Unsupervised classification applications in enhancing lithological mapping and geological understanding: a case study from Northern Ireland

Accepted Manuscript

Journal of the Geological Society

Unsupervised classification applications in enhancing lithological mapping and geological understanding – A case study from Northern Ireland

Z. Smillie, V. Demyanov, J. McKinley & M. Cooper

DOI: https://doi.org/10.1144/jgs2022-136

To access the most recent version of this article, please click the DOI URL in the line above. When citing this article please include the above DOI.

Received 27 September 2022
Revised 14 February 2023
Accepted 30 March 2023

Supplementary material at https://doi.org/10.6084/m9.figshare.c.6603098

Unsupervised classification applications in enhancing lithological mapping and geological understanding – A case study from Northern Ireland

Z. Smillie[1*], V. Demyanov[2], J. McKinley[3] & M. Cooper[4]

[1]*University of Stirling, Stirling FK9 4LA, Scotland, UK*
[2]*Heriot-Watt University, Edinburgh EH14 4AS, Scotland, UK*
[3]*Queen's University, Belfast BT7 1NN, Northern Ireland, UK*
[4] *Geological Survey of Northern Ireland, Belfast BT4 3SB, Northern Ireland, UK*
**Corresponding author (e-mail: Zeinab.smillie@stir.ac.uk)*

Abstract:

Using pattern classification algorithms can help recognise and predict patterns in large and complex multivariate datasets. Utilising competitive learning, self-organising maps (SOMs) are known unsupervised classification tools that are considered very useful in pattern classification and recognition. This technique is based on the principles of vector quantification of similarities and clustering in a high-dimensional space, where the method can handle the analysis and visualization of high-dimensional data. The tool is ideal for analysing a complex combination of categorical and continuous spatial variables, with particular applications to geological features.

In this paper, we employ the tool to predict geological features based on airborne geophysical data acquired through the Tellus project in Northern Ireland. SOMs are applied through 8 experiments (iterations), incorporating the radiometric data in combination with geological features, including elevation, slope angle, terrain ruggedness (TRI), and geochronology. The SOMs proved successful in differentiating contrasting bedrock geology, such as acidic versus mafic igneous rocks, while data clustering over intermediate rocks was not as apparent. The presence of a thick cover of glacial deposits in most of the study area presented a challenge in the data clustering, particularly over the intermediate igneous and sedimentary bedrock types.

Supplementary material: include additional tables and figures that are not accommodated within the limition of GSL paper size
[GSL will assign the doi and url unless you already have one]

Geological maps have enormous economic and societal value. They support our ability to find and develop natural resources, assess and protect the environment and identify and prepare for natural hazards. Geological maps inform ground conditions for infrastructure development and allow us to make more informed land-use and planning decisions (DeJong et al., 2019). Geological maps, however, are expensive and time-consuming to produce. So, methods that allow the valid integration of remote sensed, or other less traditionally used datasets, are of interest as they could provide rapid mapping or enhancements over large survey areas.

It is generally understood that geological maps are not exact. They contain varying levels of uncertainty, bias, and error brought about, for example, by the scale of the mapping, the experience of the geologist(s) involved, when the mapping was done, what methods, conventions or datasets were used, and the purpose of the survey. Further uncertainty may be introduced where interpretations, for example, of geophysical or geochemical datasets, are integrated with limited or even no fieldwork to check their validity (Aldiss, 2013; Lark et al., 2015; Halotel et al., 2020). It is

essential for more reliable uncertainty assessment to use rapid unbiased spatial data integration methods to come up with multiple possible mapping scenarios, while some can be overlooked in geological mapping based on manual interpretation.
In this paper, we will present the challenges facing geological mapping. The benefits of implementing machine learning (ML) are examined, and an in-depth discussion on the methods used here to approach the ML algorithms is offered. The methods section covers aspects of data preparation and visualisation, aiming to understand the nature of each dataset. The results section is dedicated to illustrating the outputs of each ML experiment used. In addition, the anatomy of the data clusters and how these are organised is presented, resulting in cluster mapping in correspondence to the geological units.

Previous Work

Since it was established in 1947, the Geological Survey of Northern Ireland (GSNI) has been engaged in mapping the region's bedrock and overlying superficial deposits. Field survey has been conducted at a scale of 1:10 000, which has then been made available, along with 1:50 000 and 1:250 000 scale derivative maps, in both paper and digital format. In addition to the geological survey, the Northern Ireland region has been mapped by the Tellus Project airborne geophysical survey, which was flown in 2005 - 2006 (Young and Donald, 2013). Geophysical information collected includes magnetic, electromagnetic and radiometric datasets. In short, there is a wealth of geological information available for Northern Ireland, which can be used to test and develop Machine Learning (ML) techniques.

Tellus radiometric data has attracted the attention of many researchers and has been discussed in a number of publications, e.g. (Jones and Scheib, 2007; Hodgson and Young, 2016). Ternary radiometric maps were applied to the Newry Igneous Complex to identify and map zonation between and within its three main plutons (Cooper et al., 2016). However, there is no established method to process and integrate these data with other geological and geomorphological features.

How machine learning can improve mapping and identifying patterns in geological data

Machine learning (ML) is an interdisciplinary field that was developed within the computer sciences. Although machine learning algorithms have been around for decades (under different names, e.g. Artificial Intelligence), there has been a growing interest in machine learning across a wide variety of disciplines. This has been motivated by the development of new robust algorithms capable of solving complex problems that are resistant to more traditional modelling, e.g. (Lin et al., 2012; Bhamare and Suryawanshi, 2018; Toulia et al., 2018; Baldini et al., 2020; Halotel et al., 2020; Elbegue et al., 2022; Kirkwood et al., 2022).

ML is progressively gaining approval and success in many domains, including the geosciences, where approaches towards geological modelling were equally welcomed by academics and businesses, e.g. (Harris and Grunsky, 2015; Demyanov et al., 2019, 2020; Halotel et al., 2020). Many of the recent attempts have substantial predictive power and proved the establishment of the ML as an exceptional tool when geological mapping is required in remote areas or areas with limited outcrop exposure can be challenging. Recent attempts include using ML to predict lithology predictions from wire logs (Gifford and Agah, 2010; Ferraretti et al., 2012; Sebtosheikh and Salehi, 2015; Silversides et al., 2015; Bhattacharya et al., 2016; Halotel et al., 2020) and superficial sediments and bedrock from remote sensing datasets (Cracknell and Reading, 2014; Harris and Grunsky, 2015; Harvey and Fotopoulos, 2016; Bachri et al., 2019).

Automation of lithofacies classification using ML tools presents an excellent opportunity to manipulate and visualise various kinds of large datasets, such as remote sensing and airborne

geophysical surveys, and appears to be a promising solution for the challenges associated with traditional manual interpretation. Furthermore, ML, such as clustering and classification techniques (Pires and Harthill, 1989; Carneiro et al., 2012), are capable of processing high-dimension, complex, multivariate datasets, where patterns and relationships of different features are hard to recognise and interpret.

In this research, we employ one of the ML algorithms known as the self-organizing maps (SOMs). We test the use of SOMs as a possible approach to analysing large multivariate datasets - the Tellus Project radiometric data (Donald, 2016) for lithological discrimination. We present a synthetic study illustrating the ability of the SOMs to recognise bedrock distribution when faced with various geological, geophysical and geomorphological datasets. Comparison with published maps, of both the bedrock and superficial deposits, provides a way of assessing the success of the techniques.

Self-organising maps (SOMs)

SOMs are data-mining tools and unsupervised neural network algorithms that use a statistical approach to cluster multivariate datasets (Kohonen, 1982; Fraser and Dickson, 2007; Kohonen et al., 2012). SOMs have the ability to reduce the high dimensionality of large datasets yet highlight subtle relationships of multivariate data by clustering the data in a 2D computational map space. SOMs have been shown to be useful for analysing, visualising and clustering multisource and high-dimensional geoscientific data (Penn, 2005; Bierlein et al., 2008; Carneiro et al., 2012) and have been employed in research related to mining and exploration data sets (Bierlein et al., 2008), as a classification tool of remote sensing imagery (Gonçalves et al., 2008; Bedini, 2009; Li and Eastman, 2010; Shivakumar and Rajashekararadhya, 2020) and airborne geophysical data sets (Carneiro et al., 2012; Fraser et al., 2012).

SOMs employ the principles of vector quantisation and vector similarity measures (Cracknell et al., 2014) using a structure of self-adapted elements, called neurones (Supplementary material Fig. S1), to group and classify data points (Demyanov et al., 2019). The networks learn to form their own classifications of the training data without external help. Class membership is broadly defined by the input patterns sharing common features, and that the network will be able to identify those features across the range of input patterns (Cracknell et al., 2014).

A self-organising map is a single-layer feedforward network, where the neurons are arranged in a two-dimensional pattern (2D grid) with a distance measure between them. This distance measure is used to control the similarity of the neurons in the network (Demyanov et al., 2019; Kanevski, 2009). The network is initialised prior to training, so all data vectors of the neurons are assigned random values. Upon input data entries, data are then compared to each of the neurons whilst training. The output neurons compete amongst themselves to be activated, i.e. to be representative of a training data sample, with the result that only one is activated at any one time, called a "winner-takes-all" neuron or simply the winning neuron (Kanevski, 2009; Cracknell et al., 2014). The neurons are forced to organise or cluster themselves in a network or "Self-Organizing Maps" (SOMs). The learning process implies that the vector of values of the winner neuron is modified to become even more similar to the data vector presented as an input (Demyanov et al., 2019).

Objectives

The objectives of this study are threefold: (1) to investigate the use of ML applied to airborne radiometric data and a digital terrain model; (2) to predict the distribution of geological units and reproduce a known geological map of the study area using SOMs, in its entirety or at least in parts; and (3) to reveal previously unrecognised and unmapped aspects of the geology; and as a result,

assist in mapping areas where modern geological maps are not available, but there is the availability of data such as airborne geophysical and topographic data.

