Novel technologies for greenfield exploration
Edited by Pertti Sarala
Geological Survey of Finland, Special Paper 57, 87–120, 2015

A NOVEL SPATIAL ANALYSIS APPROACH FOR ASSESSING REGIONAL-SCALE MINERAL PROSPECTIVITY IN NORTHERN FINLAND

by
Johanna Torppa[1)], Maarit Middleton[2)], Eija Hyvönen[2)], Jouni Lerssi[1)] and Stephen Fraser[3)]

Torppa, J., Middleton, M, Hyvönen, E., Lerssi, J. & Fraser, S. 2015. A novel spatial analysis approach for assessing regional-scale mineral prospectivity in northern Finland. *Geological Survey of Finland, Special Paper 57*, 87–120, 25 figures and 5 tables.

We applied spatial analysis techniques in investigating regional prospectivity in two study areas in Lapland, Finland. Different combinations of self-organizing maps, K-means clustering and fuzzy logic were used to analyse geophysical, geochemical and structural data from study areas in Enontekiö and Savukoski. The study demonstrated how the distribution of a suitable multidimensional dataset can be interpreted in terms of mineral prospectivity by utilizing spatial analysis, without any training data. The resulting models can help in highlighting and drawing attention to exploration targets that might be interesting for further investigations.

Keywords (GeoRef Thesaurus, AGI): self-organizing maps, fuzzy logic, spatial analysis, mineral exploration, Enontekiö, Savukoski, Finland

[1)] *Geological Survey of Finland, P.O. Box 1237, FI-70211 Kuopio, Finland*
[2)] *Geological Survey of Finland, P.O. Box 77, FI-96101 Rovaniemi, Finland, tel. +358 (0) 20950 34345*
[3)] *CSIRO Mineral Resources Flagship, Queensland Centre for Advanced Technologies, Technology Court, Pullenvale 4069 Qld, P.O. Box 883, Kenmore 4069 Qld, Australia, tel. +61 (0) 733274544*

E-mail: johanna.torppa@gtk.fi

INTRODUCTION

In mineral prospecting, regional-scale prospectivity mapping has an important role in selecting areas for targeted field observations and target-scale mineral potential mapping. In Finland, regional airborne geophysical measurements, till geochemical analysis and field gravity measurements are available for the entire country. In addition, airborne gradiometric and time-domain electromagnetic measurements have been carried out in regions of special interest. The applicability of the data in evaluating regional mineral prospectivity in northern Finland has been investigated, for instance, by Nykänen & Salmirinne (2007), Nykänen et al. (2008a) and Nykänen et al. (2008b) using supervised and knowledge-driven spatial analysis techniques. Supervised techniques, which rely on training data, are useful in well-explored brown fields, such as the Central Lapland Greenstone Belt (CLGB), where a number of known deposits exist. Usually, however, there are very few or no known deposits in the area of interest, making supervised methods inapplicable. Knowledge-driven methods, which apply expert knowledge instead of training data, can be applied when there is a basis to assume certain dependencies between mineral occurrences and measured geophysical and geochemical quantities. These dependencies are not unique, however, making the use of knowledge-driven methods often challenging. An efficient approach to begin prospectivity modelling in poorly explored and green-field areas is to use unsupervised spatial analysis methods (from here on *clustering*) to find natural clusters in the data, and later use expert knowledge to interpret the meaning of the clusters and distinguish the interesting ones. Although clustering does not directly provide information on the prospectivity of a particular mineral (if such information is not explicitly contained in the dataset), anomalous features, typical of mineralization in general, can be detected. The data distribution within clusters can also be used to make a rough division between prospective and non-prospective regions.

Partitive and hierarchical clustering methods have been widely used to analyse multidimensional datasets, but the computational effort quickly grows as the number of data points increases. We made use of self-organizing maps (SOM) (Kohonen 2001), which represent a neural network-based clustering method that performs faster than traditional methods (such as K-means) for large datasets, does not assume normally distributed data or linear relationships within the dataset and simultaneously performs clustering and dimension reduction. SOM has been compared with other clustering methods, for instance, by Waller et al. (1998), Bacao et al. (2005) and Abbas (2008), who proposed that SOM performs better than many other methods, while Flexer (1999) was more critical. The types of datasets and/or exact methodology of the studies, however, have often been poorly described and differ from ours, making it difficult to apply the results to our problem.

We followed the two-level approach described in Vesanto & Alhoniemi (2000), whereby SOM is used to carry out pre-clustering, and K-means to further cluster the SOM proto-clusters to a handful (<30) of clusters. The approach is here referred to as *SOMKm* and has been used in a variety of geological problems, such as in stream sediment studies (Fraser & Hodgkinson 2009), lithological interpretation of 3D geophysical data (Fraser et al. 2012) and analysis of the mineral deposit databases (Bierlein et al. 2008). To our knowledge, SOM has not been applied to the evaluation of regional mineral prospectivity. Some of the results are here directly derived from the SOM clustering, without the execution of K-means, but we consider this as a special case of SOMKm, and use the term *SOM* only to emphasize the SOM part of the procedure. When interpreting the meaning of clusters and anomalous regions revealed using unsupervised clustering, expert knowledge must be used, qualitatively or quantitatively, to bind the regions to certain geological features and possible prospective areas. For the quantitative interpretation of the SOMKm results, we used fuzzy logic.

The choice of datasets used in mineral prospectivity modelling is rarely unique, because we do not have unique models for relating the measured quantities to the mineral potential: numerous geophysical and geochemical quantities have been observed to be more or less vaguely related to the existence of deposits, but the exact character of signatures depends on the detailed geological setting. Based on the best of our knowledge, we used datasets that we assumed to somehow be affected by a certain deposit type. The choice of suitable datasets also depends on how effectively the available relevant datasets can be adjusted together. The

datasets usually differ in character and combining them into one set requires modifications such as interpolation, level correction and spatial transformation. Such operations result in information loss, leading to a trade-off between the number of datasets and the information content of each dataset.

In this study, we used SOMKm to cluster geophysical and geochemical data in estimating the regional 2D prospectivity of two sample areas in Enontekiö and Savukoski. In Enontekiö, prospectivity was evaluated for iron oxide copper-gold (IOCG) type deposits, which is a highly versatile group of Fe-oxide bearing deposits often associated with Au and Cu, but also with REEs. IOCGs are commonly characterized by magnetic and dense lithology and strong structural control in genesis. High radiometrics, especially U, can occasionally occur, but electromagnetic properties can be variable. Host rocks commonly exhibit extensive alteration patterns, but the alteration types vary greatly (Hitzman 2000, Niiranen 2005). In Savukoski, a prospectivity model was constructed for orogenic gold deposits, which have strong structural and lithological control revealed by gravity and magnetic gradients (Bierlein et al. 2006, Airo & Mertanen 2008). The goal of this study was to formulate an efficient workflow for estimating prospectivity without using any training data. Our emphasis was on the preprocessing of data and the ways of combining spatially different types of data, and here we discuss the most efficient way to visualize the prospectivity information contained in the dataset.

STUDY AREAS AND MATERIALS

Geology and tectonic setting of Enontekiö and Savukoski

Finnish Lapland is part of the Fennoscandian shield consisting of four main regions (Figs. 1a and b): 1) Archaean granite-gneiss complexes and greenstone belts, 2) Palaeoproterozoic greenstone belts, 3) the 1.9 Ga Lapland granulite belt and 4) Svecofennian 1.89–1.80 Ga syn- to late and 1.80–1.77 Ga post-orogenic granitoids (Hanski et al. 1997).

The Enontekiö study area (Fig. 1c) is composed of Archaean granitoids and schists on the western side and Palaeoproterozoic metasedimentary and metavolcanic rocks in the central and eastern parts (Bedrock of Finland – DigiKP). In addition, intrusive mafic rocks such as gabbros and diabases are present. Currently, the bedrock geology and stratigraphy is not well understood, but in the areas having the shallowest sediment cover, diverse mineral potential has been discovered (e.g., Fe-Cu-Au, Ni-Cu-PGE, Mo, see GTK Ore deposit database, http://www.gtk.fi/tietopalvelut/karttapalvelut/). The area is considered potential for IOCGs because of being geologically analogous to the Fe-Cu±Au deposits of northern Sweden situated in the vicinity of vulcanite belts and major crustal faults (Bergman et. al 2001), the presence of the major crustal ductile Karesuando-Arjeplog NE–SW-oriented shear zone extending from Sweden and Norway across the study area (Bergman et al. 2001), and the existence of alteration by scapolitization (personal communication with H. Hulkki, 2014).

The Savukoski study area (Fig. 1d) is geologically part of the Palaeoproterozoic CLGB, which hosts more than 30 known gold occurrences (Mineral Deposit Database of Finland). The rocks in CLGB show a history of episodic sedimentation, magmatic activity and hydrothermal alteration. Orogenic gold deposits in the region are mainly hosted by volcanogenic sedimentary and mafic to ultramafic volcanic rocks. Most of the known deposits were formed during the 1.85–1.79 Ga collision phase, some possibly during the 1.92–1.88 Ga compression (Sorjonen-Ward et al. 1992, 2003, Eilu et al. 2007). The majority of the known CLGB deposits are associated with the Sirkka shear zone (SSZ); some are within the main SSZ and some in subsidiary fault and shear zones within 3 km of the SSZ. The bedrock in the Savukoski study area is poorly studied, and there is only one known gold deposit just outside the area in the east, in the contact of the mafic volcanic rocks and the surrounding biotite paragneisses. In the southeast corner of the area is the Akanvaara layered intrusion, studied in detail by Mutanen (1997). A comprehensive overview of gold in the CLGB is provided in Ojala (2007).

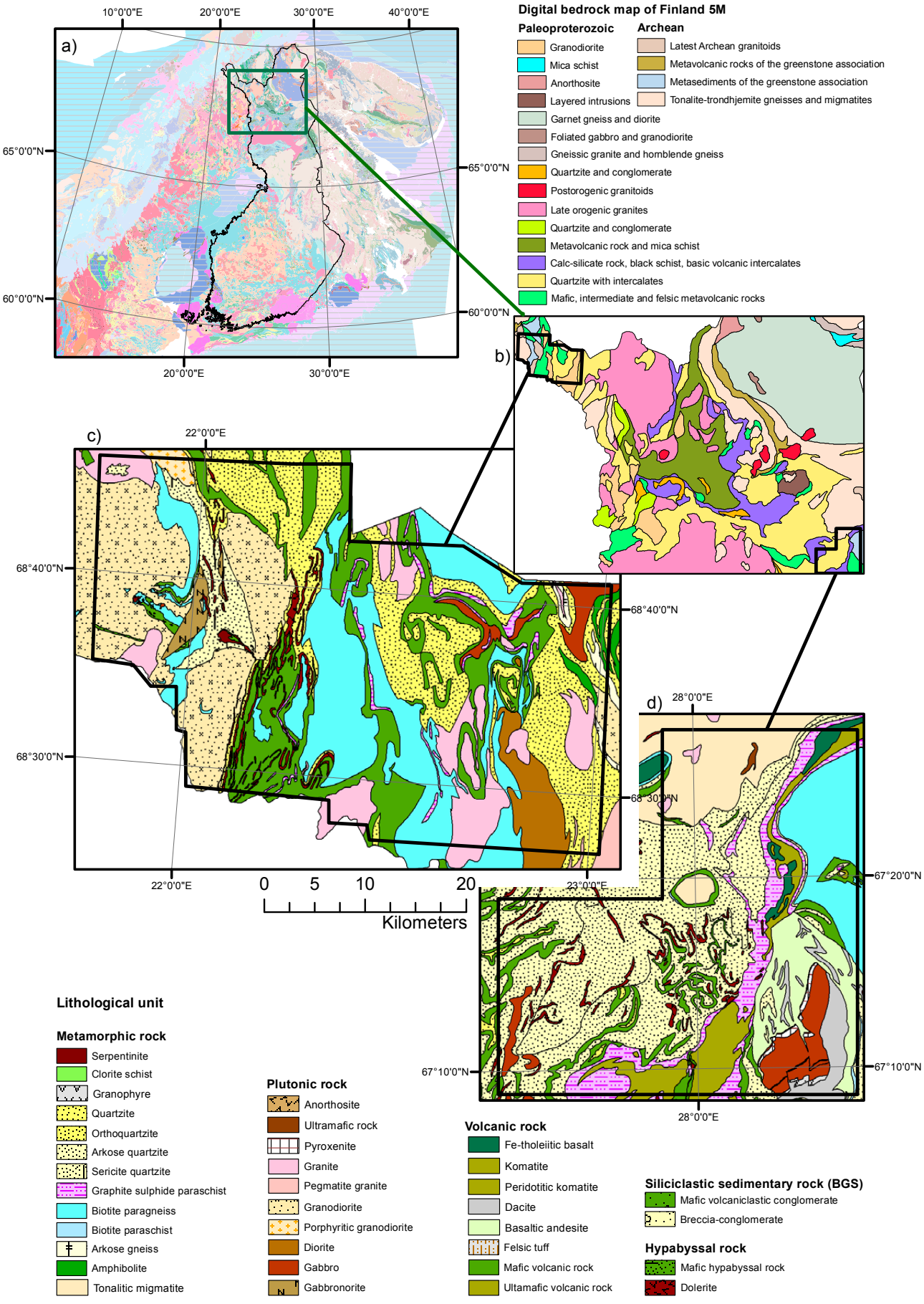

Fig. 1. Location and lithology of the study areas. Bedrock maps (Bedrock of Finland – DigiKP).

Geophysics

For both Enontekiö and Savukoski study areas, we used the nationwide airborne geophysical data collected by the Geological Survey of Finland (GTK) in 1973–2007. Simultaneous magnetic, frequency-domain electromagnetic and radiometric measurements were carried out with 200-m line spacing at a nominal altitude of 30–40 m using a fixed-wing aircraft with vertical coplanar coils (coaxial until 1979) for the electromagnetics (Hautaniemi et al. 2005). The Enontekiö study area consists of three flight areas, two of which were flown in 1983 and one in 2002, and the Savukoski study area has been flown on five occasions from 1979–1992 (Fig. 2). From the instrumentation statistics in Table 1, it is quite clear that the 2002 measurements in Enontekiö are more informative than the older ones due to the use of two frequencies, instead of one, for the electromagnetic measurements, the more frequent registration of the magnetic field and the larger number of spectral channels used in radiometric measurements. The caesium magnetometer used in 2002 and 1992 provided improved the accuracy and coverage of magnetic data. A more detailed description of each measurement system and data processing is provided in Korhonen (2005) for magnetics, Suppala et al. (2005) for electromagnetics and Hyvönen et al. (2005) for radiometrics.

In addition to the airborne national survey, airborne gradiometric gravity data were available from the entire Savukoski study area. Measurements were carried out in 2011 using a FALCON™ airborne gravity gradiometer. The line spacing was 500 m and the nominal flight altitude 107 m. FALCON™ is a partial tensor system with 8 accelerometers on a single circular spinning disk, and it measures two horizontal gravity gradients. In Enontekiö, additional field gravity data were used in areas indicated in Figure 2. Measurements were carried out by GTK in 1980–1983 (Worden) and 2011 (Scintrex CG5) with an average station density varying from 1 to 6 points per square kilometre. The 10-m digital elevation model of the National Land Survey of Finland was also used in the Enontekiö study area. The 10-m DEM was generated using various data sources, including updated contours, aerial photographs and all the individual points with known elevation.

Airborne magnetics

A few reductions were carried out for the measured total magnetic intensity to reveal the features associated with the sources in the Earth's topmost crust whose mineral deposits are within our reach. Here, we do not provide the mathematical details of the pre-processing methods, but only references to further information.

After subtracting the International Geomagnetic Reference Field from the total magnetic intensity, reduction to the northern pole (RTP) was carried out to model what the field would be like if measured at the North Pole, i.e., in a vertical inducing magnetic field. The RTP-reduced field, hereafter referred to as *RMI* (for RTP-reduced total Magnetic Intensity), was computed using the formula in Grant and Dodds (1972). In an *RMI* map, anomalies are located directly above the respective sources and the asymmetry of the field is entirely caused by the asymmetry of the sources, provided that the magnetization of the rocks is caused by induction. Magnetization of the basement rocks is, indeed, usually induction-dominated, but the remanent magnetization component can be significant in mineralized zones containing magnetite and/or monoclinic pyrrhotite, which makes the

Table 1. Statistics of instrumentation in the national airborne geophysical survey at the time of the Enontekiö and Savukoski area observations. Nm = number of magnetometers, C/P = magnetometer type (C = Caesium, P = proton), Fm = registration frequency of the magnetic field, D = coil distance, F = frequency of the primary field, M = moment, Fem = registration frequency of the electromagnetic field, V = crystal volume and Nr = number of spectral channels.

