated for each level of prospectivity, unfavourable through to highly favourable, as a sum of the area for each level and as a percentage of total area of the model. The analysis provides a quantitative assessment of the spatial distribution of the class probabilities for each model and the associated confidence. The proportion of pixels at each prospectivity level are compared to determine which model is better at discriminating prospective areas.

2.3.4. Depth evaluation

The rich mining history of SW England means that there is an extensive repository of data but the quality of digital records is highly variable. Legacy mining data is available through the British Geological Survey from the Mineral Exploration and Investigation Grants Act (MEIGA) records and published works such as Dines (1956). These resources are used to further evaluate the depth at which potential targets may occur.

3. Results and Discussion

The results of the MLA modelling using both feature extraction methods are presented below. These are assessed, based on the AUC values from ROC curve analysis, and further evaluated using the Confidence Metric, areal analysis and legacy mining data. These evaluation techniques aim, respectively, to generate robust targets, compare the spatial attributes of the model and to give an indication of whether targets are likely to reside at surface or at depth.

3.1. Tungsten prospectivity modelling results

The results of the modelling using standard and fuzzy input variables are presented in Figure 9 and Figure 10. Each figure comprises the binary classification of all prospective areas, the class probability for a cell being classified as prospective and the confidence map derived using the Equation 4.

The class map for the prospectivity model shows broad areas of prospective areas for tungsten mineralisation due to the binary classification. The Random Forest class probability map is therefore more useful as it signifies the likelihood that a location is prospective. For Figure 9 and Figure 10, the data have been categorised to show only values greater than 0.5 in colour, this is to indicate that anything below this value would have been classified as unfavourable in the binary classification.

The class probability map for the standardised variables (Figure 9) shows a good correlation with known tungsten occurrences. Areas of high favourability are constrained to areas of known deposits marked as W-Y in Figure 9b, which include the Camborne-Redruth district, the St Austell district and the east Bodmin-Kit Hill area, respectively. However, no highly favourable areas are identified that were not previously known and only limited areas have been identified as favourable.

Figure 10 shows the class probability map for the fuzzy-transformed variables that identifies highly favourable areas over known tungsten occurrences, similar to those in Figure 9b (W-Y), including the Cligga Head area (Z). Additional areas include the Breage district (A), the southern margin of the Bodmin Granite (B) and some discrete targets along the eastern margin of the Dartmoor Granite (C) which are new prospects. The map also shows broader areas of favourable prospectivity away from main targets.

The ROC curve tool was used to validate these models and generate a quantitative measure of accuracy for the binary classification. A summary of the validation results from the ROC curve analysis is included in Table 4. The average AUC values for both class probability models are very high and not significantly different. It is unsurprising that both models have such similar AUC values due to sharing the same initial evidence layers and the invariance of the Random Forest algorithm to changes in scale (but not *midpoint* and *spread*) imparted by the fuzzy membership transformation. Furthermore, the similarity in AUC values underlines that the use of training samples with the ROC curve tool during feature extraction has not overly biased the model. However, the reduction in dimensionality from 28 to 17 input variables in the fuzzy-transformed model appears to have provided no significant improvements to the modelling process.

Despite the minimal difference in AUC values, the lack of new highly prospective targets in the standardised variable model is disappointing. Nevertheless, the greater number of new targets in the fuzzy-transformed model indicates that the incorporation of user-knowledge, through fuzzy-transformed variables during feature extraction, has refined target identification within a data-driven Random Forest modelling approach.

3.2. Target confidence

The use of model variance (Equation 3) and manipulation of this metric into a measure of target confidence is novel and has demonstrated significant value for evaluating the prospectivity models. The confidence maps for each model shown in Figure 9c and Figure 10c reveal highly favourable and favourable areas that are not only significantly refined in area, but define more reliable targets. Any area shown to be >0.5 in terms of confidence should be compared to the class probability map to determine its favourability and those areas with high class probabilities and high confidence are likely to be robust. Therefore, the confidence map helps to elucidate highly favourable and favourable areas and interpret reliable exploration targets. Furthermore, it gives a greater understanding where the model has performed best and goes beyond the use of single accuracy metrics which can be misleading.

3.3. Model comparison from areal evaluation

The two Random Forest models presented here can also be assessed to determine the prospectivity by area. Models for class probability and confidence have been assessed in terms of area in Table 5. These show the total area and normalised area for each class shown in Figure 9 and Figure 10.

The total areas are similar for each model and small discrepancies are due to rounding errors. The class probability model for standardised variables shows a greater proportion of the study area having some degree of prospectivity (>0.5). In contrast, the class probability model for the fuzzy-transformed variables shows a smaller proportion of the study area to be prospective (>0.5) but the areas that are identified have a greater degree of prospectivity. The most prospective areas (>0.8) accounts for 3.7% of the total area compared to 2% when using standardised variables. Similarly, the confidence model for both methods has been assessed. If a value of >0.5 is taken as a reasonable confidence level, 3.2% and 5.2% of the models for standard variables and fuzzy-transformed variables, respectively, can be considered to be robust.

The results from this analysis would infer that the fuzzy-transformed variables give an overall greater confidence when generating exploration targets compared to the standardised variables. By revisiting Table 3, it can be seen that the combination of W, Sn, As and Na has a mutually beneficial effect on the AUC values compared to the prior values for the individual soil and stream-sediment geochemical layers. These mutually beneficial combinations are likely to improve the MLA model and enhance target delineation.

3.4. Evaluation using legacy mining data

New targets were identified from the Random Forest model using fuzzy-transformed variables. These include the Breage district, the southern margin of the Bodmin Granite and a series of discrete targets along the eastern margin of the Dartmoor Granite labelled A, B and C, respectively (Figure 10b). These are further highlighted in Figure 11 alongside additional legacy data to further assess the fuzzy-transformed variable model.

In the Breage district (Figure 11a), historic mining records indicate tungsten mineralisation was intersected at depth at Prospidnick on the SW margin of the Carnmenellis Granite and at Great Wheal Fortune on the eastern margin of the Tregonning-Godolphin Granite (Dines, 1956). Furthermore, a borehole was drilled in the area to 214.14 m that intersected the granite contact at 173.6 m where the upper 20 m showed greisen textures and reported tungsten and tin mineralization in assay (Ball et al., 1984). Note, this occurrence is missing from the BGS GeoIndex (2018) data.

Studies conducted under MEIGA are not recorded in the BGS GeoIndex (2018). The mineralisation along the southern margin of the Bodmin Granite (Figure 11b) was investigated by Consolidated Gold Fields Ltd as part of regional tungsten exploration study funded by MEIGA in 1972. Tungsten and tin anomalies were identified in streams and follow-up soil sampling was also conducted. A drilling campaign along the southern margin of the granite was conducted which intersected tungsten mineralisation but grades and tonnages were deemed uneconomic at the time.

Targets identified in Figure 11c along the eastern margin of the Dartmoor Granite require further follow-up work. No records of tungsten have been found, however, four mines are inferred by Dines (1956) to become uneconomic with depth with respect to tin and it was suggested that other “uneconomic” metals may exist but are not described further. One of these mines exists outside of the surface crop of the granite and intersects the granite margin at approximately 90 m below surface.

The use of these additional resources helps validate the mineral prospectivity model. The reference to tungsten mineralisation found in old mines and former drilling projects suggests that some of these targets may be within a few hundred metres of surface. This further supports the model for identifying blind deposits and the inclusion of the proximity-to granite in Z evidence layer is likely to be important; high resolution gravity measurements may improve the analysis significantly.

Conclusions

Mineral prospectivity modelling has been conducted using a data-driven Random Forest MLA approach for tungsten in SW England. A particular focus has been put on feature extraction and the use of initial variables that were standardised to zero mean and equal variance compared to those that were further processed using knowledge-driven fuzzy membership and fuzzy overlay functions.

The two models presented here have similar accuracies based on ROC curve analysis but show different spatial distributions of prospectivity in the region. The model that uses standardised variables only identifies areas of high prospectivity (>0.9) proximal to the training data. The second model, using fuzzy-transformed input variables, identifies three new highly prospective targets that were previously unidentified in the training data. The improvement in target generation is directly attributable to the use of knowledge-driven feature extraction techniques within a data-driven MLA framework.

These models are enhanced using model variance to derive a new Confidence Metric. The Confidence Metric is a simple calculation to infer where class probabilities are most robust. These are presented as a map that can be combined with the initial class probabilities to determine the most reliable targets. The approach results in spatially refined and robust mineral exploration targets that can allow for a more focussed follow-up field campaign.

The models have been further evaluated by an areal analysis showing that the fuzzy-transformed model is a better discriminator for prospective areas compared to the standardised variable model due to the mutually beneficial effect of combining geochemical layers such as W, Sn, As and Na during feature extraction. Also, the fuzzy-transformed model has greater confidence and generates a greater proportion of robust targets by area based on the Confidence Metric. By conducting model evaluation in this way, two models with the same statistical accuracy but different spatial distributions can be better understood. This study underlines how single accuracy metrics can be fallible when applied to spatial datasets.

Finally, the use of legacy mining data further reinforces the strength of the model where all three new target areas have potential economic mineralisation either through direct sampling or inferred from mine descriptions. Further, the legacy mining data suggests that the targets generated may be within 300 m of surface. This would indicate the “Proximity-to granite in Z” evidence layer derived from regional gravity data is valuable and that new discoveries of tungsten mineralisation in SW England may be enhanced by a new high-resolution gravity survey.

Acknowledgements

We thank two anonymous reviewers for their comments that improved the final manuscript. CMY would like to thank Dr Charles Moon for his comments on an early draft and Dr Anya Reading for constructive conversations during CMY's viva voce examination. The work was completed as part of doctoral research by CMY funded by the British Geological Survey (S267) and the Natural Environment Research Council (NERC) GW4+ Doctoral Training Partnership (NE/L002434/1) in collaboration with the Geological Survey of Finland (GTK). Data used in this study are available from the Tellus South West project website (www.tellusgb.ac.uk). PL publishes with the permission of the Executive Director, British Geological Survey (UKRI).

References

Abedi, M., Norouzi, G.H., Torabi, S.A., 2013. Clustering of mineral prospectivity area as an unsupervised classification approach to explore copper deposit. Arabian Journal of Geosciences 6, 3601–3613. https://doi.org/10.1007/s12517-012-0615-5

Agterberg, F.P., Bonham-Carter, G.F., 2005. Measuring the performance of mineral-potential maps. Natural Resources Research 14, 1–17. https://doi.org/10.1007/s11053-005-4674-0

Alexander, A.C., Shail, R.K., 1996. Late- to post-Variscan structures on the coast between Penzance and Pentewan, South Cornwall. Proceedings of the Ussher Society 9, 72–78.

Alexander, A.C., Shail, R.K., 1995. Late Variscan structures on the coast between Perranporth and St. Ives, Cornwall. Proceedings of the Ussher Society 8, 398–404.