Geological background

With an area of about 14000 km^2, Northern Ireland is a small country whose bedrock geology is known to be remarkably diverse both in terms of its age and composition (GSNI, 1997; Mitchell, 2004). The region is a host to bedrocks ranging in age from the Neoproterozoic to the Oligocene and includes metamorphic, igneous and sedimentary rock groupings. It is also known for its superficial geology, including extensive coverage of Quaternary glacial and postglacial Holocene deposits that conceal much of the bedrock geology (GSNI, 1991).

The study area (Fig. 1a) lies in the southeast of Northern Ireland. It is part of the Southern Uplands–Down–Longford terrane, formed chiefly of greywacke sandstones and mudstones of Ordovician and Silurian age (Anderson, 2004). The greywacke sandstones are represented by the Gala Group to the northwest and Hawick Group to the southeast (GSNI, 1997) and are arranged in fault-bound, NE–SW orientated tracts that have undergone both low-grade regional metamorphism and higher-grade thermal contact metamorphism. The latter contact metamorphism came about as a consequence of the intrusion of magma at two separate intervals. Older magmatism took place during the Lower Devonian and led to the formation of the Late Caledonian Newry Igneous Complex (Cooper and Johnston, 2004a; Cooper et al., 2016), whilst later magmatism during the Palaeogene resulted in the development of the Mourne Mountains and Slieve Gullion igneous complexes (Cooper and Johnston, 2004b). The Newry Igneous Complex is composed of three NE–SW aligned granodiorite plutons (Fig. 1b), with an ultramafic–intermediate composition Seeconnell Complex at its northeast end (Anderson et al., 2016). The Mourne Mountains Complex (Fig. 1c) is granite dominated but is subdivided based on composition and texture into five separate units (G1-G5) that are distributed across a western and an eastern centre (Stevenson et al., 2007). The Slieve Gullion Complex (Fig. 1a) includes a central layered unit of gabbro and granite surrounded by a circular ring dyke composed mainly of porphyritic felsite and granophyre (Cooper and Johnston, 2004b). In the northwest of the study area, there are fault-controlled areas of basalt (Fig. 1a) that belong to the Palaeogene Antrim Lava Group (Cooper, 2004). Finally, there are numerous minor intrusions (sills and dykes) of Late Caledonian (Cooper and Johnston, 2004a) and Palaeogene age (Cooper et al., 2012; Anderson et al., 2016, 2018) that traverse the region.

The geomorphology of the study area (Fig. 1d) is dominated by the high ground of the Mourne Mountains, Slieve Gullion and the Ring of Gullion, and Slieve Croob, which are underlain by Palaeogene igneous complexes with the exception of Slieve Croob, which is Late Caledonian (Fig. 1d). In such areas, rocks tend to be at or close to the surface, though maybe concealed by peat (Fig. 1e). The lower-lying ground is underlain mainly by Silurian greywackes covered by glacial till or diamicton that forms impressive swarms of drumlins and ribbed moraines (Fig 1e) that reveal how ice from the last Ice Age progressed and retreated from the area (McCabe and Dunlop, 2006). There are many glacial meltwater channels within the region, including at least two tunnel valleys that have formed along the line of the Newry and Camlough faults. Also present in the area are glacial and postglacial sand and gravel deposits occupying valleys and coastal regions.

Methods (data setup and processing)

In this section, we present how the data are processed, including visualisation techniques to understand the nature of each dataset. Also, we provide the methods to implement the SOMs techniques, including steps taken to decide the number of iterations and data clusters.

Tellus Project Radiometric Survey

The Northern Ireland Tellus Project geophysical survey was flown in 2005-2006 by the Joint Airborne-geophysical Capability, a partnership of the British Geological Survey (BGS) and the Geological Survey of Finland (Beamish et al., 2006). During this survey, the aircraft, a De Havilland DHC-6 Twin Otter, flew 86,000 line-km at a nominal height of 56 m along a network of pre-planned parallel lines, spaced 200 m apart, and captured airborne magnetic, radiometric and frequency domain electromagnetic datasets. Digital readings of the various parameters were made automatically at appropriate intervals, up to one second (for gamma radiation), equivalent to measurements at intervals of approximately 60 m.

All rocks, superficial deposits and soils contain minimal amounts of radioactive elements (IAE Agency, 2003), which produce gamma radiation detectable at the survey altitude (Beamish et al., 2006). The radiometric dataset was collected using an Exploranium GR-820 gamma-ray spectrometer, which allowed the various radioactive isotopes of uranium, thorium and potassium to be distinguished by the energy of their emissions. The aircraft's position was continuously recorded by a Global Positioning System (GPS). The data were recorded digitally, transferred for processing, and corrected for various extraneous effects and standardisation (Hodgson and Young, 2016). For full data acquisition and processing details, see (Young and Donald, 2013; Hodgson and Young, 2016).

The main data used in this research are the distribution of naturally occurring potassium (D_KAL), uranium (D_URA), and thorium (D_THO). These elements were derived from the calibrated gamma-ray amplitudes from the near ground surface, assuming no disequilibrium in the radioactive decay series of each element (Hodgson and Young, 2016).

Additional data of the recorded total count (T_CPS) was used. The latter is not exclusive to these three elements and may include other elemental recordings. Further geological features, mainly elevation data and bedrock age (geochronology), were combined with radiometry to test how these may enhance (or not) the prediction of the bedrock distribution. The elevation data were sourced from the open-access European Digital Elevation Model (EU-DEM), version 1.1, sheet E30N30, at 25m resolution with vertical accuracy: +/- 7 meters RMSE (Copernicus Land Monitoring, 2017). Slope angle and terrain ruggedness index (TRI) analyses were later extracted from the DEM data using ArcGIS Pro.

Radiometric data values below zero values, i.e. those collected over major water bodies, were removed. The dataset consists of 125275 spatial points, with 12 variables attached. These variables are X, Y, D_KAL, D_URA, D_THO, T_CPS, bedrock lithology, bedrock age, DEM, slope angle, TRI, and Superficial deposits. These variables were used in data visualisation plots and understandings prior to any ML data processing. Radiometry and elevation data were then used to test the SOMs functions in predicting the bedrock distribution. The geographic coordinates (X, Y) of the data were then removed to create "unlabelled" data. Radiometry, elevation-derived variables were trained and examined using the SOMs. Age data are not unique to specific lithological. For example, Paleogene igneous emplacement included both mafic and acidic igneous rocks. However, the nature of these data influences the natural clustering of other data and presents a degree of bias when used to predict the bedrock geology.

Data were scaled into a range of [0,1], transformed into a number of DAT files and processed using the software Machine Learning Office - MLO (Kanevski, 2009). Input datasets ranged between 2 to 6 variables; a list of iterations used in the current study is given in Table 1. The iterations are applied through adding or removing variables, beginning with radiometry data (two elements), then adding between one to three other variables to test the data clustering.

Data Overview and Visualisation

Radiometric maps and Histograms

In this section, we employed GIS and data visualisation tools to understand the patterns and relationships between the radiometric data, topography and bedrock in the study area. As the bedrock distribution is the main output of the SOMs work in this paper, their relationship with other geological features (radiometry and topography) is investigated and presented. Radiometric data are mapped (supplementary material - SII) with a standard colour scale; warm colours (red-orange-yellow) represent high values in contrast to lower values that are plotted using cool colours (blue-green). Additionally, histograms and cross-plots are used to present the feature distributions and relationships.

Unsupervised classification and Self-organising maps (SOMs)

Self-Organising Maps (SOMs) are used in this research to identify, if present, spatially contiguous and geologically meaningful classes within the bedrocks in the area from key geophysical and geological variables. First, un-labelled data sets (without spatial attributes) were trained using the Machine Learning Office software (Kanevski, 2009). Data then were transformed into .dat format and processed into 20 iterations . The latter are combinations of different sets of the variables, e.g. radiometry only, radiometry + slope, etc. (see Table 1).

Several visualization tools (maps) are used to present the trained SOMs (Kanevski, 2009): (1) Hits maps displaying how many times (number of hits) each neuron wins (depends on data set, training or testing), (2) U-matrix (unified distance matrix) showing the distances between each single neuron and all its neighbours, (3) Slices (2D slices) of the SOMs weighs (the total number of maps is equal to the dimension of the data), (4) Clusters (the recognized clusters in the SOMs).

The construction of the U-matrix map consists of two steps: (1) calculating the average distance between the nearest SOMs neuron, (2) calculating the average (or median) values of the nearest U-matrix cells, computed in step 1, (3) optional, the calculated U-matrix is treated as an image (for better visualisation) and is smoothed with a simple average (or median) filter. The resulting maps are then painted using a colour scale; high colour values (dark shades) correspond to large distances between neighbouring nodes, hence the borders of the clusters on the map.
Hexagonal 10 x 10 grids were used to define the structure of the lattice, with K-means number of trials as 100. These parameters were sustained during all the iterations in this study.

Clusters maps are not strictly speaking part of the SOMs, but it supports the visualisation of the cluster distribution. A fundamental step for any unsupervised algorithm is to determine the optimal number of clusters into which the data may be clustered, also known as the model selection problem. The optimum number of clusters was assessed using the Elbow method (Suppl. III a), one of the most popular methods to determine this optimal value of k. Applying such a method to the present data sets resulted in k= 4. However, applying such a value of k did not capture the variation in the radiometry distribution within the dataset. Therefore, a 95% variance rule is used, resulting in an optimum number of clusters of 7. The data slices clearly show that our data have 7 clusters, not 4, and are used to verify such k value (Suppl. III b).

SOMs (Number of hits, U-Matrix, Slices, Clusters)

To map the current sets of data, SOMs allocated each input data to one of the SOMs neurons. A 10 × 10 SOMs hexagonal lattice with 100 nodes was used. Figure 2 shows an example of the outcomes

of the SOMs analysis performed after training, using the radiometry data of the potassium and uranium. The number of hits (Fig. 2A) illustrates how many data records fall into each neuron on the lattice, where the blue colour corresponds to a low number of hits and brown to a high number (Fig. 2A). Fig. 2B shows the U-matrix, visualising the distances between the neurons. The data clusters are the results of K-means clustering on the 2D SOM lattice based on the U-matrix distance metric.