	Magnetometers				Electromagnetics			Radiometrics	
Year	Nm	C/P	Fm/s^{-1}	D /m	F /Hz	M	Fem/s^{-1}	V/l	Nr
1979	2	P	2	25.0	3220	127	2	27.3	54
1980	2	P	2	21.44	3222	105	4	25.0	120
1982–83	2	P	2	21.36	3113	105	4	25.0	120
1987	3	P	4	21.36	3113	105	4	25.0	120
1992	2	C	4	21.36	3113	105	4	25.0	120
2002	2	C	10	21.36	3125/14368	115/55	4	33+8	256

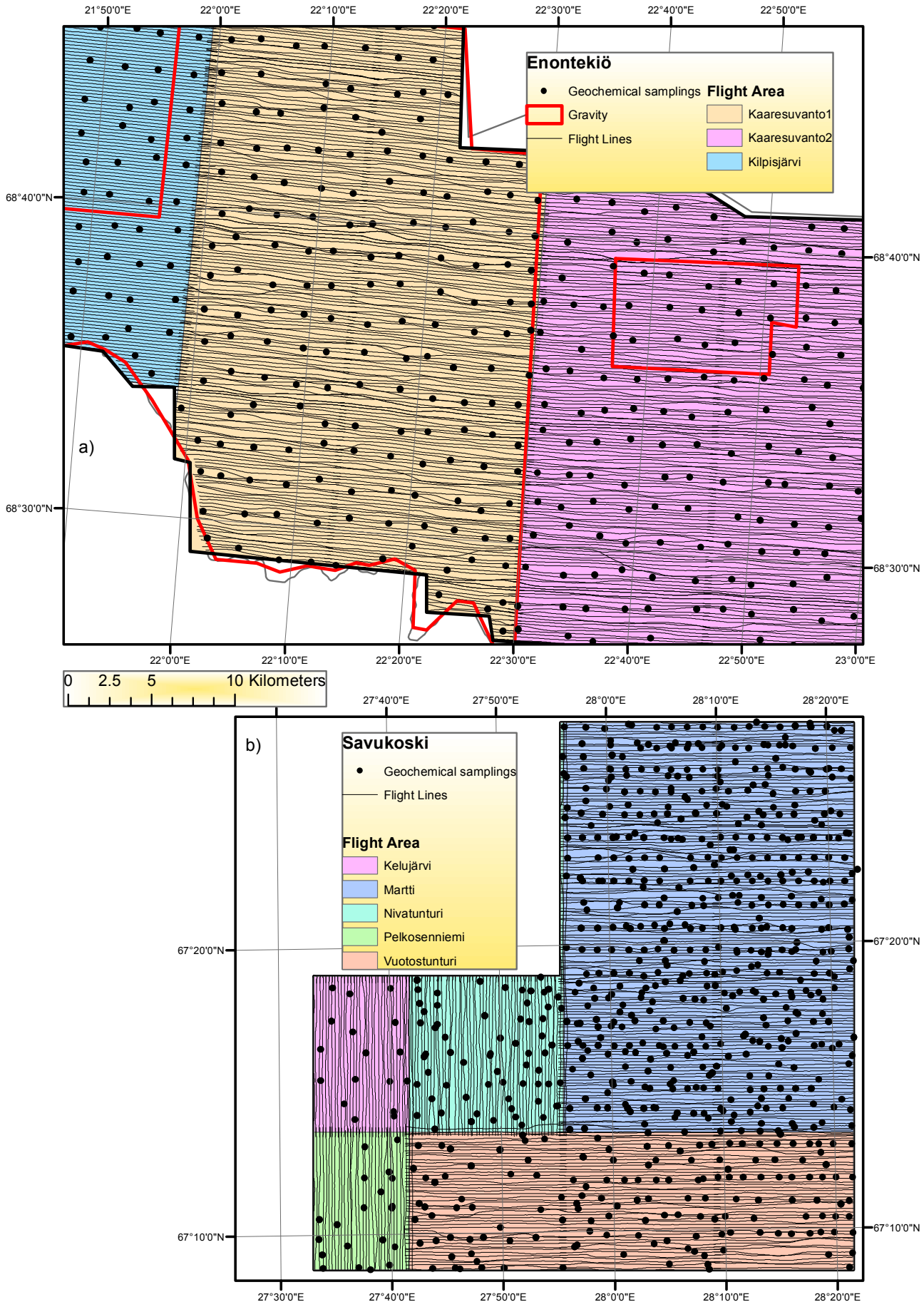

Fig. 2. Flight areas, flight lines, geochemical sampling points and Enontekiö gravity measurement areas.

use of RTP questionable in such areas. The quantities that we used in this study to enhance the local variations in the magnetic intensity were the analytic signal (RMI_{AS}), tilt derivative (RMI_{TDR}) and first vertical derivative (RMI_{1V}) of *RMI*, as well as the contrast normalized *RMI* (RMI_{CN}). The analytic signal vector $\boldsymbol{RMI_{AS}}$ (Nabighian 1984) gives the direction of maximum change of magnetic intensity, and its length gives the magnitude of the change. RMI_{AS} does not provide information on the tilt of the magnetic source, but it is efficient in determining the locations of magnetic anomalies. Maxima of RMI_{AS} are located at the edges of extensive magnetic sources, and directly above narrow ones. $RMI_{TDR,}$ is the angle between the horizontal *x,y*-plane and $\boldsymbol{RMI_{AS}}$. According to Verduzco et al. (2004), RMI_{TDR} may be more effective in discriminating between signal and noise than RMI_{1V}, which is computed for *RMI* in the wavenumber domain using Fourier transform for upward continuation of the field (e.g., Blakely 1995a). RMI_{CN} is computed using a non-linear space variant contrast stretch filter that removes long wavelength variation and enhances short wavelength variation. Preprocessing was carried out using *Intrepid* (Intrepid Geophysics, Melbourne, Australia) and *Oasis Montaj* (Geosoft, Toronto, Canada) software.

Airborne electromagnetic measurements

Electromagnetic measurements only provide information on the surface layer, and reductions typical for potential fields (gravity, magnetic) are not needed. Moreover, anomalies rather than changes are of interest, supporting the use of measured values rather than derivatives. Conductivity, calculated from the measurements, is often used as electromagnetic information, but to minimize the amount of data interpretation prior to SOMKm analysis, we used the measured in-phase (EM_{Re}, in the same phase as the primary field) and quadrature (EM_{Im}, 90° shifted to the primary field) components of the measured electromagnetic response and their ratio EM_{Re} / EM_{Im}. Values of EM_{Im} close to zero in the denominator were transformed to a constant small value to prevent the ratio from approaching infinity.

Gravity

Field gravity data were transformed to Bouguer anomalies (G_z). The field was continued up to 300 m to signify the depth of the gravity bodies and exclude the gravity variation due to the varying overburden thickness and weathered bedrock. Upward continuation of gravity data is analogous to that of magnetic data, i.e., it is carried out in the wavenumber domain using Fourier transform (Blakely 1995a). Preprocessing was carried out using *Oasis*.

FALCON™ airborne gradiometric gravity data consisted of two horizontal gravity gradients that were spatially transformed using the equivalent-source method (Dampney 1969) and corrected to a terrain density of 2.67 g/cc to produce vertical gravity conformed to regional gravity (G_{zRG}, Dransfield 2009). The vertical gravity gradient (G_{zRG_1V}) was computed according to Blakely (1995a). All the preprocessing was performed by Fugro Airborne Surveys Pty Ltd., who carried out the measurements.

Radiometry

Radiometric data consisted of K, Th and U concentrations calculated from the measured gamma-ray radiation spectrum emitted from the bedrock. The radiometric signals of crystalline bedrock in northern Finland are significantly complicated by Quaternary transportation of glacial drift, variation in the thickness of the sediment cover and weathering intensity of the bedrock, and attenuation by water and vegetation in the humid climatic conditions (Hyvönen et al. 2005). Thus, careful preprocessing was carried out by replacing data in wetlands and waterbodies with Null values, by spatial high-pass filtering to sharpen the anomalies, and by normalization of the effect of radon according to the IAEA recommendation (2003).

Enontekiö

IOCGs are often magnetite bearing (Hitzman 2000), thus exhibiting increased magnetic susceptibility. To emphasize the lateral variation in the magnetic field and to reduce the effect of remanence and magnetization, RMI_{AS} was included in the Enontekiö dataset (Fig. 3a). Electromagnetic signals of IOCSs are ambiguous and electrical properties of the deposits are poorly reported (Hitzman 2000). Magnetite-bearing deposits, which have a high magnetic permeability (susceptibility), are poor conductors and have a negative EM_{Re} response, whereas sulphide-bearing deposits are characterized by high EM_{Re}. In Enontekiö, we used the high-pass filtered level corrected in-phase component EM_{Re_HP} (Fig. 3c). Gravity values of IOCGs in the presence of metal richness are high

(Nisbet et al. 2000). The Bouguer anomaly G_z was high-pass filtered (G_{z_HP}) in order to emphasize the local-scale variation in the gravity field (Fig. 3b). IOCG mineralizations may be characterized by increased radiation from Th and/or U and alteration halos by K (Niiranen 2005, Williams 2001), and GTK's airborne radiometric data were thus included in the Enontekiö dataset as high-pass filtered concentrations K_{HP}, Th_{HP} and U_{HP} (Figs. 3d–f).

All the Enontekiö geophysical data were interpolated to grids with square pixels of size 50 m (Figs. 3 and 5, Table 3).

Savukoski

The most dominant characteristic of orogenic gold occurrences is usually their structural control and correlation with magnetic minima (e.g. Patison 2007, Airo 2007). For structural control, we used the magnetic RMI_{AS}, RMI_{1V} and RMI_{CN} (Figs. 4 a–c).

A high conductivity is typical due to the existence of, for instance, conductive sulphides and black schists, often related to orogenic gold deposits. Of the electromagnetics, we used EM_{Re}/EM_{Im} and EM_{Re}, high values of which represent conductive materials (Figs. 4f and e, respectively). The ratio is especially useful, because for good conductors $EM_{Re}/EM_{Im} > 1$ and for poor conductors $EM_{Re}/EM_{Im} < 1$, irrespective of the absolute values of EM_{Re} and EM_{Im}. High magnetic susceptibility reduces $EM_{Re,}$ but should not cause problems for good conductors. For poor conductors, EM_{Re} may turn negative for magnetic materials. In gravimetric data, orogenic gold occurrences are often evident as near-surface horizontal gradient maxima, especially in locations where rocks of contrasting densities are juxtaposed. We used the vertical gravity data conformed to regional gravity (G_{z_RG}) and the vertical gravity gradient G_{zRG_1v} (Figs. 4g and h, respectively).

All the Savukoski geophysical data were interpolated to grids with square pixels of size 100 m (Figs. 4 and 6, Table 3).

Structures

Structural information is partly derived from the airborne magnetic measurements and partly from geological maps. To detect structures, where the magnetization and/or density of the rocks usually abruptly changes, we used visual interpretation as well as numerical worming methods in *Intrepid* and in the GX toolset by USGS (Phillips 2007) implemented in *Oasis Montaj*. Numerically calculated worms were computed for various levels of upward continuation of *RMI* (Blakely 1995a, b) to mimic the field as it would be if measured from a higher altitude. In practice, upward continuation smoothens out the high-frequency features caused by near-surface sources and brings up the low-frequency features caused by deeper sources. The data were cleaned by excluding surface noise (i.e., near-surface worms) and the deepest worms, which were uncertain.

Enontekiö

The genesis of IOCGs is associated with major crustal-scale fault and shear zones (Niiranen 2005, Hitzman 2000, Sandrin et al. 2009). They do not necessarily form in the main shear zone but in intersecting first- or second-order branches, fault bends, lithological contacts, intersections of these, step-over structures or breccia bodies. Therefore, we compiled zero contours and minima worms of RMI_{TDR} as well as horizontal gradient maxima of gravity (Blakely & Simpson 1986). Worms were computed for upward continuation levels from 200 m to 1000 m and smoothed and generalized to exclude random noise. Lineaments interpreted visually from RMI_{TDR} and the horizontal gradient of gravity were merged with the worms derived from the same data origin. Line features were also visually interpreted from EM_{Re} and the digital elevation model (see Middleton et al., 2014, for details). Eventually, all vector lineaments were merged into one vector layer. The abundance of the interpreted linear structures, combined from these six different data sources, was imputed into the SOMKm as the density of the structures (S_ρ, Fig. 3g) computed to a 50-m grid size, similarly to other datasets. The line density gives the magnitude per unit area from the line features that fall within a radius of 200 m around each 50-m cell. This approach emphasizes the depth of the structures, as IOCG are related to major crustal level structures, and their intersections, which are also more potential for the origin of IOCGs.

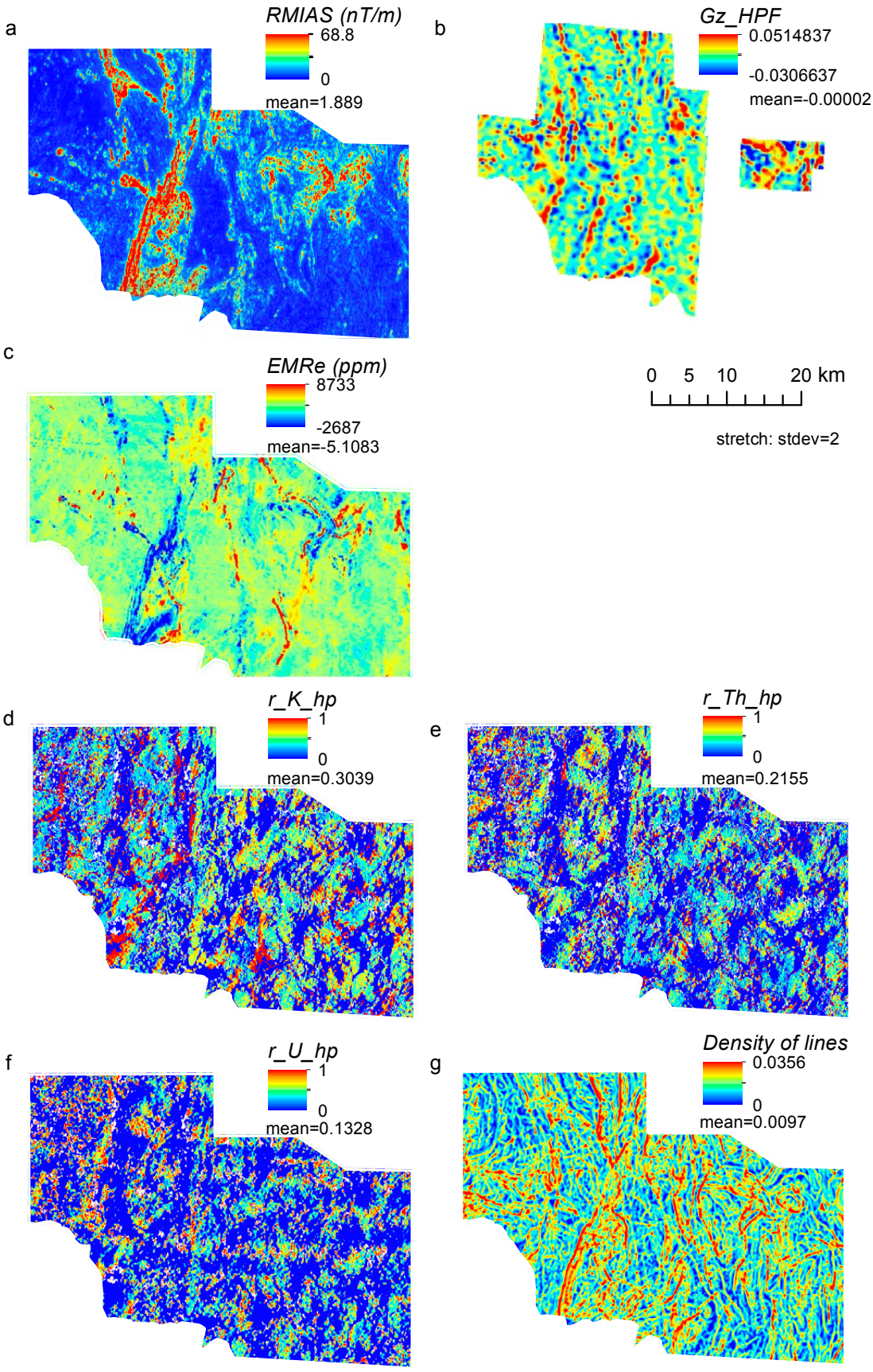

Fig. 3. Enontekiö preprocessed airborne geophysical and structural data, field gravity data and structural data. a) RM_{IAS} = analytic signal of the R_{TP} reduced magnetic field, b) Gz_{HP} = high-pass filtered gravity Bouguer anomaly, c) EM_{Re} = in-phase component of electromagnetics, d) K_{HP}, e) Th_{HP} and e) U_{HP} = high-pass-filtered potassium, thorium and uranium concentrations, respectively, and f) S_{ρ} = density of all the input structural elements. The data range and mean are shown for each dataset. Colouring is stretched over two standard deviations.

