Ore Geology Reviews 142 (2022) 104714

Contents lists available at ScienceDirect

Ore Geology Reviews

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

Semi-automated geological mapping and target generation from geochemical and magnetic data in Halkidiki region, Greece

Simo Piippo [a,b,*], Martiya Sadeghi [c], Emilia Koivisto [a], Pietari Skyttä [b], Tim Baker [d]

[a] University of Helsinki, Department of Geosciences and Geography, University of Helsinki, FI-00014, Finland
[b] University of Turku, Department of Geography and Geology, University of Turku, FI-20014, Finland
[c] Geological Survey of Sweden, Box 670, Uppsala SE-751 28, Sweden
[d] Eldorado Gold Corporation, 1188 Bentall 5-550 Burrard St., Vancouver, British Columbia V6C 2B5, Canada

ARTICLE INFO

Keywords:
SOM
PCA
Porphyry
Exploration
Geological mapping
Target generation

ABSTRACT

This study focuses on using multivariate analyses to generate semi-automated geological maps and exploration targets associated with porphyry Au-Cu mineralization within the Kassandra mining district, Greece. We use principal component analysis (PCA) and self-organizing maps (SOM) to reveal variations in geochemical and magnetic signatures within the input datasets. We visualize the results as pseudo-geological maps reflecting the associated geological processes and their end products. In specific, we utilize the potential of these two methods through an integrated interpretation and comparison of the results. We test the validity of the unsupervised PCA- and SOM-derived lithological and prospectivity models by comparing them with existing geological observations and interpretations. The results of this investigation show that both PCA and SOM are able to reproduce the key features of existing geological observations within the study area, but more importantly, also provide useful information that can be used to recognize prospective geological units and exploration targets from previously unknown locations.

1. Introduction

The primary purpose of multivariate analysis in geoscience is, based on statistical and spatial variations within a dataset, to identify specific geochemical or geophysical signatures associated with the end results of specific geological processes of interest. Multivariate analyses approach the challenge of recognizing even the subtlest relations or patterns within the data using dimensionality reduction and calculating measures of vector similarities. These methods characteristically enable integrated analysis of different types of data, such as geological, geochemical, geophysical or spectral data. Processes of interest include for example fractional crystallization, alteration, and ore-formation, which result in generation of contrasting geological units or variation of properties within the units. Moreover, the recognized patterns may represent a combination of features such as primary lithology overprinted by metamorphism, alteration, weathering and mineralization, and as such constitute the basis for semi-automated geological mapping and exploration targeting.

Principal component analysis (PCA) is a type of conventional linear multivariate method that allows the user to reduce a large number of dimensions (variables) into a relatively small number of principal components (groups of variables) which account for most of the variation within the data (e.g. Loughlin, 1991). PCA is frequently used in mineral exploration applications, especially for interpretations and anomaly detection from multi-element geochemical data (Harris et al., 1997; Carranza, 2008; Grunsky, 2010; Cheng et al., 2011; Grunsky and de Caritat, 2019; Sadeghi et al., 2013a, Sadeghi et al., 2013b), due to its relative simplicity and the availability of a wide variety of software applications and tools. Moreover, through spatial band composite visualization of three selected principal components (PC) for red, green and blue bands (RGB), PCA can be used for effective generation of pseudo-geological representations from different types of georeferenced data. This method is commonly used for creating pseudo-lithological maps by combining the information from several spectral bands in remote sensing data (e.g. Loughlin, 1991). In contrast to selecting a few characterizing elements or element associations for the RGB bands (e.g. Sadeghi et al., 2015), the use of PCs enables simultaneous visualization of the majority of variation within the data. However, the nuances are often less apparent compared to discrete classification of the data, and the applicability of PCA and other conventional linear methods is also

* Corresponding author at: University of Helsinki, Department of Geosciences and Geography, University of Helsinki, FI-00014, Finland.
E-mail address: simo.v.piippo@utu.fi (S. Piippo).

https://doi.org/10.1016/j.oregeorev.2022.104714
Received 1 December 2020; Received in revised form 14 December 2021; Accepted 11 January 2022
Available online 18 January 2022

reduced by the nonlinear character of most natural datasets (e.g. Park et al., 2003).

By contrast, the self-organizing map (SOM; e.g. Kohonen, 1982; Kohonen, 2013) is a type of unsupervised artificial neural network capable of mapping non-linear relationships within multivariate input data. SOM has earlier been shown to produce e.g. pseudo-lithological classifications resembling the mapped geology associated with diverse geological settings and datasets (Carneiro et al., 2012; Fraser et al., 2012; Kuhn et al., 2019). Overall, SOM has been successfully utilized for a wide variety of applications within mineral exploration (e.g. Fraser and Dickson, 2007) to improve understanding of the geological processes responsible for the geochemical (e.g. Bierlein et al., 2008; Cracknell et al., 2014; Leväniemi et al., 2016; Cracknell and de Caritat, 2017) and geophysical (e.g. Junno et al., 2019; Junno et al., 2020) signals. Previous studies using the SOM for pseudo-lithological maps have typically used a limited number of similar types of variables (e.g. only geophysical) as input data, or have involved simplified geological and/or pseudo-geological interpretations, i.e. a small number of clusters (e.g. Carneiro et al., 2012; Fraser et al., 2012). Consequently, SOM has rarely been applied to generate pseudo-geological maps, especially from data with extensive set of variables of different types, and in relatively complex geological settings (Carneiro et al., 2012; Fraser et al., 2012; Kuhn et al., 2019). Therefore, practical evaluation of its usability for such purpose has remained incomplete.

In this study we show that integrated use of the linear (PCA) and nonlinear (SOM) multivariate methods are successful in generating pseudo-geological representations and new exploration targets. As the principal dataset of the study, we use soil geochemistry together with high-resolution low-altitude aeromagnetic data from the Kassandra mining district of Halkidiki region, northern Greece (Fig. 1; e.g. Heinrich and Neubauer, 2002). In particular, the study focusses on the porphyry Au-Cu prospects, which are associated with Oligocene-Miocene intrusions emplaced into polydeformed metamorphic basement rocks of Permo-Carboniferous to Late Jurassic age (Siron et al., 2018). The area is largely covered by soil, which often prevents direct outcrop observations. However, similar environments in Greece show very limited sediment transportation and deposition outside the influence of fluvial processes within the river valleys (e.g. Kanellopoulos and Argyraki, 2013), and the geochemical signatures of the soil samples correlate well with the drill core samples from the underlying bedrock (Tsitsanis et al., 2016). For these reasons, the soil samples used in this study can be used for detailed geological interpretations and exploration targeting in the area.

Overall, we use PCA and SOM to reduce the dimensions of the input data from the Kassandra mining district, and to quantify and spatially visualize the similarity of the input samples: we use PCs from the PCA to spatially visualize the input geochemical soil samples as continuous RGB composite images, and clustered SOMs for spatial visualization by classifying the input geochemical soil and aeromagnetic data into discrete classes corresponding to the clusters. Moreover, we compose several selections of input datasets for the SOM analysis to visualize the results, and to compare the use of original sample points with data upscaled by interpolation. We validate the produced maps against the existing geological interpretations (Kockel et al., 1975; Siron et al., 2018 and references therein; unpublished work by Hellas Gold), and further use the PCs and SOM clusters to interpret the geological processes and

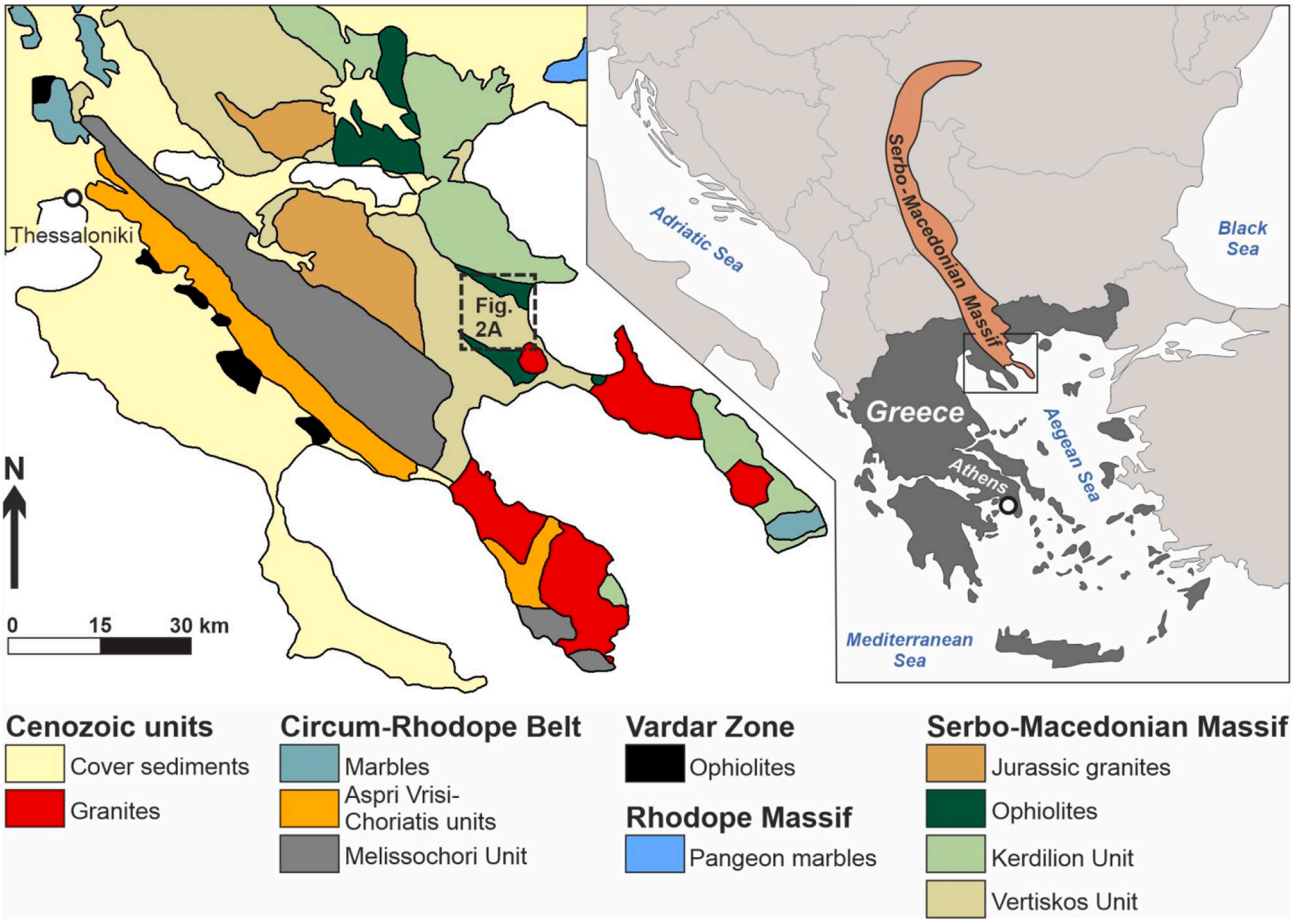

Fig. 1. The location of Serbo-Macedonian Massif (simplified from Schmid et al., 2008; and van Hinsbergen and Schmid, 2012) and the Halkidiki peninsula on the right. Geological interpretation of the Halkidiki peninsula (modified after Melfos and Voudouris, 2012) and the location of the study area in Kassandra mining district (Fig. 2a) on the left.

their end products from the distinct geochemical and/or magnetic associations recognized within the dataset.

2. Geological setting

The NW-trending Serbo-Macedonian Massif (Fig. 1) is divided into two major lithostratigraphic units of Paleozoic age, the Permo-Carboniferous to Late Jurassic Kerdilion unit and the Ordovician-Silurian Vertiskos unit (Kockel et al., 1977, according to Siron et al., 2018). At the ground level in the Kassandra mining district, the contact zone between the Kerdilion and Vertiskos units is largely deformed and displaced by the major south-dipping and E-W-striking Stratoni fault zone (SFZ, Fig. 2a), previously inferred as a part of the Middle Eocene Kerdilion detachment separating the two units (Haines, 1998, according to Siron et al., 2018; Kounov et al., 2015). Study by Siron et al. (2018) suggests that the structure between Kerdilion and Vertiskos units may be of earlier origin and not related to the Kerdilion detachment, and that the SFZ is a normal fault active since the Middle Eocene. The Kerdilion unit to the north comprises biotite gneiss interbedded with hornblende gneiss, amphibolite and marble, probably derived from an original volcano-sedimentary sequence. The Vertiskos unit, presently overlying the Kerdilion unit, consists mainly of two-mica gneiss, and in places contains *ortho*-amphibolite layers and garnet-, tourmaline- and sillimanite-bearing quartz-feldspar units. Metamorphosed mafic and ultramafic amphibolites occur broadly near the contact between the Kerdilion and Vertiskos units, mostly within the hanging wall of the SFZ (e.g. Siron et al., 2018). The amphibolites may be correlated with the Gomati ophiolite body in the southern border of the study area (Siron et al., 2018). Similar rocks are also recognized within the Vertiskos unit south from the contact zone of Vertiskos and Kerdilion units as well as the SFZ (Fig. 2a; unpublished work by Hellas Gold).