Here we provide an example using K & Th datasets. Both of the data slices of the K and Th (Figs. 2 C & E) displays seven groups of data that can be matched with 7 clusters. In general, the results demonstrate a logical grouping of the data with several significant groups of data, which can be attributed to a particular bedrock feature or group. The clusters were then linked to the original coordinates (X, Y) data and mapped using ArcGIS Pro, using the same colour codes used in the clustering process (Fig. 2 D). The results show that the features can be grouped to match some of the bedrock geology, with a clear recognition of the acidic igneous complexes (granites and felsites) in the south and mafic rocks (basalt) in the northwest (Figs. 2 F & G).
This technique was repeated using different combinations of the radiometric and geological data (Table 1). The ID of each cluster may vary from one experiment to another as the SOMs re-assign the clusters after each data training.

Spatial distribution of the clusters and comparison with bedrock geology.

To compare the SOMs results with the published maps (GSNI, 1997), the colour codes of the clusters in each experiment were manipulated. The same set of colours was used in each map, but cluster colour was modified as much as possible to present a visual representation of the bedrock that is mostly associated with. The cluster colours were set, where warm (red-pink) colours are used where the clusters are believed to match with acidic igneous rocks (associated with high radiometry), and calm (dark blue) colours are associated with the basic rocks (related to low radiometry) and peat (superficial deposits). Green-yellow and orange were assigned clusters matching other rocks in the area, e.g. granodiorite and sandstone, generally displaying intermediate radiometry).

Data Visualisation Results

Feature distributions and relationships

In this section we discuss the nature of distribution and relationships of various features, especially in connection to bedrocks. Geographically and geologically, K shows a great deal of variation (Suppl. II A1), with higher concentrations in the south and southwest associated with the more acidic lithologies of the Mourne Mountains, Slieve Gullion and Newry Igneous complexes. Interestingly in this area, many K values remain high despite the presence of overlying glacial till (diamicton). In the northwest, low concentrations of K are seen where basic igneous rocks of the Antrim Lava Group are also located under diamicton and alluvium. On high ground, low concentrations are found where there are peat deposits. K values over the sandstones show a wide variation, although most fall within a medium range. K measurements show a unimodal normal distribution (Suppl. II A2)

Both the U and Th show a similar distribution to K, with high concentrations associated with the igneous complexes and low with the Antrim Lava Group. U and Th have strong associations with the Slieve Gullion Complex (including the Ring of Gullion ring dyke), the Mourne Mountains Complex, and the outer zone of the northwest Rathfriland pluton of the Newry Igneous Complex.

There is also a clear contribution to the alluvium to the northwards draining, Upper Bann River, from the Mourne Mountains Complex (Suppl. II B1 & C1). The elements data show a unimodal distribution with skewness (tail) towards the higher values (Suppl. II B2 & C2), indicating that the majority of values are concentrated around low and medium values and with fewer high to very high values.

Very strong signals of U are recorded over the Mourne Mountains Complex, although no significant variations are observed within this rock mass. In addition, relatively high U values are recorded over Slieve Gullion and the Ring of Gullion, Slieve Croob, as well as the till and other superficial deposits south of the Mourne Mountains.

U and Th values are similar over the Newry granodioritic-dioritic pluton and sandstone, with even lower values within the granodiorite than in most sandstone areas. This is possibly attributed to the presence of peat and inland water on top of the eastern part of the Newry pluton. Areas underlain by mafic rocks are generally low in U and Th, with significantly low values of Th especially.

The structure of the Ring of Gullion is clearly reflected through the distribution of potassium and thorium, with less representation when mapping the uranium distribution on its own. Both the uranium and thorium values are significantly high, where alluvial deposits are recorded near or sourced from either the Mourne complex or the Newry granodiorite. Whereas the alluvial deposits sourced from other bedrock, e.g. the mafic rocks, they show significantly low U and Th values. In the meantime, areas covered by peat or inland water show similarly low values of radiometry (blue colours (Suppl. II B1 & C1).

In general, high radiometry values are associated with acidic bedrock compositions, while the basic bedrocks show low values. The effect of superficial peat deposits is displayed by lowering the radiometry recordings. Total counts (T_CPS) generally follow a similar pattern to the K, U and Th distribution, with high values over the Mourne rock mass and lower values over the mafic rocks (Suppl. II D1). However, values over other bedrocks are moderate, with no clear differentiation among the signals over the granitoids, sandstones, felsite and granophyre of the Ring of Gullion.

Terrains are highly controlled by the bedrock underneath as well as the tectonic history and structural elements, e.g. uplifted areas and fault scarps. The area is generally dominated by intermediate terrains around 100 metres above sea level (Suppl. II E1-2). High altitudes are observed where parts of the Mourne Mountains, Slieve Gullion, the Ring of Gullion, and Slieve Croob are exposed. Ground elevation in such areas is up to 854 metres above sea level. Intermediate to low terrains are observed where the Newry granodiorite and sandstone are present. At the same time, areas underlain by mafic plutons in the northwest or occupied by valleys, rivers and coastal lines show the lowest altitudes in the area.

The area is characterised by dominant gentle slopes, mostly less than 5 degrees (Suppl. II F1-2). Steeper terrains are observed around the Mourne Mountains, Slieve Gullion and the Ring of Gullion, and Slieve Croob.
Terrain ruggedness index (TRI) data display a similar pattern to the slope data, with highly rugged terrain at the Mourne Mountains, Slieve Gullion and the Ring of Gullion, and Slieve Croob (Suppl. II G1-2). Furthermore, TRI emphasizes the trend in terrain irregularities within the areas covered by the diamicton with NE-SW small-scale irregularities that rotate around the uplands in the area as an effect of moraine migrations (Suppl. II G1). This pattern is possibly caused by moraine movements around obstacles of high mountains, e.g. Mourne granite and the Ring of Gullion.

The influence of the superficial deposits on the radiometry recordings

Superficial deposits cover more than three quarters (~ 77 %) of the bedrock area. Hence, one would expect that the presence of such vast and locally thick deposits may influence the radiometry signals produced by the bedrocks beneath. Although, enhanced signals can be locally produced through the development of placer deposits where a significant concentration of heavy, durable radioactive-bearing minerals (e.g. zircon and monazite) are expected to be sourced from nearby granitoid masses, e.g. the thin red line near the centre of the area (Suppl. II B1 & C1). Some of these alluvial deposits are well recognised through their high U and Th signals. Those recorded near the Mourne granite complex show higher radiometry compared to the alluvial sediments sourced from the Newry granodiorite. Other alluvial deposits are present near the northwestern parts of the area and show very low radiometry where they are sourced from mafic rocks (Suppl. II A-D).

Diamicton deposits cover vast terrains, around 67.5 % of the area. The radiometry recordings above these deposits are generally low to intermediate. Potassium is an exception where it shows considerably high values over large areas where the granodiorite exists.

Peat deposits significantly cause a reduction of the radiometry readings, and their presence is marked by patchy blue colour over the Mourne and Newry granitoids and sandstone bedrocks (Suppl. II A-C). The weak radiometry signals from these deposits are overwhelming those very high signals from the underlying Mourne granites in the high grounds. Both the peat and alluvial deposits substantially affect the radiometry due to their presence at or near the ground surface and present a challenge in mapping the bedrocks beneath.

Cross-plots

In general, there is a positive relationship between all three elements investigated in this study. Supplementary material IV shows the relationship between the elements K, U, and Th with the data points colour-coded based on elevation, slope, or TRI values. In general, blue (cool) colour indicates low values of altitude, slope or TRI, green-yellow signifies intermediate values, whereas yellow and red (warm) colours indicate high values.

Radiometric elements show a considerable variation among different terrains. High radiometry recordings are generally associated with high altitude, slope angles and TRI (Suppl. IV A-I). However, such a relationship is less clear using potassium distribution. Both the low and high k values are spread over a broader range of altitudes. On the other hand, there is a general tendency towards high k values associated with high slope and TRI values.

Both U and Th tend to increase in highly elevated terrains. A similar pattern is shown in relation to the slope with a more significant increase in potassium with increasing slope values. This relation is not clear when comparing to the TRI values, as it seems that the TRI has a minimum control on the K distribution. Low terrains are associated with medium K and uranium and low thorium recordings.

U-Th plots show a relatively strong positive relationship between the two elements compared to K-U or K-Th relationships. Also, the U-Th plot shows a significant differentiation in their values between low (blue) and high terrains (yellow and red).

Here we present cross-plots (Suppl. V) to visualise the relationship between the radiometry (X-axis), topography-related data (Y-axis) and the bedrock distribution (colour distribution of the data points). There is a clear connection between the topography and bedrocks in the area. Younger hard acidic rocks (resistant to weathering and erosion) tend to form high terrains. Mourne complex and

the Ring of Gullion felsites are excellent examples of such rocks. On the other hand, younger mafic-rich volcanic bodies (contain more hydrous minerals and relatively weaker resistance to erosion) and old rocks form low grounds.

Young (Paleogene) granites and felsite rocks show high radiometry values, especially U and Th, and are associated with a higher altitude, steep slope and rugged grounds. In comparison, the (older) Newry granodiorite-diorite complex shows intermediate radiometry and occupies low grounds with more gentle and smoother grounds. Sandstones occupy most of the studied area and display a widespread variation in both radiometry and topography. Although the mafic rocks are associated with a relatively wide range of altitudes and slope profiles, they display a relatively low and narrow range of ruggedness values (Suppl. V I).