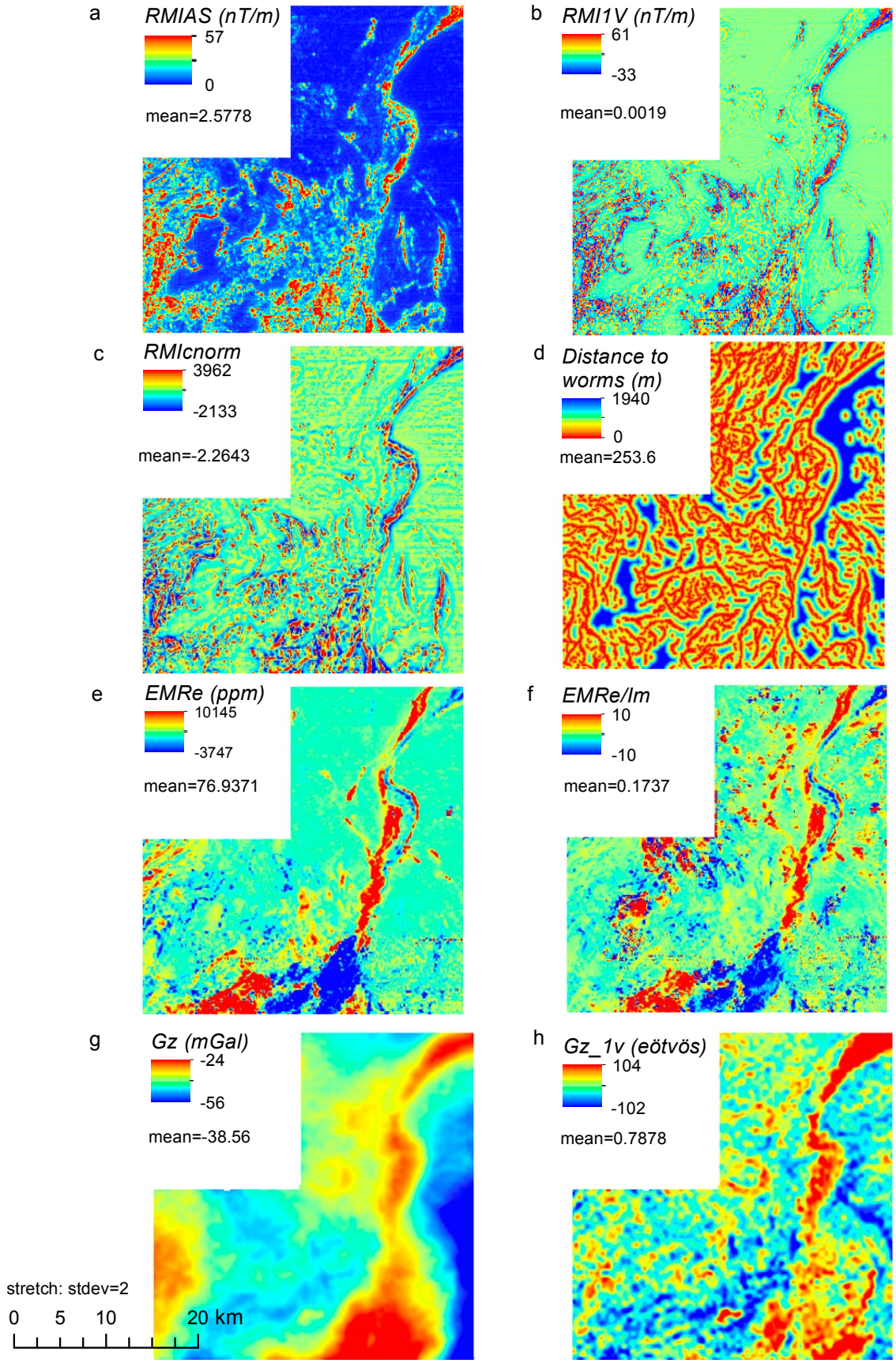

Fig. 4. Savukoski preprocessed airborne geophysical and structural data: a) RMI_{AS} = analytic signal of the RTP reduced magnetic field, b) RMI_{1V} = first vertical derivative of RMI, c) RMI_{CN} = contrast normalized RMI, d) RMI_{W_D} = distance to the pseudogravimetric gradient maxima, e) EM_{Re} = in-phase component of electromagnetics, f) EM_{Re}/Im = ratio of the in-phase and quadrature components of the electromagnetic measurements, g) G_{zRG} = airborne gravimetric data conformed to regional gravity and h) G_{z_1V} = first vertical derivative of G_{zRG}. The data range and mean are shown for each dataset. Colouring is stretched over two standard deviations.

Savukoski

Similar to IOCGs, orogenic gold deposits also are structurally controlled. In the Savukoski study area, the structures were represented as pseudo-gravimetric worms, i.e., the gradient maxima computed from the surface integrated *RMI* using the multi-scale edge detection method presented by Hornby et al. (1999). Upward continuation was carried out up to 300 m, and worms below 50 m were removed to exclude random noise. As the input to SOMKm, we used a grid representing the distance to the closest worm (RMI_{W_D}, Fig. 4d). The grid was the same 100 m grid that was used to represent the geophysical data. Worms were also used as line elements in one representation of prospectivity to show the exact location of the modelled structures (Fig. 22).

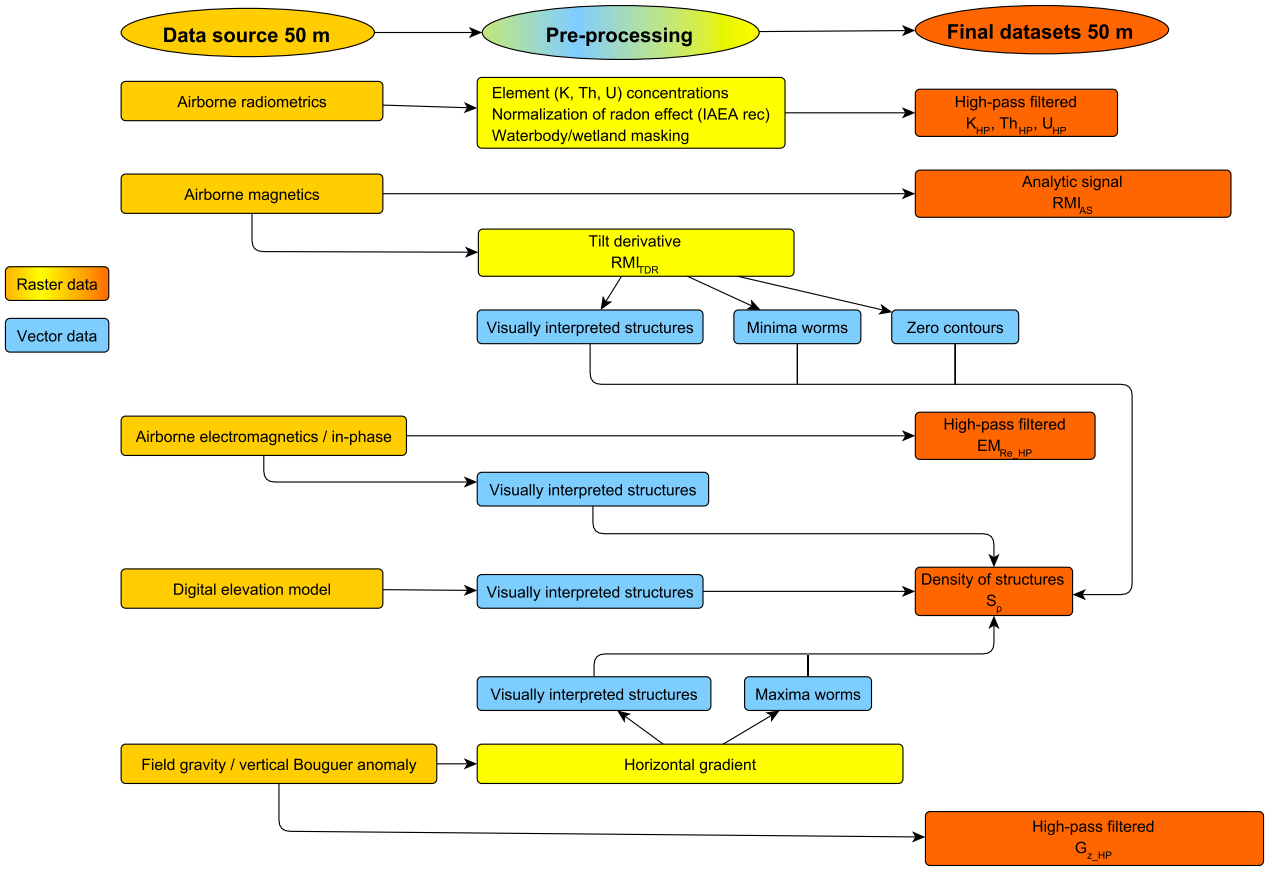

Fig. 5. Preprocessing steps for Enontekiö geophysical and structural data.

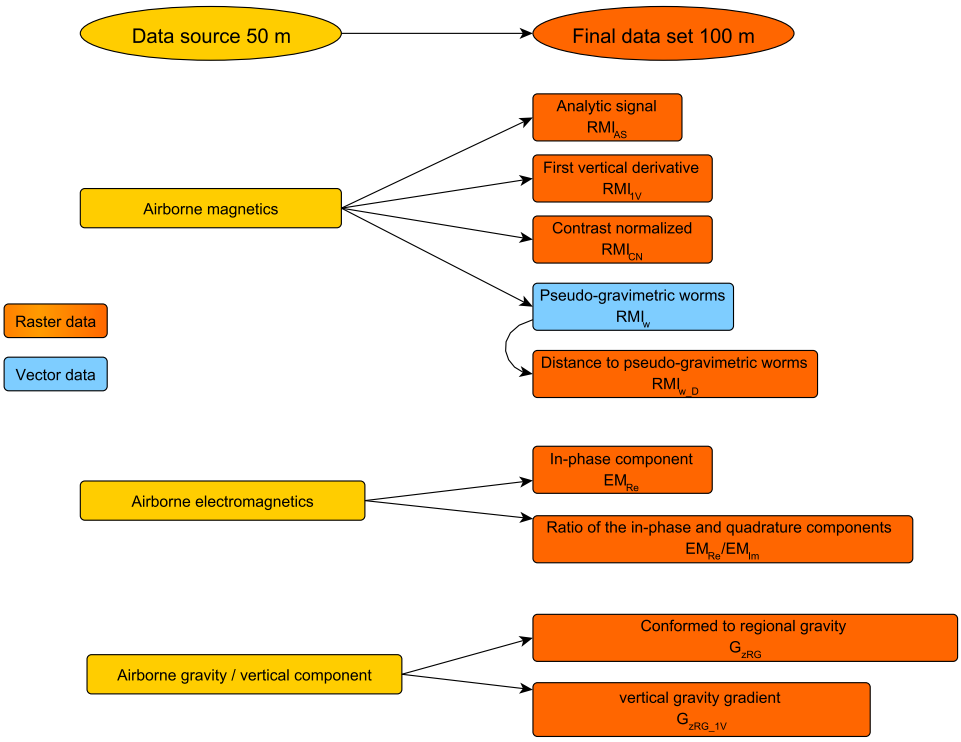

Fig. 6. Preprocessing steps for Savukoski geophysical and structural data.

Till geochemistry

For both study areas, data from the GTK's national geochemical survey of glacial tills were used. The survey data, collected in the 1970s and 1980s, mostly consisted of surficial till-sample analyses, with sampling conducted at a depth from 1–2 m. The fine fraction (<0.06 mm) of each sample was hot aqua regia leached and assayed using ICP-AES. One sample was a combination of 3–5 separate samples taken at a density of approximately 1 sample per km^2. The geochemical mapping and databases are described by Salminen and Tarvainen (1995). In the Savukoski study area, we also had new till geochemical data, described in more detail below, in the *Savukoski* section.

For the SOMKm analysis, we used both the original sampling points that were analysed independently, and interpolated grids that were combined with geophysical data (Fig. 7). Due to the undersampled nature of geochemical data and the sparseness compared to geophysical data, combining geochemical and geophysical data by simple interpolation easily causes either a loss of geophysical information (if interpolated to the resolution of the geochemical data) or the computing of geochemistry for points where it is not known to an acceptable level of accuracy (if interpolated to the resolution of geophysical data). We chose to use the same grid resolution as for geophysics, but performed interpolation of geochemistry so as to represent only large-scale variations with no single sampling points showing up as anomalous points. In this way, all the grid points had the same character, displaying the regional average of concentrations. We used the inverse distance weighting method with the 12 closest points and a distance power of 0.5. We further smoothed the interpolated solution using a mean filter with a circular radius of 10 pixels (Figs. 7, 8 and 10).

Enontekiö

In Enontekiö, the transportation distance and directions are very variable (Kujansuu 1967) and poorly studied. In addition, the overburden thickness is highly variable, ranging from 0–30 m (personal communication H. Hulkki, 2014). The shallowest glacial deposits are located on the western side, where the landscape is mountainous. The central part of the study area is topographically flat, indicating thicker glacial deposits. The shallow sampling depth of the till and the sparseness of the data (Fig. 2a) pose considerable limitations for evaluating the mineral prospectivity in Enontekiö based on the till geochemical data (only national

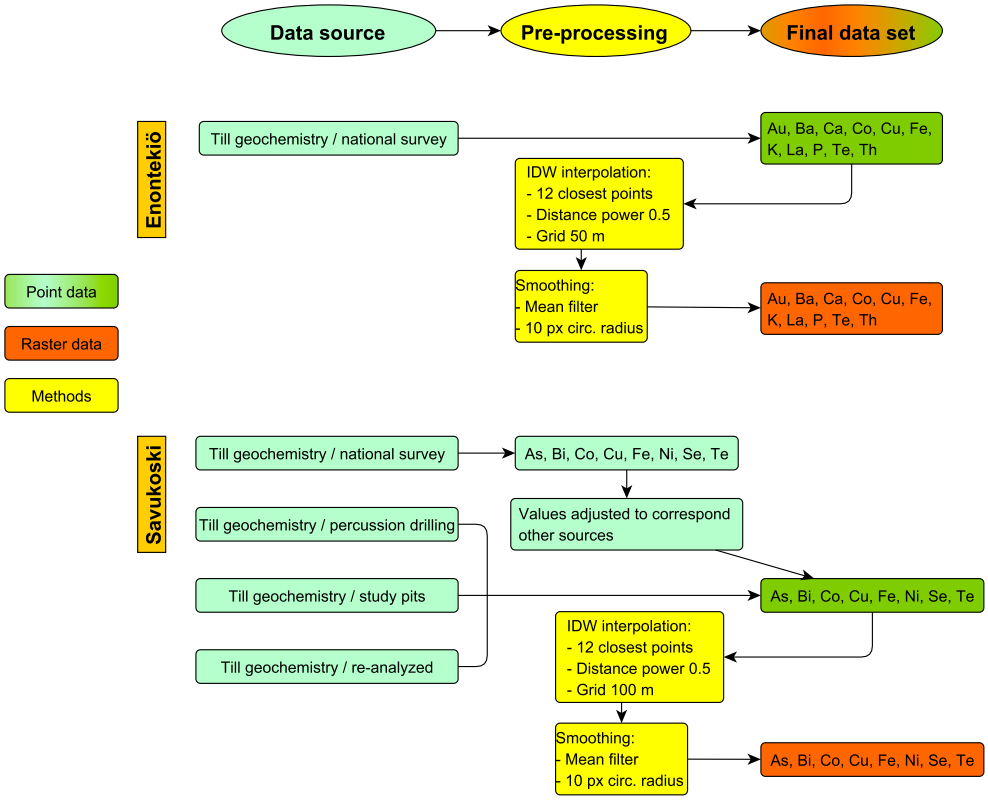

Fig. 7. Preprocessing steps of Enontekiö and Savukoski geochemical data.

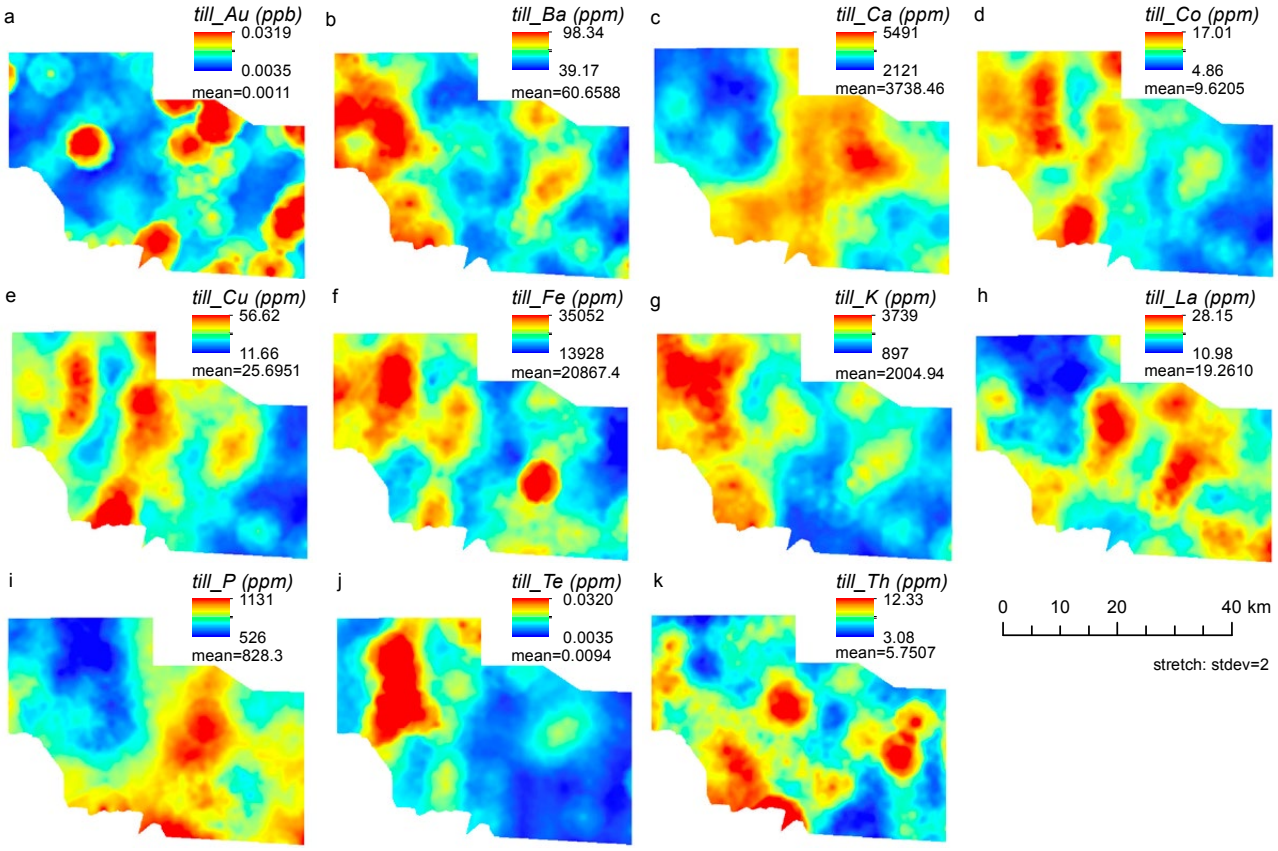

Fig. 8. Interpolated Enontekiö geochemical data. The data range and mean are shown for each dataset. Colouring is stretched over two standard deviations.

survey data were available). Visual inspection of the data indicates a spatial correlation of till geochemistry with lithological units rather than mineral occurrences (compare Fig. 1c and Fig. 8). Despite these facts, we included Au, Ba, Co, Cr, Cu, Fe, La (correlated to LREEs), Mg, P, Th and Te as indicators of IOCG mineralizations, and K of alteration (Niiranen 2005, Williams et al. 2005).