Two main types of economically significant ore deposits associated with Oligocene-Miocene intrusions (e.g. Heinrich and Neubauer, 2002; Arvanitidis, 2003) have been recognized within or close to the metamorphosed and deformed country rocks of the study area: porphyry Au-Cu deposits and polymetallic carbonate replacement deposits. The single known porphyry Au-Cu deposit of the area (Skouries deposit: Frei, 1995; Sullivan et al., 2018), a high-K calc-alkaline to shoshonitic porphyry stock associated with zonal alteration from inner potassic zone to outer propylitic zone (e.g. Frei, 1995; Siron et al., 2016), is located outside the study area close to its western border. The 25 to 27 Ma diorite and granodiorite stocks of the area, associated with quartz-sericite-pyrite alteration (Siron et al., 2018), are inferred as possible igneous sources for the mineralizing fluids of the nearby carbonate replacement deposits (e.g. Mavres Petres deposit: Siron et al., 2016).

The Skouries ("Sk" in Fig. 2a) Au-Cu porphyry deposit (Frei, 1995; Sullivan et al., 2018) is a Miocene age, pipe-shaped, high-K calc-alkaline to shoshonitic porphyry stock intruding amphibolite and biotite-chlorite schist country rocks. The deposit is characterized by concentric alteration zones encompassing an inner potassic zone, with stockwork quartz veinlets and an outer propylitic zone, affecting mostly the host schists. Weak phyllic and argillic alteration is confined to vein haloes and faults. Mineralization within the potassic zone primarily comprises chalcopyrite veinlets with subordinate bornite and disseminated chalcopyrite and bornite. Mineralization within the propylitic zone contains disseminated pyrite, and rare molybdenite and chalcocite. Gold mineralization occurs commonly as blebs within sulphides and less often as native gold associated with gangue minerals. An oxide zone occurs from surface to 30–50 m depth and includes malachite, cuprite, secondary chalcocite and minor azurite, covellite, digenite and native copper.

An example of the carbonate replacement deposits of the Kassandra mining district is the Mavres Petres deposit ("M" in Fig. 2a), where the sulphide orebodies are hosted in the complexly deformed lenses of marble within the SFZ (Fig. 2a; Arvanitidis, 1993; Siron et al., 2018). The ore is gold-bearing, mostly associated with the arsenian pyrite and less commonly with arsenopyrite. Quartz, calcite and minor rhodochrosite/kutnohorite occur as gangue minerals containing a measured and indicated resource of 0.55 Mt at 212 g/t Ag, 8.1% Pb, and 11.0% Zn (Eldorado Gold Corporation, 2017). The coeval 25 Ma (Siron et al., 2016) Fisoka and Stratoni diorite and granodiorite stocks ("F" and "St" in Fig. 2a, respectively) are inferred as possible igneous sources for the mineralizing fluids of the carbonate replacement deposits occurring within the SFZ (e.g. Siron et al., 2016). Therefore, the Fisoka prospect and its proximity is seen as a potential exploration target within the study area. In south, the 27 Ma Tsikara monzogabbros and granodiorites ("T" in Fig. 2a) may also have such potential as sources for mineralizing fluids. Both Fisoka and Tsikara intrusions are enclosed within quartz-sericite-pyrite altered rocks of the Vertiskos unit (Kockel et al., 1975; Siron et al., 2018). The intrusions and the alteration zone around the Tsikara area comprise a large part of the study area, covering the majority of the SW part and extending beyond the boundary of the study area and the Gomati Fault Zone (Fig. 2a) in the south. Since the marbles in the Kassandra mining district have mostly been linked to the SFZ and/or the Kerdilion formation in the north (e.g. Siron et al., 2018), there may be only minor potential for carbonate replacement deposits within our study area.

3. Data and methods

3.1. Geochemical and magnetic data

The original geochemical dataset includes 2151 soil samples in a roughly rectangular grid from an area of 80 km^2 and with an average of 200 m between adjacent sampling locations. In addition, a 0.6 km^2 area in the NW part of the study area near the town of Neochori has been sampled at an interval of 50 m ("N" in Fig. 2a). The soil sampling was conducted by Hellas Gold, and all samples were assayed by aqua regia digestion followed by inductively coupled plasma mass spectrometry at an accredited laboratory, ALS Limited. This set of geochemical soil samples with their GPS-measured sampling locations is referred to as "original samples" in the following sections.

Low-altitude aeromagnetic survey was carried out in the Halkidiki region between October 29th, 2007 and January 14th, 2008 (Fugro Airborne Survey, 2008). Except for the northernmost parts, the used aeromagnetic data covers the geochemical sampling grid area (Fig. 2b and c). The flight altitude was 60 m with 100 m in-line measurement spacing. The direction of the main flight lines was NS with a line spacing of 100 m, surveyed approximately perpendicular to the dominant geological strike direction. Additionally, EW-directed tie lines were surveyed with a spacing of 1000 m. The data received for this work was total field intensity data which had been corrected for the diurnal variation and leveling errors. Furthermore, the data had been interpolated to a regular grid with 50 m cells using kriging interpolation method.

3.2. Preprocessing of the geochemical data

Since variations in geochemical data are commonly influenced by null values, i.e. values below analytical detection limits, we replaced such values in the original samples by censored values one half of the corresponding detection limits. Because of high percentage of censored values in some elements, those elements have been removed from the analyses: for each variable, we used 33% as a cutoff limit for the allowed maximum portion of censored data, which resulted in 30 accepted independent variables (chemical elements: Au, Cu, Pb, Zn, Al, As, Ba, Bi, Ca, Cd, Co, Cr, Fe, Ga, Hg, K, La, Mg, Mn, Mo, Na, Ni, P, Sb, Sc, Sr, Th, Ti, Tl and V). In the end, each of the accepted variables has less than 15% of the values censored due to null values. Typical for geochemical data, the distributions within each element originally resembled inverse Gaussian distributions. Therefore, we transformed the distributions into more normal Gaussian-like distributions using base-ten logarithm function, and Z-score standardization to transform the data to the same scale for

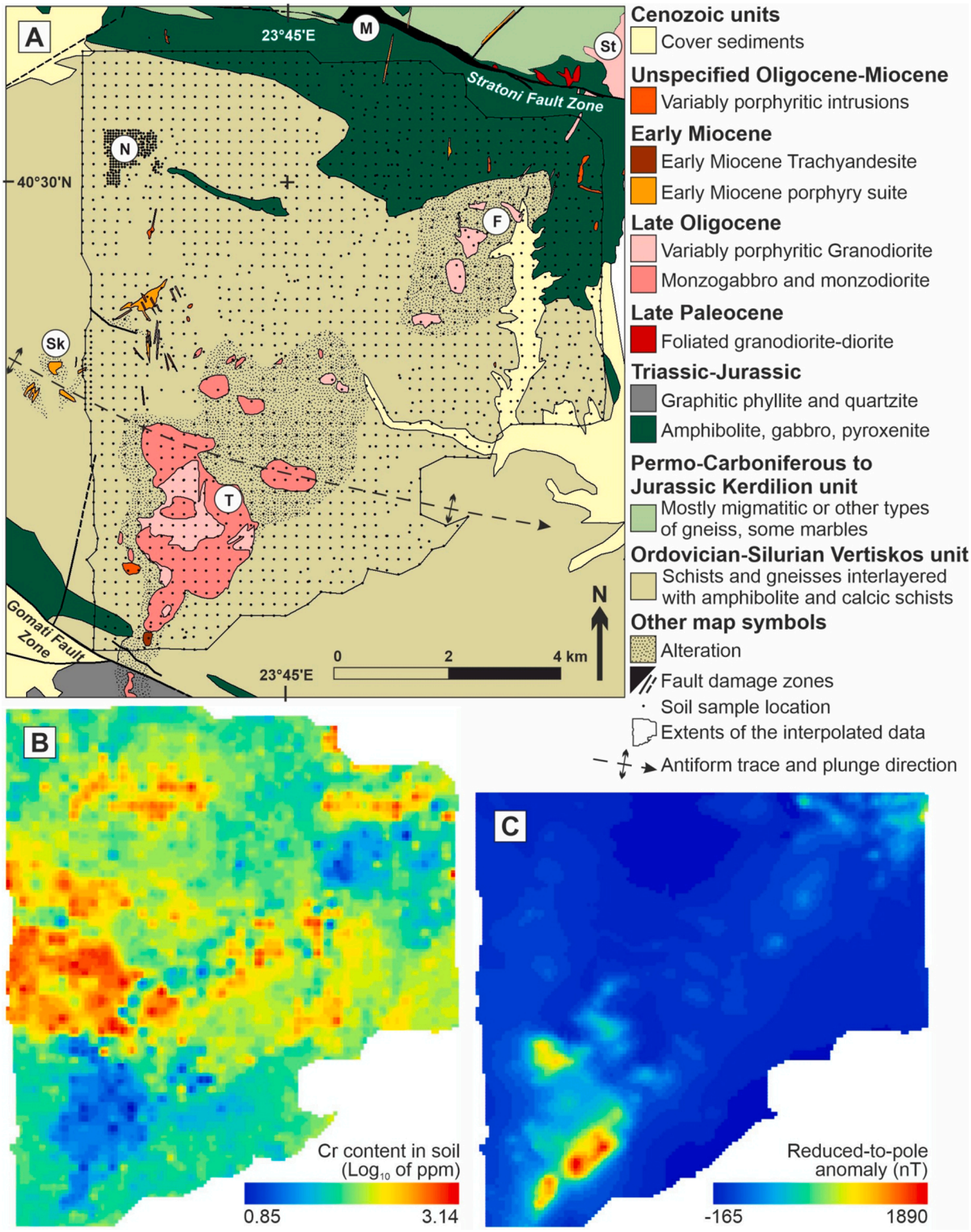

Fig. 2. A) Geological interpretation of the study area within the Kassandra mining district, modified after Kockel et al. (1975), Siron et al. (2018) and unpublished work by Hellas Gold. Original soil sample locations and the extents of the interpolated grid are displayed as dots and a polygon, respectively. Locations referenced in the text: Sk) Skouries deposit, M) Mavres petres mine, St) Stratoni diorite/granodiorite, F) Fisoka diorite/granodiorite, T) Tsikara monzogabbro/granodiorite, N) detailed sampling area near the town of Neochori. B) A single example out of the 30 chemical elements within the geochemical dataset, selected to highlight some of the distinct variation associated with different geological units: interpolated Cr content in soil displayed in logarithmic scale. C) Reduced-to-pole low-altitude magnetic data at the locations of interpolated points (northern part of the study area lacks the magnetic data). Here the magnetic anomaly has been shown, after removal of the contribution of the Earth's main field to the magnetic total field intensity, based on the IGRF model (Finlay et al., 2010).

SOM analyses. For PCA, all values were first converted to ppm before using centered log-ratio (Aitchison, 1986) to transform the data (see Section 3.4 for more details).

For the SOM analyses, we created an additional dataset, referred to in the following sections as "interpolated samples", by interpolation from the base-ten logarithmic values individually for each of the accepted variables of the original samples. This was done because the discrete output classes (i.e. clusters) from SOM cannot be interpolated, unlike in the case of floating-point PC scores obtained from PCA. We used inverse distance weighting (IDW) method with a power value of 1 and a variable search radius to include 25 nearest neighbors for the interpolations. IDW and the power parameter were selected for their simplicity, since the main purpose of this step is visualization of already relatively densely-sampled point data, and therefore their impact on results and interpretations is arguably negligible. In order to maintain uniformity within the data, we excluded the samples from the additional high-density sampling in the NW part of the study area ("N" in Fig. 2a) from the interpolation. For the interpolation, we generated a rectangular grid with a point spacing along the × and y axes set at 67 m, which is roughly a third of the original sampling interval. To avoid unnecessary and unreliable extrapolation, we set the extents of the interpolation to tightly encompass the original dataset, leaving out a single outlying sample in NE. This resulted in 17,407 samples (or pixels) with variables corresponding to the preprocessed original samples. Due to occasional differences between the original sample locations and the closest points within the generated interpolation grid, some peak values of the interpolated data appear attenuated compared to the original values. Nevertheless, we used the better spatial coverage and resolution of the interpolated points to enhance visualizations as well as the inter-layer coupling between the soil geochemistry and the low-altitude magnetic data. Same as with the original sample points, Z-score standardization was used for scaling the interpolated data.