SOMs Results

Supplementary material VI illustrates the spatial distribution of the clusters, also compared to the K-mean cluster maps produced by the MLO software (Left side of the figures). The same set of colours was then used to map the spatial distribution of these clusters (the map sets are shown on the right side of the figure). The general distribution of the clusters is plotted using inverted pyramids of numbers (middle of the figures).
Although, the clusters showed a similar pattern to the general bedrock distribution in all the experiments, with the largest class representing between 25.66 - 58.38 % of the area (Suppl. Table 1 and Suppl. IV). In most cases, the largest class can be matched with the sandstone and the Newry granodiorite, while the smallest class is matched with the Mourne granite +/- the alluvial deposits (Suppl. Table 1). Some experiments are more successful than others in reflecting the detailed bedrock geology.

The 20 iterations are provided in supplementary figure VI. However, here we discuss key findings to demonstrate the influence of different features on the SOM.

Using SOMs of radiometry only

The SOMs of the radiometry data proved to be successful in capturing the difference between the acidic and basic igneous rocks in all the experiments used. Also, a significant differentiation between the Mourne and the Ring of Gullion is well observed. In most cases, the Newry Granodiorite is slightly segregated from the adjacent sandstones, especially using Th as an input of the experiment.

The presence or absence of the superficial deposits affects the radiometric signals, hence causing the re-arrangement of the data into different clusters. For example, the exposed outcrops of the sandstones (i.e. no superficial sediment cover) show lower radiometry compared to the concealed parts of the same bedrock; hence they are sometimes combined with the low-radiometry mafic rocks into one cluster (deep blue colours in Fig. 3 A). In the meantime, the exposed parts of the Newry granodiorite show higher radiometry than the concealed parts and are incorporated in the same clusters as the felsites of the Ring of Gullion.

The ring of Gullion outlines are well recognised when using K as input, while these boundaries became less significant when using U and Th alone. Using the three elements together as inputs enhances the recognition of the granitoids plutons outlines (both the Newry and Mourne plutons) and the ring of the Gullions (experiment IV).

Using K+U+Th as inputs, the Mourne granite can be differentiated into two parts: The eastern part is expressed by 2-3 clusters, usually dominated with high radiometry, while the western part shows lower radiometry values and is expressed by up to 5 clusters. Such classification is not observed using experiment II (K+U) or III (U+Th), see supplementary material VI.

Using SOMs of radiometry combined with geological features:

Adding the T_CPS to the radiometry (Fig. 3b) provides similar, yet enhanced, results compared to the cluster map where only elements are used. This experiment reveals high resolution within the Mourne granite (with different eastern and western lobes), good outlines of the Ring of Gullion and a fair outline of the Newry granitoid. There is also a good differentiation of alluvial deposits.
The alluvial deposits show a wide range of radiometry and are usually controlled by their derivation from specific bedrocks. The alluvial deposits near or derived from the mafic rocks and sandstone bedrocks show very low radiometry (here given the acronym A1, a dark blue strip near the northwest of most of the maps), see Fig. 3B. While the alluvial deposits sourced from the granodiorite (A2, yellow colour thin strip near the centre (Fig. 3B), over the Newry granodiorite) show intermediate radiometry compared to those of high radiometry derived from the granite (A3, pink lines south of the Mourne Granite). The latter shows high radiometric recording, similar to some parts of the Mourne granite itself (Fig. 3B). Details of the glacial deposits south of the Mourne granites show high radiometry and clustered together with the A2 alluvial and the felsite of the Ring of Gullion. These glacial deposits are underlain by the Hawick sandstone bedrocks.

When added, elevation data (DEM) dominates the clustering at intermediate radiometry levels, while DEM has a more negligible effect in areas showing high radiometry, i.e., over the Mourne Granite (Fig. 3c). The Newry complex and the western lobe of the Mourne are not well defined when using the topographic data. Highland peat has a diminishing effect on the radiometry signals when DEM data are incorporated.

Adding the slope enhances the capturing of more variations within the Mourne granites and the Ring of Gullion (Fig. 3D). Differentiation between the eastern and western lobe of the Mourne granites, also clear Newery eastern boundaries, can be observed when slope (and TRI) are incorporated.

Combining the age data with the radiometry as inputs have smoothened the bedrocks' outlines (Fig. 3E). However, fewer clusters were recognised within the plutons in the area but some variations in the clusters can be matched with different rock phases, e.g. the microgabbro within the Ring of Gullion. In the meantime, A2 and A3 alluvial sediments are not segregated, hence grouped in one cluster (yellow colour). Generally, a fair recognition of the alluvial deposits is observed when adding geological features to the radiometry, with a strong observation using experiment shown in Fig. 3E.

By incorporating the T_CPS and slope data (Fig. 3F), more details within the Mourne granites can be observed with a clear outline of the pluton and also the Ring of the Gullion. However, no obvious outline of the Newry complex is observed using this experiment.

Various aspects of bedrock and superficial geology are well represented using experiments showin in Fig. 3F. In the southwest, the circular Slieve Gullion Complex is well defined, shown in yellow colour. Antrim Lava Group basalts are displayed in dark blue in the northwest areas, clustered together with the NNW-SSE to N-S orientated fault-controlled tunnel valleys (alluvium deposits A1). In the northeast, the outermost zone of the Rathfriland pluton (the eastern part), the Newry

Igneous Complex, is differentiated in yellow and green, with an orange colour associated with the ultramafic-intermediate Seeconnell Complex, the eastern lobe of the Newry granite (Anderson et al. 2016). However, for the most part on this cluster map, there is no clear differentiation of Newry Igneous Complex granodiorite from its greywacke country rock, all of which appears grey and dark blue. In the centre of the map, an area of river alluvium appears yellow and purple, matching the colour of superficial deposits surrounding the Mourne Mountains Complex granites from which it was derived. The granites of the Mourne Mountains Complex appear as purple and red, particularly in the eastern lobe of the complex, whilst in the western lobe, the granites are known to be more mafic in composition and appear in orange colours. Finally, areas with peat or large water bodies appear dark blue. They include the tunnel valleys, inter-drumlin and ribbed moraine deposits and alluvium, and reservoirs in the Mournes.

Further iterations are provided in supplementary material VII, where 3 or 4 features are added to the radiometry. These generally demonstrate the effect of T_CPs in reducing the dominant impact of the DEM while the age may provide a better outline of the bedrocks. However, the age may cause adverse clustering of bedrock compositions together that are unlikely to be similar in composition.

Anatomy of SOMs clusters

Applying experiment XI, SOMs seem to have successfully represented both a reasonable delineation of and good details within the igneous masses. Figure 4 shows the anatomy of the data clustering where the relative percentages of these seven clusters are expressed by the pie chart (A), while the histograms (B) show the distribution of the elements, T_CPS and slope values within each cluster. A comparison to both granite phases and superficial sediment cover is provided in C, D and E.

Generally, data clustering reflects the differences among various phases of the granites (Fig. 4), with the eastern lobe being dominated by two smooth clusters and the more complex western lobe is expressed by a patch pattern of 5 clusters (B). Where superficial sediment cover is absent (orange colour on map E), high radiometry values (cluster #7) are prominent, especially in the eastern part of the plutons, while this is not well observed in the western region.

For the Mourne Mountains Complex (Fig 4I-C), there is a precise match between G2 and G3 granite types, particularly of the eastern lobe and cluster 7, which appear red (Fig. 4I-D). Areas of granite in both the western and eastern lobes that are overlain by diamicton (Fig. 4I-E) appear to fall into cluster 6, which appears purple. In the western lobe, an area mapped as mafic-rich G5 granite appears orange, which is cluster 3, though this cluster may also be controlled by changes in slope. Areas covered by peat or that have water reservoirs are represented by the dark blue cluster 2. The light blue of cluster 4 is associated with changes in slope, and this is observed in the western lobe where NNE-SSW orientated, fault-controlled valleys are present.

The dark blue of cluster 2 is related to areas of peat, inter-drumlin and ribbed moraine deposits and alluvium. In the northeast, where there is a significant exposure or close to the surface presence of the bedrocks (Fig. 4II-E), patchy multi-clustered appearances are observed. In this part, cluster 5 (yellow colour) shows the highest radiometry readings within the area underlain by the Newry complex. The ultramafic-intermediate Seeconnell Complex area (NEGD-DI) appears mostly in the orange of cluster 3, showing low radiometry values. However, this cluster may also be strongly influenced by the slope. The light blue of cluster 4 seems to be closely associated with changes in slope present around the northeast end of the Newry Igneous Complex and the Ring of Gullion in

the southwest. Although, a similar pattern is observed on maps 3 I-IV, where the radiometry used on their own (with no topography data) and the high radiometry recording in these areas are caused by the direct exposure of these rocks, i.e. no sediment cover to weaken the radiometry signals collected.

Discussion

Understanding the nature and distribution of the input data is critical when aiming to assess the SOMs clustering. For example, the elevation data are high-resolution datasets with almost continuous gradational coverage, unless data are collected over fault scarps or cliffs. The elevation data tend to dominate the clustering and, in many cases, overcome the effect of the radiometry and age on the data clustering. However, this is not the case where the radiometry is significantly high (strong enough to overwhelm the DEM data), resulting in a good representation of the Mourne granite in all the experiments.

Contrasting values of low and high radiometry contribute to a good (consistent) segregation and clustering of data, expressed by excellent differentiation between the acidic (Mourne) granite pluton and the basic (basalt) bedrocks. Differentiation between these two bedrock categories was observed in all the iterations (see supplementary material VII). In some cases, when the age is used, parts (but not all) of the Mourne granites and felsite are clustered together with the mafic-rich rocks.

Age indirectly influences the clustering, as the young acidic rocks in the area are resistant to weathering and tend to occupy high altitudes. Being in areas of high altitudes, these rocks are mostly uncovered or thinly covered by superficial sediments. Hence the radiometry signals collected are stronger than those from concealed acidic rocks nearby. Also, when elevation data is incorporated, the high topography, slope angles and raggedness contribute to the segregation of these exposed rocks into one or two clusters, distinguishing them from other bedrocks in the area.