Savukoski

In addition to the national geochemical survey data, here termed *Old*, three more recently (2012–13) acquired/assayed datasets were available for the Savukoski study area. The dataset referred to here as *Perc* was collected using percussion drilling, the dataset *Pit* is from a programme involving the excavation of an extensive number of study pits and the dataset *Rean* consists of reanalysed samples from the old national survey. The density of the data points is of the same order as for the dataset *Old*. We considered As, Bi, Co, Cu, Fe, Ni, Se and Te as indicative elements for orogenic gold deposits, and obtained their concentrations in the Savukoski study area from *Old* as well as *Perc*, *Pit* and *Rean* datasets. As, Bi and Se were only available from the new survey. Based on the sampling and analysis methods, all the new datasets, *Perc*, *Pit* and *Rean*, were comparable and could be combined without further processing. To make the *Old* concentrations comparable to *Perc*, *Pit* and *Rean*, we were able to study their frequency distributions, because both datasets cover the entire study area. We assumed that the frequency distributions of the real concentrations at sampling locations coincide, and defined a linear transformation coefficient to adjust the measured concentrations by fitting the histograms of *Old* and the combination of *Perc*, *Pit* and *Rean* by randomly sampling the transformation coefficient (Fig. 9). In Table 2, it can be seen that the relative difference between the median concentrations (difference between the medians divided by the mean of the medians) of the *Old* dataset and the combination of *Perc*, *Pit* and *Rean* datasets varies from 1.3 % (Ni) to 9.2 % (Te). As an example, we also show plots of the difference between the distributions of *Old* and the combination of *Perc*, *Pit* and *Rean* before and after the transformation of *Old* and the difference between the distributions as a function of the transformation coefficient in Figure 9.

Table 2. Δm = percentual difference between the medians of the *Old* and the combined *Perc*, *Pit* and *Rean* datasets of Savukoski, *K* = linear transformation coefficient for relating *Old* concentrations to those of *Perc*, *Pit* and *Rean*.

	Co	Cu	Fe	Ni	Te
Δm	7.1%	8.4%	7.2%	1.3%	9.2%
K	0.90	0.86	1.16	0.98	0.79

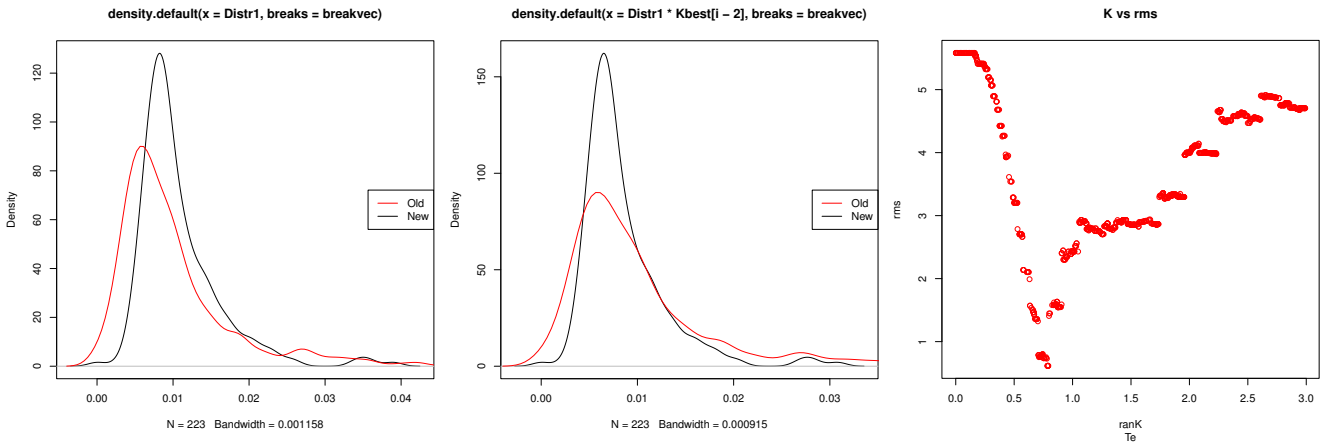

Fig. 9. Left: frequency distributions of the original *Old* and the combination of *Perc, Pit* and *Rean* datasets for Cu in Savukoski. Middle: frequency distributions of the transformed *Old* and the combination of *Perc, Pit* and *Rean* concentrations of Cu. Right: Plot of the difference between the distributions (rms) vs the coefficient *K* for Cu datasets.

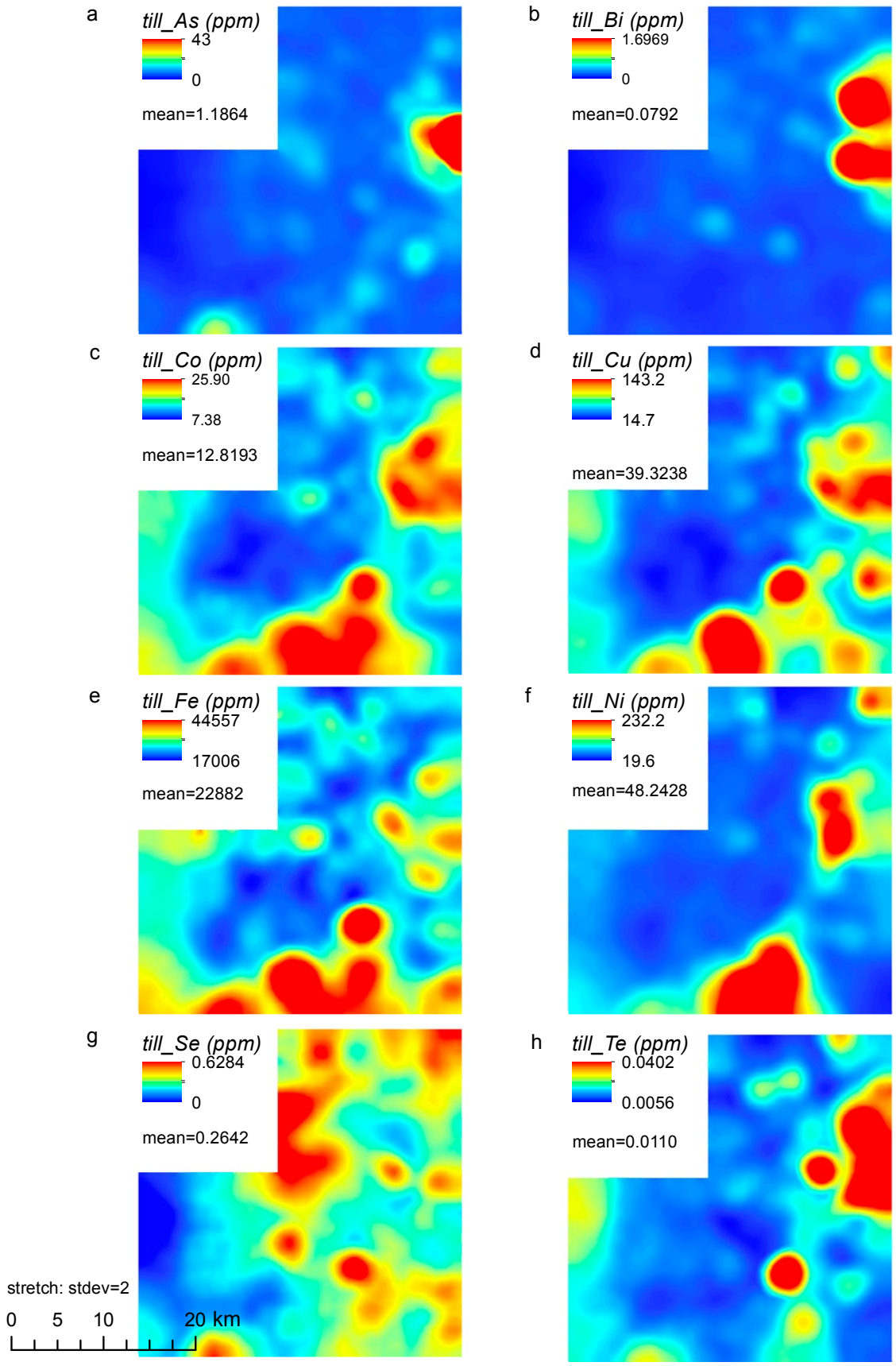

Fig. 10. Interpolated Savukoski geochemical data. The data range and mean are shown for each dataset. Colouring is stretched over two standard deviations.

Table 3. Datasets used for IOCG and orogenic gold prospectivity modelling in the Enontekiö and Savukoski study areas. *Dataset name* is the dataset name referred to in this document and *Quantity* is the description of the dataset. The datasets used in Enontekiö (E) and Savukoski (S) study areas are marked with 'x' in the respective columns.

Dataset name	Quantity	E	S
GEOPHYSICS			
Magnetic			
RMI_{CN}	Contrast normalized *RMI* *		x
RMI_{1V}	Vertical derivative of *RMI* *		x
RMI_{AS}	Analytical signal of *RMI* *	x	x
Electromagnetic			
EM_{Re}	In-phase component of *EM* **		x
EM_{Re_HP}	High-pass-filtered EM_{Re}	x	
$EM_{Re/Im}$	Ratio of the in-phase and quadrature components of *EM* **		x
Gravity			
G_{z_HP}	High-pass-filtered Bouguer anomaly	x	
G_{zRG}	Vertical component of the airborne gravity field, conformed to regional gravity		x
G_{zRG_1v}	Vertical derivative of G_{zRG}		x
Radiometric			
K_{HP}, U_{HP}, Th_{HP}	High-pass filtered potassium, uranium and thorium concentrations from airborne radiometric measurements	x	
STRUCTURES			
RMI_{W_D}	Distance to the locations of the wavelet edges of the pseudo-gravimetric field computed for various levels of upward continuation from *RMI* *		x
S_p	Density of all the interpreted structures in Enontekiö	x	
GEOCHEMISTRY			
Au, Ba, Co, Cr, Cu, Fe, La, Mg, P, Th, Te	Element concentrations in the till fine fraction (point data and interpolated grid)	x	
As, Bi, Co, Cu, Fe, Ni, Se, Te	Element concentrations in the till fine fraction (point data and interpolated grid)		x

*) *RMI* = magnetic field reduced to pole

**) *EM* = measured electromagnetic response

METHODS

Self-organizing map (SOM)

Method. The computation of an SOM (Kohonen 2001) is an unsupervised data-mining method that clusters similar data vectors and represents them as single vectors termed *BMUs* (best-matching units). The product of the SOM computation is a SOM map, or only *SOM*, which is an *n*-dimensional array, the elements of which are the BMU vectors. The beauty of SOM is in its ability to project the original data vectors to a subspace where similar data vectors are spatially close to each other, and still explicitly retain the original data vector components. SOM can be computed in any number of dimensions, but 2D is most commonly used due to its efficiency in visualization. If a toroid surface is used in the computation, the final 2D-surface map is continuous from one edge to the opposing one. SOM computation is an iterative process, where the dataset is scanned through a large number of times. The initial values for the SOM BMU vectors are determined using the data statistics. Briefly, SOM computation proceeds as follows:

1. The starting *N*-dimensional BMU vectors ($\mathbf{BMU}_k$, k = 1,K) are initialized for each SOM map element. *N* is the dimension of the data vectors, i.e., the number of input datasets, and K is the number of elements in the SOM map.
2. The data vectors $\boldsymbol{D}_m$ (m = 1,M) are scanned through, where *M* is the number of data points.

a. Each data vector is assigned the index *k* of the closest $\boldsymbol{BMU}_k$, i.e., the index *k* for which

$$d_{k,m} = \sum_{n=1}^{N} \left(BMU_{k,n} - D_{m,n} \right)^2$$

is the minimum.

b. $\boldsymbol{BMU}_k$ is changed to be more similar to $\boldsymbol{D}_m$.
c. The BMU values of the SOM elements in the neighbourhood of $\boldsymbol{BMU}_k$ are changed to be more similar to $\boldsymbol{D}_m$.
3. Steps 2–4 are repeated a certain number of times or until the convergence criterion is fulfilled.

Parameters and interpretation. We used the CSIRO's SOM implementation, *SiroSOM*, to generate a 2D SOM. Parameters that are inputted to the program are: the shape (hexagonal/rectangular, sheet/toroid/cylinder) and size of the SOM grid, i.e. the number of best matching units, the type of initiation (linear/random), the initial and final search radii and the number of iterations of the fine and rough tuning phases, and the neighbourhood function (Gaussian / cut Gaussian / Epanechikov function / rectangle). We used random initialization for the map, i.e., initial BMU vectors were randomly sampled within the range of the data values, a hexagonal SOM grid, where each element has six neighbouring elements, and toroid geometry, which means that when opened up, the map is continuous from one edge to the other.

In SiroSOM, SOM can be used for pre-clustering using a large SOM map and K-means for further clustering the SOM clusters into a smaller number of clusters. A descriptive parameter obtained from the SOM pre-clustering is the quantization error (QE), which is computed for each data vector as its deviation from the corresponding BMU vector. Large QE values thus indicate anomalous data vectors (Fig. 11). K-means clusters are often used to indicate larger scale structures in the data. The loop over data points as well as over the SOM map in SiroSOM is systematic. One drawback in SiroSOM is that it scales all the data to the range [0,1], so there is no possibility to set weights for the importance of different datasets, which would sometimes be useful.

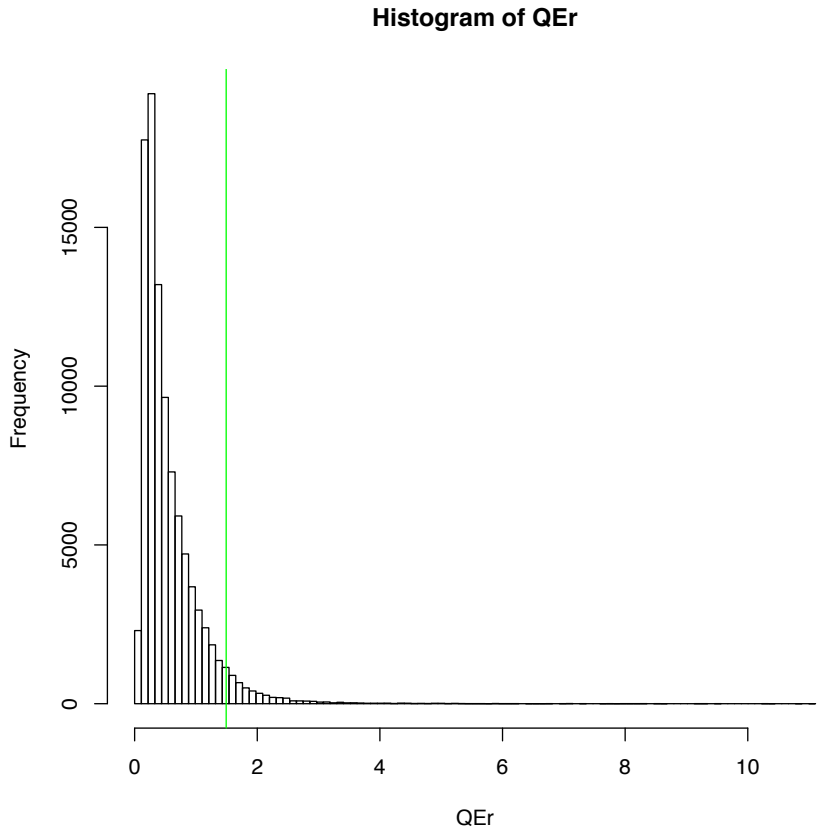

Fig. 11. A typical frequency distribution histogram of SOM QE. The 95% quantile, used in this study as the limit of anomalous values, is in this example at the QE value 1.49.

K-means

Method. K-means is a partitive clustering method, many closely related variations of which have been developed. We followed the procedure described below:

1. The K N-dimensional vectors $\boldsymbol{C}_k$ representing the cluster centres were initialized. N is the dimension of the data vectors, i.e., the number of different datasets.
2. The data vectors $\boldsymbol{D}_m$ (m=1,M) were scanned through, where M is the number of data points. Each data vector was assigned the index k of the closest $\boldsymbol{C}_k$, i.e., the index k for which

$$d_{k,m} = \sum_{n=1}^{N} \left(C_{k,n} - D_{m,n}\right)^2$$

   is at the minimum.
3. New vectors $\boldsymbol{C}_k$ were calculated as the means of the data vectors assigned to each cluster centre.
4. Steps 1–3 were repeated a certain number of times or until the convergence criterion was fulfilled.

Parameters. K-means is implemented into SiroSOM, and the user cannot affect the input parameters but only provides the maximum number of clusters (K_{max}) computed. SiroSOM performs clustering from 1 to K_{max} clusters and computes the Davies-Bouldin index (DB, Davies & Bouldin 1979) for each set of clusters. The user then determines which cluster number is used. Usually, the one with the lowest DB is chosen. To investigate the stability of the K-means procedure, clustering has to be run several times.