3.3. Preprocessing of the low-altitude magnetic data

For the SOM analyses, we used reduction to the pole (RTP; Baranov, 1957; Baranov and Naudy, 1964) to remove the dependence of the magnetic anomalies on the local inclination of the Earth's magnetic field. RTP converts the total field intensity data to what would have been observed at the magnetic pole with vertical magnetization direction, directed downwards. This locates the magnetic anomalies above the causative bodies which allows for more accurate spatial aligning of the geochemical and magnetic data. As the sources of the magnetic anomalies are in the bedrock, the depths of the sources are not constrained, and as the geochemical data is from soil samples, joint usage of such datasets does not necessarily produce meaningful results. However, in this case the joint analysis is based on the observation, and underlying assumption behind the concept of using the geochemical soil data for semi-automated geological mapping, that the geochemical samples reflect local bedrock properties and can be used as proxies for lithologies below. Furthermore, since visually the magnetic data seem to mostly display relatively sharply-confined short-wavelength anomalies, we make the further underlying assumption that the sources of the magnetic anomalies are shallow and that the magnetic data also represent the local bedrock properties (e.g. Airo et al., 2014).

It should be also noted, that RTP relies on the assumption that induced magnetization is the main cause of magnetization. If remanent magnetization components in other directions than parallel to the direction of the induced magnetization are significant, the magnetic anomalies get distorted. In this case there is no remanent magnetization data available to assess the possible effect on the results, and we simply assume that the remanent magnetization is negligible compared to the induced magnetization. Furthermore, in low latitudes, RTP operator becomes unstable and produces irregular noise. Analytic Signal (AS; Nabighian, 1972; Nabighian, 1974; Roest et al., 1992) is an alternative for RTP, and has been used in similar analyses utilizing magnetic data in SOM analyses, in particular in the cases of low-latitude magnetic data (e.g. Carneiro et al., 2012). The amplitude of AS of the total magnetic field produces maxima over the magnetic contacts regardless of the magnetization direction. Our data are not really low-latitude data (40°30′N), but we also ran tests with the AS regardless. However, as the AS is calculated from derivatives, it is sensitive to short-wavelength noise. In this case, we judged RTP to produce results with better spatial resolution. We also ran tests with other derivative products, such as tilt derivative and total horizontal derivative, but as these amplify the relatively very short-wavelength anomalies, and do not represent the properties of volumes such as lithological units but rather the planar structures such as contacts and faults (e.g. Verduzco et al., 2004), we deemed these inapplicable for the aim of this study.

To get the RTP-corrected magnetic field intensity values (nT) in the correct sample locations (for both original and interpolated points), we created a 16-bit georeferenced raster interpolation with a 10-meter cell size from the RTP magnetic data. We used a "nearest neighbor" method to extract cell values from the raster to the original and interpolated sample points described in Section 3.2. From this method of value extraction alone, the 10-meter cell size of the raster translates to a maximum spatial error of roughly 7 m for any given sample point. Z-score standardization was used to transform the values to the same scale as in geochemical data. Due to different extents of the used magnetic and geochemical data, some 20% of the original geochemical sample points and 16% of interpolated points lack the RTP magnetic data in the northern side of the study area (as seen in Fig. 2b-c). Since the missing portions are relatively small and SOM can handle missing values by simply ignoring them in the analysis (e.g. Cottrell and Letrémy, 2005), we deemed the spatial coverage of RTP values acceptable.

3.4. Principal component analysis

When mapping spatial distribution of variables in mineral exploration or environmental geochemistry in PCA, different transformation methods, additive log-ratio (alr: Aitchison, 1986), centered log-ratio (clr: Aitchison, 1986) and isometric log-ratio (ilr: Egozcue et al., 2003), give varying results. The clr-transformation (Aitchison, 1986) is a useful transformation for evaluating geochemical data. The principal components of clr-transformation are orthonormal (statistically independent) and can reflect linear processes associated with stoichiometric constraints (Grunsky and de Caritat, 2019). In this study, we used clr-transformation to identify multi-element associations related to Au-Cu mineralization that account for a substantial proportion of the data variability.

The foundation of PCA is the correlation (or covariance) matrix, which measures the interrelationships among multiple variables, allowing the rearrangement of the data into vectors within uncorrelated dimensions, i.e. principal components (PC). The PC scores of the samples can then be used to recognize similarities and differences between the samples, and loadings of the variables (defining the "orientation" of their respective PC within the multidimensional space) to study the effects those variables have on the variance within a given PC.

We calculated the covariance matrices to understand how the variables are varying from the mean with respect to each other and to see if there is any relationship between them. Next we calculated the eigenvectors and eigenvalues of the covariance matrix to identify the principal components (PC). The first PC accounts for the largest possible variance in the dataset and each following PC accounts for a progressively smaller portion of variance. We studied the loadings of the input variables to identify geochemical associations related to various geological processes.

Unlike the discrete output classes of clustered SOMs (see Section 3.5), continuous floating-point PC scores can be used in interpolation. Thus to create a visual presentation of the PCA results for the purposes of correlating with SOM results, we interpolated the PC scores of transformed data into grids with spatial extents and shape corresponding to

the geometry of the interpolated input points used in SOM models II and IV (see Section 3.5). We used IDW method with a power value of 1 to create interpolated layers for the PCs with eigenvalues above 1 (i.e. the first seven PCs, see Section 4.1). IDW and the power value were selected on the basis that same interpolation method and parameters were used as with the creation of "interpolated samples" input dataset for SOM, described in Section 3.2. We created two spatial visualizations by first normalizing the ranges of PC values to a similar range for every PC (e.g. 32-bit floating-point value in the range of 0–1), and by then making two suitable selections of three normalized PCs to represent the color bands in RGB composites. We selected the first three PCs of the analysis for the first visualization since they explain the largest portion of the total variance (and also most of the variance, as is apparent from the results described in the Section 4.2 and in Table 2) within the input data, relative to the other PCs. For the second visualization, we selected three PCs based on their apparent ability to explain variance related to the three known exploration/mining sites in the study area.

3.5. Self-organizing maps

The foundation of SOM lies in a competitive learning process where a set of node vectors are trained to represent the n-dimensional input data vectors projected onto a topology-preserving two-dimensional "map", where each sample is also considered an n-dimensional vector. This allows several possible means of further analysis and interpretation, such as identifying anomalous samples using quantization error or cluster-normalized values (e.g. Fraser and Hodgkinson, 2009), studying the properties of individual nodes, and clustering based on similarities between the nodes (e.g. Kohonen, 1982; Kohonen, 2013; Fraser and Dickson, 2007). Mostly depending on the used map size (e.g. Park et al., 2003), the self-organization process often reduces the issues related to noise and overfitting, and thus enhances the feasibility of clustering algorithms. Furthermore, each sample vector can be classified by their closest map node, i.e. "best matching unit" (BMU). A single node typically represents a BMU for several sample vectors that are relatively similar to each other in the n-dimensional space. It is important to have large enough map to not lose any useful details of the input data, but at the same time avoid too large map sizes since they come with problems related to reduced recognizable differences between the nodes and with possible over-fitting issues (Park et al., 2003).

We tested the applicability of two methods to optimize the map sizes used in the SOM analyses. Initially, we tried optimizing by minimizing the mean quantization error and topological error by finding local minima for those values. Quantization error (QE) is calculated from the distance between the sample vector and its BMU, whereas the mean QE within the model is the average of all these individual values. Topological error (TE; sometimes also called topographic error) is a measure of topological preservation within the model: this error is the proportion of the sample vectors for which the first and second BMUs are not adjacent nodes in the model. Mean QE and TE are typically reduced with increasing map sizes, but due to the aforementioned issues, merely increasing the map size may cause unwanted outcome for the analysis. In this study, no local minimum for mean QE or TE could be determined, and therefore the map sizes were selected according to the heuristic rule (Vesanto, 1999), where the number of nodes is determined as five times the square root of the number of samples. We used a toroidal map geometry and hexagonal lattice of 18 rows and 14 columns for the original sample points, and 30 rows and 22 columns for the interpolated samples (Table 1). We created two different SOMs for both of these datasets, resulting in four models in total (Table 1): one SOM from each set of points, original and interpolated, using only geochemical data (Models I and II), followed by SOMs with magnetic data included in the models (Models III and IV).

For SOM analyses in this work, we used SiroSOM application developed by The Commonwealth Scientific and Industrial Research Organization (CSIRO). SiroSOM is based on the SOM Program Package of Kohonen et al. (1996) and the SOM Matlab Toolbox of Vesanto et al. (2000). We used SiroSOM's built-in tool for Davies Bouldin analysis (DB analysis: Davies and Bouldin, 1979) to identify the optimal number of clusters and *k*-means clustering (*k*-means algorithm: Lloyd, 1957; MacQueen, 1967) to classify nodes into such number of clusters. We selected a large number of clusters (around 15) with the lowest DB index to make a more detailed separation for the classes. We used the resulting clusters to assign discrete class numbers for each corresponding sample (based on their BMUs) within the input datasets. Finally, we described the cluster characteristics in detail based on their average BMU values and the spatial relationships of the newly classified samples and the existing geological interpretations.

4. Results

4.1. Principal component analysis of the original geochemical samples

The PCA using the clr-transformed dataset resulted in seven PCs with eigenvalues above 1 (Table 2 and Fig. 3). Eigenvalues of the PCs are decreasing rapidly, indicating that the majority of variation in the data is accounted for by the first five PCs. The remaining PCs can be interpreted to represent random or under-sampled processes (see Section 5.2 and Grunsky and De Caritat, 2019). This is because a factor with an eigenvalue of 1 accounts for as much variance as a single variable, and therefore only factors that explain at least the same amount of variance as a single variable may be useful for further analysis.

PC1 (Fig. 4a) of the clr-transformed data explains about 26% of total variance and in this PC the highest loaded elements in order of factor-loading values are: Fe-Co-Ni-Cr-Al-Mg. The PC2 (Fig. 4a and b) of the clr-transformed data explains about 21% of total variance, and it depicts an association of Pb-As-Hg-Mo-Sb. The PC3 (Fig. 4b) shows association of Ba-Al-Ga-K-La-Th-Tl, explaining 14% of total variance. The PC4 represents 7% of the variance and shows an association of elements Co-Cr-Mg-Ni. The PC5 of the clr-transformed data explains about 6% of the total variance, and possibly depicts two associations, Au-Cu-Bi-Mo and Ba-Ca-Cd-Co-Mn-P-Zn. PC6 and PC7 with the eigenvalues close to 1 should be dismissed or at least used with caution, considering the relatively low signal-to-noise ratio. PC2 and PC5 show the most interesting element associations in regard to mineralized systems and are therefore represented relative to each other in Fig. 5.

The first three PCs of the clr-transformed data used in the first RGB composite (Fig. 6a) explain most of the total variance within the data (61%). The second RGB composite comprises the PC2, PC4 and PC5 (Fig. 6b), since their element associations explain different types of mineralizations and their extremities coincide spatially with the known target sites: Fisoka, Tsikara and Skouries ("F", "T" and "Sk" in Fig. 6c, described in detail in Section 5.1). The latter three PCs together explain 34% of total variance within the dataset.

Table 1
Main attributes of the four created SOM models (QE = quantization error, TE = topological error).

Model	Included data	No. samples	No. variables	Map size	Mean QE	TE	No. clusters
I	Original samples (geochemistry only)	2151	30	18×14	2.73	0.0934	16
II	Interpolated samples (geochemistry only)	17,407	30	30×22	2.31	0.0757	15
III	Original samples with RTP magnetic	2151	31	18×14	2.79	0.0953	14
IV	Interpolated samples with RTP magnetic	17,407	31	30×22	2.33	0.0921	16

Table 2
Principal components and the loadings of the clr-transformed data. The first seven PCs explain more than 81% of total variance. Blue and red colors show positive and negative element associations in their respective PCs. The blue elements are selected based on the values greater than Median + std for each PC and red elements are selected based on the values less than Median-std for each PC.