In general, the SOMs maps discussed above reflect various aspects of the bedrock and superficial geology as well as topography. Superficial deposit signals mainly conceal the signals from moderate radiometry values, while the superficial deposits' effect is weak in areas with very high radiometry.

Due to the vast superficial deposit cover, the maps presented are most influenced by and reflect the nature and distribution of superficial deposits. These are less representative of the bedrock geology, and in most cases, they do not clearly distinguish the Newry Igneous Complex (showing moderate radiometry readings) from its surrounding greywacke country rock or differentiate the Gala from Hawick greywacke groupings, as these all show moderate radiometry values. Where radiometry data are either used on their own or combined with the T_CPS with or without additional geological feature, , the outline of the Newry granodiorite can be well-recognised.

The limited ability of the SOMs in differentiating the bedrock discussed above can be attributed to 1) the maps make use of radiometric data, which only 'sees' into the top few metres of the subsurface (Jones and Scheib, 2007); 2) across much of the region there is almost continuous coverage of superficial deposits concealing bedrocks.

Very strong signals from some alluvial (placer) deposits are clearly observed, usually expressed by the cluster colour yellow +/- pink (e.g. Fig. 3). These signals even overcome the signals from other features, including age. This is clear where radiometry data is combined with elevation and age data,. This again indicates that other geological features may have a diminishing effect in areas of high radiometry.

The age of bedrocks data was tested here in this work, and some experiments showed that the radiometry signals are locally strong and can overcome those from the age data. This raised the question of how the nature of the data can influence clustering, i.e. continuous versus discrete. Continuous data may hinder the clustering pattern compared to discrete data. Where the elevation and age data were added to the radiometry as inputs (see supplementary data VII), the altitude dominated the clusters within the Mourne granite and the Ring of Gullion, preventing the display of compositional variation of these igneous bodies.

The age and composition of the bedrocks may influence their appearances and their expression on surface geology and nowadays ground profile. For example, acidic volcanic rocks contain a high percentage of anhydrous, framework silicate minerals and tend to have an intact microcrystalline texture that was developed under relatively lower temperatures compared to their basic rock equivalents. Harder rocks tend to be more resistant to erosion and weathering, while younger rocks are less exposed to those superficial processes. Hence, hard young bedrocks commonly form high mountains with steep and rugged outcrops. In contrast, older rocks are exposed to a possibly hundreds of million years history of tectonics, uplifting, weathering and erosions. Such a long history could leave its marks through smoothening the surface, eroding much of the rock away and covering the rocks with younger superficial sediments. Such processes would significantly affect the hydrous ferromagnesian minerals within the old mafic rocks, being more susceptible to weathering.

Altitudes are generally homogenous continuous data that may be modified or interrupted by tectonic movement in the area. e.g. uplifting old mafic rocks, while slope degrees and ruggedness can be more related to the rock composition. Hence, slope angle and ruggedness data can be more associated with high-resolution variations in the bedrocks and potentially represent the mineralogical and geochemical signatures of the bedrocks. Hence, slope and TRI are good candidates to reflect the bedrock variations in our experiments compared to the altitude (DEM) data.

Different levels of bedrock differentiation could be aimed using the SOMs clustering: 1) differentiation between acidic, intermediate and basic rocks, 2) igneous versus sedimentary (no metamorphic rocks identified in the area), 3) various phases and groups within the same lithology type.

Depending on the data input combination and the nature of these data, one or two aspects of the above could be achieved at a time. For example, using the K, U and Th radiometric signals on their own seems to reflect the differences between the granite and granodiorite, but the basic rocks are not clearly seen. In the meantime, using U and Th resulted in a clustering that reflects internal variations in the Mourne granite, yet the pluton was well differentiated from the basic rock while the granodiorite was not distinguished from the sandstone. In most cases, when including the age and/or topography, the compositional variations within the Mourne granite are absent, yet the granite outline is well retained.

Conclusions

This work demonstrated how unsupervised classification with SOM can generate insights on bedrock mapping and make more efficient use of the multivariate data by integrating remote sensed radiometric survey with other soft geological information. SOM classification has provided a range of bedrock maps that depict the variation between spatial distributions and help approach its uncertainty. Furthermore, the analysis of SOM classification mapping revealed the impact of

different geological and radiometric features on the spatial distribution of the bedrock classes and revealed some inherent dependencies.

In general, the experiments in this research were able to distinguish and map distinctive geophysical signatures related to the various geological units that were identified on the published geological maps, especially acidic versus mafic bedrocks. However, fewer achievements were reached regarding the intermediate igneous and sedimentary bedrocks.

Using the radiometric elements on their own does not reveal the details of all the bedrocks in one experiment, while various combinations of experiments and the addition of further geological features enhance the data clustering and differentiation of bedrock geology. Incorporating K can clearly differentiate the acidic from basic bedrocks. However, it does not reveal details within each category.

Contrasting (low and high) values of radiometry supported better clustering, while intermediate values showed a weaker expression. SOMs were able to identify and enhance very subtle signatures from the radiometry that could be related to different phases of the granitoids, and a useful clustering was achieved using a combination of radiometry (elements + TCPS) and slope data. While elevation (DEM) data can adversely dominate the clustering (due to their continuous and gradual nature), slope and TRI data provided better results and enhanced clustering and bedrock mapping.

Adding T_CPs data reduced the adverse effect of the DEM and enhanced the representation of the bedrock geology. At the same time, the less continuous slope data (compared to the DEM) allows the radiometry to reflect more changes in lithology.

Consideration of the nature, resolution and homogeneity of the data is essential. For example, the continuous high-resolution elevation data can dominate the clustering; some of the clusters cannot be correlated to any specific lithology, especially when these data are solely added to the radiometry. Meanwhile, adding slope or TRI enhanced the details within the acidic igneous bedrocks as well as their outline, e.g., the Ring of Gullion and the Mourne granite.

Acquiring age data requires prior access to bedrock geology and knowledge of the bedrock. The choice of the rock age measurements can vary depending on the rock composition and nature, i.e., suitable techniques for the granites may not be suitable for black shale. This reduces the reliability of such data for unsupervised classification. Also, due to the discrete nature of the age data, they may result in coarse clustering and poor results in some cases.

The distribution of superficial deposits, including their absence, especially the glacial, peat, and alluvial deposits, greatly influenced the data clustering and the identification of bedrock signals.

Although the superficial covers may generally conceal the signals from the bedrock below, they can locally strengthen these signals, e.g. alluvial deposits A3, where radioactive minerals can be concentrated.

The recognition of specific superficial lithologies can present environmental importance. For example, peat coverage is important in calculating soil carbon storage in terrestrial environments (JNCC, 2011; McKinley et al., 2018). Generally, upland peat in the study area is usually well-differentiated compared to the lowland ones. This could be attributed to their contrasting low radiometry compared to the surrounding granites (high radiometry), i.e. high contrast leads to a

better clustering. At the same time, lowland peat is surrounded and underlain by rocks that show low-intermediate radiometry (sandstone or granodiorite). Also, upland peatlands form blankets that are more continuous in coverage than the lowland raised bogs (McKinley et al., 2018).

It is worth noting that the exposed sandstone shows lower values than those covered by superficial deposits. This is attributed to the patchy cover of the diamicton deposits. The latter could be sourced initially from different rocks nearby; hence the radiometry readings from the diamicton are not unique to the sandstone bedrocks but generally low to moderate.

The paper presents evidence of the ability of SOMs to integrate multiple data types and characterise bedrock and superficial deposits, especially peatlands and alluvial deposits. However, characterising the glacial deposits presents a challenge due to the mixed nature of these sediments. More research is required into using the technique to characterise glacial deposits, considering that glacial processes shaped the landscape of most of the UK. Incorporating additional Tellus data may be key and help better mapping of inaccessible areas. Also, the technique could be applied to resource exploration, such as mineralised contact zones and hydrothermal deposits that possess a unique geochemical signature, but that are problematic to distinguish using traditional field techniques.

Captions

Fig. 1: A) The southeast of Northern Ireland, part of the Southern Uplands–Down–Longford terrain (GSNI, 1997). B) Newry granodiorite-diorite complex. DI = ultramafic-intermediate complex (Seeconnell complex), GD1 = Northeast Rathfriland granodiorite pluton, GD2 = Central Newry granodiorite pluton, GD3 = Southwest Clohoge granodiorite pluton (Cooper et al., 2016). C): Mourne granitic igneous complex (GSNI, 1997). D): Digital elevation model (DEM) of the study area, draped with a semi-transparent hillshade overlay to show a near 3-dimensional appearance of the terrain. Dashed lines mark fault lines. E): Superficial sediment cover of the southeast of Northern Ireland (GSNI, 1997). Areas of exposed bedrock are mainly associated with the high ground in the southern and northeastern parts of the area.

Fig. 2: SOM layout using the potassium and thorium data sets. A) Hit map showing the number of hits of each neuron, B) U-matrix (distance matrix), C) the data slices of the potassium, D) the clusters with specific colour coding, E) data slices of the uranium, F), the spatial distribution of the clusters using the same colour codes from E, G) geological map of the bedrock geology for comparison with the cluster's spatial distribution.

Fig. 3: Examples of combining the radiometry data with geological features. Colours are designed to be consistent as much as possible and to match bedrock and superficial rock features. K = potassium, U = Uranium, Th = Thorium, TCPs = total counts per second, DEM = altitude, Age = bedrock age, Slope = slope angle.

Fig. 4: I) Anatomy of the SOM clusters using experiment XI, compared to the bedrock and superficial geology in the area underlain by the Mourne Granite. A) percentage of clusters 1-7; B) Distribution of radiometry, T_CPS, and slope data among various data clusters; C) geological map of Mourne granite after (GSNI, 1997), D) SOM of the data over the Mourne Granite; E) superficial sedimentary deposits over the Mourne granite (GSNI, 1991).