Fuzzy logic

Fuzzy logic is a knowledge-driven data-mining technique. Unlike SOM, expert knowledge is used in fuzzy logic methods when determining the fuzzy membership function for each component of the multidimensional dataset. The membership function transforms the original data values to values that should reflect the probability of the hypothesis set down (e.g. "mineral deposit exists") being true at the data point. Fuzzy operators are then used to combine the information in the membership grids. The result is strongly affected by the form of the membership function, and hence there should be good scientific grounds for selecting a particular function and parameters.

The fuzzy operators that were used here were the OR, AND and PRODUCT operators. OR takes the highest membership value of the input grids and thus outputs a grid where high membership values are present if even one of the inputs values is high. The AND operator takes the lowest of the membership values at each point, requiring that all the input values must be high to produce a high fuzzy value. PRODUCT simply multiplies the corresponding input membership values.

Parameters. We used the ArcGIS Fuzzy Logic tool contained in the Spatial Analyst Tools. Necessary fuzzy membership functions are *Large* and *Small*, which give large membership values to large and small grid values, respectively. The functions are of the form

$$\text{Large: } \mu(x) = \frac{1}{1 + \frac{x^{-s}}{k}}$$

$$\text{Small: } \mu(x) = \frac{1}{1 + \frac{x^{s}}{k}}$$

where k is the midpoint, defined as the grid value corresponding to the membership value of 0.5, and s is the spread, defining the steepness of the membership function near k (the larger the spread, the steeper the function).

RESULTS

After intensive testing, we chose to present three different approaches for both study areas. Each model provides a different type of representation of the data and a different type of approach for interpreting the mineral prospectivity of the study area. The main difference in the approaches is the amount of expert knowledge used on the one hand in prospectivity modelling and on the other hand in the interpretation of the resulting prospectivity map. SOMKm and fuzzy logic were used to compute the final models, and the parameters used in the computations for both study areas are summarized in Tables 4–5 and Figure 12.

Table 4. SOM parameters for the SOMKm computations for both study areas. In all cases, SOM initialization was random, the shape was toroid, the grid was hexagonal and the neighbourhood function was Gaussian. Subscripts *phy* refer to geophysical and *che* to geochemical datasets.
SOM size = the size of the SOM matrix
Rough i = initial and *f* = final radius, and *l* = number of iterations for the rough tuning
Fine i = initial and *f* = final radius, and *l* = number of iterations for the fine tuning
QE = quantization error and *TE* = topographical error of the entire SOM
N K-m = number of clusters and *D-B* = Davis-Bouldin index for the K-means clustering

Model/ dataset	SOM size	Rough i/f/l	Fine i/f/l	QE	TE	N K-m	D-B
E1a	56 x 42	70/18/20	18/1/400	1.50	0.835	24	1.3
E1b	56 x 42	70/18/20	18/1/400	0.559	0.128	22	1.2
$E3_{phy}$	56 x 42	70/18/20	18/1/400	0.351	0.018	-	-
$E3_{che}$	16 x 10	19/5/20	5/1/400	1.35	0.122	10	1.0
S1	74 x 67	100/25/20	25/1/400	0.882	0.0664	23	1.1
$S2_{phy}$	74 x 67	100/25/20	25/1/400	0.341	0.0743	-	-
$S2a_{che}$	22 x 14	27/7/20	7/1/400	0.804	0.126	14	0.92

Table 5. Fuzzy membership parameters for the fuzzy operations for both study areas.

Dataset	Midpoint	Spread	Membership F
RMI_{AS}	7.425 µT/m	3	Large
G_{z_HP}	0.03063	3	Large
QE_{E1a}	2.593	3	Large
QE_{E3}	1.039	3	Large
S_{ρ}	0.009723	5	Large
QE_{S2}	1.49	8	Large
RMI_{W_D}	100 m	8	Small
DE_{S2}	1 km	5	Small

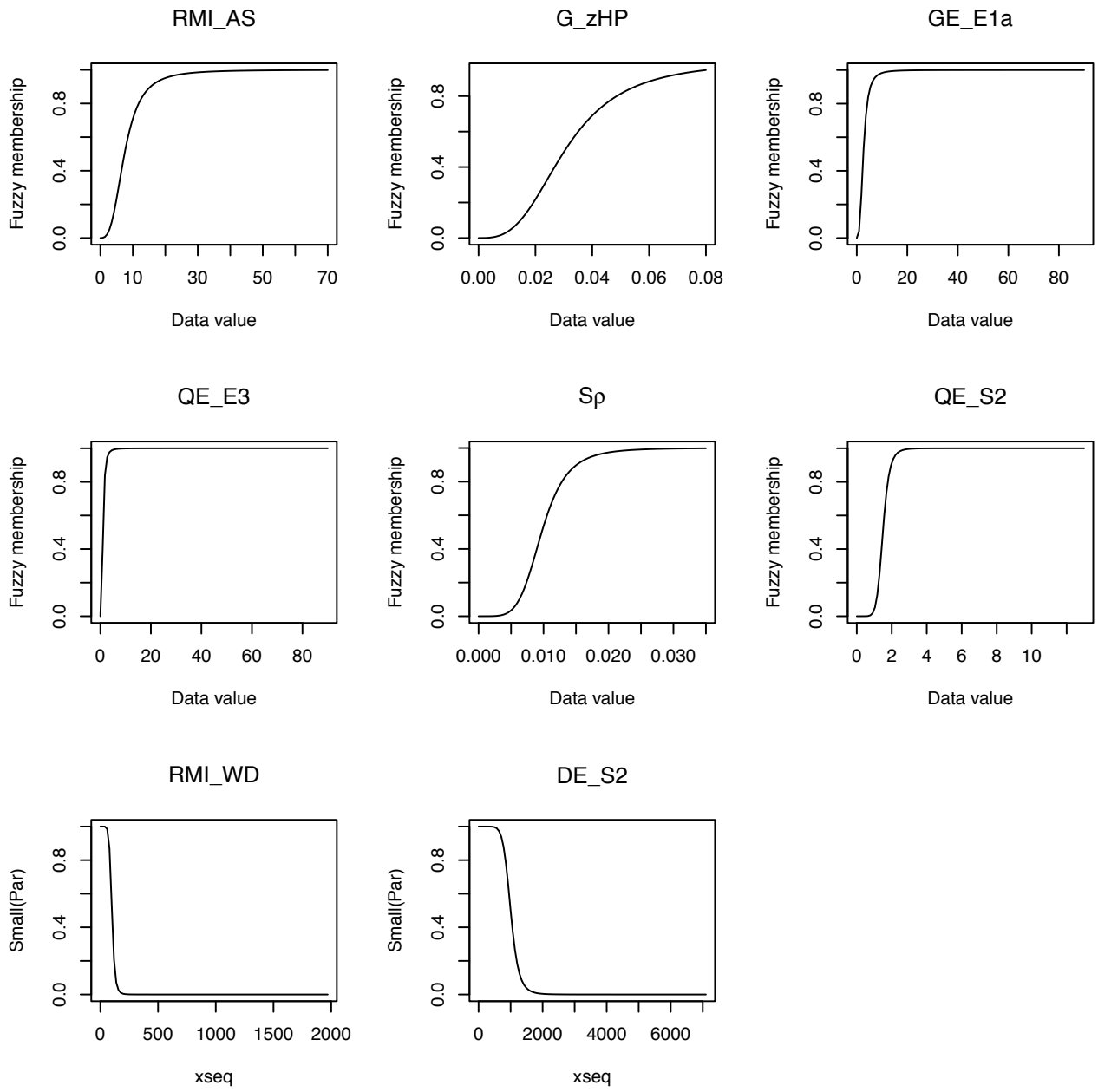

Fig. 12. Plots of the membership functions for the datasets used in fuzzy logic analysis (Table 5).

Approach 1, model E1 (a and b)

Two alternative datasets were input to SOMKm, and models E1a and E1b represent the spatial distribution of SOM QE and K-means clusters (Fig. 13). For model E1a, all the Enontekiö data were used with the logarithm taken from RMI_{AS}, EM_{Re_HP}, G_{z_HP}, Au, Cu and Fe. For model E1b, only geophysical and structural data were used with the logarithm taken from RMI_{AS}, EM_{Re_HP} and G_{z_HP}.

The mean QE of E1a (Fig. 14a) is higher than that of E1b (Fig. 14c), but when stretched differently, the images look similar. The difference in means is due to the geochemical data, but anomalously large values are still mainly caused by the geophysical data, i.e., the range of QE values is the same in both models. Both QEs give a general overview of the anomalous features in the data. However, they cannot be directly connected to the existence of IOCG in particular, but have to be further processed, because not all the anomalies are located in regions typical for the existence of IOCG deposits.

In model E1a, geochemistry has a dominating effect on the SOMKm classes by splitting the study area into regional clusters (Fig. 14b). However,

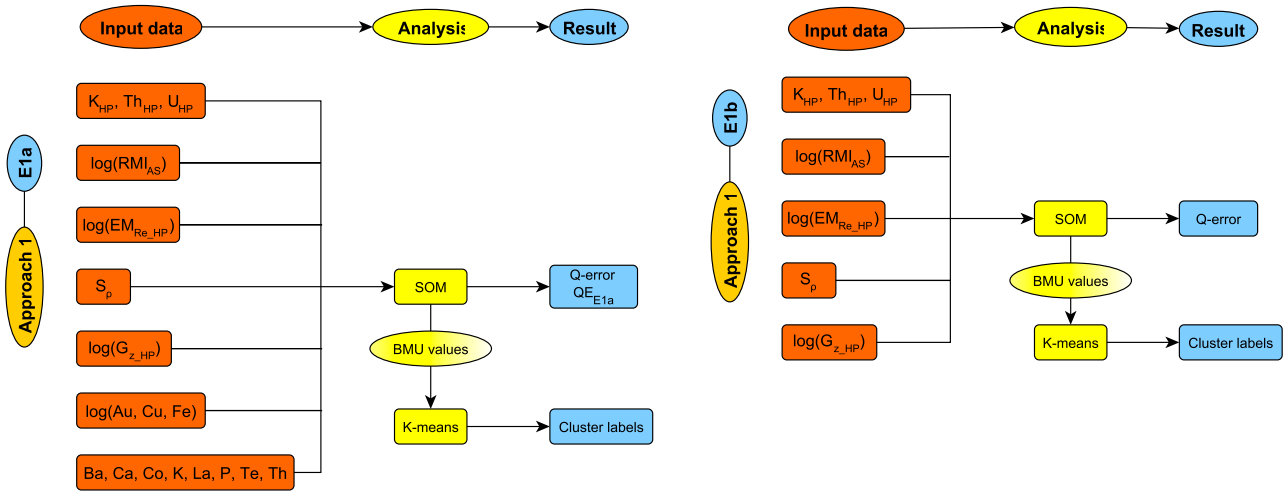

Fig. 13. Flowchart for Approach 1 in Enontekiö data analysis.

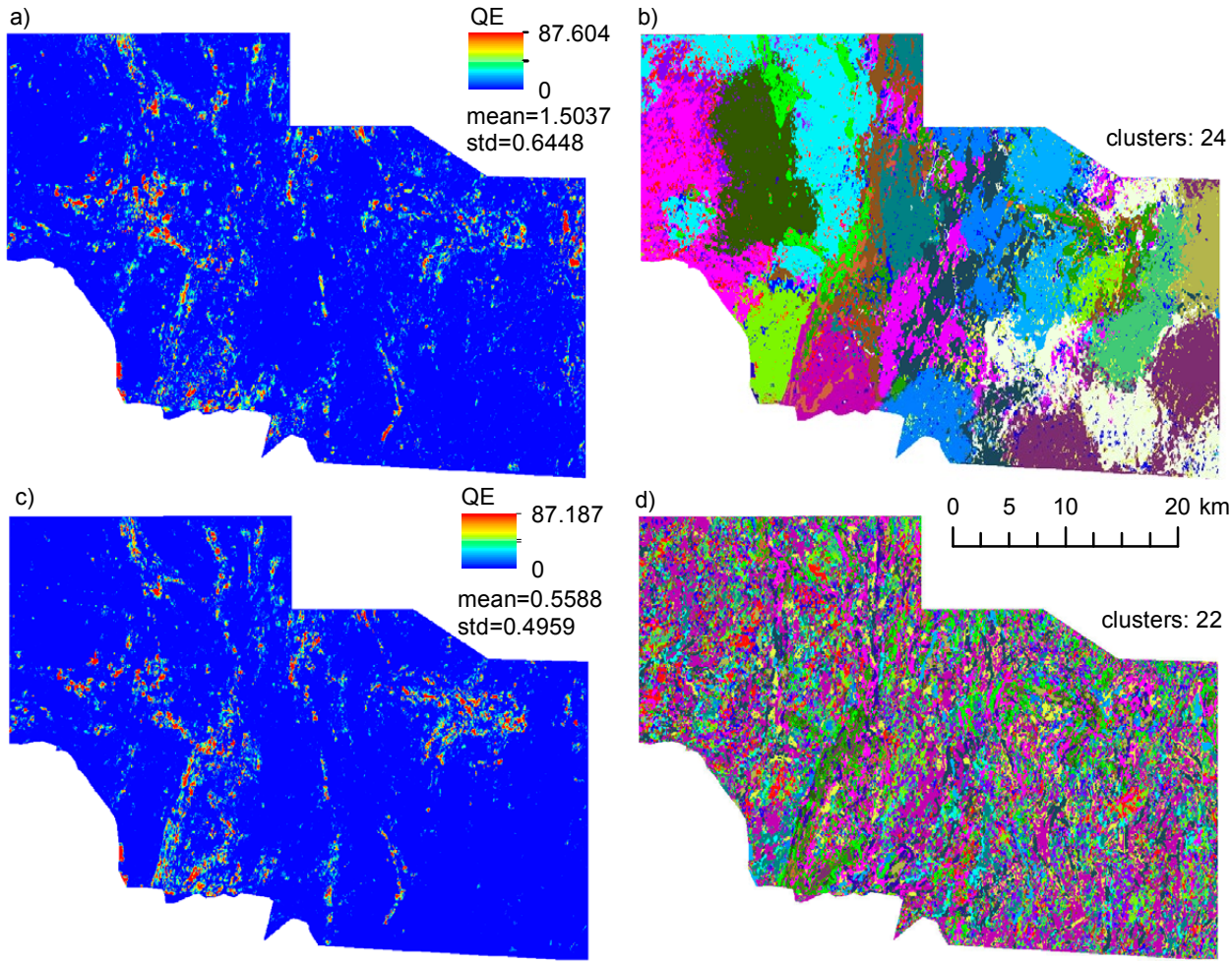

Fig. 14. Results of Approach 1, E1a and E1b, for the Enontekiö dataset. (a) The quantization error of model E1a, stretched with 90% clip from the small end and 1% clip from the high end of values. (b) The K-means clusters for model E1a. (c) The quantization error of model E1b, stretched with 0% clip from the small end and 20% clip from the high end of values. (d) K-means clusters for model E1b.

some of the clusters having a high magnetic response, density and electromagnetic response can still be seen in both models E1a and E1b (Fig. 14d). If K-means clusters are used for mineral exploration, model E1b is more suitable than E1a, because it reveals more small-scale features.

Approach 2, model E2

Fuzzy overlay analysis using the PRODUCT operator was used to combine the QE of model E1a (QE_{E1a}) with RMI_{AS} and G_{z_HP} to give weight to those high QE values that also have high magnetic susceptibility and are dense according to gravity measurements (Figs. 15 and 16). The fuzzy membership function *Large* was used for all the data with midpoints and spreads listed in Table 5. The midpoints for RMI_{AS} and G_{z_HP} are defined as the mean of the datasets, and for QE_{E1a} as the 95% quantile of QE values. The area with anomalously high QE is significantly restricted compared to E1, and can be assumed to better indicate the potential IOCG-derived anomalies due to the exclusion of low gravity and magnetic data values. Missing G_{z_HP} membership values were replaced with unity in order to not affect the fuzzy PRODUCT computation at locations with no gravity data.

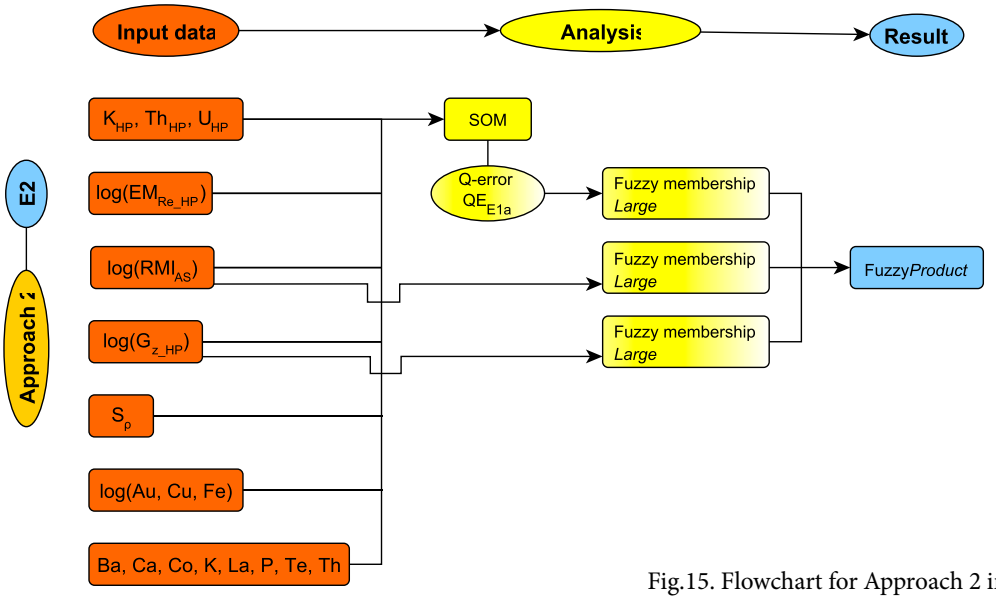

Fig.15. Flowchart for Approach 2 in Enontekiö data analysis.