Element	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Au	0.31	0.60	-0.39	0.11	-0.34	-0.13	-0.07
Cu	0.58	0.12	-0.36	0.07	-0.45	-0.01	0.15
Pb	0.36	0.72	-0.26	-0.17	0.07	0.01	-0.32
Zn	0.69	0.39	-0.18	-0.15	0.34	-0.13	-0.15
Al	0.70	-0.43	0.37	0.01	-0.10	-0.06	-0.11
As	0.41	0.72	-0.27	-0.18	-0.03	-0.09	-0.21
Ba	0.45	0.25	0.63	0.32	0.30	0.08	-0.07
Bi	0.10	0.59	-0.11	0.00	-0.43	0.28	-0.05
Ca	0.58	-0.35	-0.41	0.30	0.20	0.33	-0.01
Cd	0.58	0.52	-0.38	-0.07	0.30	0.01	-0.09
Co	0.75	-0.25	-0.10	-0.36	0.24	0.05	0.21
Cr	0.64	-0.20	0.22	-0.50	-0.30	0.28	0.07
Fe	0.75	-0.04	-0.11	0.06	-0.17	-0.34	0.30
Ga	0.68	-0.35	0.40	0.10	-0.18	-0.21	-0.14
Hg	0.14	0.64	-0.16	0.13	0.08	0.19	0.15
K	0.50	0.08	0.65	0.27	0.05	0.00	-0.27
La	0.21	0.45	0.71	0.24	0.05	-0.14	0.15
Mg	0.70	-0.47	0.18	-0.27	-0.04	0.23	0.00
Mn	0.64	0.23	-0.05	-0.14	0.60	-0.09	0.17
Mo	0.07	0.63	0.15	0.24	-0.39	0.06	0.25
Na	0.29	-0.48	-0.42	0.47	-0.02	0.12	-0.20
Ni	0.71	-0.09	0.27	-0.51	-0.10	0.28	0.18
P	0.47	0.21	-0.25	0.39	0.26	-0.04	0.33
Sb	0.25	0.79	-0.34	-0.07	-0.14	-0.03	-0.22
Sc	0.64	-0.43	-0.33	0.04	-0.21	-0.28	0.15
Sr	0.41	-0.09	-0.17	0.70	-0.01	0.39	0.10
Th	0.11	0.42	0.80	0.16	-0.09	-0.10	0.07
Ti	0.47	-0.63	0.08	0.11	-0.08	0.05	-0.42
Tl	0.43	0.51	0.51	0.02	-0.10	0.01	-0.08
V	0.54	-0.55	-0.31	0.16	-0.15	-0.34	-0.07
Eigenvalue	7.93	6.32	4.12	2.18	1.76	1.05	1.04
% of variance explained	26.44	21.06	13.75	7.27	5.88	3.51	3.48
Cumulative % of variance	26.44	47.50	61.25	68.52	74.40	77.91	81.39

4.2. *Self-organizing maps from the four selections of datasets*

The main attributes of the SOM analyses are presented in Table 1, including the final mean QE and TE of each model, as well as the number of clusters determined by finding a local minimum of DB index. Even though the spatial information of the samples was not included in the analysis, similarly classified samples often form clear groups or segments of spatially adjacent samples (Fig. 8). Most of the variation between the results of cluster analyses of the four models appear around the spatial boundaries of the aforementioned segments, or between the clusters consisting of the least anomalous samples (discussed in detail in Section 5.1). The models from the interpolated points (Models II and IV) show results comparable with the corresponding models from the original sample points (Models I and III, respectively). The models where RTP magnetic data was included in the analysis (Models III and IV) display mostly similar results to the models from geochemical data only (Models I and II), yet some samples in several areas are classified differently, especially between Models I and III (see Section 5.1). Since the variations between the models and between individual clustering attempts for each model all appear small, two models were used as reference for characterizing the cluster attributes. To describe all of the recognized characteristics exhaustively, Models III and IV, with all of the 31 input variables included in the analyses, were selected as reference (Table 3 and Fig. 7). In every model, the few most distinct (i.e. anomalous) clusters or groups of clusters were classified repeatedly similarly (cluster names refer to those in Table 3): *"Anthropogenic"*; *"Fisoka 1"* and *"Fisoka 2"*; *"North 1"* and/or *"North 2"* and *"North 3"*; and *"South 1"*. The BMUs within these clusters are often characterized by anomalously high or low values of several variables, relative to the values observed within other nodes (e.g. see the anomalously low Ba, La, Th, K and Tl in clusters "North 1" and "North 2" in Fig. 7). The justification for selection of representative colors and names for the clusters in Table 3 and in Figs. 7 and 8 is explained in Section 5.1.

5. Discussion

5.1. *Comparison between the SOM clusters and PC scores, and evaluating model validity with the known geological features*

Methodology behind the existing geological interpretations has to be considered when using those interpretations to evaluate the results of this study. Kockel et al. (1975) do not describe the used methods for their geological map interpretations, but at least we can be certain that they have not used the same input data as we have. The interpretations by Siron et al. (2018) and unpublished work by Hellas Gold closely resemble those by Kockel et al. (1975), and mostly present fine modifications and small additions to the previous interpretations in the scale of our study area. The same or very similar input data we have utilized, may have been used in these interpretations (Siron et al., 2018; unpublished work by Hellas Gold) e.g. to delineate the boundaries of observed lithological units, altered zones, etc., and that may explain why we arrive at very similar results in many cases. However, since this study focuses on new targets and both used methods produced similarly classified outputs in previously unknown locations, these said cases are not just redundant results, but encouraging findings as they represent the identification of "type localities" of these units within the study area.

The results of both PCA and SOM give detailed information of the bedrock below the soil, as can be seen in the map presentations of the sample points classified based on their BMU clusters in the SOM analyses (Figs. 7 and 8, and Table 3) and RGB composite images of the selected interpolated PC scores (Fig. 6). The most striking geochemical differences can be seen from the belt-scale features, where the study area is divided into three main sections: amphibolites in the north and distinctly different schist/gneiss zones in the middle and south parts of the study area (Figs. 6 and 8a). These most distinct features are accompanied by the highly anomalous element associations in Fisoka area, clearly seen in the RGB composite even when using the first three PCs as input bands ("F" in Fig. 6). The geochemical responses of the other two intrusion sites, Tsikara and Skouries ("T" and "Sk" in Fig. 6), are less obvious in the results of both methods compared to the aforementioned units. However, Tsikara and Fisoka sites are still distinctly different from each other based on the PCA (especially in PC5), and only a few samples within Tsikara area are classified into "Fisoka type" clusters, and vice versa, in any of the SOM models. The samples close to the Skouries deposit are also distinctly different from the other two sites (i.e. Fisoka and Tsikara) in results of both methods (especially in PCs 1 to 4 in Fig. 3, see also "Sk" in Figs. 6 and 8).

PC biplots can reveal and visualize the structure in the data related to lithological changes or weathering processes. Identifying the mineral host or host rocks for the elements selected for this study helps in understanding the association of elements related to Cu ± Au mineralization. In this study mineral hosts have not been determined and possible host rocks noted in the regional geology have been used for the interpretation. In Fig. 4a, a biplot of PC1 and PC2 discriminates a group of elements that contain PC1 and PC2 values with a meaningful loading score (see Section 4.1). Pb-As-Mo association likely reflects hypogene-related elements or the weathering of supergene zones, while a group of elements with positive PC1 and negative PC2 (Fe-Ni-Co-Ga-Al-Mg-Ca) are commonly associated with iron-rich accessory minerals, such as

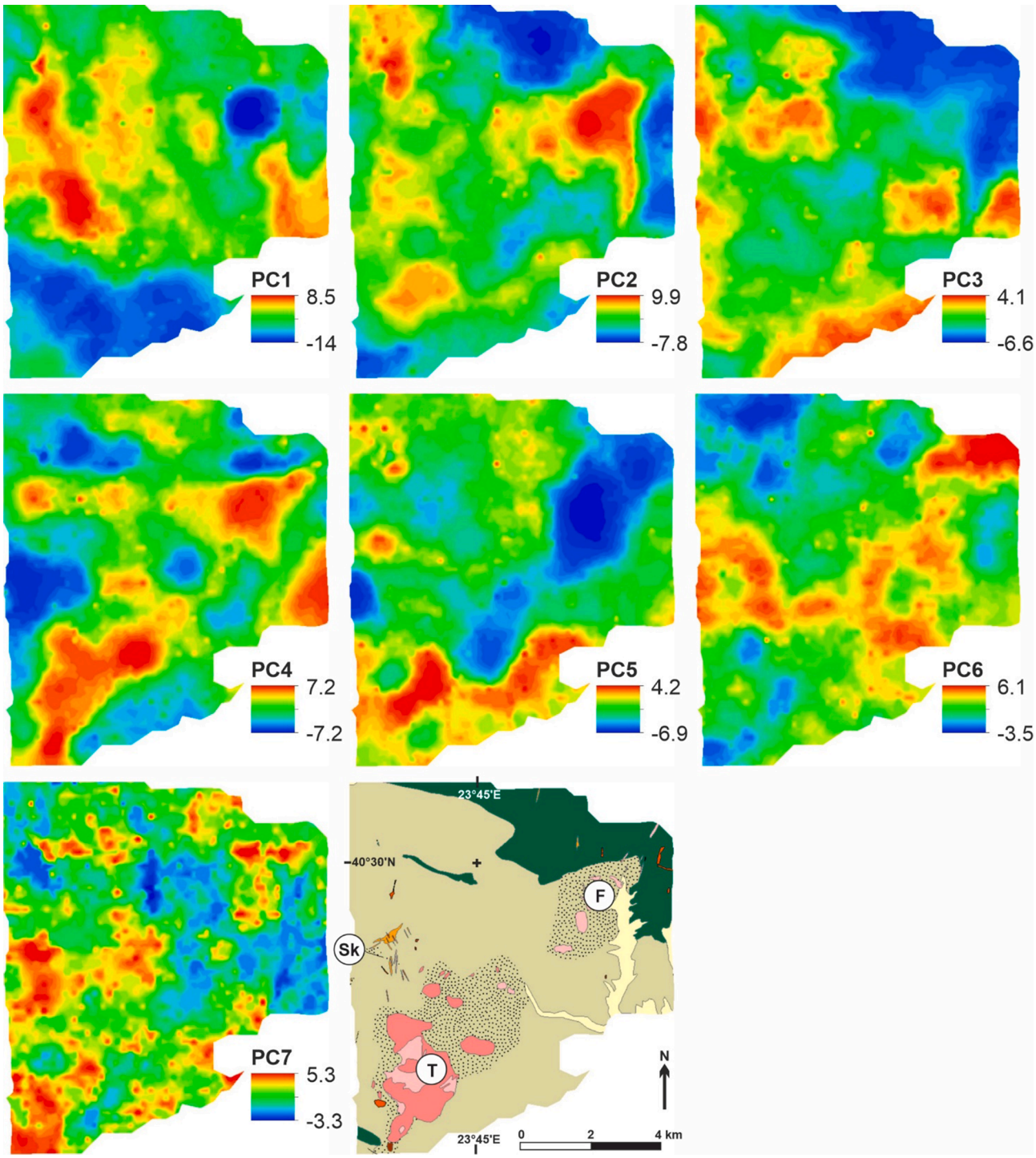

Fig. 3. Interpolated maps of the first seven PCs (PC1-7), where sample scores are represented by a color scale defined individually for each PC based on their range of values. Bottom right: Geological interpretation of the same area for comparison (see also Fig. 2a). Locations referenced in the text: Sk) Skouries, F) Fisoka and T) Tsikara.

amphibolite, and may reflect biotite-amphibolite carbonaceous schist in the study area. Biplot of PC2 and PC3 (Fig. 4b) shows association of Mg-Al-Ga in Zone I which probably reflects schist. In Zone II, the association of K-Ba-Th-La-TI-Mo may related to porphyry rocks and granitoids. In Zone III, Ca with less loading of Sr-Fe is probably related to calcareous rocks in the study area, and finally association of Pb-Hg-As with less loading of Cu-Au-Bi-Sb reflects weathering or supergene zones with or without Au-Cu mineralization. The group of elements in PC4, Co-Cr-Mg-Ni, represent ferride elements associated with mafic minerals or rocks. Na-P-Sr are a group of elements associated in the PC4 that may represent silicate minerals in the soil and lithology of granitoids. The Au-Cu-Bi-Mo depicted by PC5 reflects parts of the porphyry copper gold enrichment in these precious metals (Fig. 5b), while association of Ba-Ca-Cd-Co-Mn-P-Zn may reflect different common minerals in amphibolite or in mafic rocks. PC6 and PC7 with the eigenvalues close to 1 may reflect the small portion of the samples representing different lithologies, weathering process or noise in the samples.