II: Anatomy of the SOM clusters using experiment XI, compared to the bedrock and superficial geology in the area underlain by the Newry granodiorite-diorite complex. A) percentage of clusters 1-7; B) Distribution of radiometry, T_CPS and slope data among various clusters; C) geological map of Newry igneous complex after (GSNI, 1997), D) SOM of the data over the Newry igneous complex; E) superficial sedimentary cover over the Newry igneous complex (GSNI, 1991).

References

Aldiss, D.T., 2013. Under-representation of faults on geological maps of the London region: reasons, consequences and solutions. Proceedings of the Geologists' Association 124, 929–945. https://doi.org/10.1016/j.pgeola.2013.03.001

Anderson, H., Walsh, J.J., Cooper, M.R., 2018. The development of a regional-scale intraplate strike-slip fault system; Alpine deformation in the north of Ireland. Journal of Structural Geology 116, 47–63. https://doi.org/10.1016/j.jsg.2018.07.002

Anderson, P.E., Cooper, M.R., Stevenson, C.T., Hastie, A.R., Hoggett, M., Inman, J., Meighan, I.G., Hurley, C., Reavy, R.J., Ellam, R.M., 2016. Zonation of the Newry Igneous Complex, Northern Ireland, based on geochemical and geophysical data. Lithos 260, 95–106. https://doi.org/10.1016/j.lithos.2016.05.009

Anderson, T.B., 2004. Southern Uplands – Down-Longford Terrane, in: Mitchel, W.I. (Ed.), The Geology of Northern Ireland: Our Natural Foundation. Geological Survey of Northern Ireland, Belfast, pp. 41–60.

Bachri, I., Hakdaoui, M., Raji, M., Teodoro, A.C., Benbouziane, A., 2019. Machine Learning Algorithms for Automatic Lithological Mapping Using Remote Sensing Data: A Case Study from Souk Arbaa Sahel, Sidi Ifni Inlier, Western Anti-Atlas, Morocco. IJGI 8, 248. https://doi.org/10.3390/ijgi8060248

Baldini, D., Piazza, L., Barbanotti, L., 2020. Artificial Intelligence and Machine Learning Techniques Provide Operations Geologists With an Automated and Reliable Lithology-Fluid Pattern Recognition Assistant: A Case History in a Clastic Reservoir in West Africa, in: Day 2 Tue, January 14, 2020. IPTC, Dhahran, Kingdom of Saudi Arabia, p. D021S031R001. https://doi.org/10.2523/IPTC-19701-MS

Beamish, D., Cuss, R.J., Lahti, M., Scheib, C., Tartaras, E., 2006. The Tellus airborne geophysical survey of Northern Ireland : final processing report [WWW Document]. URL http://www.bgs.ac.uk (accessed 3.31.21).

Bedini, E., 2009. Mapping lithology of the Sarfartoq carbonatite complex, southern West Greenland, using HyMap imaging spectrometer data. Remote Sensing of Environment 113, 1208–1219. https://doi.org/10.1016/j.rse.2009.02.007

Bhamare, D., Suryawanshi, P., 2018. Review on Reliable Pattern Recognition with Machine Learning Techniques. Fuzzy Information and Engineering 10, 362–377. https://doi.org/10.1080/16168658.2019.1611030

Bhattacharya, S., Carr, T.R., Pal, M., 2016. Comparison of supervised and unsupervised approaches for mudstone lithofacies classification: Case studies from the Bakken and Mahantango-Marcellus Shale, USA. Journal of Natural Gas Science and Engineering 33, 1119–1133. https://doi.org/10.1016/j.jngse.2016.04.055

Bierlein, F.P., Fraser, S.J., Brown, W.M., Lees, T., 2008. Advanced methodologies for the analysis of databases of mineral deposits and major faults. Australian Journal of Earth Sciences 55, 79–99. https://doi.org/10.1080/08120090701581406

Carneiro, C. de C., Fraser, S.J., Crósta, A.P., Silva, A.M., Barros, C.E. de M., 2012. Semiautomated geologic mapping using self-organizing maps and airborne geophysics in the Brazilian Amazon. GEOPHYSICS 77, K17–K24. https://doi.org/10.1190/geo2011-0302.1

Cooper, M., Anderson, P., Condon, D., Stevenson, C., Ellam, R., Meighan, I., Crowley, Q., STEWART, I., 2016. Shape and intrusion history of the Late Caledonian Newry Igneous Complex, Northern Ireland, in: YOUNG, M. (Ed.), Unearthed, Impacts of the Tellus Surveys of the North of Ireland. Royal Irish Academy, pp. 145–156.

Cooper, M.R., 2004. Palaeogene extrusive igneous rocks., in: Mitchell, W.I. (Ed.), The Geology of Northern Ireland - Our Natural Foundation. Geological Survey of Northern Ireland, Belfast, pp. 167–178.

Cooper, M.R., Anderson, H., Walsh, J.J., Van Dam, C.L., Young, M.E., Earls, G., Walker, A., 2012. Palaeogene Alpine tectonics and Icelandic plume-related magmatism and deformation in Northern Ireland. Journal of the Geological Society 169, 29–36. https://doi.org/10.1144/0016-76492010-182

Cooper, M.R., Johnston, T.P., 2004a. Late Palaeozoic Intrusives, in: Mitchell, W.I. (Ed.), The Geology of Northern Ireland - Our Natural Foundation. Geological Survey of Northern Ireland, Belfast, pp. 167–178.

Cooper, M.R., Johnston, T.P., 2004b. Palaeogene intrusive igneous rocks, in: Mitchell, W.I. (Ed.), The Geology of Northern Ireland - Our Natural Foundation. Geological Survey of Northern Ireland, Belfast, pp. 167–178.

Copernicus Land Monitoring, 2017. Copernicus Land Monitoring Service - EU-DEM — European Environment Agency [WWW Document]. URL https://www.eea.europa.eu/data-and-maps/data/copernicus-land-monitoring-service-eu-dem (accessed 9.26.20).

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 & Geosciences 63, 22–33. https://doi.org/10.1016/j.cageo.2013.10.008

Cracknell, M.J., Reading, A.M., McNeill, A.W., 2014. Mapping geology and volcanic-hosted massive sulfide alteration in the Hellyer–Mt Charter region, Tasmania, using Random Forests™ and Self-Organising Maps. Australian Journal of Earth Sciences 61, 287–304. https://doi.org/10.1080/08120099.2014.858081

DeJong, M.J., Schooling, J.M., Viggiani, G.M.B., Cambridge Centre for Smart Infrastructure and Construction (Eds.), 2019. Driving data-informed decision-making: International Conference on Smart Infrastructure and Constructions 2019 (ICSIC). Presented at the International Conference on Smart Infrastructure and Construction, ICE Publishing, London.

Demyanov, V., Gloaguen, E., Kanevski, M., 2020. A Special Issue on Data Science for Geosciences. Math Geosci 52, 1–3. https://doi.org/10.1007/s11004-019-09846-0

Demyanov, V., Reesink, A.J.H., Arnold, D.P., 2019. Can machine learning reveal sedimentological patterns in river deposits? Geological Society, London, Special Publications 488, 221–235. https://doi.org/10.1144/SP488-2018-84

Donald, A., 2016. Portal | GSNI Tellus Regional Airborne Geophysical Survey - Radiometrics. The Geological Survey of Northern Ireland (GSNI).

Elbegue, A.A., Allek, K., Zeghouane, H., 2022. Geological mapping using extreme gradient boosting and the deep neural networks: application to silet area, central Hoggar, Algeria. Acta Geophysics, Springer. https://doi.org/10.1007/s11600-022-00814-7

Ferraretti, D., Gamberoni, G., Lamma, E., 2012. Unsupervised and supervised learning in cascade for petroleum geology. Expert Systems with Applications 39, 9504–9514. https://doi.org/10.1016/j.eswa.2012.02.104

Fraser, S., Dickson, B.L., 2007. A New Method for Data Integration and Integrated Data Interpretation : Self-Organising Maps, in: 5th Decennial International Conference on Mineral Exploration, Expanded Abstracts. pp. 907–910.

Fraser, S.J., Wilson, G.A., Cox, L.H., Čuma, M., Zhdanov, M.S., Vallée, M.A., 2012. Self-organizing maps for pseudo-lithological classification of 3D airborne electromagnetic, gravity gradiometry and magnetic inversions. ASEG Extended Abstracts 2012, 1–4. https://doi.org/10.1071/ASEG2012ab086

Gifford, C.M., Agah, A., 2010. Collaborative multi-agent rock facies classification from wireline well log data. Engineering Applications of Artificial Intelligence 23, 1158–1172. https://doi.org/10.1016/j.engappai.2010.02.004

Gonçalves, M.L., Netto, M.L.A., Costa, J.A.F., Zullo Júnior, J., 2008. An unsupervised method of classifying remotely sensed images using Kohonen self-organizing maps and agglomerative hierarchical clustering methods. International Journal of Remote Sensing 29, 3171–3207. https://doi.org/10.1080/01431160701442146

GSNI, G.S. of N.I., 1997. Geological Map of Northern Ireland (Solid).

GSNI, G.S. of N.I., 1991. Geological Map of Northern Ireland (Quaternary).

Halotel, J., Demyanov, V., Gardiner, A., 2020. Value of Geologically Derived Features in Machine Learning Facies Classification. Math Geosci 52, 5–29. https://doi.org/10.1007/s11004-019-09838-0

Harris, J.R., Grunsky, E.C., 2015. Predictive lithological mapping of Canada's North using Random Forest classification applied to geophysical and geochemical data. Computers & Geosciences 80, 9–25. https://doi.org/10.1016/j.cageo.2015.03.013

Harvey, A.S., Fotopoulos, G., 2016. GEOLOGICAL MAPPING USING MACHINE LEARNING ALGORITHMS. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. XLI-B8, 423–430. https://doi.org/10.5194/isprsarchives-XLI-B8-423-2016

Hodgson, J., Young, M., 2016. The Tellus airborne geophysical surveys and results, in: Young, M. (Ed.), Unearthed: Impacts of the Tellus Surveys of the North of Ireland. Royal Irish Academy. https://doi.org/10.2307/j.ctt1g69w6r

IAE Agency, 2003. Guidelines for radioelement mapping using gamma ray spectrometry data. International Atomic Energy Agency, Vienna.