An, P., Moon, W.M., Rencz, A., 1991. Application of fuzzy set theory to integrated mineral exploration. Canadian Jounral of Exploration 27, 1–11. https://doi.org/10.3997/2214-4609.201410970

Andrews, M.J., Ball, T.K., Fuge, R., Nicholson, R.A., Peachey, D., 1987. Trace elements in soils around the Hemerdon tungsten deposit, Devon; implications for exploration. Proceedings of the Ussher Society 6, 536–541.

Bahiru, E.A., Woldai, T., 2016. Integrated geological mapping approach and gold mineralization in Buhweju area, Uganda. Ore Geology Reviews 72, 777–793. https://doi.org/10.1016/j.oregeorev.2015.09.010

Ball, T.K., Basham, I.R., Charoy, B., 1984. Petrogenesis of the Bosworgey granitic cusp in the SW England tin province and its implications for ore mineral genesis. Mineralium Deposita 19, 70–77. https://doi.org/10.1007/BF00206599

Ball, T.K., Fortey, N.J., Beer, K.E., 2002. Aspects of the lithogeochemistry of arsenic, antimony and bismuth in South West England. Geoscience in South-West England 10, 352–357.

Ball, T.K., Fortey, N.J., Beer, K.E., 1998. Alkali metasomatism from Cornubian granite cupolas. Geoscience in South-West England 9, 171–177.

Beamish, D., Howard, A., Ward, E.K., White, J., Young, M.E., 2014. Tellus South West airborne geophysical data.

Beer, K.E., Ball, T.K., Bennett, M.J., 1986. Mineral investigations near Bodmin, Cornwall. Part 5 - The Castle-an-Dinas wolfram lode. Mineral Reconnaissance Programme Report, British Geological Survey. No.82; Mineral Reconnaissance Programme Report, British Geological Survey. No.82.

Beer, K.E., Burley, A.J., Tombs, J.M., 1975. The concealed granite roof in south-west Cornwall. Mineral Reconnaissance Programme Report, Institute of Geological Sciences, No.1 [Unpublished]; Mineral Reconnaissance Programme Report, Institute of Geological Sciences, No.1 [Unpublished].

Belgiu, M., Drăguţ, L., 2016. Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing 114, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Bennett, M.J., Beer, K.E., Jones, R.C., Turton, K., Rollin, K.E., Tombs, J.M.C., Patrick, D.J., 1981. Mineral investigations near Bodmin, Cornwall. Part 3 - The Mulberry and Wheal Prosper area. Mineral Reconnaissance Programme Report. Institute of Geological Sciences, No. 48; Mineral Reconnaissance Programme Report. Institute of Geological Sciences, No. 48.

BGS GeoIndex, 2018. Mineral Occurrence Database. https://www.bgs.ac.uk/mineralsuk/data/mineocc.html

Bivand, R., Keitt, T., Rowlingson, B., 2019. rgdal: Bindings for the ’Geospatial’ Data Abstraction Library. https://cran.r-project.org/package=rgdal

Bonham-Carter, G.F., 1994. Geographic information systems for geoscientists: modelling with GIS, First. ed. Elsevier Science Ltd, Kidlington, UK.

Breiman, L., 2001. Random forests. Machine Learning 45, 5–32. https://doi.org/10.1023/A:1010933404324

British Geological Survey, 2016. G-BASE for Southwest England.

Cameron, J., 1951. The Geology of the Hemerdon wolfram mine, Devon. Transactions of the Institution of Mining and Metallurgy 6L, 1–14.

Campbell, A.R. and Panter, K.S., 1990. Comparison of fluid inclusions in coexisting (cogenetic?) wolframite, cassiterite, and quartz from St. Michael's Mount and Cligga Head, Cornwall, England. Geochimica et Cosmochimica Acta, 54, 673-681

Camps-Valls, G., Bandos Marsheva, T.V., Zhou, D., 2007. Semi-supervised graph-based hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing 45, 3044–3054. https://doi.org/10.1109/TGRS.2007.895416

Carranza, E. J. M. (2009). Objective selection of suitable unit cell size in data-driven modeling of mineral prospectivity. Computers and Geosciences, 35(10), 2032–2046. https://doi.org/10.1016/j.cageo.2009.02.008

Carranza, E.J.M., 2010. Mapping of anomalies in continuous and discrete fields of stream sediment geochemical landscapes. Geochemistry: Exploration, Environment, Analysis 10, 171–187. https://doi.org/10.1144/1467-7873/09-223

Carranza, E.J.M., Laborte, A.G., 2016. Data-Driven Predictive Modeling of Mineral Prospectivity Using Random Forests: A Case Study in Catanduanes Island (Philippines). Natural Resources Research 25, 35–50. https://doi.org/10.1007/s11053-015-9268-x

Carranza, E.J.M., Laborte, A.G., 2015a. Data-driven predictive mapping of gold prospectivity, Baguio district, Philippines: Application of Random Forests algorithm. Ore Geology Reviews 71, 777–787. https://doi.org/10.1016/j.oregeorev.2014.08.010

Carranza, E.J.M., Laborte, A.G., 2015b. Random forest predictive modeling of mineral prospectivity with small number of prospects and data with missing values in Abra (Philippines). Computers and Geosciences 74, 60–70. https://doi.org/10.1016/j.cageo.2014.10.004

Chen, Y., Clark, A. H., Farrar, E., Wasteneys, H. A. H. P., Hodgson, M. J., & Bromley, A. V. (1993). Diachronous and independent histories of plutonism and mineralization in the Cornubian Batholith, southwest England. Journal of the Geological Society, London, 150, 1183–1191

Coward, M.P., Smallwood, S., 1984. An interpretation of the Variscan tectonics of SW Britain. Geological Society, London, Special Publications 14, 89–102. https://doi.org/10.1144/GSL.SP.1984.014.01.08

Cracknell, M.J., Reading, A.M., 2015. Spatial-Contextual Supervised Classifiers Explored: A Challenging Example of Lithostratigraphy Classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 8, 1371–1384. https://doi.org/10.1109/JSTARS.2014.2382760

Cracknell, M.J., Reading, A.M., 2014. Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. Computers and Geosciences 63, 22–33. https://doi.org/10.1016/j.cageo.2013.10.008

Cracknell, M.J., Reading, A.M., 2013. The upside of uncertainty: Identification of lithology contact zones from airborne geophysics and satellite data using random forests and support vector machines. Geophysics 78, 113––126. https://doi.org/10.1190/GEO2012-0411.1

Dearman, W.R., 1970. Some aspects of the tectonic evolution of South-West England. Proceedings of the Geologists' Association 81, 483–491. https://doi.org/10.1016/S0016-7878(70)80009-8

Dearman, W.R., 1963. Wrench-faulting in Cornwall and south Devon. Proceedings of the Geologists' Association 74, 265–287. https://doi.org/10.1016/S0016-7878(63)80023-1

Dines, H.G., 1956. The Metalliferous mining region of south-west England. Economic Memoirs of the Geological Survey of Great Britain.

Dominy, S.C., Camm, G.S., Bussell, M.A., Scrivener, R.C., Halls, C., 1995. A review of tin stockwork mineralization in the south-west England orefield. Proceedings of the Ussher Society 8, 368–373.

Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recognition Letters 27, 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Ferraccioli, F., Gerard, F., Robinson, C., Jordan, T., Biszczuk, M., Ireland, L., Beasley, M., Vidamour, A., Barker, A., Arnold, R., Dinn, M., Fox, A., Howard, A., 2014. LiDAR based Digital Terrain Model (DTM) data for South West England. https://doi.org/10.5285/e2a742df-3772-481a-97d6-0de5133f4812

Gao, Y., Zhang, Z., Xiong, Y., & Zuo, R. (2016). Mapping mineral prospectivity for Cu polymetallic mineralization in southwest Fujian Province, China. Ore Geology Reviews, 75, 16–28. https://doi.org/10.1016/j.oregeorev.2015.12.005

Gislason, P.O., Benediktsson, J.A., Sveinsson, J.R., 2006. Random forests for land cover classification. Pattern Recognition Letters 27, 294–300. https://doi.org/10.1016/j.patrec.2005.08.011

GTK, 2019. ArcSDM. https://github.com/gtkfi/ArcSDM

Hall, A., 1971. Greisenisation in the granite of Cligga Head, Cornwall. Proceedings of the Geologists’ Association 82, 209–230.

Hariharan, S., Tirodkar, S., Porwal, A., Bhattacharya, A., & Joly, A. (2017). Random Forest-Based Prospectivity Modelling of Greenfield Terrains Using Sparse Deposit Data: An Example from the Tanami Region, Western Australia. Natural Resources Research, 26(4), 489–507. https://doi.org/10.1007/s11053-017-9335-6

Harris, J.R., Grunsky, E., Behnia, P., Corrigan, D., 2015. Data- and knowledge-driven mineral prospectivity maps for Canada’s North. Ore Geology Reviews 71, 788–803.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning, Springer series in statistics. Springer New York, New York, NY. https://doi.org/10.1007/978-0-387-84858-7

Henery, R.J., 1994a. Classification, in: Michie, D., Spiegelhalter, D.J., Taylor, C.C. (Eds.), Machine Learning, Neural and Statistical Classification. Ellis Horwood, New York, pp. 6–16.

Henery, R.J., 1994b. Methods for Comparison, in: Michie, D., Spiegelhalter, D.J., Taylor, C.C. (Eds.), Machine Learning, Neural and Statistical Classification. Ellis Horwood, New York, pp. 107–124.

Hijmans, R.J., 2019. raster: Geographic Data Analysis and Modeling. https://cran.r-project.org/package=raster

Hobson, D.M., Sanderson, D.J., 1983. Variscan Deformation in Southwest England, in: Hancock, P.L. (Ed.), The Variscan Fold Belt in the British Isles. Adam Hilger Ltd, Bristol, pp. 108–129.

Hosking, K.F.G., Trounson, J.H., 1959. The mineral potential of Cornwall, in: Non-Ferrous Mining in Great Britain and Ireland. Institution of Mining; Metallurgy Symposium, London, pp. 335–369.

Hughes, G.F., 1968. On the Mean Accuracy of Statistical Pattern Recognizers. IEEE Transactions on Information Theory 14, 55–63. https://doi.org/10.1109/TIT.1968.1054102

IAEA, 2003. Guidelines for radioelement mapping using gamma ray spectrometry data, IAEA-TECDOC-1363. International Atomic Energy Agency, Vienna, Austria.

Jackson, N.J., Willis-Richards, J., Manning, D.A.C., Sams, M.S., 1989. Evolution of the Cornubian ore field, Southwest England; Part II, Mineral deposits and ore-forming processes. Economic Geology 84, 1101–1133.

James, J.M., Moore, J.M., 1985. Multi-seasonal imagery studies for geological mapping and prospecting in cultivated terrain of S.W. England, in: Fourth Thematic Conference: "Remote Sensing for Exploration Geology", San Francisco, California, April 1-4, 1985. San Francisco, California, pp. 475–484.