The characterizing/anomalous features of the clusters in SOM

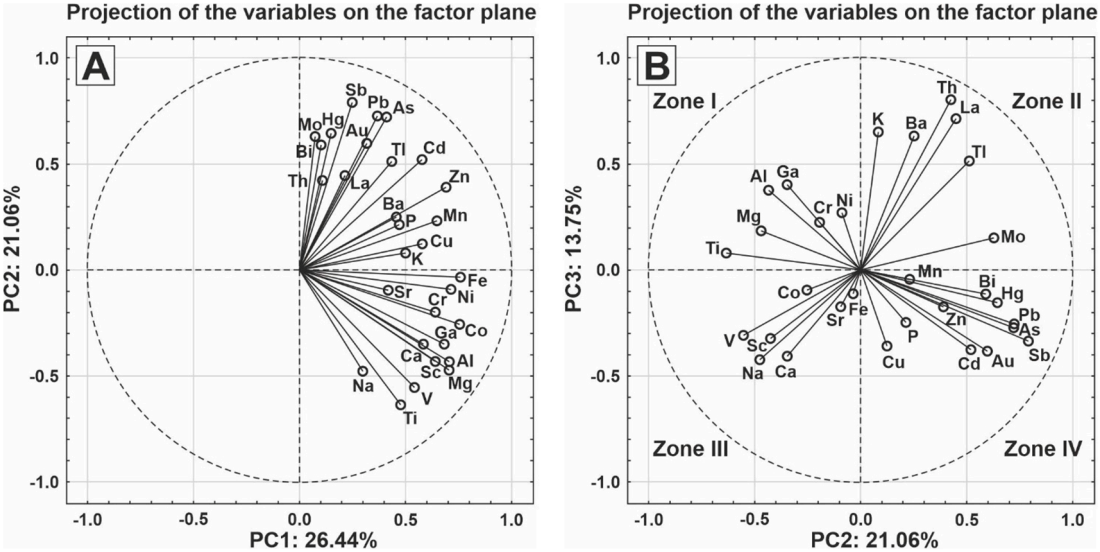

Fig. 4. A) Biplot of the variable loadings in clr-transformed PC2 vs. PC1. B) Biplot of the variable loadings in clr-transformed PC3 vs. PC2. See text for an explanation of the element associations (Sections 4.1 and 5.1).

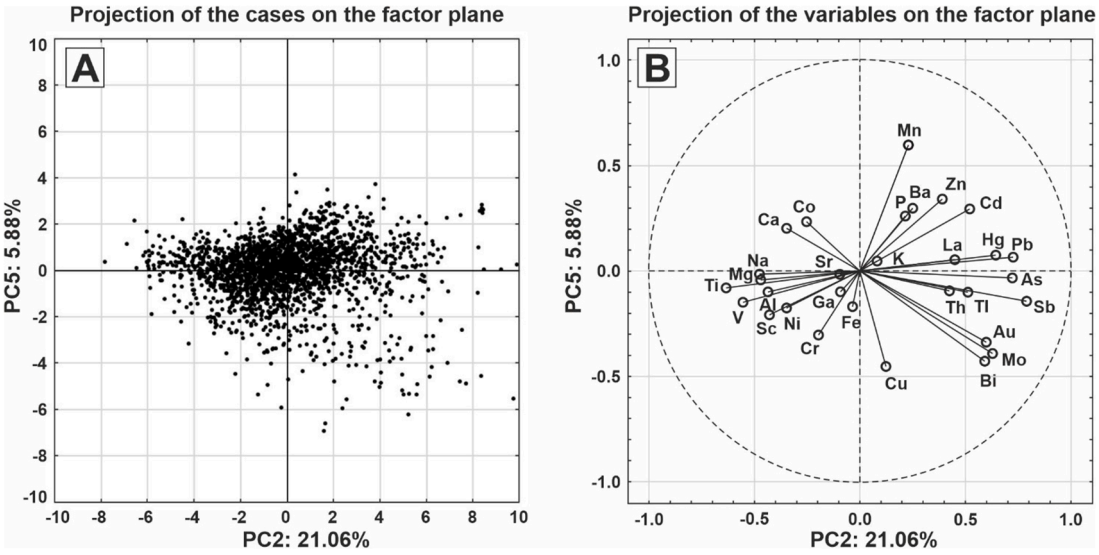

Fig. 5. Biplots of PC5 vs. PC2 in the clr-transformed data from the original geochemistry A) for the scores of cases (samples) that show the data has been successfully opened using clr-transformation, and B) for the loadings of variables that depict the element associations within the two PCs.

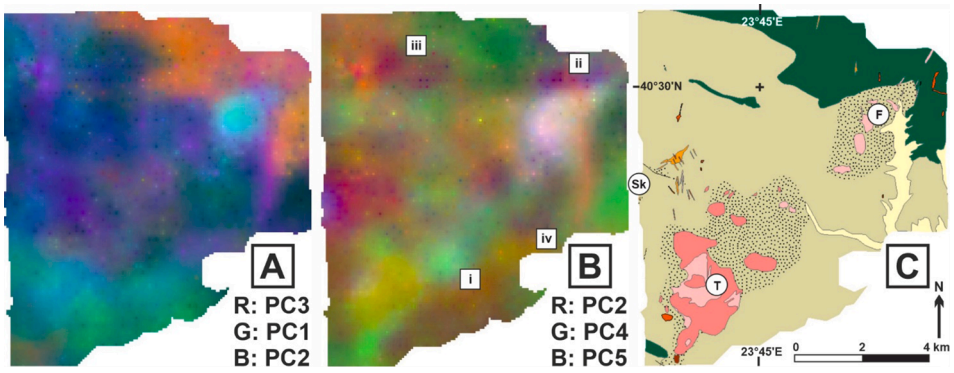

Fig. 6. Examples of visualized PCA results using RGB composites. Principal components corresponding to the RGB bands for both maps are displayed in the figure: A) the first three principal components (PCs 1, 2 and 3), B) three principal components (PCs 2, 4 and 5) selected to highlight the three target sites (Skouries labeled "Sk", Fisoka labeled"F" and Tsikara labeled "T") of the area most distinctly, and C) geological interpretation of the same area (see also Fig. 2a). The labels "i-iv" denote potential exploration targets discussed in Section 5.2.

Table 3
Color codes for clusters, the geochemical characteristics, and spatial correlations between the defined clusters and the previous geological interpretations. The characteristics are based on the cluster analyses of SOM Models III and IV.

Name	Color	Geochemical characteristics	Description
Skouries 1		Very high Cr, Mg, Ni; High Al, Co, Fe, Ga, La, Th, Tl	Spatially related to the intrusions and alteration around Skouries deposit.
Skouries 2		Very high Ba, Co, Tl; High Al, Cr, Fe, Ga, La, Mg, Mn, Ni, P, Th	Geochemically similar to the "Skouries 1" cluster.
Fisoka 1		Very high Mo; High Bi, Hg, La, Th; Very low Zn, Al, Ca, Cd, Co, Cr, Fe, Ga, Mg, Mn, Ni, Sc, Ti, V	Spatially related to the intrusions and alteration in Fisoka. Moderately positive magnetic anomaly.
Fisoka 2		Very high Cu, Bi, Th; High Au, Hg, La	Spatially related to the intrusions and alteration in and around Fisoka prospect.
Tsikara 1		Very high La; High Ba, Th; Very low Cu, Cr	Spatially related to the intrusions and alteration in Tsikara prospect. Strong positive magnetic anomaly.
Tsikara 2		Very high Sr; High Al, Ga, La, Th, Ti	Spatially related to the alteration and intrusions between Tsikara and Fisoka prospects.
Tsikara 3		High Al, Ca, Fe, Ga, Sc, Sr, Ti, V; Very low Pb, As, Hg, Sb	Spatially related to the alteration around Tsikara prospect.
Anthropogenic		Very high Au, Pb, Zn, As, Cd, Fe, Hg, Mn, P, Sb; High Cu, Ba, Bi, Co, La, Th, Tl	Represents the alluvial soil of a river valley in the east, probably an anthropogenic anomaly. May represent ore-forming processes elsewhere.
North 1		Very high Ca, Na, Sc, V; High Co; Very low Ba, Bi, K, La, Mo, Th, Tl	Spatially related to the amphibolitic gneisses especially in the north.
North 2		High Ca, Co, Sc, Sr, V	Occurs around the boundaries of "North 1" cluster, and in some models is not a separate cluster, but included within the "North 1".
North 3		Mostly similar to "North 1"	Spatially related to the amphibolitic gneisses around Fisoka. May represent altered counterpart to "North 1" cluster. Strongly positive magnetic anomaly.
Middle 1		Very high Al, Ba, Ga, K, La, Ti; High Co, Fe, Mg, Th	Spatially related to unspecified schists and gneisses in the middle of the study area.
Middle 2		High Ba, Co, La, Th, Tl	Spatially related to unspecified schists and gneisses in the middle of the study area.
Middle 3		High La, Th; Very low Na	Spatially related to unspecified schists and gneisses in the middle of the study area.
South 1		High Th; Very low Au, P, Sr	Spatially related to unspecified schists and gneisses in the southern part of the study area.
South 2		High La, Th	Spatially related to unspecified schists and gneisses mostly in the southern part, or occurring sporadically around the whole study area.

models, accompanied by the spatial correlations to the previously interpreted geology (after Kockel et al., 1975; Siron et al., 2018; and unpublished work by Hellas Gold), are given in Table 3. The geochemically most distinct location of the study area is around Fisoka prospect, where several porphyry stocks intrude the country rock. The “Fisoka 1” SOM cluster is defined by clearly anomalous values for several elements, both positive and negative in relation to other clusters, and some samples are also spatially associated with a strongly positive magnetic anomaly. Both “Fisoka 1” and “Fisoka 2” seem to represent the element groups associated with mineralized porphyry intrusions, also depicted in the PC5 (Table 2 and Fig. 3). Another strongly anomalous but spatially limited SOM cluster, named “Anthropogenic”, is interpreted to mostly relate to some soil anomalies due to anthropogenic activity in the area. This is probably mostly restricted to a single river valley on the eastern side of the study area, and similarly clustered samples may well represent geological processes elsewhere. The element associations depicted by PC3 (Table 2 and Fig. 3) are represented in the northern SOM clusters correlated with the amphibolites (“North 1” to “North 3”), typically characterized by anomalously positive association of Ca-Sc-V-Co in all of the SOM models. The samples classified into “North 3” are also spatially associated with a strongly positive magnetic anomaly, although this cluster did not appear in the Model III, which is one of the two models where RTP magnetic data was included in the analysis. In Model III, the positively anomalous samples in that area were classified into clusters “North 1”, “North 2” and “Skouries 2” instead. The SOM clusters concentrated in the middle schist/gneiss zone of the study area (“Middle 1” to “Middle 3”) usually have a positive La-Th association, with Ba often included. The differences of the southern and middle schist/gneiss zones appear to depend on a complex association of elements defined by PC1 (Table 2 and Fig. 3). The geochemically less distinct Tsikara- and Skouries-related SOM clusters still have their own key characteristics, rendering the three intrusion sites clearly distinguishable by their geochemical signatures. From the PCs, the characteristics of Tsikara intrusion appear to be mostly related to associations in PC4 (Table 2 and Fig. 3), and the SOM cluster “Tsikara 1” is mostly characterized by anomalously low Cu and Cr (as can be seen from the example in Fig. 2b), as well as strongly positive magnetic anomaly. SOM clusters “Tsikara 2” and “Tsikara 3” are defined by complex element associations usually with positive Sr-Al-Ga-Th-Ti, and their classifications differ from model to model. The “Skouries 1” and “Skouries 2” clusters are mostly characterized by positively anomalous association Cr-Ni-Co-Tl, which may relate to distal alteration and/or presence of mafic intrusive component around Skouries deposit. There appear to be more variations between the classifications of these two clusters between the different models, when compared to e.g. “Fisoka 1” and “Tsikara 1” clusters.