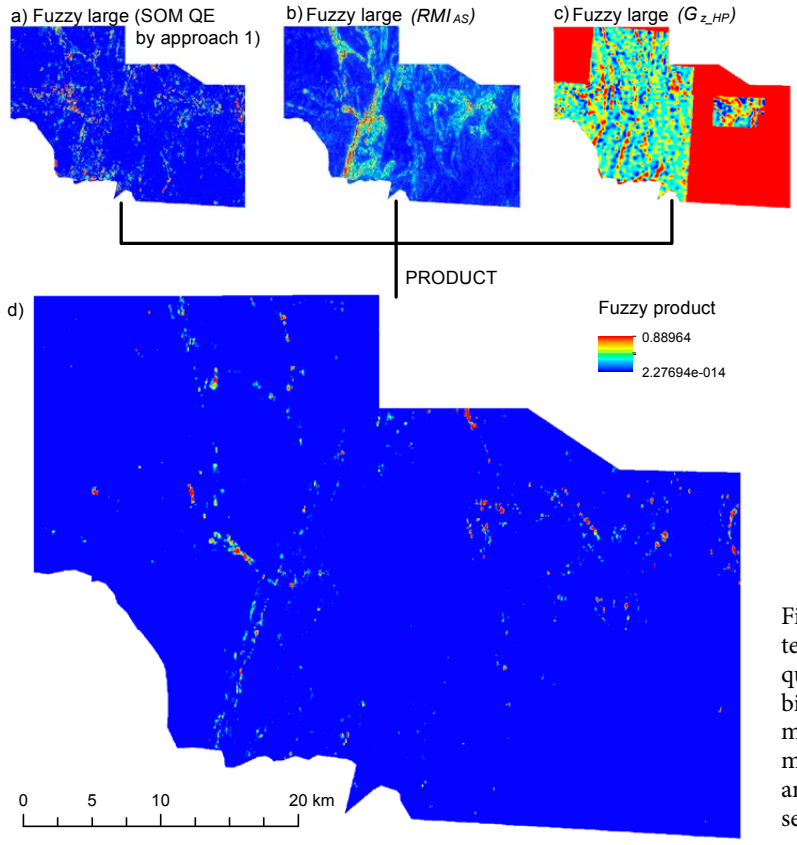

Fig. 16. Results of Approach 2, E2, for the Enontekiö data set. Membership grids of the SOM quantization error of model E1a (a) were combined using the fuzzy PRODUCT operator with membership grids of the analytical signal of magnetics (b) and high-pass-filtered Bouguer anomaly data (c). The final fuzzy model is presented in image (d).

Approach 3, model E3

For model E3 (Fig. 18f), SOMKm was computed separately for the Enontekiö geophysics and for geochemical sampling points (Fig. 17). The logarithm was taken from RMI_{AS}, EM_{Re_HP} and G_{z_HP}. K-means clusters were not computed for geophysics, but the geophysics QE (QE_{E3}) was combined with RMI_{AS} and G_{z_HP} using the fuzzy AND operator. The fuzzy membership function *Large* was used for all the data with midpoints and spreads listed in Table 5. The midpoints for RMI_{AS} and G_{z_HP} were defined as the means of the datasets, and for QE_{E3} as the 95% quantile of QE values. The fuzzy AND of QE_{E3}, RMI_{AS} and G_{z_HP} was further combined with the density of structures (RMI_{W_ρ}) using the fuzzy PRODUCT operator. The fuzzy membership midpoint of RMI_{W_ρ} was set to the mean of RMI_{W_ρ}. Missing G_{z_HP} membership values were replaced with unity in order to not affect the fuzzy AND computation at locations with no gravity data.

K-means clustering was carried out for the geochemical data, and the clusters representing elevated concentrations were overlain on the final fuzzy PRODUCT of geophysics (Fig. 18f). The Student's t-test was used to define the clusters with an elevated concentration as those clusters whose members have 90% probability of showing higher values than 90% of the whole population.

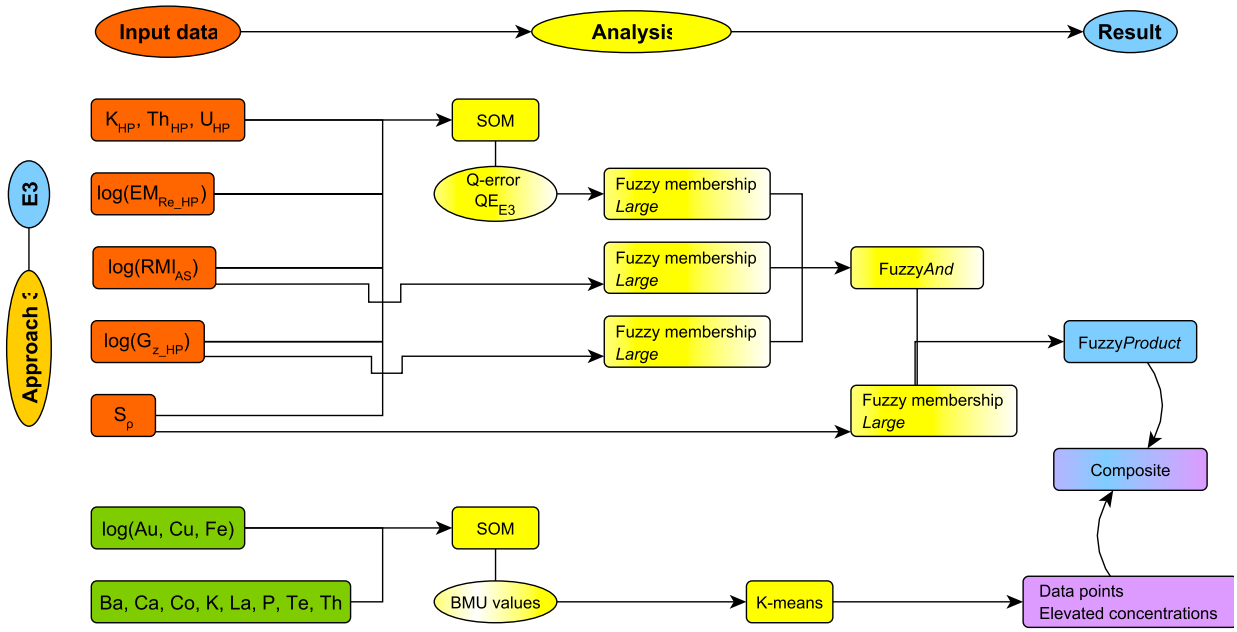

Fig. 17. Flowchart for Approach 3 in Enontekiö data analysis.

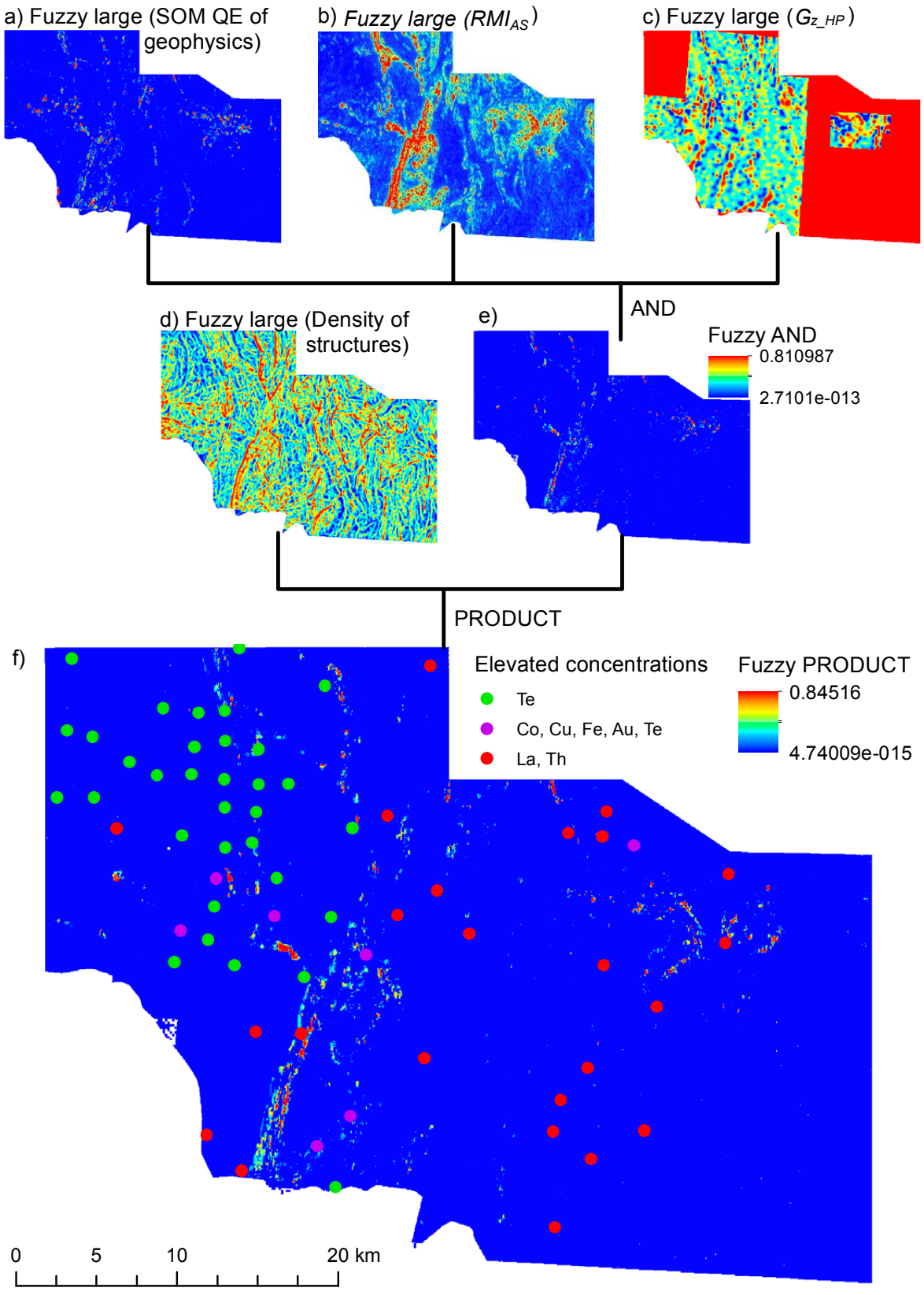

Fig. 18. Results of Approach 3, E3, for the Enontekiö dataset. SOM QE (a) was only based on geophysics, and the corresponding membership grid was combined with the fuzzy AND operator with membership grids of the analytical signal of magnetics (b) and high-pass-filtered Bouguer anomaly data (c). This intermediate result (e) was further combined with the fuzzy PRODUCT operator with the membership grid of the density of the structures (d). The final fuzzy product (f) is stretched with a percentage clip with min = 90% and max = 1%.

Savukoski

Approach 1, model S1

Model S1 was computed by SOMKm clustering all the available interpolated geophysical, geochemical and structural data (Fig. 19). Because frequency distributions of many data components had long tails, Winsorizing (limiting extreme values) was carried out for EM_{Re}, $EM_{Re/Im}$, RMI_{CN}, RMI_{AS}, RMI_{VD}, *As*, *Bi*, *Ni* and *Te* to bring up variation near the mean of the data. QE of S1 (Fig. 20a) is a general overview of the anomalous features in the data, but cannot directly, without further processing, be connected to the existence of orogenic gold in particular. The 23 K-means clusters (Fig. 20b) can be examined to show locations where data values are typical for orogenic gold deposits. Large geochemical features tend to dominate the K-means cluster distribution, but small-scale anomalies can also be seen. K-means clusters that represent anomalously large or small values (Fig. 20c, d) are spatially correlated with elevated QE values, and provide information on the source of the anomalies.

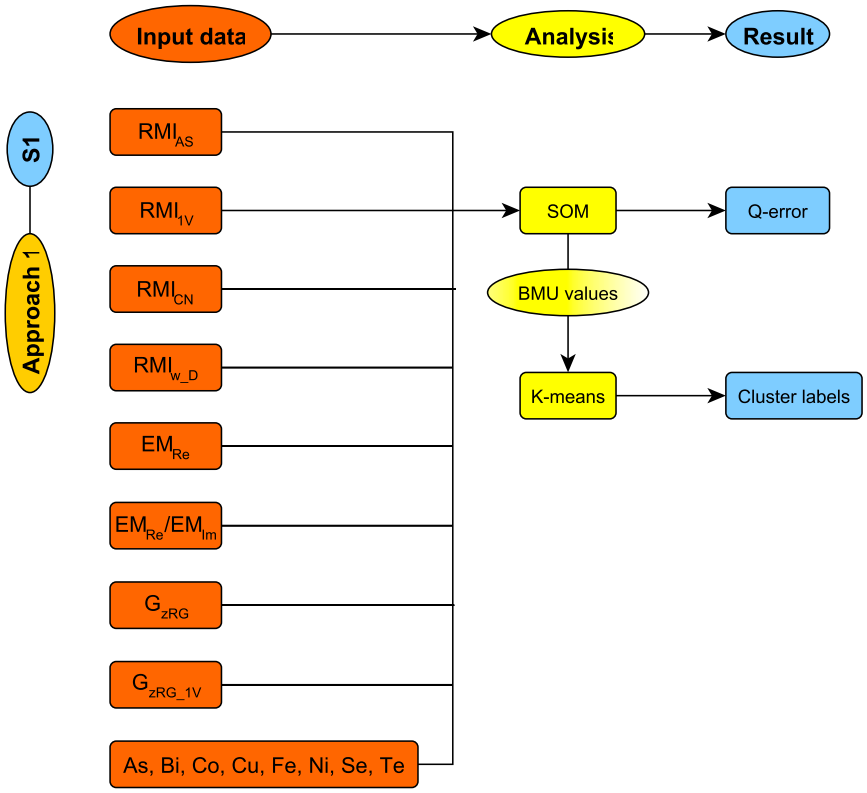

Fig. 19. Flowchart for Approach 1 of Savukoski data analysis.

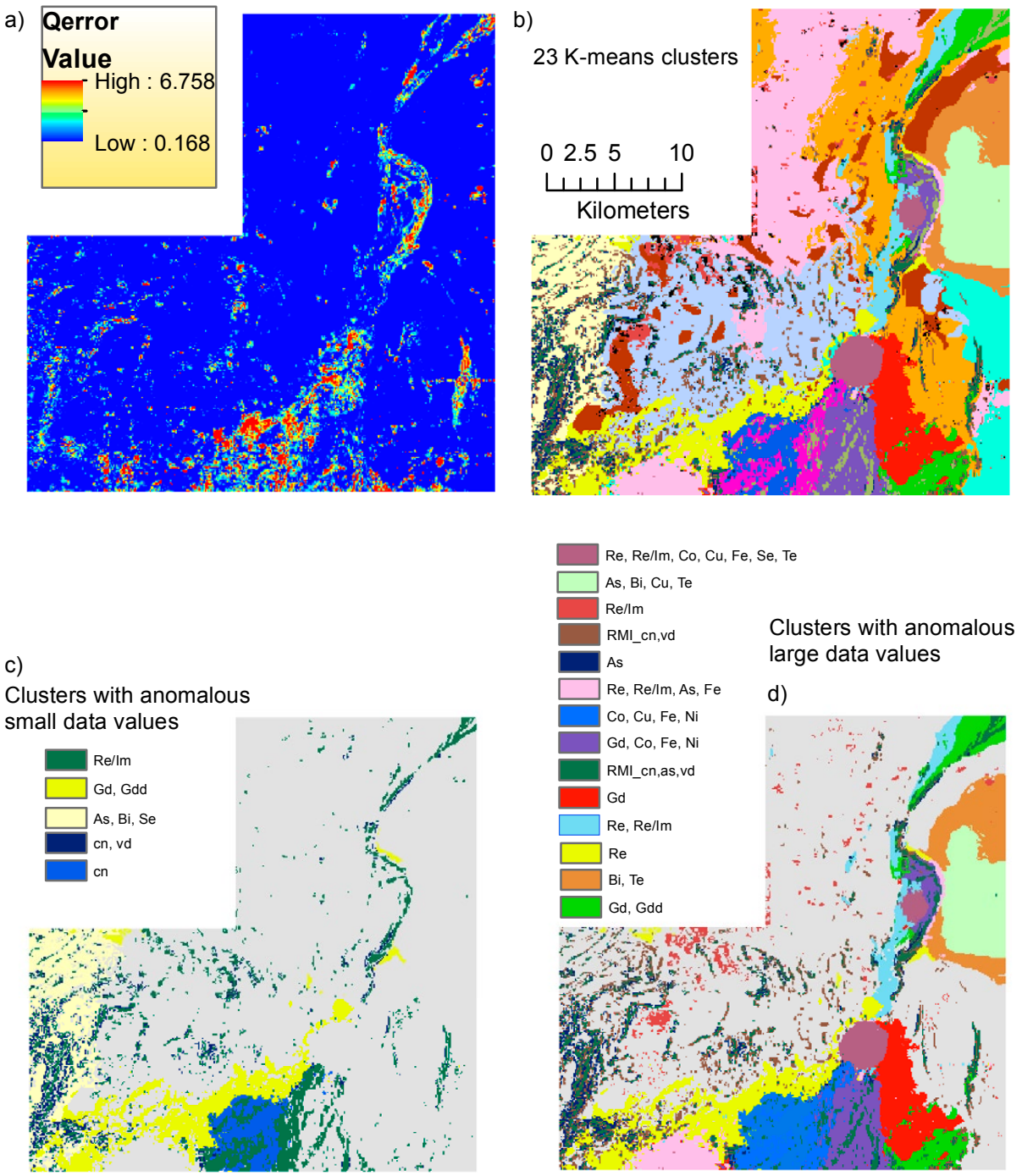

Fig. 20. Results of Approach 1, S1, for the Savukoski dataset. a) SOM QE of S1, stretched with 90% clip from the small end and 1% clip from the high end of values, b) 23 K-means clusters, c) K-means clusters representing reduced data values according to the Student's t-test and d) K-means clusters representing elevated data values according to the Student's t-test.