Kohavi, R., Wolpert, D.H., 1996. Bias plus variance decomposition for zero-one loss functions, in: Proceedings of the 13th International Conference on Machine Learning (Icml96), Bari, Italy. pp. 275–283.

Kreuzer, O.P., Markwitz, V., Porwal, A.K., McCuaig, T.C., 2010. A continent-wide study of Australia’s uranium potential. Part I: GIS-assisted manual prospectivity analysis. Ore Geology Reviews 38, 334–366. https://doi.org/10.1016/j.oregeorev.2010.08.003

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., the R Core Team, Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C., Hunt., T., 2019. caret: Classification and Regression Training. https://cran.r-project.org/package=caret

Kuhn, S., Cracknell, M.J., Reading, A.M., 2018. Lithologic mapping using Random Forests applied to geophysical and remote-sensing data: A demonstration study from the Eastern Goldfields of Australia. Geophysics 83, B183–B193. https://doi.org/10.1190/geo2017-0590.1

Leveridge, B.E., Hartley, A.J., 2006. The Varisan Orogeny: the development and deformation of Devonian/Carboniferous basins in SW England and South Wales, in: Brenchley, P.J., Rawson, P.F. (Eds.), The Geology of England and Wales. The Geological Society, London, pp. 225–256.

Manning, D.A.C., Hill, P.I., 1990. The petrogenetic and metallogenetic significance of topaz granite from the southwest England orefield. Geological Society of America Special Paper 246, 51–69.

Mellor, A., Boukir, S., Haywood, A., & Jones, S. (2015). Exploring issues of training data imbalance and mislabelling on random forest performance for large area land cover classification using the ensemble margin. ISPRS Journal of Photogrammetry and Remote Sensing, 105, 155–168. https://doi.org/10.1016/j.isprsjprs.2015.03.014

Moore, J.M., Camm, S., 1982. Interactive enhancement of Landsat Imagery for structural mapping in tin-tungsten prospecting: a case history of the S.W. England Orefield (U.K.), in:

International Symposium on Remote Sensing of Environment, Second Thematic Conference, Remote Sensing for Exploration Geology. Fort Worth, Texas, December 6 - 10, 1982, pp. 727–740.

Moore, J.M., Jackson, N., 1977. Structure and mineralization in the Cligga granite stock, Cornwall. Journal of the Geological Society, London 133, 467–480. https://doi.org/10.1144/gsjgs.133.5.0467

Moscati, R. J., & Neymark, L. A. (2020). U–Pb geochronology of tin deposits associated with the Cornubian Batholith of southwest England: Direct dating of cassiterite by in situ LA-ICPMS. Mineralium Deposita, 55(1), 1–20. https://doi.org/10.1007/s00126-019-00870-y

Müller, A., Seltmann, R., Halls, C., Siebel, W., Dulski, P., Jeffries, T., Spratt, J., Kronz, A., 2006. The magmatic evolution of the Land's End pluton, Cornwall, and associated pre-enrichment of metals. Ore Geology Reviews 28, 329–367. https://doi.org/10.1016/j.oregeorev.2005.05.002

Newall, P.S., 1994. An integrated geochemical approach to investigate the concealed mineralization at the Redmoor Sn/W sheeted vein deposit, east Cornwall, England. Journal of Southeast Asian Earth Sciences 10, 109–130. https://doi.org/10.1016/0743-9547(94)90013-2

Newall, P.S., Newall, G.C., 1989. Use of lithogeochemistry as an exploration tool at Redmoor sheeted-vein complex, east Cornwall, southwest England. Transactions of the Institution of Mining and Metallurgy 98, B162–B174.

Nykänen, V., 2008. Radial Basis Functional Link Nets Used as a Prospectivity Mapping Tool for Orogenic Gold Deposits Within the Central Lapland Greenstone Belt, Northern Fennoscandian Shield. Natural Resources Research 17, 29–48. https://doi.org/10.1007/s11053-008-9062-0

Nykänen, V., Lahti, I., Niiranen, T., Korhonen, K., 2015. Receiver operating characteristics (ROC) as validation tool for prospectivity models - A magmatic Ni-Cu case study from the Central Lapland Greenstone Belt, Northern Finland. Ore Geology Reviews 71, 853–860. https://doi.org/10.1016/j.oregeorev.2014.09.007

Nykänen, V., Niiranen, T., Molnhár, F., Lahti, I., Korhonen, K., Cook, N., Skyttä, P., 2017. Optimizing a Knowledge-driven Prospectivity Model for Gold Deposits Within Peräpohja Belt, Northern Finland. Natural Resources Research 26, 571–584. https://doi.org/10.1007/s11053-016-9321-4

O'Brien, J. J., Spry, P. G., Nettleton, D., Xu, R., & Teale, G. S. (2014). Using Random Forests to distinguish gahnite compositions as an exploration guide to Broken Hill-type Pb-Zn-Ag deposits in the Broken Hill domain, Australia. Journal of Geochemical Exploration, 149, 74–86. https://doi.org/10.1016/j.gexplo.2014.11.010

Rattey, P.R., Sanderson, D.J., 1984. The structure of SW Cornwall and its bearing on the emplacement of the Lizard Complex. Journal of the Geological Society, London 141, 87–95.

R Core Team, 2019. R: A Language and Environment for Statistical Computing. https://www.r-project.org

Robinson, G.R., Larkins, P.M., 2007. Probabilistic prediction models for aggregate quarry siting. Natural Resources Research 16, 135–146. https://doi.org/10.1007/s11053-007-9039-4

Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., Chica-Rivas, M., 2015. Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. Ore Geology Reviews 71, 804–818. https://doi.org/10.1016/j.oregeorev.2015.01.001

Sanderson, D.J., Dearman, W.R., 1973. Structural zones of the Variscan fold belt in SW England, their location and development. Journal of the Geological Society, London 129, 527–536. https://doi.org/10.1144/gsjgs.129.5.0527

Schetselaar, E., 2002. Petrogenetic interpretation from gamma-ray spectrometry and geological data: the Arch Lake zoned peraluminous granite intrusion, Western Canadian shield. Exploration Geophysics 33, 35–43. https://doi.org/10.1071/EG02035

Scrivener, R.C., 2006. Cornubian granites and mineralization of SW England, in: Brenchley, P.J., Rawson, P.F. (Eds.), The Geology of England and Wales. The Geological Society, London, pp. 257–268.

Shail, R.K., Alexander, A.C., 1997. Late Carboniferous to Triassic reactivation of Variscan basement in the western English Channel: evidence from onshore exposures in south Cornwall. Journal of the Geological Society, London 154, 163–168. https://doi.org/10.1144/gsjgs.154.1.0163

Shail, R.K., Leveridge, B.E., 2009. The Rhenohercynian passive margin of SW England: Development, inversion and extensional reactivation. Comptes Rendus Geoscience 341, 140–155.

Shail, R.K., Wilkinson, J.J., 1994. Late-to Post-Variscan extensional tectonics in south Cornwall. Proceedings of the Ussher Society 8, 262–270.

Shail, R., McFarlane, J., Hassall, L., Thiel, H., Stock, T., Smethurst, M., Tapster, S., Scrivener, R., Leveridge, B., Simons, B., 2017. The geological setting of the Hemerdon W–Sn deposit. Transactions of the Institutions of Mining and Metallurgy, Section B: Applied Earth Science 7453, 1. https://doi.org/10.1080/03717453.2017.1306292

Simons, B., Andersen, J.C., Shail, R.K., Jenner, F., 2017. Fractionation of Li, Be, Ga, Nb, Ta, In, Sn, Sb, W and Bi in the peraluminous Early Permian Variscan granites of the Cornubian Batholith: precursor processes to magmatic-hydrothermal mineralisation. Lithos 278-281, 491–512. https://doi.org/10.1016/j.lithos.2017.02.007

Simons, B., Shail, R.K., Andersen, J.C., 2016. The petrogenesis of the Early Permian Variscan granites of the Cornubian Batholith: Lower plate post-collisional peraluminous magmatism in the Rhenohercynian Zone of SW England. Lithos 260, 76–94. https://doi.org/10.1016/j.lithos.2016.05.010

Smith, M., Banks, D.A., Yardley, B.W. and Boyce, A., 1996. Fluid inclusion and stable isotope constraints on the genesis of the Cligga Head Sn-W deposit, SW England. European Journal of Mineralogy, pp.961-974

Sukumar, M., Venkatesan, N., Babu, C.N.K., 2014. A review of various lineament detection techniques for high resolution satellite images. International Journal of Advanced Research in Computer Science and Software Engineering 4, 72–78.

Sun, T., Chen, F., Zhong, L., Liu, W., Wang, Y., 2019. GIS-based mineral prospectivity mapping using machine learning methods: a case study from Tongling ore district, eastern China. Ore Geology Reviews 109, 26–49. https://doi.org/10.1016/j.oregeorev.2019.04.003

Tapster, S. R. and Bright, J. W. G.: High-precision ID-TIMS Cassiterite U-Pb systematics using a low-contamination hydrothermal decomposition: implications for LA-ICP-MS and ore deposit geochronology, Geochronology Discuss., https://doi.org/10.5194/gchron-2019-22, in review, 2020

Wang, C., Rao, J., Chen, J., Ouyang, Y., Qi, S., Li, Q., 2017a. Prospectivity mapping for “Zhuxi-type” copper-tungsten polymetallic deposits in the Jingdezhen region of Jiangxi Province, South China. Ore Geology Reviews 89, 1–14. https://doi.org/10.1016/j.oregeorev.2017.05.022

Wang, J., Zuo, R., Caers, J., 2017b. Discovering geochemical patterns by factor-based cluster analysis. Journal of Geochemical Exploration 181, 106–115. https://doi.org/10.1016/j.gexplo.2017.07.006

Wang, W., Cheng, Q., Zhang, S., Zhao, J., 2018. Anisotropic singularity: A novel way to characterize controlling effects of geological processes on mineralization. Journal of Geochemical Exploration 189, 32–41. https://doi.org/10.1016/j.gexplo.2017.07.019

Willis-Richards, J., Jackson, N.J., 1989. Evolution of the Cornubian Ore Field, Southwest England: Part I. Batholith Modeling and Ore Distribution. Economic Geology 84, 1078–1100.

Witten, I.H., Frank, E., Hall, M.A., Pal, C.J., 2017. Data Mining: Practical Machine Learning Tools and Techniques, Fourth. ed. Morgan Kaufmann (Elsevier).