JNCC, 2011. Towards an assessment of the state of UK peatlands (No. 445). Joint Nature Conservation Committee.

Jones, D.G., Scheib, C., 2007. A preliminary interpretation of Tellus airborne radiometric data [WWW Document]. URL http://nora.nerc.ac.uk/id/eprint/16872/ (accessed 9.27.20).

Kanevski, M., 2009. Machine Learning for Spatial Environmental Data: Theory, Applications, and Software, 0 ed. EPFL Press. https://doi.org/10.1201/9781439808085

Kirkwood, C., Economou, T., Pugeault, N., Odbert, H., 2022. Bayesian Deep Learning for Spatial Interpolation in the Presence of Auxiliary Information. Math Geosci 54, 507–531. https://doi.org/10.1007/s11004-021-09988-0

Kohonen, T., 1982. Self-organized formation of topologically correct feature maps. Biol. Cybern. 43, 59–69. https://doi.org/10.1007/BF00337288

Kohonen, T., Huang, T.S., Schroeder, M.R., 2012. Self-Organizing Maps. Springer Berlin / Heidelberg, Berlin, Heidelberg.

Lark, R.M., Lawley, R.S., Barron, A.J.M., Aldiss, D.T., Ambrose, K., Cooper, A.H., Lee, J.R., Waters, C.N., 2015. Uncertainty in mapped geological boundaries held by a national geological survey:eliciting the geologists' tacit error model. Solid Earth 6, 727–745. https://doi.org/10.5194/se-6-727-2015

Li, Z., Eastman, J.R., 2010. Commitment and typicality measures for the Self-Organizing Map. International Journal of Remote Sensing 31, 4265–4280. https://doi.org/10.1080/01431160903246725

Lin, J., Williamson, S., Borne, K., Debarr, A., 2012. Pattern Recognition in Time Series, in: Srivastava, A. (Ed.), Advances in Machine Learning and Data Mining for Astronomy, Chapman & Hall/CRC Data Mining and Knowledge Discovery Series. Chapman and Hall/CRC. https://doi.org/10.1201/b11822-36

McCabe, A.M., Dunlop, P., 2006. The last glacial termination in Northern Ireland. Geological Survey of Northern Ireland, Belfast.

McKinley, J.M., Grunsky, E., Mueller, U., 2018. Environmental Monitoring and Peat Assessment Using Multivariate Analysis of Regional-Scale Geochemical Data. Math Geosci 50, 235–246. https://doi.org/10.1007/s11004-017-9686-x

Mitchell, W.I. (Ed.), 2004. The geology of Northern Ireland: our natural foundation, 2. ed. ed. Geological Survey of Northern Ireland, Belfast.

Penn, B.S., 2005. Using self-organizing maps to visualize high-dimensional data. Computers & Geosciences 31, 531–544. https://doi.org/10.1016/j.cageo.2004.10.009

Pires, A.C.B., Harthill, N., 1989. Statistical analysis of airborne gamma-ray data for geologic mapping purposes: Crixas-Itapaci area, Goias, Brazil. GEOPHYSICS 54, 1326–1332. https://doi.org/10.1190/1.1442592

Sebtosheikh, M.A., Salehi, A., 2015. Lithology prediction by support vector classifiers using inverted seismic attributes data and petrophysical logs as a new approach and investigation of training data set size effect on its performance in a heterogeneous carbonate reservoir. Journal of Petroleum Science and Engineering 134, 143–149. https://doi.org/10.1016/j.petrol.2015.08.001

Shivakumar, B.R., Rajashekararadhya, S.V., 2020. Classification of Landsat 8 Imagery Using Kohonen's Self Organizing Maps and Learning Vector Quantization, in: Kalya, S., Kulkarni, M., Shivaprakasha, K.S. (Eds.), Advances in Communication, Signal Processing, VLSI, and Embedded Systems, Lecture Notes in Electrical Engineering. Springer Singapore, Singapore, pp. 445–462. https://doi.org/10.1007/978-981-15-0626-0_35

Silversides, K., Melkumyan, A., Wyman, D., Hatherly, P., 2015. Automated recognition of stratigraphic marker shales from geophysical logs in iron ore deposits. Computers & Geosciences 77, 118–125. https://doi.org/10.1016/j.cageo.2015.02.002

Stevenson, C.T.E., Owens, W.H., Hutton, D.H.W., Hood, D.N., Meighan, I.G., 2007. Laccolithic, as opposed to cauldron subsidence, emplacement of the Eastern Mourne pluton, N. Ireland: evidence from anisotropy of magnetic susceptibility. Journal of the Geological Society 164, 99–110. https://doi.org/10.1144/0016076492006-008

Toulia, E., Kokinou, E., Panagiotakis, C., 2018. The contribution of pattern recognition techniques in geomorphology and geology: the case study of Tinos Island (Cyclades, Aegean, Greece). European Journal of Remote Sensing 51, 88–99. https://doi.org/10.1080/22797254.2017.1405716

Young, M., Donald, A., 2013. A guide to the Tellus data. Geological Survey of Northern Ireland, Belfast, UK.