Approach 2, model S2

For computing model S2, SOMKm clustering of geophysics grid and geochemistry sampling points was carried out separately (Fig. 21). The Student's t-test was applied to identify the geochemistry K-means clusters that represent elevated concentrations. The criterion for being elevated was that cluster members must have at least a 90% probability of showing higher values than 90% of the whole population. The prospectivity information is represented as a composite 3-component map consisting of 1) the QE raster of geophysics, 2) geochemical sampling points that represent elevated concentrations and 3) a line-element layer showing the structures (Fig. 22).

This approach best retains the information content of geophysics and gives a possibility for expert interpretation of the resulting map, provided that the character and limitations of the geophysics QE as well as the structural and geochemical data are well understood.

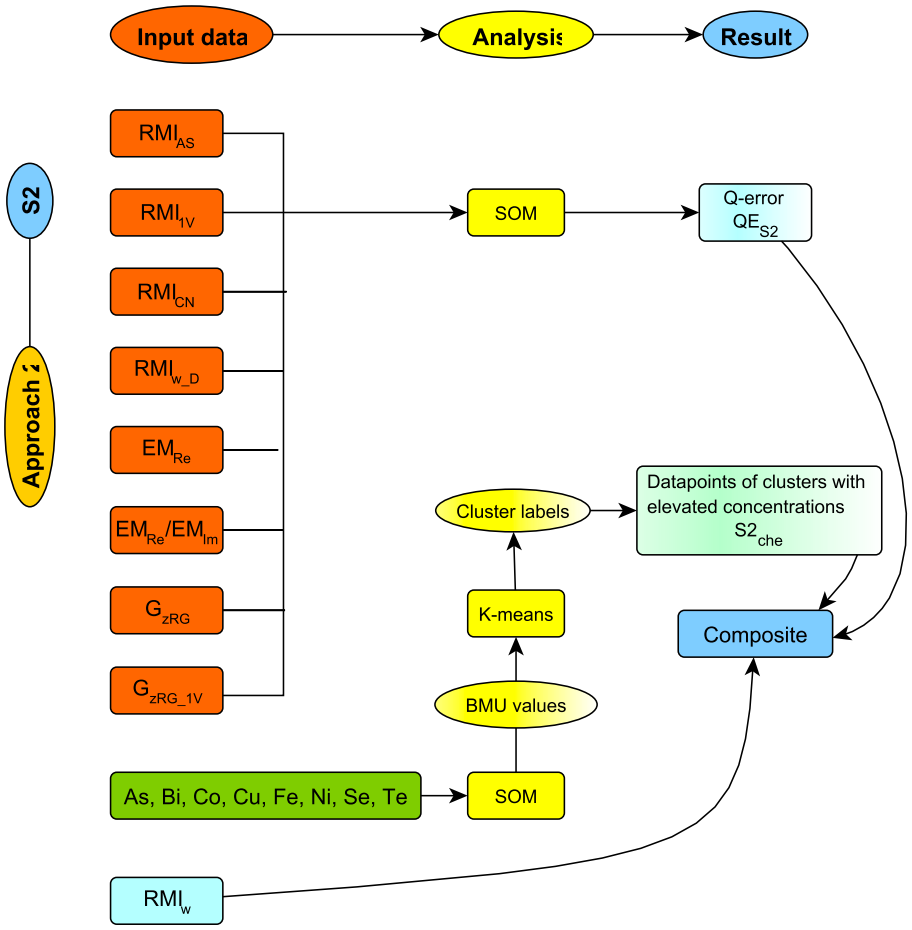

Fig. 21. Flowchart for Approach 2 of Savukoski data analysis.

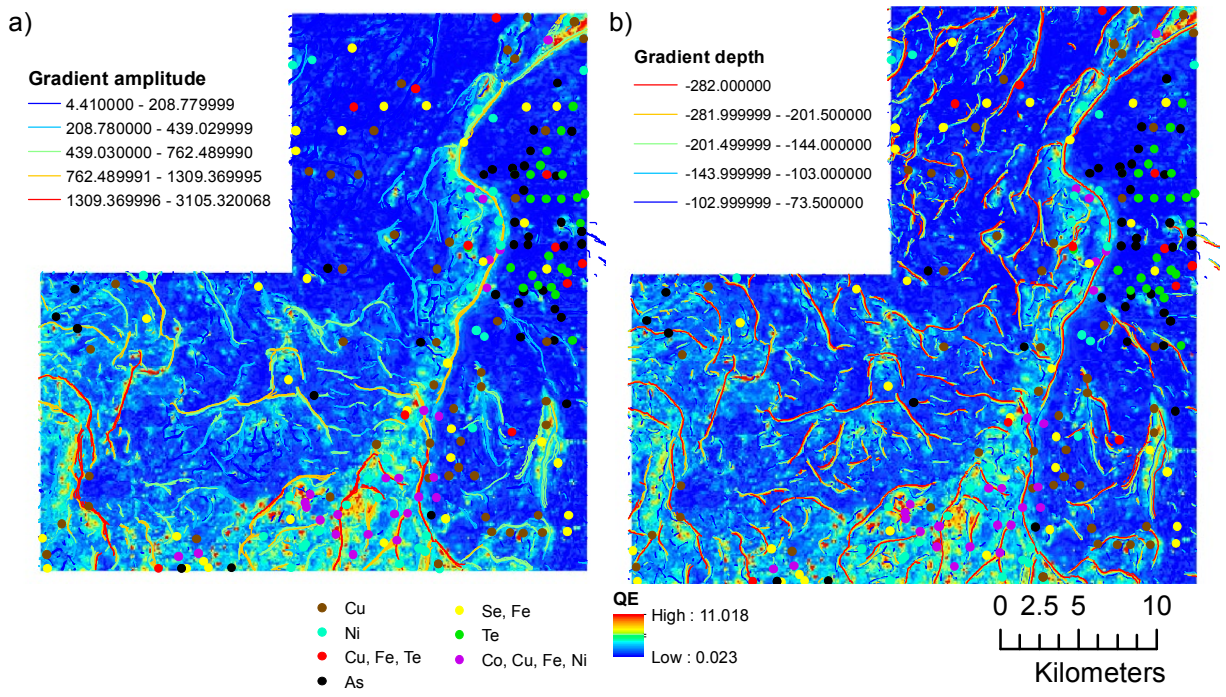

Fig. 22. Results of Approach 2, S2, for the Savukoski dataset. Composite image of the quantization error raster from geophysics SOM, sampling points representing elevated concentrations from geochemistry, and polylines representing pseudogravimetric worms coloured by the steepness of the structure (a) and the depth of the structure (b).

Approach 3, model S3 (a and b)

For computing models S3a and S3b, SOMKm clustering was independently performed on the geophysics grids and geochemical sampling points similarly to Approach 2 (Fig. 23). Also in accordance with Approach 2, QE of the geophysics SOM was computed and clusters representing elevated geochemical concentrations were identified and included in the further analysis.

For model S3a, the grid of the distance to the included geochemical points was combined with the distance-to-structures grid using the fuzzy OR operator in order to not give false negative signals from the geochemical membership, which is

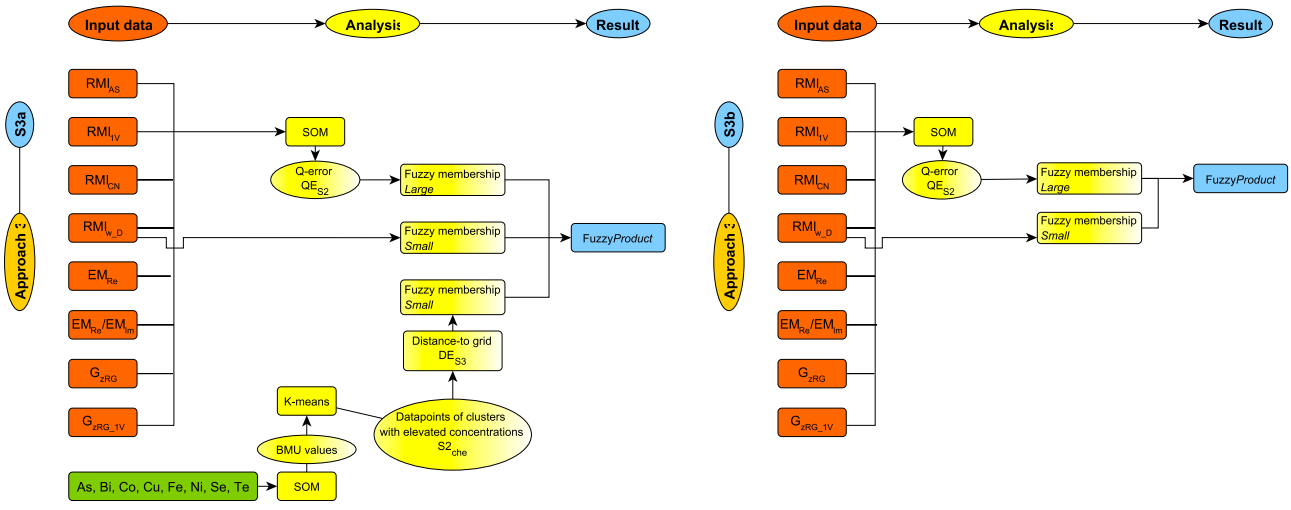

Fig. 23. Flowchart for Approach 3 of Savukoski data analysis.

naturally low in those areas far from the sampling points. Membership values for the distance grids were defined using the function *Small*. The midpoint for geochemistry was set to 1 km and spread to 5 to allow for uncertainty in the location of the respective bedrock source and to not give false negative signals. The midpoint for the distance to structures was set to 100 m and spread to 8, to restrict large membership values to only occur near the structures.

K-means clusters were not computed for geophysics, but geophysics SOM QE (QE_{S2}) was combined with the fuzzy combination of geochemistry and structures for model S3a and with the distance-to-structures grid for model S2b using fuzzy PRODUCT. The geophysics QE 95% quantile (QE = 1.49) was used as the midpoint and spread $s = 10$ in the *Large* fuzzy membership function to clearly distinguish between non-anomalous and anomalous values (Fig. 11). S3a (Fig. 24e) differs little from the geophysics QE (Fig. 24a), because large geochemical fuzzy membership values cover a major part of the image where high QE values exist. In S3b (Fig. 25), geochemical data have been omitted, and only structures were used to restrict the prospectivity indicated by QE.

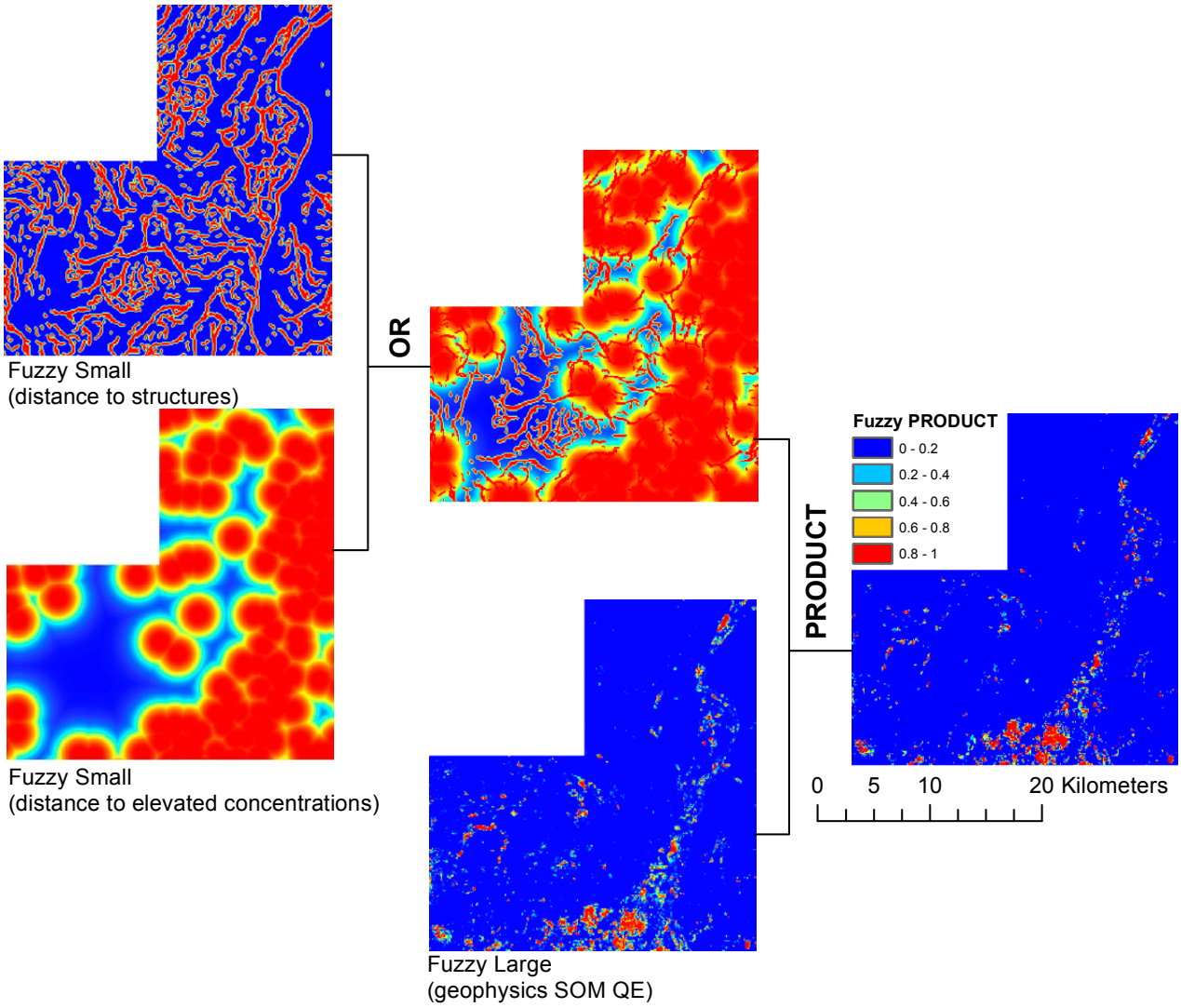

Fig. 24. Results of Approach 3, S3a, for the Savukoski dataset. Fuzzy process tree: 1) Fuzzy membership grids of "distance to structures" and "distance to elevated concentrations" are combined with the fuzzy operator OR, 2) the result of step 1 is combined with the "geophysics SOM quantization error" grid using the fuzzy operator PRODUCT.

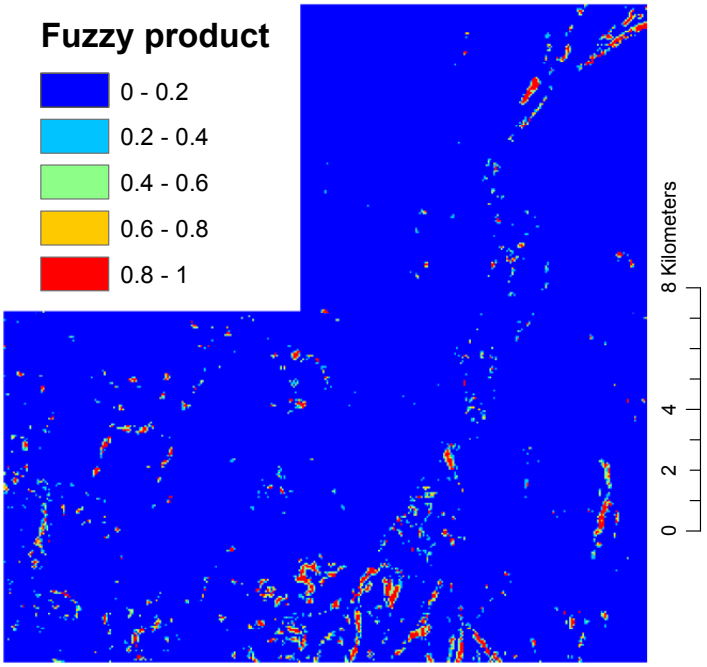

Fig. 25. Results of Approach 1, S3b, for the Savukoski dataset. Fuzzy PRODUCT of memberships of geophysics QE and distance to structures.

DISCUSSION AND PROPOSED WORKFLOW

A combination of two unsupervised data-driven clustering methods, SOM and K-means, and the knowledge-driven fuzzy-logic method were used in regional mineral prospectivity modelling of two regions in northern Finland: for IOCG in Enontekiö and for orogenic gold in Savukoski. The aim was to apply different methodological approaches and datasets to generate a general workflow for the modelling from data selection to the interpretation of the results. Because different study areas and deposit types should be investigated using different datasets, which in turn require different methodological approaches, the proposed workflow is flexible in terms of selecting the most suitable combination of analysis methods for the available dataset. As it is not necessarily obvious which choice of methods and datasets provides the most accurate results, analysis should be carried out in a few alternative ways.