Wragg, J., Cave, M., Hamilton, E., Lister, T., 2018. The Link between Soil Geochemistry in South-West England and Human Exposure to Soil Arsenic. Minerals 8, 570. https://doi.org/10.3390/min8120570

Yeomans, C.M., Middleton, M., Shail, R.K., Grebby, S., Lusty, P.A.J., 2019. Integrated Object-Based Image Analysis for semi-automated geological lineament detection in southwest England. Computers & Geosciences 123, 137–148 [Available Online November 2018]. https://doi.org/10.1016/j.cageo.2018.11.005

Zadeh, L.A., 1965. Fuzzy sets. Information and Control 8, 338–353. https://doi.org/10.1016/S0019-9958(65)90241-X

Zhao, J., Zuo, R., Chen, S., Kreuzer, O.P., 2015. Application of the tectono-geochemistry method to mineral prospectivity mapping: A case study of the Gaosong tin-polymetallic deposit, Gejiu district, SW China. Ore Geology Reviews 71, 719–734. https://doi.org/10.1016/j.oregeorev.2014.09.023

Figure Captions

Figure 1: Summary geology of SW England showing Devonian-Carboniferous sedimentary host rock in grey, granite outcrop in red and depth-to granite contours based on the granite surface model by Willis-Richards and Jackson (1989). Black lines represent regional lineaments derived by Yeomans et al. (2019) from Tellus South West airborne geophysical data.

Figure 2: Schematic illustrations of the kinematics and structures generated during Permian-Triassic extension (D3-D6). After Shail and Alexander (1997).

Figure 3: Schematic outline of extractive areas in SW England showing tin, copper and tungsten. Data from BGS GeoIndex (2018) are based on historic production values from known mines, deposit and prospect localities as well as reported mineral showings and panned concentrates. Important tungsten producers are labeled based on data from Dines (1956) and Jackson et al. (1989). Key mining areas are highlighted on the map: a = St Just, b = Camborne-Redruth, c = Breage, d = St Austell, e = Bodmin, f = Tamar Valley.

Figure 4: Mineral prospectivity modelling workflow for combining knowledge-based feature extraction into a data-driven machine learning approach to generate spatially refined and robust targets for mineral exploration.

Figure 5: Conceptual deposit model for tungsten mineralisation in SW England showing the main geological phenomena targeted by the prospectivity modelling.

Figure 6: Granite geochemistry showing the distribution of granite types based on the classification by Simons et al. (2016). The G2 granite is distinct having a low Zr/Eu ratio and high K, however, the G1a granite shows a similar signature.

Figure 7: (A) interpolated stream-sediment geochemical data for tungsten that have been transformed using the fuzzy membership function. (B) interpolated soil geochemical data for tungsten that have been transformed using the fuzzy membership function. (C) resulting tungsten geochemical data that have been combined using the fuzzyOR operator to emphasis key anomalies.

Figure 8: Schematic Random Forest diagram illustrating the interaction of decision trees in determining a classification value. Where randomly generated trees attempt to resolve the class value for a single instance through a majority vote system based on the leaf nodes (based on Belgiu & Drăguţ, 2016).

Figure 9: (A) Classification map (B) Class probability map and (C) confidence map for the standardised variables Random Forest prospectivity model. Classes show the two class scenario where 1 is unprospective and 2 is prospective. The class probability and confidence models are categorised to show 0.9 to 1 as highly favourable (red), 0.8 to 0.9 as favourable (amber), 0.65 to 0.8 as less favourable (turquoise), 0.5 to 0.65 as possibly favourable (blue) and <0.5 as unfavourable (grey).

Figure 10: (A) Classification map (B) Class probability map and (C) confidence map for the fuzzy-transformed variables Random Forest prospectivity model. Classes show the two class scenario where 1 is unprospective and 2 is prospective. The class probability and confidence models are categorised to show 0.9 to 1 as highly favourable (red), 0.8 to 0.9 as favourable (amber), 0.65 to 0.8 as less favourable (turquoise), 0.5 to 0.65 as possibly favourable (blue) and <0.5 as unfavourable (grey).

Figure 11: Key target locations based on the class probability map from the fuzzy-transformed variables model. The Breage district is shown in (A) where drilling projects and mining legacy data are shown to validate the targets. Targets around the Bodmin Granite are shown in (B) with new areas validated by a drilling report. The eastern margin of the Dartmoor Granite is shown in (C) where mining legacy data are proximal to favourable targets.

Table Captions

Table 1: Geochemical data included as evidence for tungsten mineralisation. The geochemistry are grouped into three phenomena describing the mineralisation, granite aureole and granite type.

Table 2: AUC values for evidence layers transformed using fuzzy membership functions. The AUC values are calculated from ten ROC curve analyses using randomly generated false occurrences.

Table 3: AUC values for combined geochemical elements and ratios, calculated from ten ROC curve analyses using randomly generated false occurrences. These are compared to the geochemical values for original datasets from soil and stream-sediment (SS) data. In some cases (W, Sn, As, Na) the combination is mutually beneficial.

Table 4: AUC values for each Random Forest™ prospectivity model. Calculated from ten ROC curve analyses using randomly generated false occurrences. The key parameters have been included for each model.

Table 5: Area assessment for both standardised and fuzzy-transformed models. The data have been calculated in a GIS to show the area accounted for by each class as a sum and a percentage for both the class probability (Prob) map and confidence (Conf) maps. Small discrepancies are attributed to rounding errors.

Figure 1

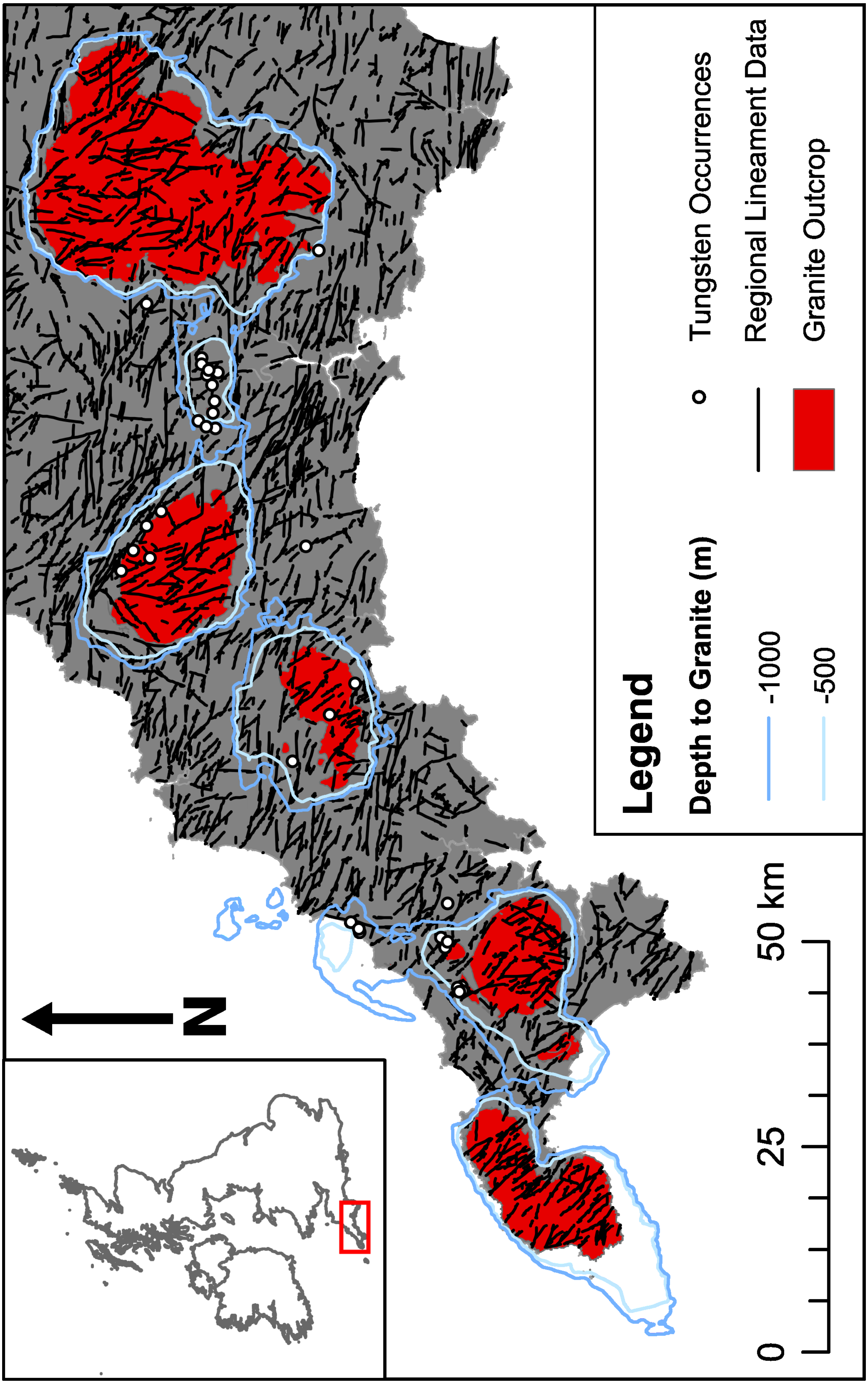

Legend
Depth to Granite (m)
-1000
-500
Tungsten Occurrences
Regional Lineament Data
Granite Outcrop
N
0
25
50 km