Fig. 1

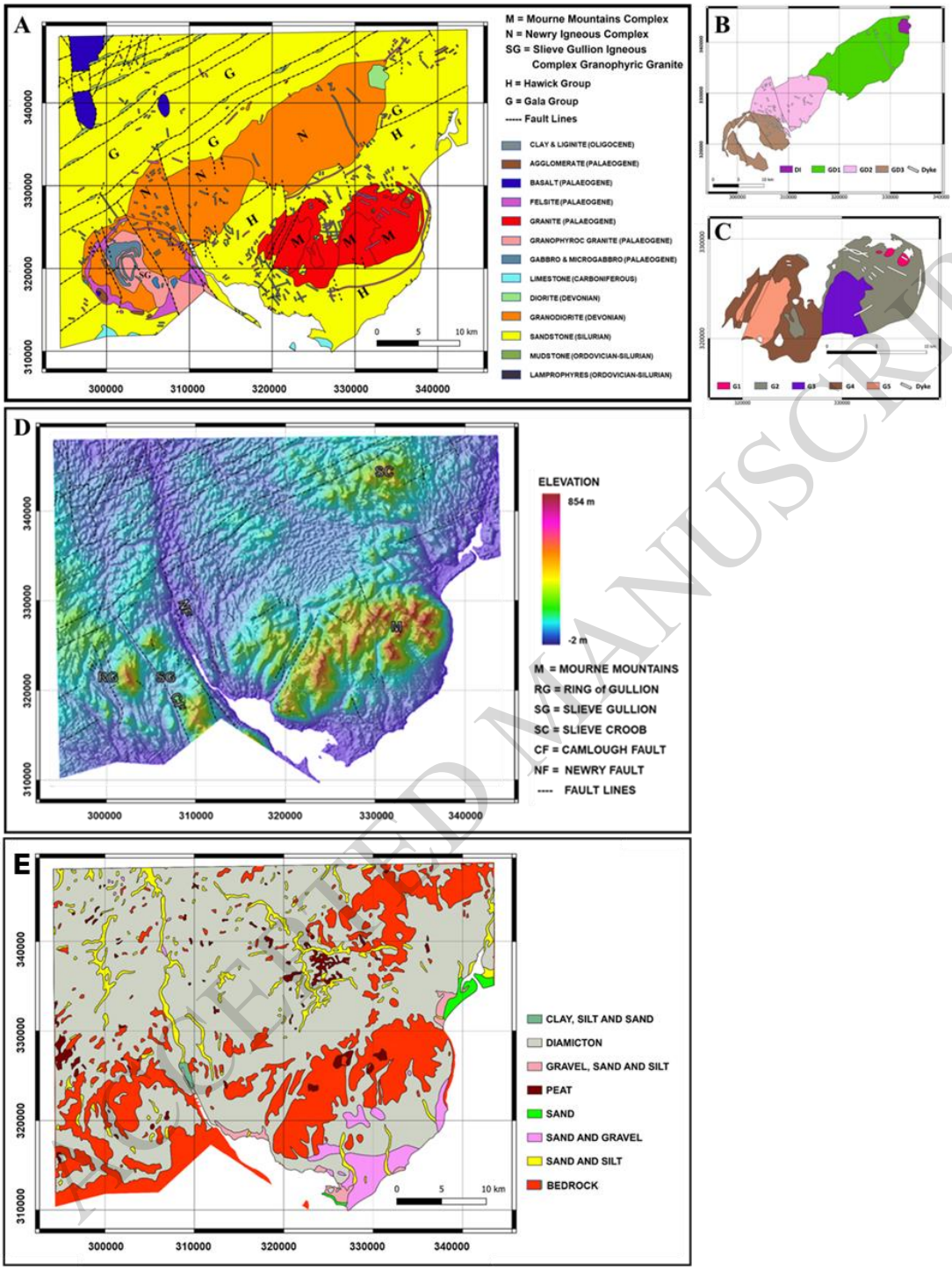

Fig. 2

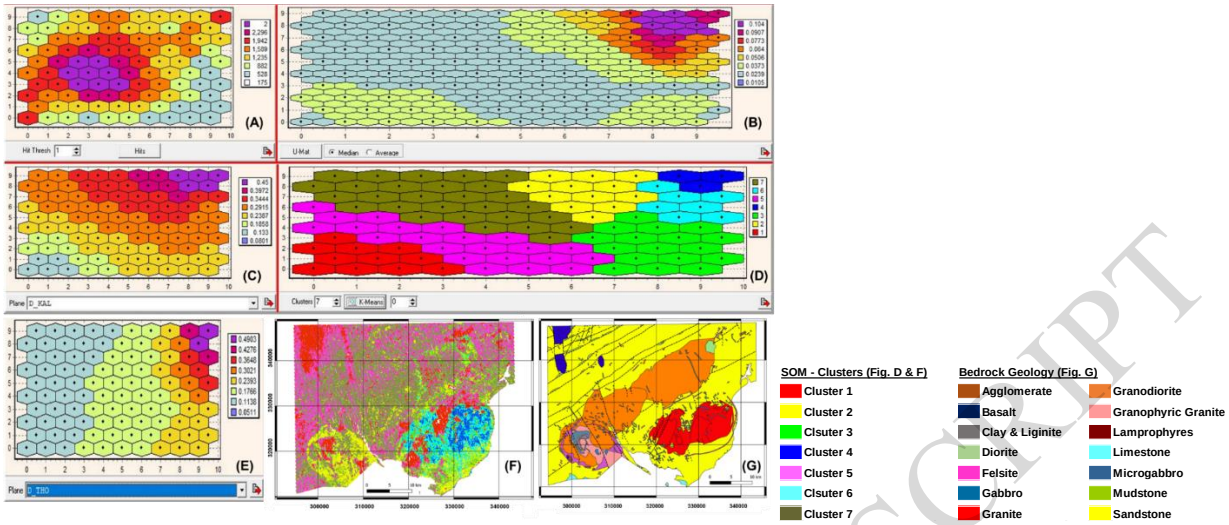

Fig. 3

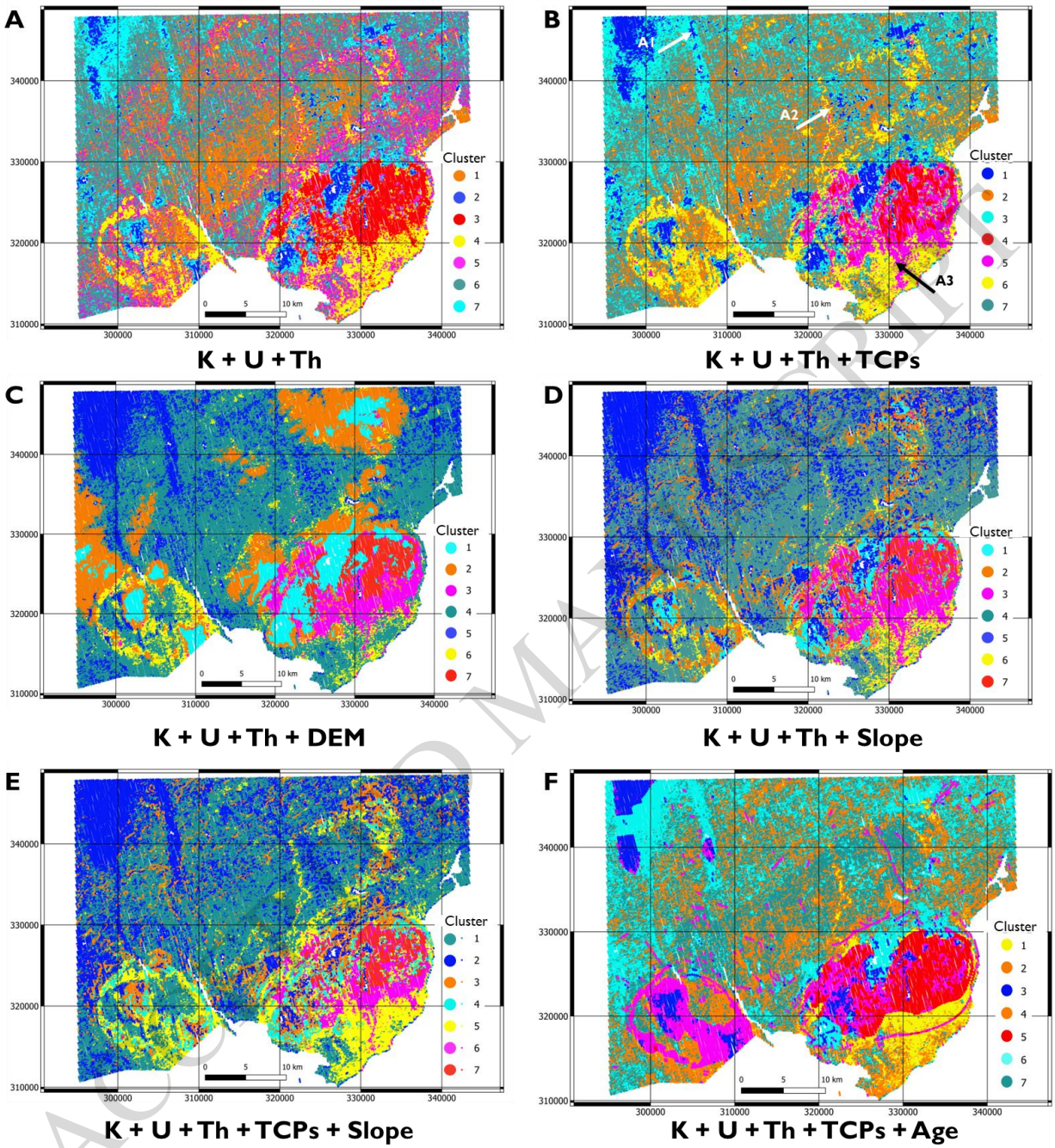

Fig. 4 - I

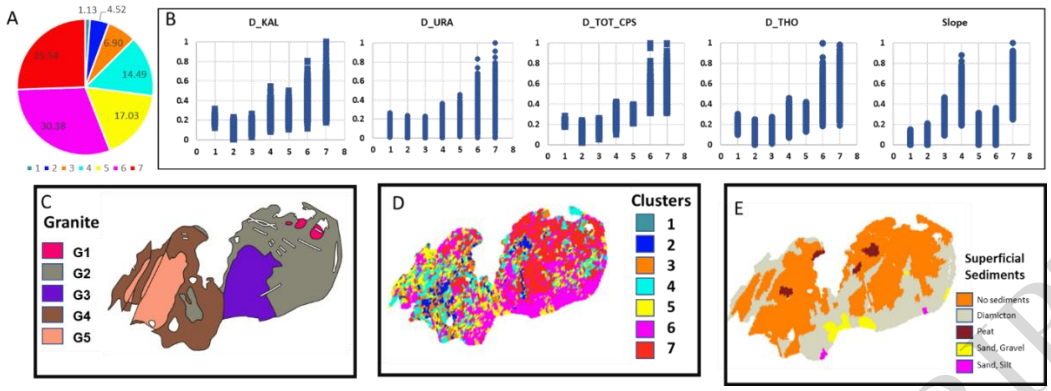

Fig. 4 - II

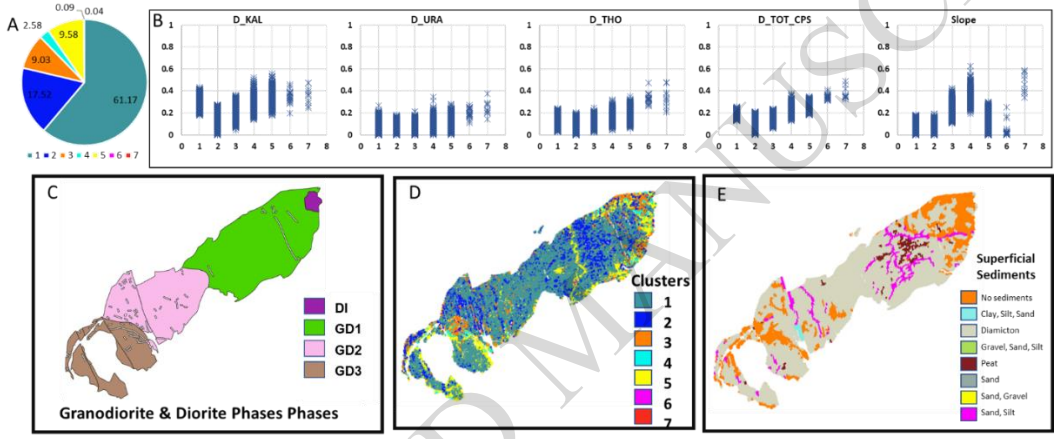

Table 1. *List of the experiments used to test the contribution of various radiometric and geological features in reflecting the bedrock distribution. D_KA L: potassium counts, D_URA= uranium counts, D_THO: thorium counts, TCPS: total counts of elements per second, DEM: digital elevation modelling, TRI: terrain ruggedness index, Slope: slope angle of the terrains, Age: age of the bedrocks.*

Type	No	Input #1	Input #2	Input #3	Input #4	Input #5	Input #6	Input #7
Radiometry Only	1	D_KAL	D_URA					
	2	D_KAL	D_THO					
	3	D_URA	D_THO					
	4	D_KAL	D_URA	D_THO				
Radiometry + 1 Feature	5	D_KAL	D_URA	D_THO	TCPS			
	6	D_KAL	D_URA	D_THO	DEM			
	7	D_KAL	D_URA	D_THO	Slope			
	8	D_KAL	D_URA	D_THO	TRI			
	9	D_KAL	D_URA	D_THO	Age			
Radiometry + 2 Features	10	D_KAL	D_URA	D_THO	TCPS	DEM		
	11	D_KAL	D_URA	D_THO	TCPS	Slope		
	12	D_KAL	D_URA	D_THO	TCPS	TRI		
	13	D_KAL	D_URA	D_THO	DEM	Age		
	14	D_KAL	D_URA	D_THO	TRI	Age		
	15	D_KAL	D_URA	D_THO	Slope	Age		
	16	D_KAL	D_URA	D_THO	TCPS	Age		
Radiometry + 3 Features	17	D_KAL	D_URA	D_THO	TCPS	DEM	Age	
	18	D_KAL	D_URA	D_THO	TCPS	Slope	Age	
	19	D_KAL	D_URA	D_THO	TCPS	TRI	Age	
Radiometry + 4 features	20	D_KAL	D_URA	D_THO	TCPS	DEM	Slope	Age