Many of the interesting sites in the study area are not easily defined

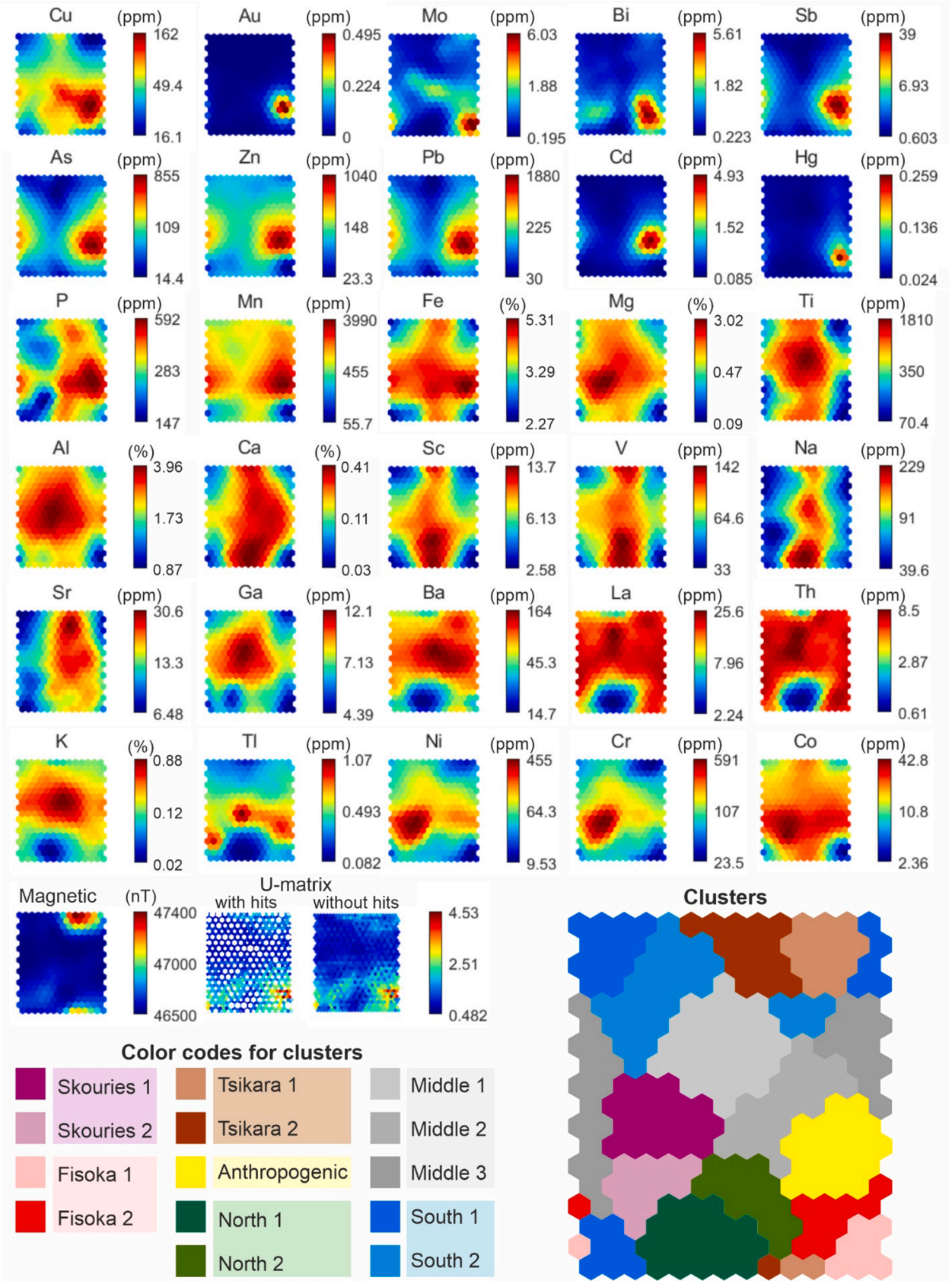

Fig. 7. The input data, U-matrix and clusters plotted on a 2D presentation of the nodes in SOM model III, which was derived from the original geochemical soil samples and the reduced-to-pole magnetic data. The main attributes of all SOM models are presented in Table 1. Note that a toroidal SOM topology was used for the model. See Table 3 for detailed characteristics of the clusters, and Fig. 8 for the spatial distribution of the samples and clusters in a map view.

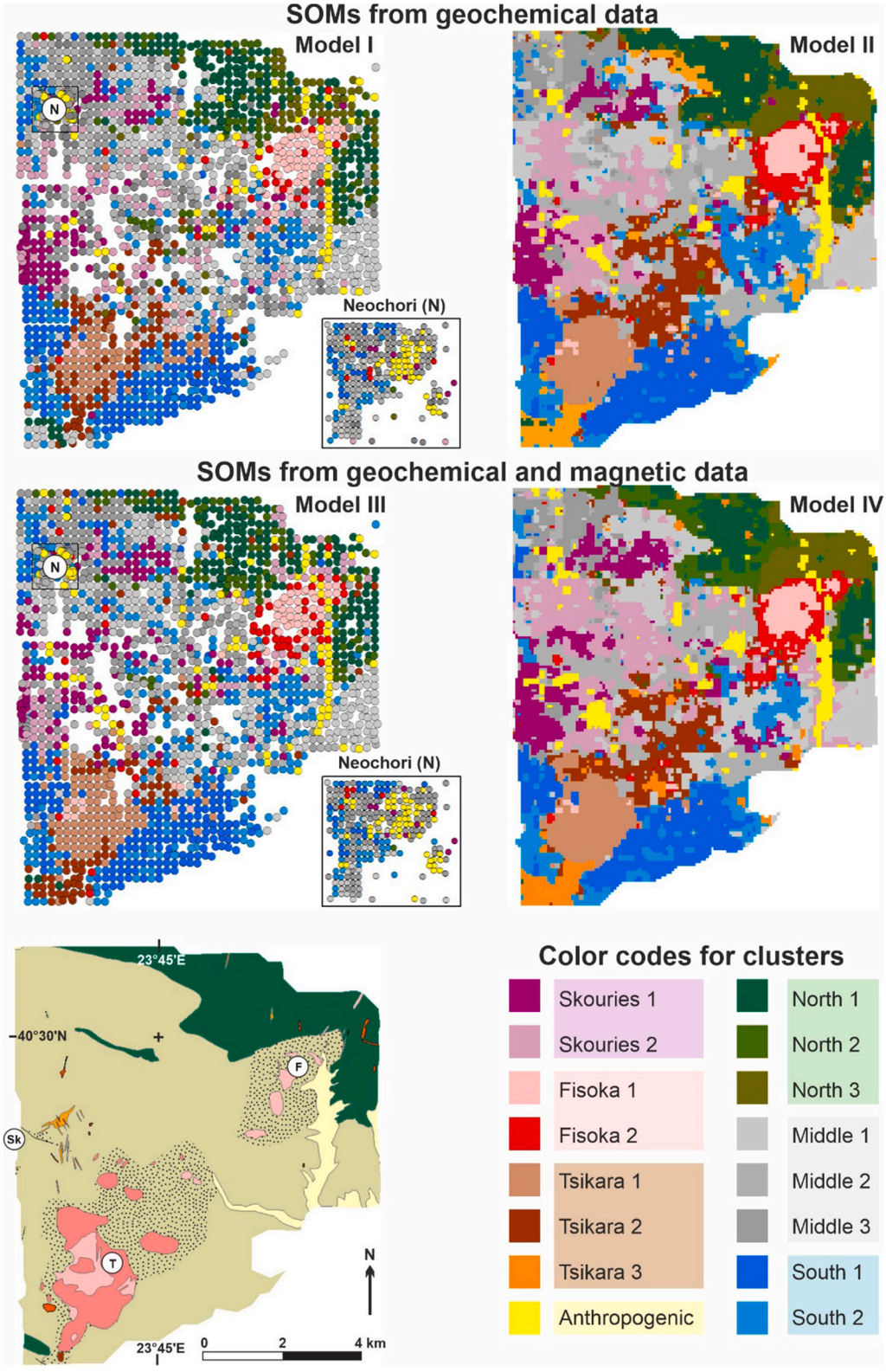

Fig. 8. Results from the SOM analyses: k-means clustered BMUs from the original sample locations on the left and from the interpolated grid on the right, where each round symbol or square-shaped point/pixel represents an input data point. Only the geochemical data were used in the Models I and II (upper figures), while the geochemical and reduced-to-pole magnetic data were used in combination in the Models III and IV (middle figures). The detailed sampling area near the town of Neochori (N) is displayed separately for the original sample points. Colors were assigned to clusters (lower right) based on the apparent spatial coincidence of similarly clustered BMUs between the four models. See Table 3 for the detailed characteristics of the clusters. Lower left: Geological interpretation of the same area for comparison (see also Fig. 2a). Locations referenced in the text: Sk) Skouries, F) Fisoka and T) Tsikara.

by any single PC, but rather by several of them. This issue is most notable from the Skouries site, which can only be distinctly visualized by including all three RGB bands appropriately in the composite image (Figs. 3 and 6b). This is one of the reasons why the selection of RGB bands, and therefore the whole method of using RGB composites, depends heavily on the choices of the user. It can even be argued, that three simple similarity analyses, guided by selected representative and interesting samples or "target compositions", could largely reproduce the results of Fig. 6b, possibly with less effort. Clustered SOM, especially when using the typical analysis tools (e.g. DB index), thus needs less user input and is therefore probably more data-driven from these two methods. This also makes it more straightforward to use, since it lacks

the time-consuming trial and error for selecting the appropriate inputs.

The interpolated datasets aid in visualization of the results and also improve the joint use of lower-resolution geochemical data with high-resolution RTP magnetic data. Overall, the original points used in SOM Models I and III give comparable results to the Models II and IV, where the interpolated dataset was used. However, in some cases the interpolated datasets (Models II and IV) seem to depict features that are not as clearly apparent from the original data (Models I and III), e.g. the cluster "Tsikara 2" between Fisoka and Tsikara target sites coincides with the interpreted extents of altered country rocks (Fig. 9d, see also Siron et al., 2018). It is difficult, if not virtually impossible, to reliably determine if the clustering within that area is affected by a more or less random association produced by the used methods, or if they in fact depict a signal from geological features, possibly amplified as a side effect of under-fitting due to the interpolation and SOM training. Spatially neighboring original sample points often show similar properties, as is apparent also from the results of this study (Fig. 6, and Models I and III in Fig. 8), but can still be considered independent measurements from each other. Due to the underlying process of IDW interpolation, the interpolated points between the original sample locations are a mixture of those samples from different locations. This may be the reason for some of the rim-shaped features of Models II and IV that are not expressed similarly within the original samples (e.g. the clusters "Fisoka 1" and "Fisoka 2", compare Models III and IV in Fig. 8). Curiously enough, in some cases the model from interpolated data appears to be lacking similar rims, although they exist in the clusters of the original points (e.g. the clusters associated with the Tsikara target site in the models I and II around the location labeled "T" in Fig. 8).

5.2. Potential for previously unidentified porphyry-type occurrences

Both methods delineate the geological differences previously recognized (e.g. Siron et al., 2018) between the intrusions in Fisoka, Tsikara and Skouries areas ("F", "T" and "Sk" in Figs. 6 and 9). These distinguishable geochemical fingerprints may be used as proxies for known types of porphyry intrusions and associated mineralization, even if such outcrops are not observed around all occurrences. In addition to the clusters from SOM models and RGB composite maps, we use cluster-normalized values (e.g. Fraser and Hodgkinson, 2009) for Cu and Au to discuss the potential exploration targets not recognized prior to this study.

Generally, the most significant PCs provide maximum distinction between the lithologies (Grunsky and de Caritat, 2019). As demonstrated in Section 4.1, PCs 1–5 from geochemical data using clr-transformation discriminate the lithologies and represent other signatures such as alteration or mineralization in the study area. The lesser PCs may represent random variation, or as described by Grunsky and de Caritat (2019), under-sampled processes, such as alteration and mineralization associated with a specific type of deposit occurring in a spatially narrow area, from which the signal could not be sufficiently picked up by the sampling density of the geochemical survey. Due to the presence of porphyry Au-Cu mineralization (Skouries) and Supergene Cu-Au mineralized porphyry (Fisoka) in the study area, simultaneous consideration of mineralized-related elements is extremely important for discerning geochemically favorable areas.

Geochemical association in PC2 achieved by clr-transformation shows associations of hypogene-related elements for weathering processes or supergene zones (element association Pb-As-Hg-Mo-Sb). PC5 also shows an association of Au-Cu-Mo-Bi elements demonstrating geochemical signature probably related to Au-Cu mineralization in the area. However, the multi-element signature from the dominant geological features of the region, represented by the most significant PCs (PC1-5) as well as the pseudo-lithological classification (i.e. the clusters) of the SOM models, can have a very different signature from the possibly under-sampled and unrecognized mineralization processes. The differences between the observed values of elements related to mineralization and estimated values derived from e.g. a linear regression model can define the anomalous residual that may represent under-sampled processes that are possibly associated with mineralization (Grunsky et al., 2014 and references therein). Instead of the linear regression method described by Grunsky et al. (2014), we use cluster-normalized element anomaly maps derived from SOM Model III, in order to investigate the possibly under-sampled processes represented by samples fortuitously overlying a zone of alteration and/or mineralization in the bedrock. This method described by e.g. Fraser and Hodgkinson (2009) is mathematically simple to use after the cluster analysis, and displays deviation of the sample values from their "expected" clusters mean values. The normalized value of a sample is calculated by subtracting the mean of the sample values within that cluster, and by dividing that result by the standard deviation of the sample values within that cluster. High normalized Cu and Au values may reflect processes that are potentially associated with under-sampled porphyry mineralization. These values plotted on a map show similarity with the area highlighted by the PC5 (negative values of PC5 associated with high normalized Cu and Au values in Fig. 10).