Figure 2

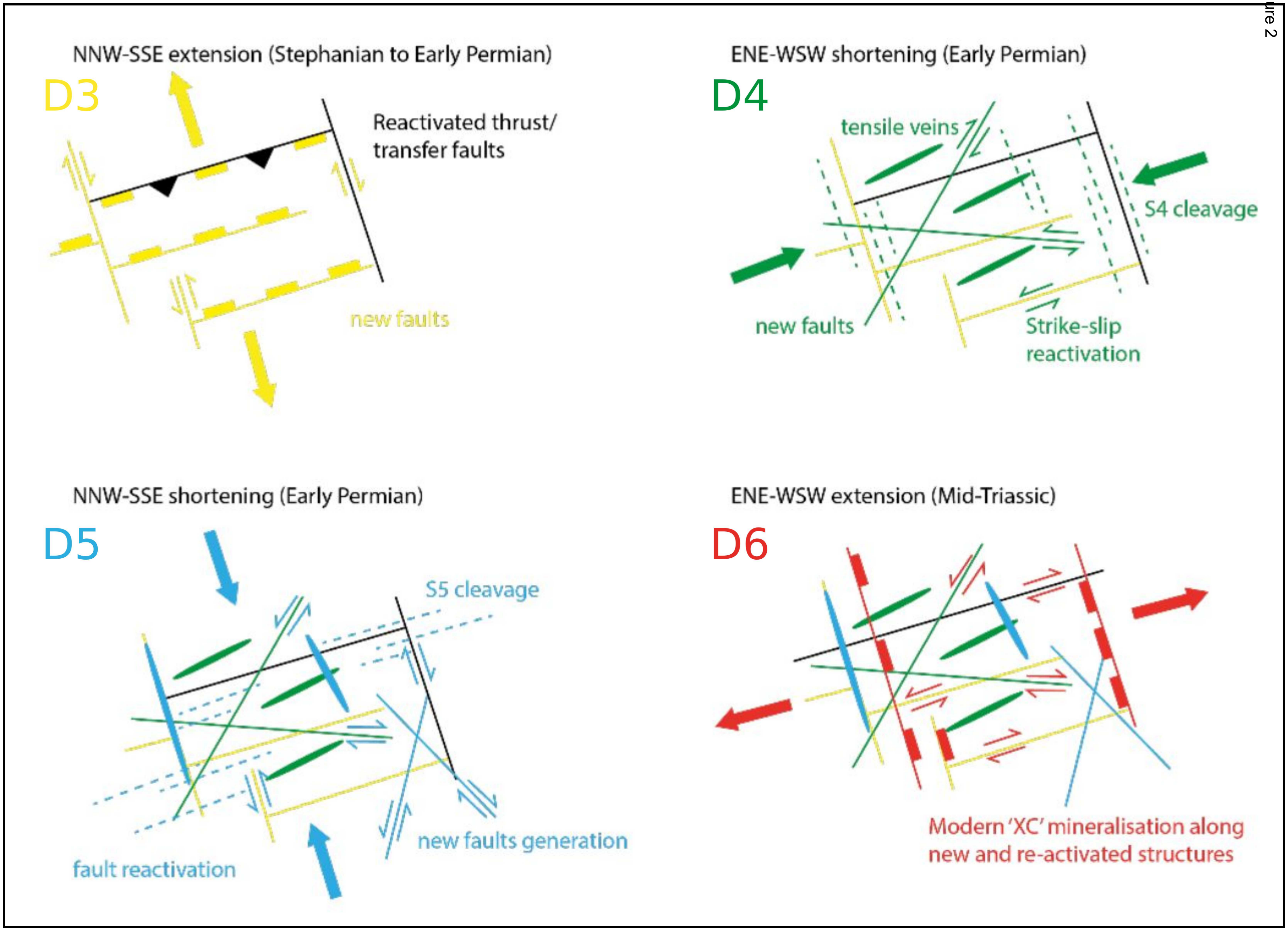
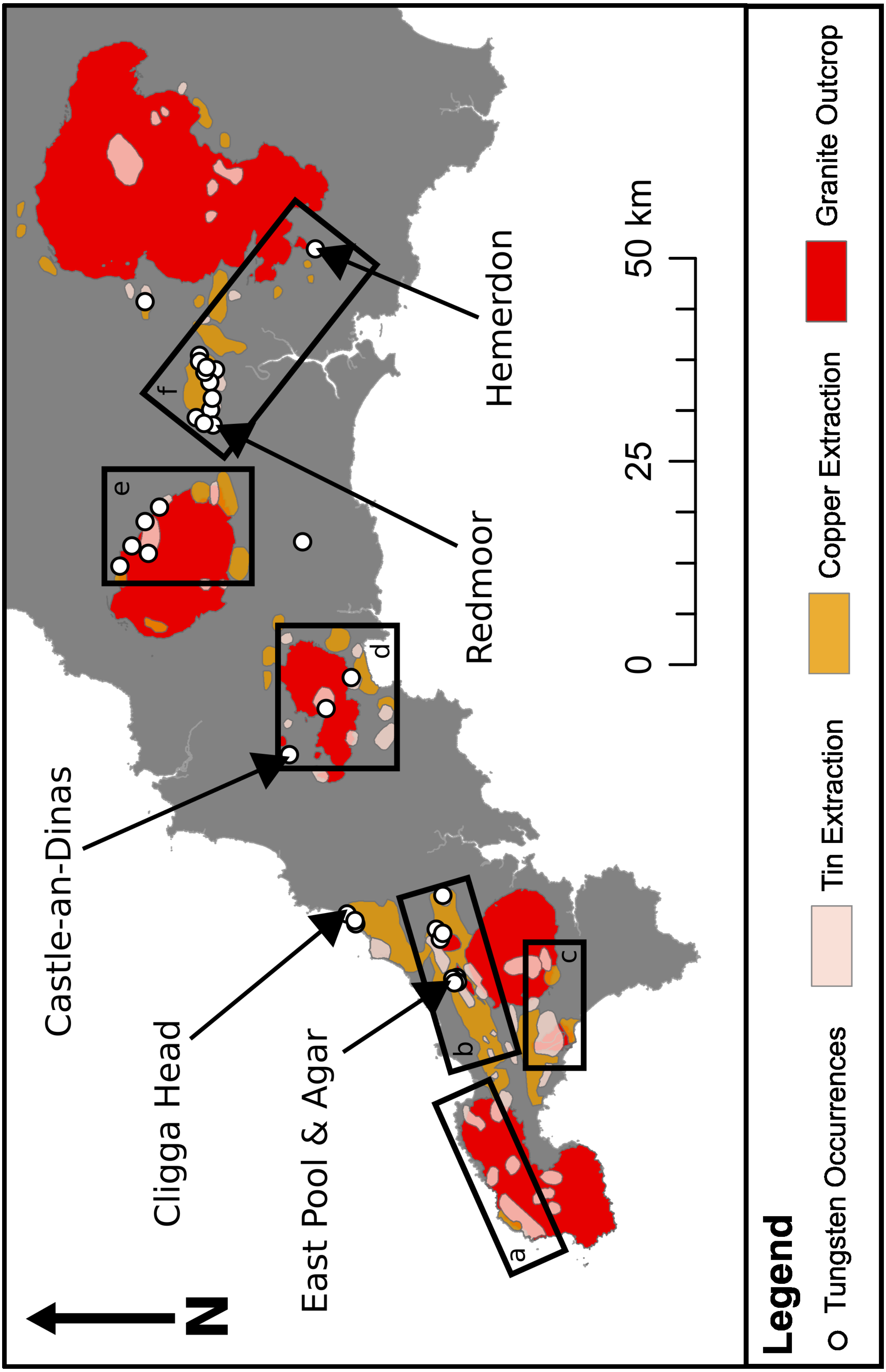
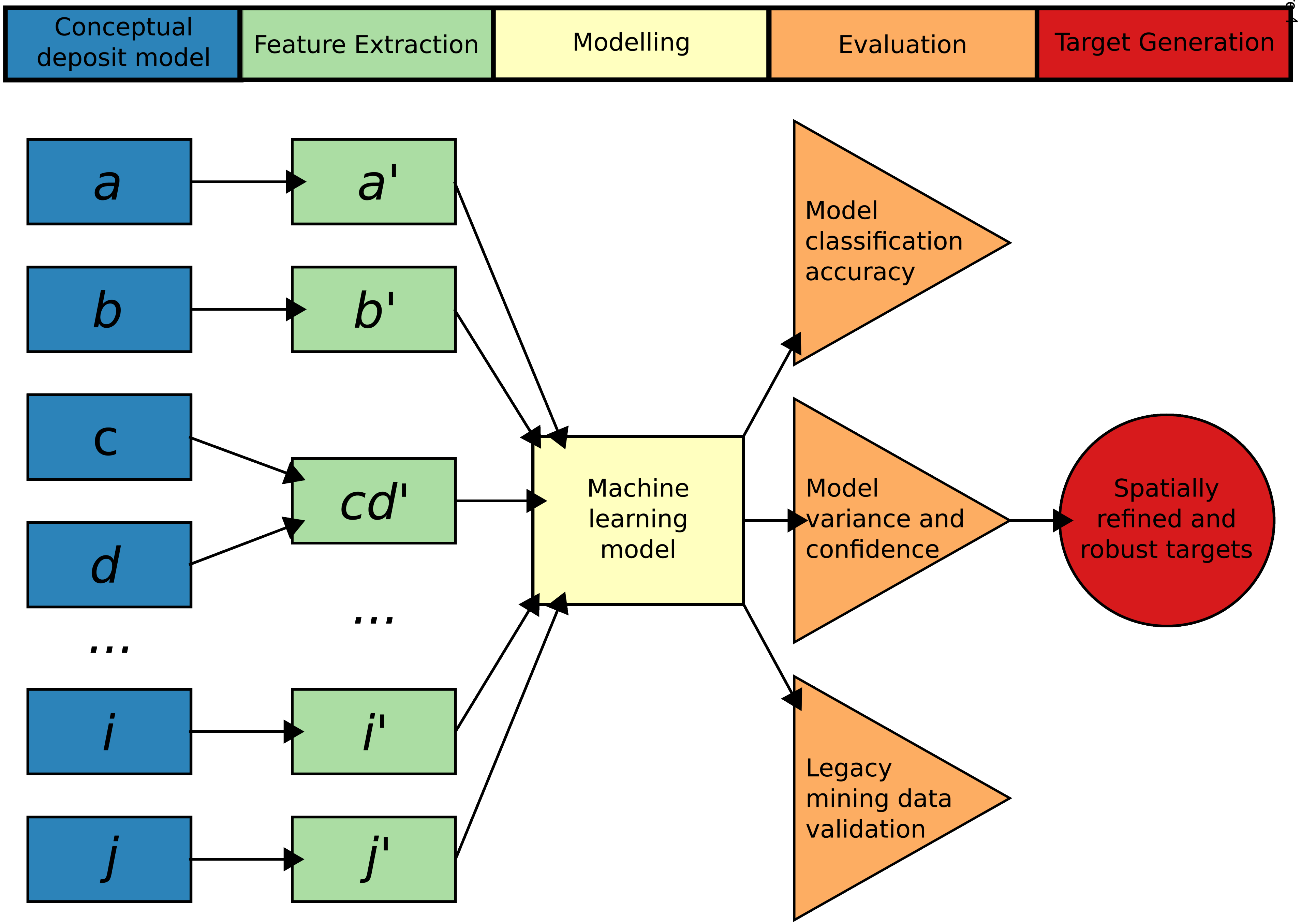

NNW-SSE extension (Stephanian to Early Permian)
D3
Reactivated thrust/
transfer faults
new faults
ENE-WSW shortening (Early Permian)
D4
tensile veins
S4 cleavage
new faults
Strike-slip
reactivation
NNW-SSE shortening (Early Permian)
D5
S5 cleavage
new faults generation
fault reactivation
ENE-WSW extension (Mid-Triassic)
D6
Modern 'XC' mineralisation along
new and re-activated structures

Figure 3



N
East Pool & Agar
Cligga Head
Castle-an-Dinas
Redmoor
Hemerdon
a
b
c
d
e
f
0
25
50 km
Legend
Tungsten Occurrences
Tin Extraction
Copper Extraction
Granite Outcrop

Figure 4



Conceptual deposit model
Feature Extraction
Modelling
Evaluation
Target Generation
a
b
c
d
...
i
j
a'
b'
cd'
...
i'
j'
Machine learning model
Model classification accuracy
Model variance and confidence
Legacy mining data validation
Spatially refined and robust targets