The available relevant datasets often differ in character, and it is not clear that they can all be combined for simultaneous spatial analysis: although not combining all the data produces information loss, so does interpolation of all the data to the same grid resolution. Geochemistry, the sparsest of the datasets, turned out to be the most cumbersome in our case to be combined with the other datasets. The interpolation of geochemical data (2-km sampling distance) to the resolution of geophysics (sampling distance *ca*. 15–200 m) was challenging, mainly for two reasons. First, spatial variation in element concentrations was much greater than the sampling density, and there was no correlation between the difference in measured element concentrations and the distance between sampling points, i.e., the data were undersampled. Thus, if interpolation of the geochemical data was carried out so as to respect the data values at the sampling points (large power for the inverse distance), the interpolated grid points near sampling points would represent the local concentration and the points further from the sampling points an average of the closest sampling points. These two do not coincide in the case of undersampled data, which would be a problem for data mining methods that expect that all the points of the same data set to represent the same property. In fact, the grid points between sampling points would represent a relatively homogeneous, very slowly varying value, while sampling points would show up as local anomalies. The other complicating factor in processing geochemical data is the dispersive nature of till, which leads to the fact that individual data points do not often represent the average of bedrock sources at the sampling point, but are somewhat shifted. Transportation distances of till can be highly variable, even at regional scales. The direction and length of the transport has not been

mapped accurately enough to be utilized in regional-scale modelling. Due to till transportation, the local anomalies would not be connected to the correct geophysical properties when geochemistry and geophysics are combined for data mining. In Central Lapland, transportation distances of till are poorly known, but according to Salonen (1986) and Sarala & Ojala (2008), for example, they can be assumed to be of the order of 1 km, i.e., less than the sampling density of the national geochemical survey (1 sample/4 km^2).

As it is not necessarily obvious which choice of methods and datasets provides the most accurate results, analysis should be carried out in a few alternative ways. We chose to use geochemical data both as an interpolated grid and as the original sampling points. Interpolation was carried out so as to show regional average concentrations with no local anomalies caused by single sampling points. Different combinations of spatial analysis methods were also used, and the results presented here were obtained using three different approaches in both study areas. Results from the different approaches can be interpreted together to complement each other, or also independently of one another. In approaches 2 and 3, we used the SOM QE result, without K-means clustering, to indicate all anomalous data vectors in the given dataset, and post-processed it using fuzzy logic to reveal the most relevant anomalous features. In approach 1, we presented a smaller number of K-means clusters, which indicates larger features, similarly to a lithological map. When searching for small-scale features and local anomalies, the SOM proto-clusters and QE are useful, because details are lost in the coarse K-means clustering, which is applicable in determining the characteristics of the environment surrounding anomalous features. The models computed by using solely unsupervised data-driven methods (SOM, K-means) are primitive, only providing information on the data distribution in the study area, while models further processed with knowledge-driven fuzzy logic provide maps showing areas with different levels of prospectivity. However, because no training data are used to compute absolute probabilities for prospectivity, the levels should be considered qualitatively, for instance as "least prospective", "medium prospectivity" or "most prospective".

In visualizing the information content of a model, there is always a trade-off between simplicity and informativeness. The human eye and a highly trained brain with high-quality background knowledge is a combination whose capabilities are difficult to mimic with numerical computing. However, because the amount of visual information the brain can effectively handle depends on the background information stored in it, there is no single answer concerning how much freedom should be given to the user to interpret the model. Perhaps the simplest representation for prospectivity is a raster with ordered colour coding for the prospectivity. This is the result of our approach 2, and does not leave much space for further interpretation. The other extreme is our approach 3, where much more freedom was given for the user to visually interpret the prospectivity map by showing the different data types, such as a raster, points and lines, on the same map. Having a thorough knowledge of the different data types shown on the map, the user should be able to perform a higher quality interpretation than the computer in approach 2. Which approach is the most suitable, and how much freedom there should be for the user to interpret the result depends on the expertise of the user.

In addition to the choice of the model values presented as a result map, the colouring of the map can also be done in various ways, depending on its further use. In maps with a continuous colour range, the colour scale and stretch is usually more or less randomly chosen to provide a good looking map. This is acceptable if information on the colour scale and stretch are available for the map user. Sometimes, a contour map is more suitable for further visual inspection and comparison with other maps.

Proposed workflow

1) Formulation of a conceptual model for the deposit type, i.e., selection of the measurable/observable quantities that are assumed to be affected by the existence of mineral deposits.
2) Preprocessing of the original measurements in order to enhance desired features and to make different datasets spatially comparable. If not all datasets can be spatially combined, the data should be divided into subsets with similar properties.
3) Visualization of the inputs and assessment of their uncertainties.
4) Preclustering of multidimensional data subsets using SOM with a large number of nodes

(usually <<100 data points/node). If anomalous regions are of interest, QE should be saved for further use. Running of K-means clustering on the SOM protoclusters using a handful (>30) of cluster centres.

5) Use of fuzzy logic to combine the single component data subsets and the QE of the clustered multidimensional data subsets. Original components of the multidimensional datasets can also be used to rule out uninteresting high QE values.
6) Investigation of the data distribution within K-means clusters to define which clusters represent interesting data values. Presentation of the interesting clusters on a map.
7) Representation of the results as:
   – a map of QE values or fuzzy results;
   – a map of clusters indicating the interesting ones;
   – a combination of the above two maps for a combination of raster, point-like and line-element data.

ACKNOWLEDGEMENTS

We thank the Finnish Funding Agency for Innovation and the K.H. Renlund foundation for funding the project. We also thank Pertti Sarala for advice concerning the geochemical data and Peter Sorjonen-Ward for geological interpretation of the geophysical data. With Vesa Nykänen, Helena Hulkki, Tuomo Karinen and Jukka Konnunaho, we had fruitful discussions on spatial modelling and they provided local expertise on bedrock geology. Tero Niiranen's expertise on IOCGs and Heikki Salmirinne's on geophysics was of utmost importance for completing this study.

REFERENCES

Abbas, O. 2008. Comparisons between data clustering algorithms. The International Arab Journal of Information Technology 5 (3), 320–325.

Airo, M.-L. 2007. Application of aerogeophysical data for gold exploration: implications for the Central Lapland Greenstone Belt. In: Ojala V. J. (ed.) Gold in the Central Lapland Greenstone Belt. Geological Survey of Finland, Special Paper 44, 187–208.

Airo, M.-L. & Mertanen, S. 2008. Magnetic signatures related to orogenic gold mineralization, Central Lapland Greenstone Belt, Finland. Journal of Applied Geophysics 64, 14–24.

Bação, F., Lobo, V. & Painho, M. 2005. Self-organizing maps as substitutes for K-means clustering. In: Sunderam, V. S., van Albada, G., Sloot, P. & Dongarra, J. J. (eds) International Conference on Computational Science 2005. Lecture Notes in Computer Science, Vol. 3516. Berlin Heidelberg: Springer-Verlag, 476–483.

Bedrock of Finland – DigiKP. Digital map database [Electronic resource]. Espoo: Geological Survey of Finland [accessed 31.3.2014]. Version 1.0. Available at: http://www.geo.fi/en/bedrock.html

Bergman, S., Kübler, L., & Martinsson, O. 2001. Description of regional geological and geophysical maps of northern Norrbotten country (east of the Caledonian orogen). Sveriges Geologiska Undersökning, Ba 56, p. 101.

Bierlein, F., Murphy, F., Weinberg, R., & Lees, T. 2006. Distribution of orogenic gold deposits in relation to fault zones and gravity gradients: targeting tools applied to the Eastern Goldfields, Yilgarn Craton, Western Australia. Mineralium Deposita 41, 107–126.

Bierlein, F., Fraser, S., Brown, W. & Lees, T. 2008. Advanced methodologies for the analysis of databases of mineral deposits and major faults. Australian Journal of Earth Sciences 55:1, 79–99.

Blakely R. 1995a. Directional derivates. In: Blakely, R. (ed.) Potential theory in gravity & magnetic applications. Cambridge University Press, 324–327.

Blakely, R. 1995b. Level Surface to Level Surface. In: Blakely, R. (ed.) Potential theory in gravity & magnetic applications. Cambridge University Press, 315–319.

Blakely, R., & Simpson, R. 1986. Approximating edges of source bodies from magnetic or gravity anomalies. Geophysics 51, 1494–1498.

Dampney, C. 1969. The equivalent source technique. Geophysics 34, 39–53.

Davies, D. & Bouldin, D. 1979. A cluster separation measure. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-1 (2), 224–227.

Dransfield, M. 2009. Conforming Falcon gravity and the global gravity anomaly. Geophysical Prospecting 58 (3), 469–483.

Eilu, P., Pankka, H., Keinänen, V., Kortelainen, V., Niiranen, T. & Pulkkinen, E. 2007. Characteristics of gold mineralisation in the greenstone belts of northern Finland. In: Ojala, V. J. (ed.) Gold in the Central Lapland Greenstone Belt. Geological Survey of Finland, Special Paper 44, 57–106.

Flexer, A. 1999. On the use of self-organizing maps for clustering and visualization, Principles of data mining and knowledge discovery, Lecture notes in artificial intelligence 1701, 80–88.

Fraser, S. & Hodgkinson, J. 2009. An investigation using SiroSOM for the analysis of QUEST stream-sediment and sake-sediment geochemical data. Geoscience BC - Report 2009-14, CSIRO Report Number: EM MDU P2009/983.

Fraser, S., Wilson, G., Cox, L., Cuma, M., Zhdanov, M. & Vallée, M. 2012. Self-organizing maps for pseudo-lithological classification of 3D airborne electromagnetic, gravity gradiometry and magnetic inversions. Proceedings of the 22nd ASEG conference, Feb 2012, Brisbane, Australia.

Grant, F. & Dodds, J. 1972. MAGMAP FFT processing system development notes. Paterson Grant and Watson Limited.

Hanski, E., Perttunen, V. & Sorjonen-Ward, P. 1997. Overview of the geology of northern Finland. In: Korkiakoski, E. & Sorjonen-Ward, P. (eds) Research and exploration - where do they meet? 4th Biennial SGA Meeting, August 11-13, 1997, Turku, Finland. Excursion guidebook B1: ore deposits of Lapland in northern Finland and Sweden. Geological Survey of Finland, Guide 43, 7−9.

Hautaniemi, H., Kurimo, M., Multala, J., Leväniemi, H. & Vironmäki, J. 2005. The "Three In One" aerogeophysical concept of GTK in 2004. In: Airo, M.-L. (ed.) Aerogeophysics in Finland 1972-2004: Methods, System Characteristics and Applications. Geological Survey of Finland, Special Paper 39, 21−74.

Hitzman, M. W. 2000. Iron oxide-Cu-Au deposits: What, where, when, and why. In: Porter, T. M. (ed.) Hydrothermal iron oxide copper-gold and related deposits a global perspective. Adelaide: Australian Mineral Foundation, 9–26.

Hornby, P., Boschetti, F. & Horowitz, F. 1999. Analysis of potential field data in the wavelet domain. Geophysical Journal International 137, 175−196.

Hyvönen, E., Turunen, P., Vanhanen, E., Arkimaa, H. & Sutinen, R. 2005. Airborne Gamma-ray Surveys in Finland. In: Airo, M.-L. (ed.) Aerogeophysics in Finland 1972-2004: Methods, System Characteristics and Applications. Geological Survey of Finland, Special Paper 39, 119−134.

Kohonen, T. 2001. Self-Organizing Maps. Third Extended Edition, Springer Series in Information Sciences 30.

Korhonen, J. 2005. Airborne Magnetic Method: Special Features and Review on Applications. In: Airo, M.-L. (ed.) Aerogeophysics in Finland 1972-2004: Methods, System Characteristics and Applications. Geological Survey of Finland, Special Paper 39, 77−102.

Kujansuu, R. 1967. On the deglaciation of Finnish Lapland. Bulletin de la Commission Géologique de Finlande 232. 96 p.

Middleton, M., Schnur, T., Sorjonen-Ward, P. & Hyvönen, E. 2015. Geological lineament interporetation by object based image analysis approach – Results of semi-automated analyses versus visual interpretation. In: Sarala, P. (ed.) Novel technologies for green field exploration. Geological Survey of Finland, Special Paper 57, 135–154. (this journal)

Mineral Deposit Database of Finland. Digital map database [Electronic resource]. Espoo: Geological Survey of Finland [referred 12.09.2014]. Available at: http://gtkdata.gtk.fi/MDaE/index.html.

Mutanen, T. 1997. Geology and petrology of the Akanvaara and Koitelainen mafic layered intrusions and the Keivitsa-Satovaara layered complex, northern Finland. Geological Survey of Finland, Bulletin 395.

Nabighian, M. N. 1984. Toward a three-dimensional automatic interpretation of potential field data via generalized Hilbert transforms: Fundamental relations. Geophysics 49 (3), 780−786.

Niiranen, T. 2005. Iron oxide-copper-gold deposits in Finland: case studies from the Peräpohja schist belt and the Central Lapland greenstone belt. Publications of the Department of Geology, D 6. Helsinki: University of Helsinki. 27. 92 p.

Nisbet, B., Cooke, J., Richards, M. & Williams, C. 2000. Exploration for iron oxide copper gold deposits in Zambia and Sweden: Comparison with the Australian experience. In: Porter, T. M. (ed.) Hydrothermal Iron Oxide Copper-Gold and Related Deposits: A Global Perspective. Adelaide: PGC Publishing, 297−308.

Nykänen, V. & Salmirinne, H. 2007. Prospectivity analysis of gold using regional geophysical and geochemical data from the Central Lapland Greenstone Belt, Finland. In: Ojala, V. J. (ed.) Gold in the Central Lapland Greenstone Belt. Geological Survey of Finland, Special Paper 44, 251−269.

Nykänen, V., Groves, D., Ojala, V. & Gardoll, S. 2008a. Combined conceptual/empirical prospectivity mapping for orogenic gold in the northern Fennoscandian Shield, Finland. Australian Journal of Earth Sciences 55, 39−59.

Nykänen, V., Groves, D., Ojala, V. & Gardoll, S. 2008b. Reconnaissance-scale conceptual fuzzy-logic prospectivity modeling for iron oxide copper-gold deposits in the northern Fennoscandian Shield, Finland. Australian Journal of Earth Sciences 55, 25−38.

Ojala, V. J. (ed.) 2007. Gold in Central Lapland Greenstone Belt. Geological survey of Finland, Special Paper 44.

Patison, N. 2007. Structural controls on gold mineralization in the Central Lapland Greenstone Belt. In: Ojala, V. J. (ed.) Gold in the Central Lapland Greenstone Belt. Geological Survey of Finland, Special Paper 44, 107−124.

Phillips, J. D. 2007. Geosoft eXecutables (GX's) Developed by the U.S. Geological Survey, Version 2.0, with Notes on GX Development from Fortran Code. U.S. Department of the Interior, U.S. Geological Survey, Open -File Report 2007−1355. 111 p. Available at http://pubs.usgs.gov/of/2007/1355/pdf/of07-1355.pdf.

Salminen, R. & Tarvainen, T. 1995. Geochemical mapping and databases in Finland. Journal of Geochemical Exploration 55, 321−327.

Salonen, V.-P. 1986. Glacial transport distance distributions of surface boulders in Finland. Geological Survey of Finland, Bulletin 338.

Sandrin, A., Edfelt, Å., Waight, T. E., Berggren, R. & Elming, S. 2009. Physical properties and petrologic description of rock samples from an IOCG mineralized area in the northern Fennoscandian Shield, Sweden. Journal of Geochemical Exploration 103, 80−96.

Sarala, P. & Ojala, V. J. 2008. Implications of complex glacial deposits for till geochemical exploration: Examples from the central Fennoscandian ice sheet. In: Stefánsson, Ó. (ed.) Geochemistry Research Advances. Chapter 1. New York: Nova Publishers, 1−29.

Sorjonen-Ward, P., Nurmi, P., Härkönen, I. & Pankka, H. 1992. Epigenetic gold mineralization and tectonic evolution of a lower Proterozoic greenstone terrane in the northern Fennoscandian Shield, in Metallogeny related to tectonics of mobile belts. In: Sarkar, S. C. (ed.) Metallogeny Related to Tectonics of the Proterozoic Mobile Belts. New Delhi: Oxford and IBH Publishing, 37–52.

Sorjonen-Ward, P., Ojala, V. & Airo, M.-L. 2003. Structural modeling and magmatic expression of hydrothermal alteration in the Paleoproterozoic Lapland greenstone belt, northern Fennoscandian Shield. In: Eliopoulos, D. G. et al. (eds) Mineral Exploration and Sustainable Development. Proceedings of the Seventh Biennial SGA Meeting, Athens, Greece, 24-28 August 2003. Rotterdam: Millpress, 1107–1110.

Suppala, I., Oksama, M., & Hongisto, H. 2005. GTK airborne EM system: characteristics and interpretation guidelines. In: Airo, M.-L. (ed.) Aerogeophysics in Finland 1972-2004: Methods, System Characteristics and Applications. Geological Survey of Finland, Special Paper 39, 103–118.

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. & Alhoniemi, E. 2000. Clustering of the self-organizing map, IEEE transactions on neural networks, 11(3).

Waller, N., Kaiser, H., Illian, J. & Manry, M. 1998. A comparison of the classification capabilities of the 1-dimensional Kohonen neural network with two partitioning and three hierarchical cluster analysis algorithms. Psychomatrika 63 (1), 5–22.

Williams, P. J. 2001. Overview of Geochemistry of Fe Oxide-Cu-Au Deposits. In: Mark, G. & Williams, P. J. (eds) Short course on geochemistry of the hydrothermal alteration systems associated with Fe oxide-copper-gold deposits, Jupiters Hotel, Townsville, 20th May 2001. Economic Geology Research Unit, James Cook University, Part B, B-1–B-6.