Most of the possible locations for porphyry-type intrusions or related alteration underneath the soil indicated by SOM clusters, RGB composite interpretations, and cluster-normalized Cu and Au values are already mapped and recognized as potential targets, but they also point to previously unknown intrusions or porphyry-related alteration in many locations of the study area. These locations comprise a set of possible subjects for future research and exploration. The list indices (i-iv) for the targets point to the corresponding labels in Figs. 6, 9 and 10, and the related target/mine sites are defined based on clusters from SOM models III and IV (see e.g. Fig. 8): i) Fisoka-related samples east from Tsikara intrusion, and Skouries-related samples ii) north from the Fisoka area, iii) in the northern parts of the schists and iv) in the southeastern part of the study area. From these, the sites i-ii may be the most promising targets, since they can be defined by all three applications (Figs. 6, 9 and 10), although the samples around the site ii display elevated cluster-normalized values only for Au. The sites iii-iv are most evident from SOM models (Fig. 9b) but less clear from RGB composites (Fig. 6b) and do not display elevated cluster-normalized Cu or Au values (Fig. 10). As discussed in Section 5.1, these Skouries-related clusters may represent some distal alteration or other processes not directly related to mineralization. The rest of the samples with strongly elevated cluster-normalized values of Cu and Au (e.g. all other samples with values above 3, in Fig. 10) are classified varyingly into different clusters in different models. The varying classifications may impact the cluster-normalized values of these samples and introduce higher uncertainty to the results, since the mean values used as a reference depend on the samples in the cluster. However, these values may still be anomalous enough to justify further investigation around their sampling locations.

6. Conclusions

1. Both PCA and SOM were able to classify the samples in groups that can be correlated to the existing interpretations of the local geology. The spatial extents and shapes of many of these groups closely resemble those of the interpreted lithological units and alteration zones.
2. The geochemical soil samples around the previously known three porphyry intrusions, Skouries, Fisoka and Tsikara, were all successfully recognized and further classified into separate groups, highlighting the known compositional differences of these intrusions.
3. Elemental associations recognized with both methods reveal specific geochemical signatures and spatial variations associated with mineralizing processes within the study area, most importantly with respect to porphyry Au-Cu type processes and occurrences.
4. Integrated analysis of the PCA and SOM outputs provide concordant and promising information on sites that may represent previously unknown occurrences of mineralization, and can therefore be

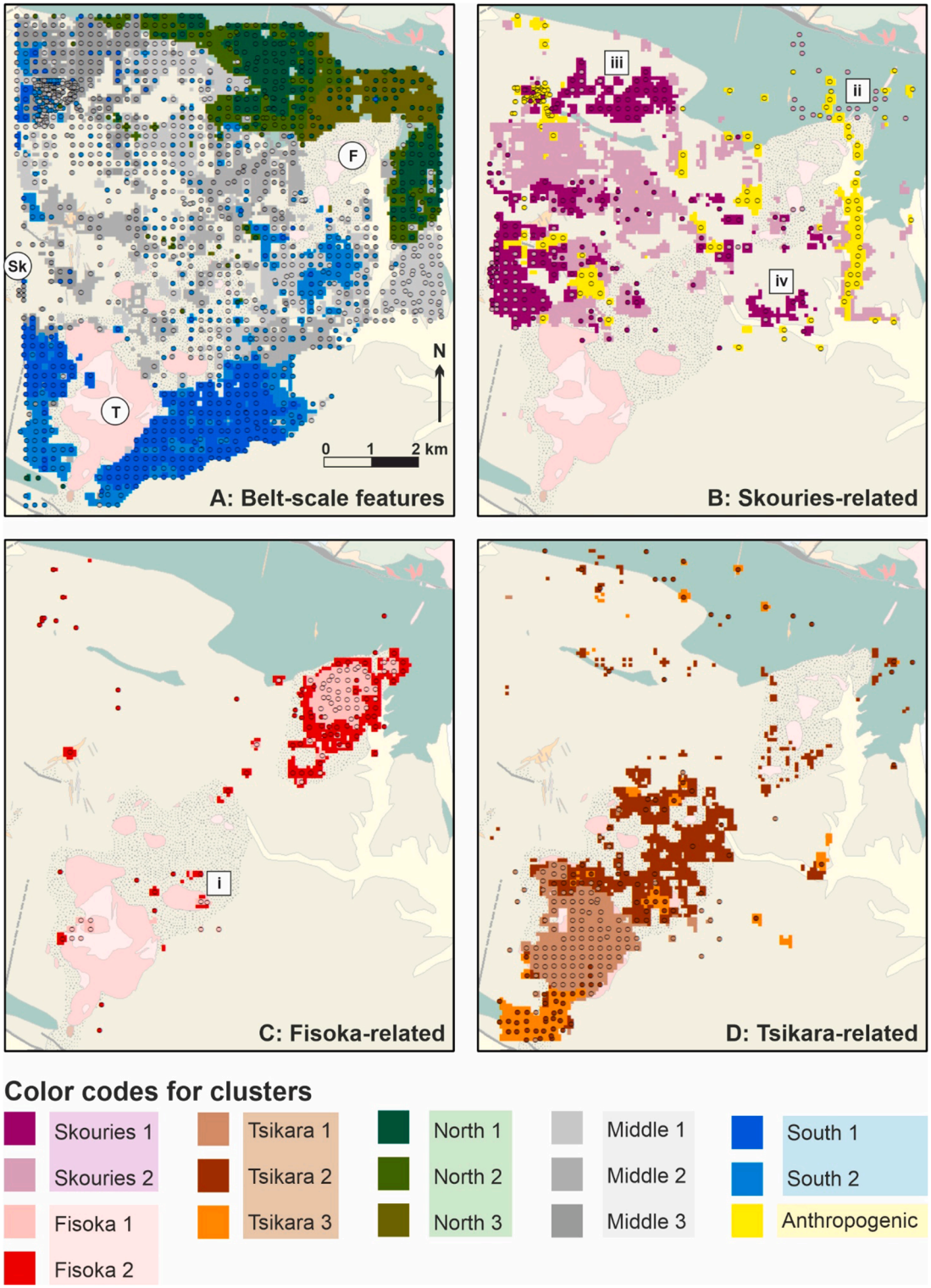

Fig. 9. Pseudo-geology based on overlaid samples from clusters of Models III and IV: A) belt-scale features and the locations of Skouries (Sk), Fisoka (F) and Tsikara (T); B) Skouries-related clusters and the cluster (yellow) at least partly representing anthropogenic anomalies (river valley in the eastern part); C) Fisoka-related clusters; and D) Tsikara-related clusters. Original sample points are displayed as colored circles, whereas the interpolated data are represented by squares/pixels. See Figs. 3 and 6 to compare these SOM clusters to PCA results, and Table 3 for detailed characteristics of the clusters. The underlying geological interpretations of the area are modified after Kockel et al. (1975), Siron et al. (2018) and unpublished work by Hellas Gold. The labels "i-iv" denote potential exploration targets discussed in section 5.2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

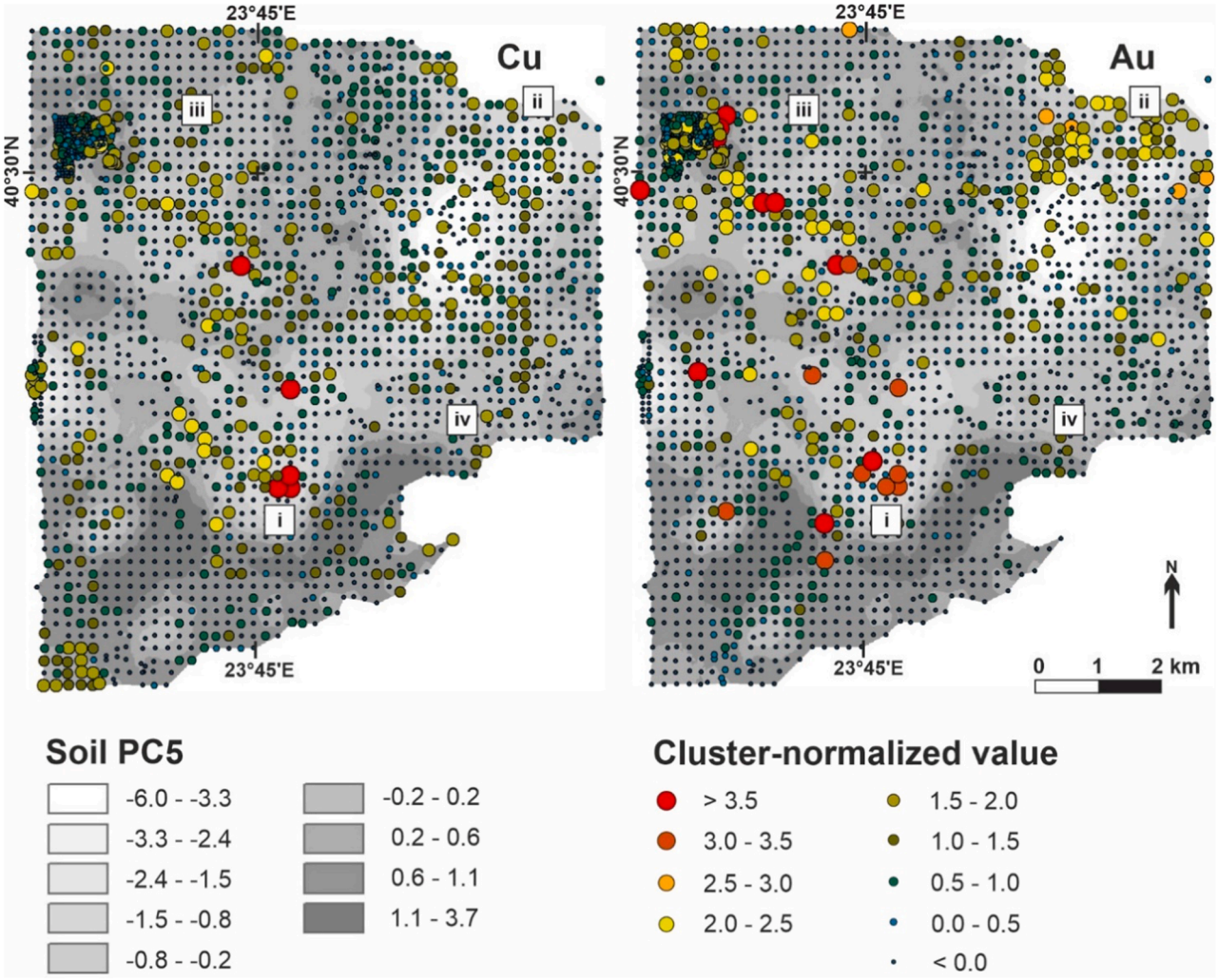

Fig. 10. Map of the cluster-normalized values of Cu and Au, depicting the deviation of observed sample values from their clusters "expected" mean values. The labels "i-iv" denote potential exploration targets discussed in Section 5.2.

directly used in the selection of possible future exploration targets that were not recognized from geological mapping or exploration work conducted in the area.

Declaration of Competing Interest

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..

Acknowledgements

This work was done as a part of the Smart Exploration project. Smart Exploration has received funding from the European Union's Horizon 2020 research and innovation programme under Grant agreement No. 775971. The SOM analysis was done using CSIRO's SiroSOM software. We thank Mehrdad Bastani for his contribution in preparing and organizing the magnetic data used in this work. We also thank the editor Franco Pirajno and the two anonymous reviewers for considerably improving the paper with their thorough critical reviews.

References

Airo, M.L., Hyvönen, E., Lerssi, J., Leväniemi, H., Ruotsalainen, A., 2014. Tips and tools for the application of GTK's airborne geophysical data. Geological Survey of Finland, Report of Investigation 215, 33 pp.

Aitchison, J., 1986. In: The Statistical Analysis of Compositional Data. Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London, UK, p. 416.

Arvanitidis, N.D. 1993. Regional ore geologic studies setting controls and distribution of Metallic ore deposit types in the Serbo-Macedonian and Western Rhodope zone. I.G. M.E. technical report.

Arvanitidis, N.D. 2003. Gold deposits in Greece genetic types and economic perspectives, Proceedings of the 7th Biennial SGA Meeting on "Mineral Exploration and Sustainable Development" Athens, Millpress Rotterdam, 941–943.

Baranov, V., 1957. A new method for interpretation of aeromagnetic maps: pseudo-gravimetric anomalies. Geophysics 22 (2), 359–382.

Baranov, V., Naudy, H., 1964. Numerical calculation of the formula of reduction to the magnetic pole. Geophysics 29, 67–79.

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.