Figure 5

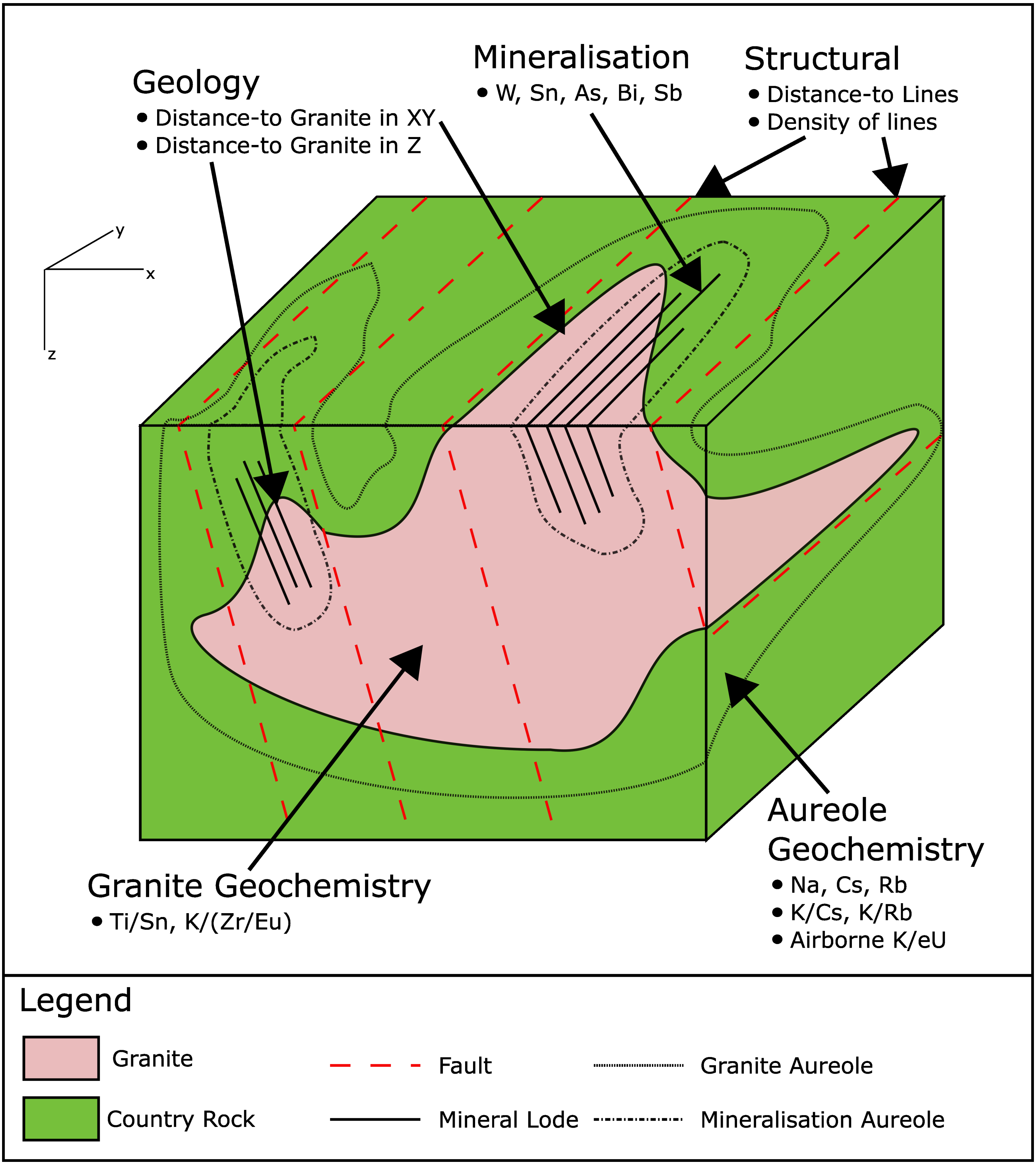

Geology
• Distance-to Granite in XY
• Distance-to Granite in Z
Mineralisation
• W, Sn, As, Bi, Sb
Structural
• Distance-to Lines
• Density of lines
y
x
z
Granite Geochemistry
• Ti/Sn, K/(Zr/Eu)
Aureole
Geochemistry
• Na, Cs, Rb
• K/Cs, K/Rb
• Airborne K/eU
Legend
Granite
Country Rock
Fault
Mineral Lode
Granite Aureole
Mineralisation Aureole

Figure 6

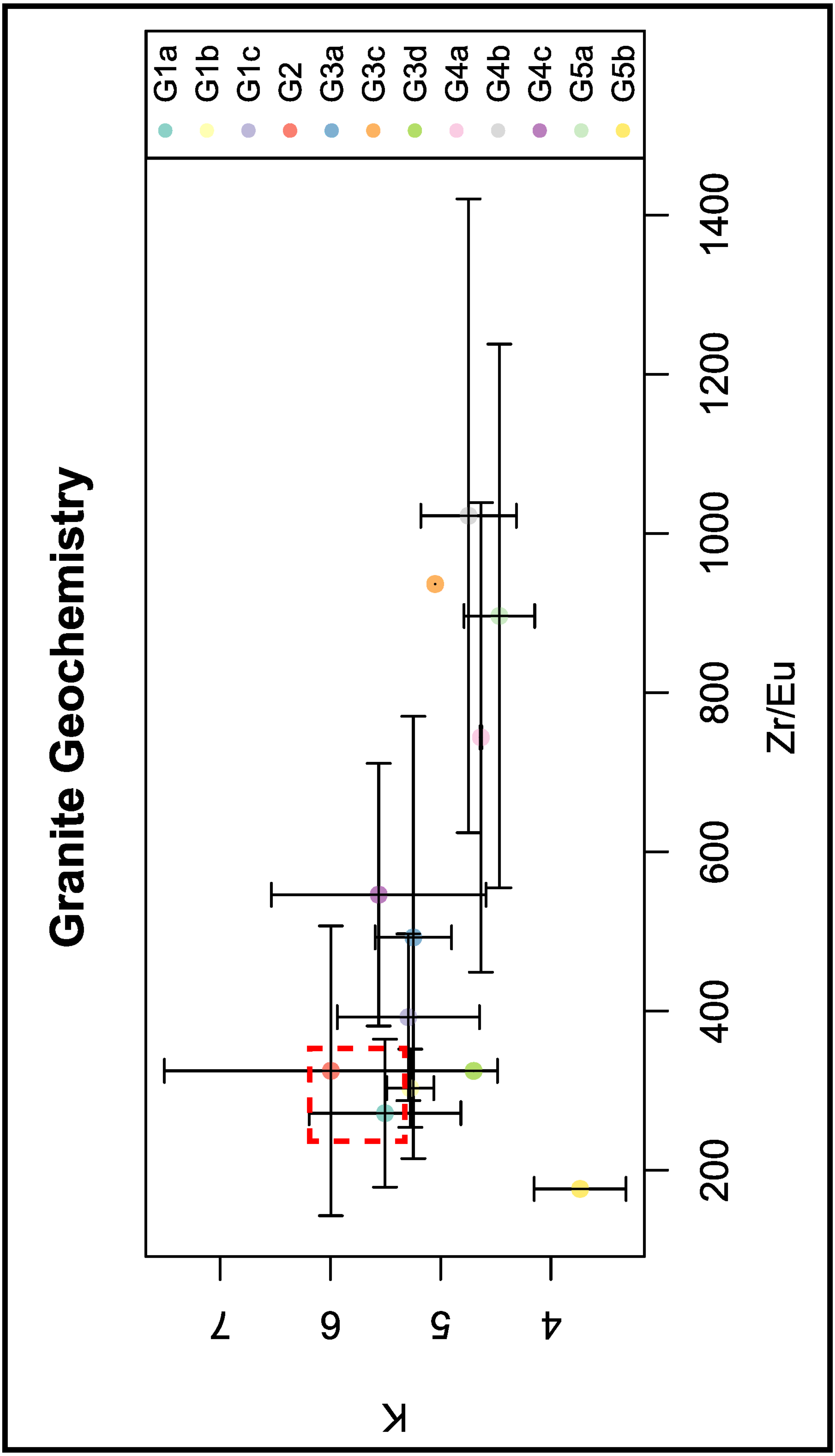

Granite Geochemistry
G1a
G1b
G1c
G2
G3a
G3c
G3d
G4a
G4b
G4c
G5a
G5b
K
4
5
6
7
Zr/Eu
200
400
600
800
1000
1200
1400

Figure 7

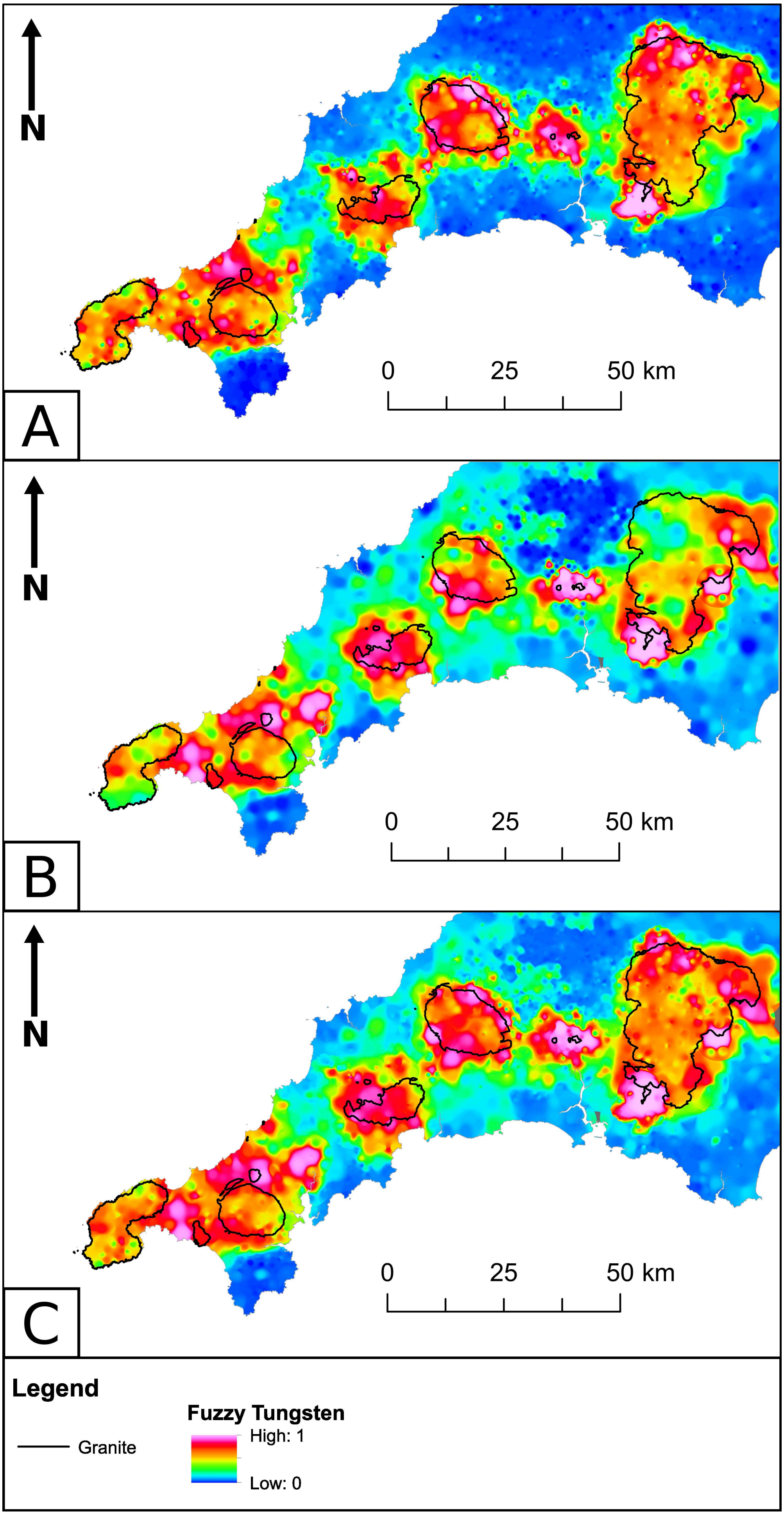

N
0
25
50 km
A
N
0
25
50 km
B
N
0
25
50 km
C
Legend
Granite
Fuzzy Tungsten
High: 1
Low: 0

Figure 8

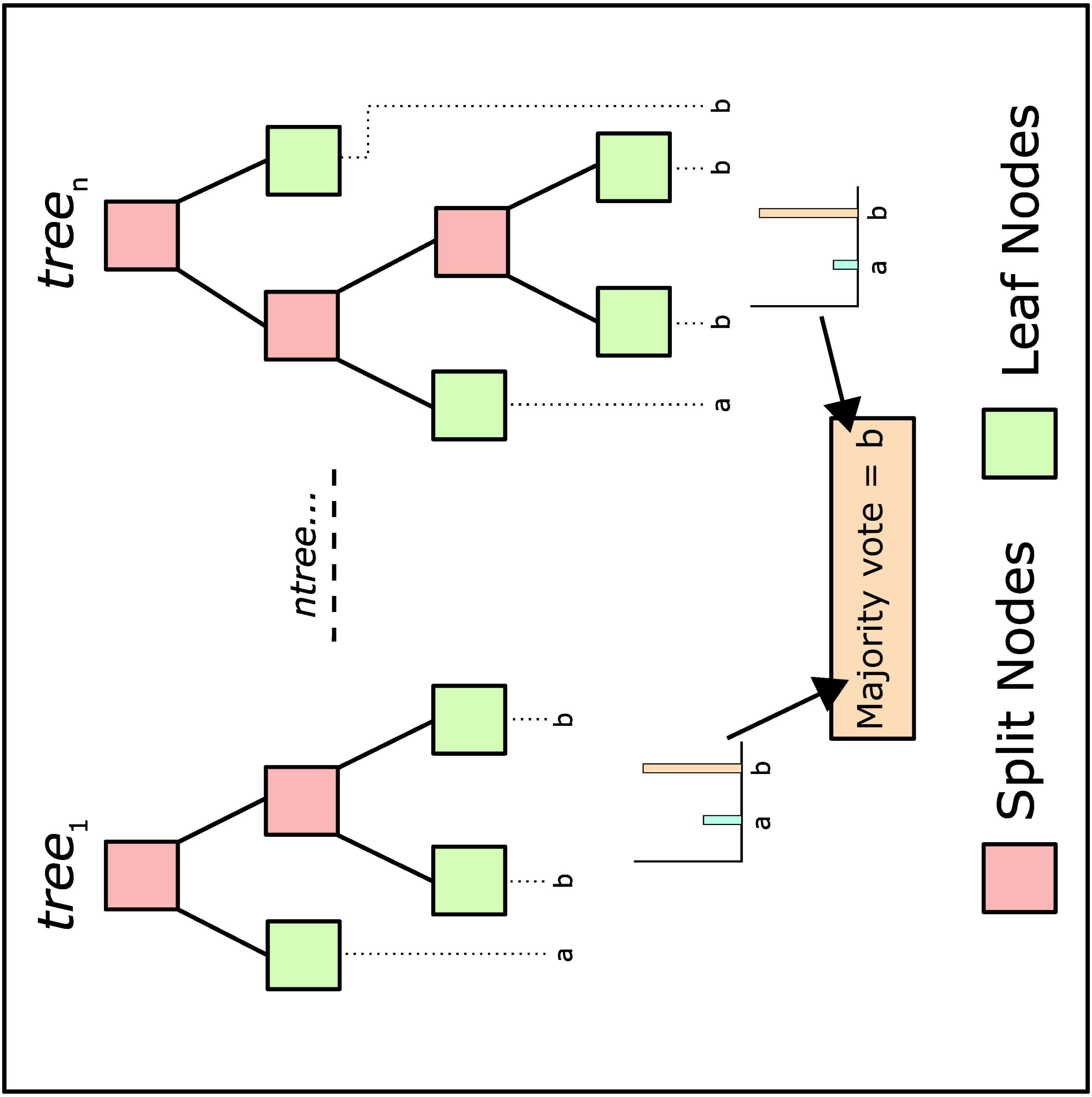


Figure 9

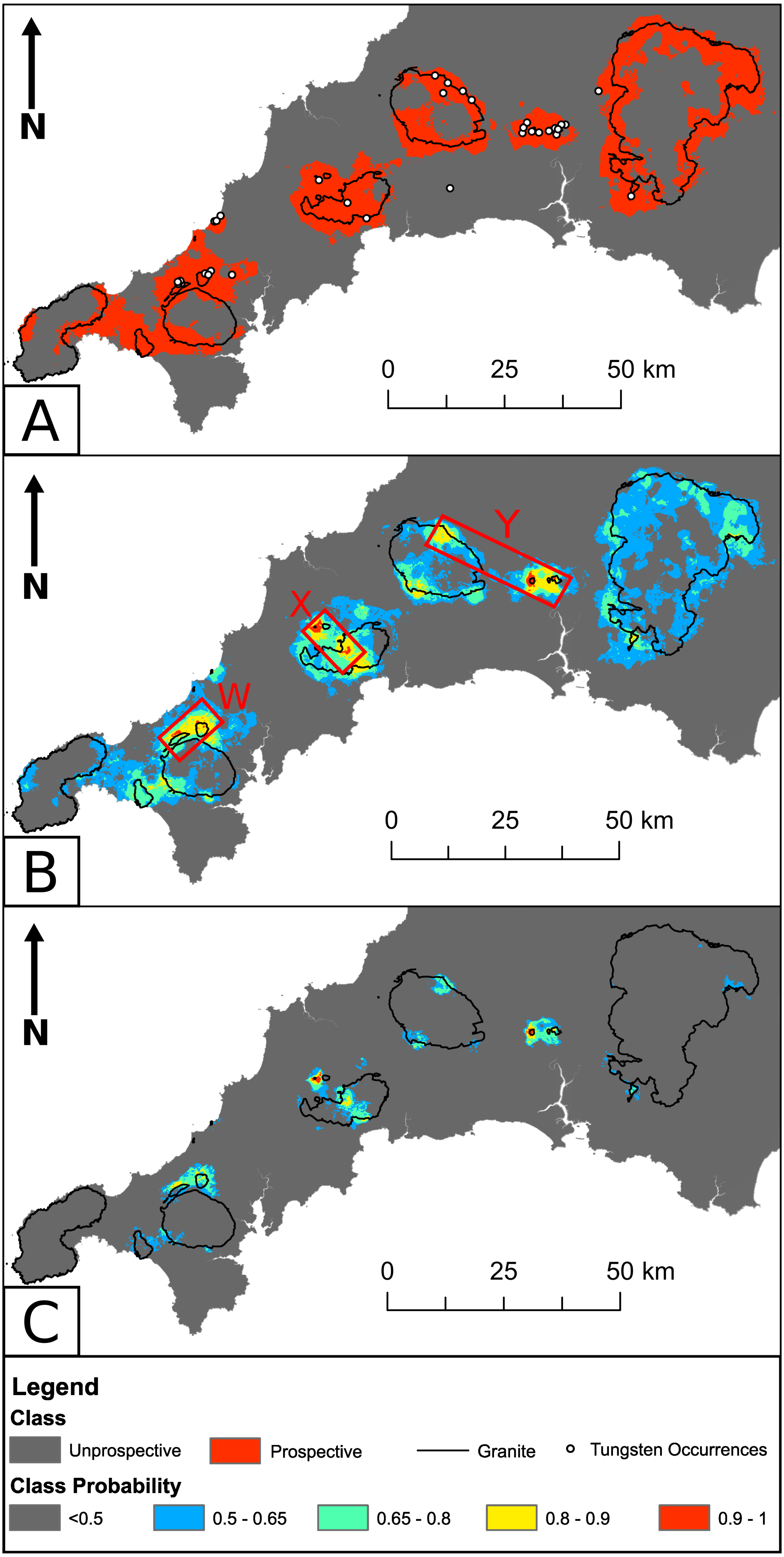
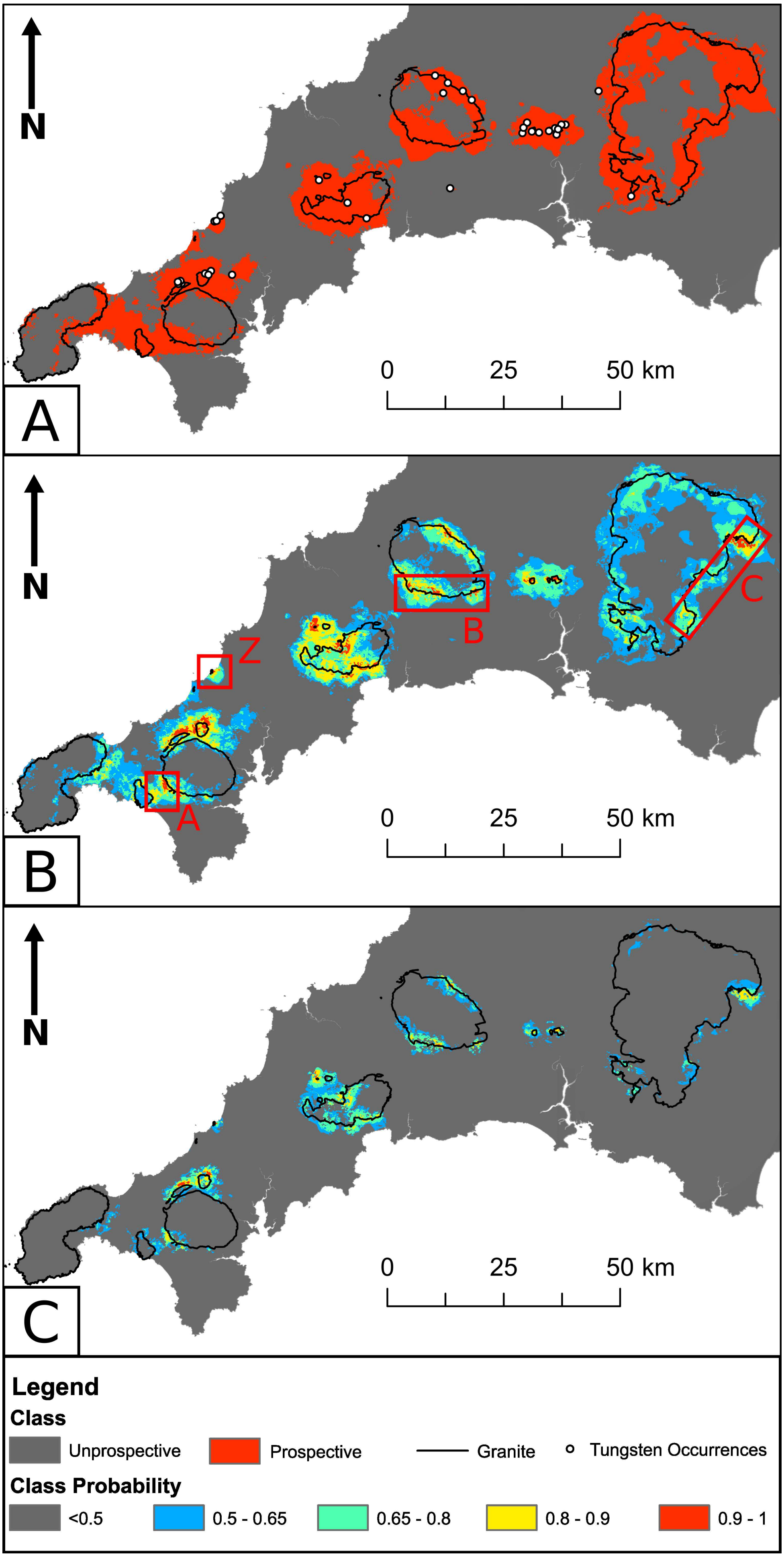