Carneiro, C.C., Fraser, S.J., Crósta, A.P., Silva, A.M., Barros, C.E.M., 2012. Semiautomated geologic mapping using self-organizing maps and airborne geophysics in the Brazilian Amazon. Geophysics 77, K17–K24.

Carranza, E.J.M., 2008. Geochemical Anomaly and Mineral Prospectivity Mapping in GIS. Handbook of Exploration and Environmental Geochemistry. Elsevier, Amsterdam.

Cheng, Q., Bonham-Carter, G., Wang, W., Zhang, S., Li, W., Qinglin, X., 2011. A spatially weighted principal component analysis for multi-element geochemical data for mapping locations of felsic intrusions in the Gejiu mineral district of Yunnan, China. Computers & Geosciences 37 (5), 662–669.

Cottrell, M., Letrémy, P., 2005. Missing values: processing with the Kohonen algorithm. Applied Stochastic Models and Data Analysis 489–496.

Cracknell, M.J. and de Caritat, P. 2017. Catchment-based gold prospectivity analysis combining geochemical, geophysical and geological data across northern Australia. Geochemistry: Exploration, Environment, Analysis, 204–216.
Cracknell, M.J., Reading, A.M., McNeill, A.W., 2014. Mapping geology and volcanic-hosted massive sulphide alteration in the Helley-Mt Charter region, Tasmania, using Random ForestsTM and Self-Organizing Maps. Australian Journal of Earth Sciences 61, 287–304.
Davies, D.L., Bouldin, D.W., 1979. A cluster separation measure. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-1, 224–227.
Egozcue, J.J., Pawlowsky-Glahn, V., Mateu-Figueras, G., Barcelo-Vidal, C., 2003. Isometric logratio transformation for compositional data analysis. Mathematical Geology 35, 279–300.
Eldorado Gold Corporation. 2017. Resources and reserves: www.eldorado-gold.com/assets/resources-and-reserves/, accessed October 29, 2017.
Finlay, C.C., Maus, S., Beggan, C.D., Bondar, T.N., Chambodut, A., Chernova, T.A., Chulliat, A., Golovkov, V.P., Hamilton, B., Hamoudi, M., Holme, R., Hulot, G., Kuang, W., Langlais, B., Lesur, V., Lowes, F.J., Lühr, H., Macmillan, S., Mandea, M., McLean, S., Manoj, C., Menvielle, M., Michaelis, I., Olsen, N., Rauberg, J., Rother, M., Sabaka, T.J., Tangborn, A., Tøffner-Clausen, L., Thébault, E., Thomson, A.W.P., Wardinski, I., Wei, Z., Zvereva, T.I., 2010. International Geomagnetic Reference Field: the eleventh generation. Geophysical Journal International 183, 1216–1230. https://doi.org/10.1111/j.1365-246X.2010.04804.x.
Fraser, S.J. and Dickson, B.L. 2007. A new method for data integration and integrated data interpretation: selforganizing maps. 5th Decennial International Conference on Mineral Exploration, 907–910.
Fraser, S.J. and Hodgkinson, J.H. 2009. An investigation using SiroSOM for the analysis of QUEST stream-sediment and lake-sediment geochemical data. Geoscience BC Report 2009-14 – CSIRO E&M Report 2009/983, 64 pp.
Fraser, S.J., Wilson, G.A., Cox, L.H., Cuma, 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), 1–4.
Frei, R., 1995. Evolution of mineralizing fluid in the porphyry copper system of the Skouries deposit, northeast Chalkidiki (Greece): evidence from combined Pb-Sr and stable isotope data. Economic Geology 90 (4), 746–762.
Fugro Airborne Survey. 2008. Resolve electromagnetic and stinger mounted magnetic and radiometric geophysical survey for Hellas Gold S.A., Thessaloniki, Greece. Report # 07077, 118 pp.
Grunsky, E.C., 2010. The interpretation of geochemical survey data. Geochemistry: Exploration. Environment, Analysis 10, 27–74.
Grunsky, E.C., de Caritat, P., 2019. State-of-the-art analysis of geochemical data for mineral exploration. Geochemistry: Exploration, Environment, Analysis 20 (2), 217–232. https://doi.org/10.1144/geochem2019-031.
Grunsky, E.C., Mueller, U.A., Corrigan, D., 2014. A study of the lake sediment geochemistry of the Melville Peninsula using multivariate methods: applications for predictive geological mapping. Journal of Geochemical Exploration 141, 15–41. https://doi.org/10.1016/j.gexplo.2013.07.013.
Haines, H.S., 1998. In: A Structural Synthesis for Sector Vb of the Madem Lakkos Polymetallic Sulfide Deposit – Northeast Greece. University of London, London, p. 81. Unpublished M.Sc. thesis.
Harris, J.R., Grunsky, E.C. and Wilkinson, L. 1997. Developments in the effective use of lithogeochemistry in regional exploration programs: application of GIS technology. In: Gubins A.G. (ed.), Proceedings of the Exploration'97, Fourth Decennial International Conference on Mineral Exploration, 285–292.
Heinrich, C.H., Neubauer, Z.F., 2002. Cu – Au – Pb – Zn – Ag metallogeny of the Alpine – Balkan– Carpathian – Dinaride geodynamic province. Mineralium Deposita 37, 533–540.
van Hinsbergen, D.J.J., Schmid, S.M., 2012. Map view restoration of Aegean-West Anatolian accretion and extension since the Eocene. Tectonics 31 (5), n/a–n/a. https://doi.org/10.1029/2012TC003132.
Junno, N., Koivisto, E., Kukkonen, I., Malehmir, A., Montonen, M., 2019. Predicting missing seismic velocity values using Self-Organizing Maps to aid the interpretation of seismic reflection data from the Kevitsa Ni-Cu-PGE deposit in northern Finland. Minerals 9 (9), 529. https://doi.org/10.3390/min9090529.
Junno, N., Koivisto, E., Kukkonen, I., Malehmir, A., Wijns, C., Montonen, M., 2020. Data mining of petrophysical and lithogeochemical borehole data to elucidate the origin of seismic reflectivity within the Kevitsa Ni-Cu-PGE bearing intrusion, northern Finland. Geophysical Prospecting 68 (1), 82–102.
Kanellopoulos, C., Argyraki, A., 2013. Soil baseline geochemistry and plant response in areas of complex geology. Application to NW Euboea. Greece. Chemie der Erde 73, 519–532.
Kockel, F., Mollat, H., Gundlach, H., 1975. Hydrothermally Altered and (Copper) Mineralized Porphyritic Intrusions in the Serbo-Macedonian Massif (Greece). Mineralium Deposita 10, 195–204.
Kockel, F., Mollat, H., Walther, H.W., 1977. In: Erläuterungen Zur GEOLOGISCHEN KARTE DER CHALKIDIKI und Angrenzender Gebiete 1:100,000 (Nord-Griechenland). Bundesanstalt für Geowissenschaften und Rohstoffe, Hannover, pp. 1–119.
Kohonen, T., Hynninen, J., Kangas, J. and Laaksonen, J. 1996. SOM_PAK: The Self-Organizing Map Program Package, Report A31, Laboratory of Computer and Information Science, Helsinki University of Technology, Espoo, Finland.
Kohonen, T., 1982. Self-organized formation of topologically correct feature maps. Biological Cybernetics 43, 59–69.
Kohonen, T., 2013. Essentials of the self-organizing map. Neural Networks 37, 52–65.
Kounov, A., Wüthrich, E., Seward, D., Burg, J.P., Stockli, D., 2015. Low-temperature constraints on the Cenozoic thermal evolution of the Southern Rhodope Core Complex (Northern Greece). International Journal of Earth Science 104, 1337–1352.
Kuhn, S., Cracknell, M.J., Reading, A.M., 2019. Lithological mapping in the Central African Copper Belt using Random Forests and clustering: Strategies for optimised results. Ore Geology Reviews 112, 103015. https://doi.org/10.1016/j.oregeorev.2019.103015.
Leväniemi, H., Hulkki, H., Tiainen, M., 2016. SOM guided fuzzy logic prospectivity model for gold in the Häme Belt, south-western Finland. Journal of African Earth Sciences 128, 72–83.
Lloyd, S.P., 1957. Least Squares Quantization in PCM. Technical note, Bell Laboratories. Published 1982. IEEE Transactions on Information Theory 28, 129–137.
Loughlin, W.P., 1991. Principal component analysis for alteration mapping. Photogrammetric Engineering and Remote Sensing 57, 1163–1169.
MacQueen, J. 1967. Some methods for classification and analysis of multivariate observations. Proceedings of the Fifth Symposium on Mathematical Statistics and Probability, 1, 281–297.
Melfos, V., Voudouris, P.C., 2012. Geological, mineralogical and geochemical aspects for critical and rare metals in Greece. Minerals 2012, 300–317.
Nabighian, M.N., 1972. The analytic signal of two-dimensional magnetic bodies with polygonal cross-section: its properties and use for automated anomaly interpretation. Geophysics 37 (3), 507–517.
Nabighian, M.N., 1974. Additional comments on the analytic signal of two-dimensional magnetic bodies with polygonal cross-section. Geophysics 39, 85–92.
Park, Y., Céréghino, R., Compin, A., Lek, S., 2003. Applications of artificial neural networks for patterning and predicting aquatic insect species richness in running waters. Ecological Modelling 160, 265–280.
Roest, W.R., Verhoef, J., Pilkington, M., 1992. Magnetic interpretation using the 3D analytic signal. Geophysics 57, 116–125.
Sadeghi, M., Morris, G.A., Carranza, E.J.M., Ladenberger, A., Andersson, M., 2013a. Rare earth element distribution and mineralization in Sweden: An application of principal component analysis to FOREGS soil geochemistry. Journal of geochemical exploration 133, 160–175.
Sadeghi, M., Petrosino, P., Ladenberger, A., Albanese, S., Andersson, M., Morris, G., Lima, A., De Vivo, B., 2013b. Ce, La and Y concentration in agricultural and grazing soils of Europe- an exercise for mapping and exploration. Journal of geochemical exploration 133, 202–213.
Sadeghi, M., Billay, A., Carranza, E.J.M., 2015. Analysis and mapping of soil geochemical anomalies: Implications for bedrock mapping and gold exploration in Giyani area, South Africa. Journal of Geochemical Exploration 154, 180–193. https://doi.org/10.1016/j.gexplo.2014.11.018.
Schmid, S.M., Bernoulli, D., Fügenschuh, B., Matenco, L., Schefer, S., Schuster, R., Tischler, M., Ustaszewski, K., 2008. The Alpine-Carpathian-Dinaridic orogenic system: correlation and evolution of tectonic units. Swiss Journal of Geosciences 101, 139–183.
Siron, C.R., Rhys, D., Thompson, J.F.H., Baker, T., Veligrakis, T., Camacho, A., Dalampiras, L., 2018. Structural controls on porphyry Au-Cu and Au-rich polymetallic carbonate-hosted replacement deposits of the Kassandra mining district, northern Greece. Economic Geology, Bulletin of the Society of Economic Geologists 113, 309–345.
Siron, C.R., Thompson, J.F.H., Baker, T., Friedman, R., Tsitsanis, P., Russel, S., Randal, S. and Mortensen, J. 2016. Magmatic and metallogenic framework of Au-Cu prophyry and polymetallic carbonate-hosted replacement deposits of the Kassandra mining district, northern Greece. In: J. Richards (ed.) Tectonics and metallogeny of the Tethyan orogenic belt, Special Publication of the Society of Economic Geologist 19, 29-55. Structural Controls on Porphyry Au-Cu and Au-Rich Polymetallic Carbonate-Hosted Replacement Deposits of the Kassandra Mining District, Northern Greece. Economic Geology, 113, 309–345.
Sullivan, N.A., Zajacz, Z., Brenan, J.M., 2018. The solubility of Pd and Au in hydrous intermediate silicate melts: The effect of oxygen fugacity and the addition of Cl and S. Geochimica et Cosmochimica Acta 2018 (231), 15–29.
Tsitsanis, P., Siron, C.R., Baker, T., Kyriakoglou, I., 2016. The application of SWIR and multi-element geochemistry in exploration at the Fisoka Cu ± Au porphyry system, Halkidiki. N. Greece. Society of Economic Geologists. SEG-MJD 2016 Conference abstract.
Verduzco, B., Fairhead, J.D., Green, C.M., MacKenzie, C., 2004. New insights into magnetic derivatives for structural mapping. The Leading Edge 23, 116–119.
Vesanto, J., 1999. SOM-Based Data Visualization Methods. Intelligent Data Analysis 3, 111–126.
Vesanto, J., Himberg, J., Alhoniemi, E., Parhankangas, J., 2000. SOM Toolbox for Matlab 5. Laboratory of Computer and Information Science Report A57. Helsinki University of Technology, Espoo, Finland.