N
0
25
50 km
A
N
Y
X
W
0
25
50 km
B
N
0
25
50 km
C
Legend
Class
Unprospective
Prospective
Granite
Tungsten Occurrences
Class Probability
<0.5
0.5 - 0.65
0.65 - 0.8
0.8 - 0.9
0.9 - 1

Figure 10



N
0
25
50 km
A
Z
B
C
Legend
Class
Unprospective
Prospective
Granite
Tungsten Occurrences
Class Probability
<0.5
0.5 - 0.65
0.65 - 0.8
0.8 - 0.9
0.9 - 1

Figure 11

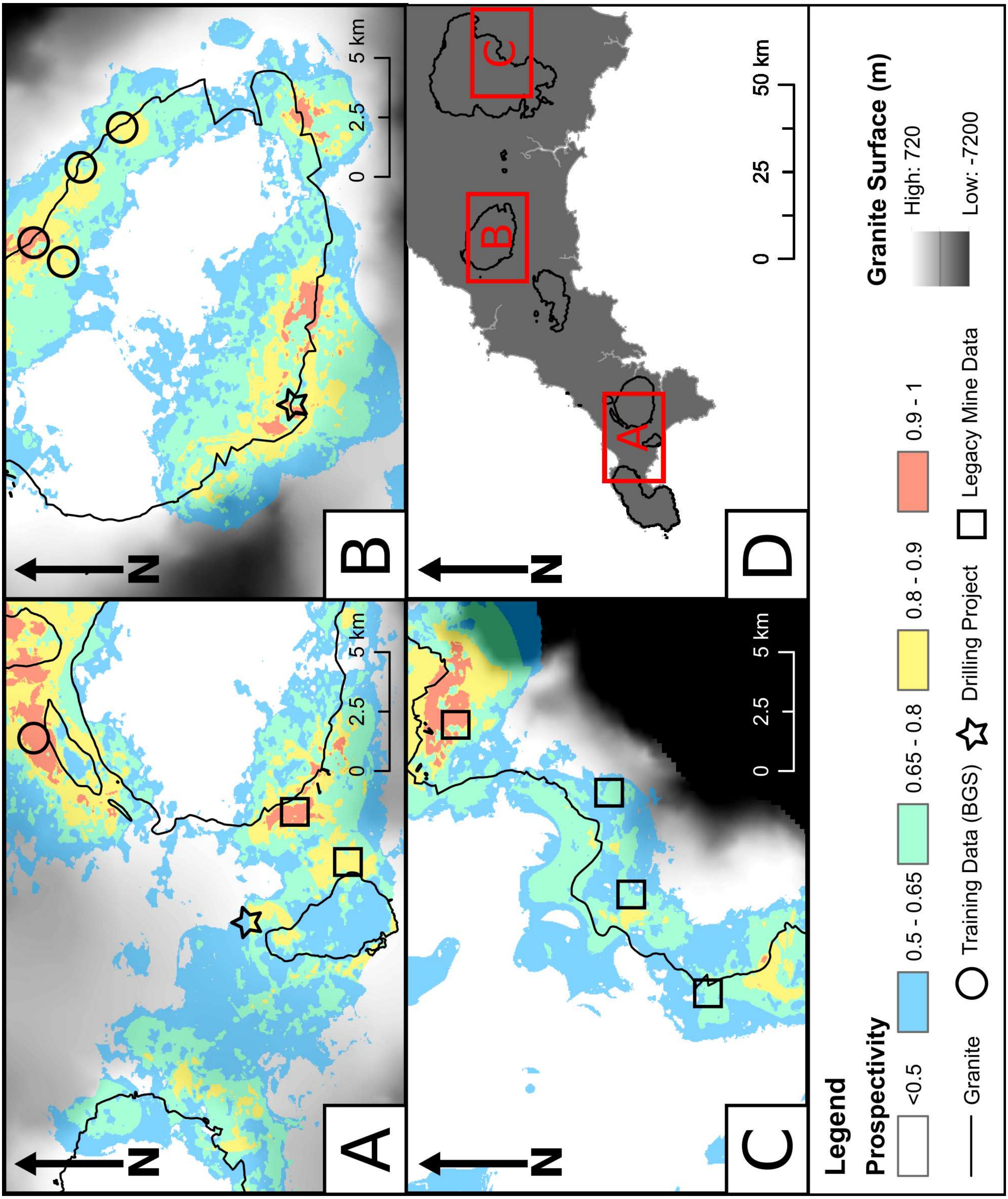

A
B
C
D
N
0 2.5 5 km
0 25 50 km
Legend
Prospectivity
<0.5
0.5 - 0.65
0.65 - 0.8
0.8 - 0.9
0.9 - 1
Granite
Training Data (BGS)
Drilling Project
Legacy Mine Data
Granite Surface (m)
High: 720
Low: -7200

Table 1

Phenomenon	Elements	Sources
Mineralisation	W, Sn, As, Bi, Sb	(Andrews et al., 1987; Ball et al., 2002; Newall, 1994; Newall and Newall, 1989)
Aureole Alteration	Rb, Cs, Na*, K/Rb*, K/Cs*, K/eU*	(Ball et al., 1984, 1998; Newall, 1994; Newall and Newall, 1989)
Granite Composition	Ti/Sn*, K/(Zr/Eu)	(Ball et al., 1984, 1998; Simons et al., 2016)

Table 2

Evidence Layer	Midpoint	Spread	Func.	Mean	SD
Proximity-to Granite in Z	N/A	N/A	TOC	0.814	0.039
Proximity-to Granite in XY	2750	2	Small	0.887	0.03
Density all lines	0.478	4	Large	0.638	0.062
Proximity-to lines	2713.41	2	Small	0.577	0.055
Airborne K/eU ratio	0.7	10	Small	0.666	0.055
Geochem Soil W	7.08	2	Large	0.887	0.032
Geochem Soil Sn	57.57	3	Large	0.829	0.034
Geochem Soil As	55.08	2	Large	0.819	0.038
Geochem Soil Bi	1.4	2	Large	0.819	0.032
Geochem Soil Sb	2.83	2	Large	0.49	0.052
Geochem Soil Rb	159.46	3	Large	0.708	0.051
Geochem Soil Cs	16.36	3	Large	0.749	0.035
Geochem Soil Na	0.83	6	Small	0.701	0.057
Geochem Soil K/Cs	0.22	3	Small	0.764	0.029
Geochem Soil K/Rb	0.02	5	Small	0.751	0.051
Geochem Soil Ti/Sn	0.08	2	Small	0.824	0.037
Geochem Stream-sediment W	27.47	1	Large	0.874	0.031
Geochem Stream-sediment Sn	636.63	1	Large	0.722	0.057
Geochem Stream-sediment As	117.68	1	Large	0.824	0.032
Geochem Stream-sediment Bi	2.86	2	Large	0.809	0.032
Geochem Stream-sediment Sb	2.69	1	Large	0.594	0.036
Geochem Stream-sediment Rb	176.41	4	Large	0.644	0.045
Geochem Stream-sediment Cs	20.35	3	Large	0.69	0.047
Geochem Stream-sediment Na	6359.1	5	Small	0.709	0.052
Geochem Stream-sediment K/Cs	1813	3	Small	0.533	0.042
Geochem Stream-sediment K/Rb	157.63	5	Small	0.668	0.058
Geochem Stream-sediment Ti/Sn	387.78	2	Small	0.706	0.064
Geochem Stream-sediment K/(Zr/Eu)	136.02	2	Small	0.739	0.044

Table 3

Element or Ratio	Func.	Mean	SD	Soil	SS	Improvement in AUC
W	OR	0.901	0.026	0.887	0.874	INCREASE
Sn	OR	0.816	0.034	0.829	0.722	INCREASE
As	OR	0.851	0.033	0.819	0.824	INCREASE
Bi	OR	0.819	0.032	0.819	0.809	NO CHANGE
Sb	OR	0.537	0.085	0.49	0.594	DECREASE
Rb	OR	0.657	0.13	0.708	0.644	DECREASE
Cs	OR	0.71	0.037	0.749	0.69	DECREASE
Na	OR	0.758	0.048	0.701	0.709	INCREASE
K/Cs	OR	0.676	0.04	0.764	0.533	DECREASE
K/Rb	OR	0.713	0.055	0.751	0.668	DECREASE
Ti/Sn	OR	0.724	0.061	0.824	0.706	DECREASE

Table 4

Model Type	Input Layers	Key Parameters	Mean	SD
Random Forest (standardised variables)	All evidence layers with zero mean and equal variance	mtry = 5; ntree = 20 000	0.959	0.03
Random Forest (fuzzy-transformed variables)	All fuzzy evidence layers, including geochemical data merged using the *fuzzy OR* operator	mtry = 4; ntree = 20 000	0.96	0.04

Table 5

	Fuzzy-transformed model				Standardised model			
Class	**Σ Prob**	**Prob (%)**	**Σ Conf**	**Conf (%)**	**Σ Prob**	**Prob (%)**	**Σ Conf**	**Conf (%)**
<0.5	4597.3	76.58	5693.2	94.83	4526.6	75.4	5811.73	96.81
0.5-0.65	723.88	12.06	174.02	2.9	969.72	16.15	106.61	1.78
0.65-0.8	460.3	7.67	104.73	1.74	386.5	6.44	67.89	1.13
0.8-0.9	188.33	3.14	28.74	0.48	108.59	1.81	14.1	0.23
0.9-1.0	33.67	0.56	2.82	0.05	12.07	0.2	3.21	0.05
Total	6003.47	100	6003.52	100	6003.47	100	6003.54	100

Supplementary Information

Click here to access/download
e-component
Supplementary_Information.pdf

Declaration of Interest Statement

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

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

Nothing to declare