Drill-Core Hyperspectral and Geochemical Data Integration in a Superpixel-Based Machine Learning Framework

Isabel Cecilia Contreras Acosta [ID], Mahdi Khodadadzadeh [ID], Raimon Tolosana-Delgado [ID], and Richard Gloaguen

Abstract—The analysis of drill-core samples is one of the most important steps in the mining industry for the exploration and discovery of mineral resources. Geochemical assays are a common approach to represent the abundance of different chemical elements and aid at quantifying the concentrations of the important ore accumulations. However, their acquisition is time-consuming and usually averages of long core portions. Hyperspectral data are increasingly being used in the mining industry to complement the analysis of drill-cores due to their efficiency and fast turn-around time. Moreover, hyperspectral imaging is a technique able to provide data with high spatial resolution. In this article, we propose to integrate the complementary information derived from hyperspectral and geochemical data via a superpixel-based machine learning framework. This framework considers the difference in spatial resolution through segmentation. We extract labels from the geochemical assays and select, from the hyperspectral data, representative samples for each measurement. A supervised machine learning classification (composite kernel support vector machine) is then used to extrapolate the elements relative abundance to the entire core length. We propose an innovative integration of hyperspectral data covering different regions of the electromagnetic spectrum in a kernel-based framework to facilitate the identification of a larger amount of elements. A qualitative and quantitative evaluation of the results demonstrates the capabilities of the proposed method, which provides approximately 20% more accurate results than the pixel-based approach. Results also imply that the method could be beneficial for the reduction of geochemical assays needed for the detailed analysis of the cores.

Index Terms—Data integration, drill-cores, geochemical data, hyperspectral data, machine learning (ML), superpixel segmentation.

I. Introduction

MINING companies dedicate part of their revenues (about 10 billion USD in 2018 [1]) to the discovery and exploration of new mineral resources in commercially viable concentrations. Drilling is a required procedure for the exploration of subsurface mineral deposits. For this purpose, drilling machines

extract cylindrical rock samples (of few centimeters in diameter and several meters in length) known as drill-cores. Drill-cores are considered as unaltered specimen of rocks, which provide valuable geological information to characterize potential ore accumulations [2].

The initial analysis of the drill-cores is performed visually by geologists when logging the core characteristics (e.g., rock type, texture, alteration facies, ore-forming minerals, and structures). However, this task is time-consuming and relies on the knowledge and experience of the geologist for consistent results. Considering the cost and importance of drilling, extracting quantitative and unbiased information is crucial. Thus, the logging is always supported by other analytical techniques performed on specific and small sections of few selected samples. Optical microscopy [3], X-ray diffraction [4], X-ray fluorescence [5], scanning electron microscopy (SEM) [6], and laser ablation inductively coupled plasma mass spectrometry [7] are examples of the analytical techniques employed. These analyses provide information on the model mineralogy and mineralogical texture (quantified mineralogy) and geochemistry of the samples (element concentrations), among others.

Minerals have different spectral responses in specific regions of the electromagnetic spectrum. These responses are mainly related to the fundamental electronic and vibrational processes of the different molecular bonds; the mineral assemblages and grain sizes also influence the responses [8], [9]. Hyperspectral imaging is an emerging technique in the mining industry, which offers a noninvasive and nondestructive tool to scan a large number of cores in a fast turn-around time [10]–[14]. Hyperspectral data are recorded in several tens of spectral bands to capture spectral information over a wide range of wavelengths. This rich spectral information can be effectively exploited for the characterization and determination of the spatial distribution of different minerals, so-called mineral mapping [8], which support the geologists on the delineation of the deposits and to target important ore accumulations.

The analysis of drill-core hyperspectral data has often been described as a chain of techniques available in well-established software [10], [11], [15]–[20]. A manual or semiautomatic analysis of the spectrum absorption properties (e.g., depth, width, and wavelength position) has also been applied in the literature to map specific minerals [9], [21]–[27]. All the aforementioned strategies require a great number of human interactions and expertise, particularly visual interpretations and comparisons

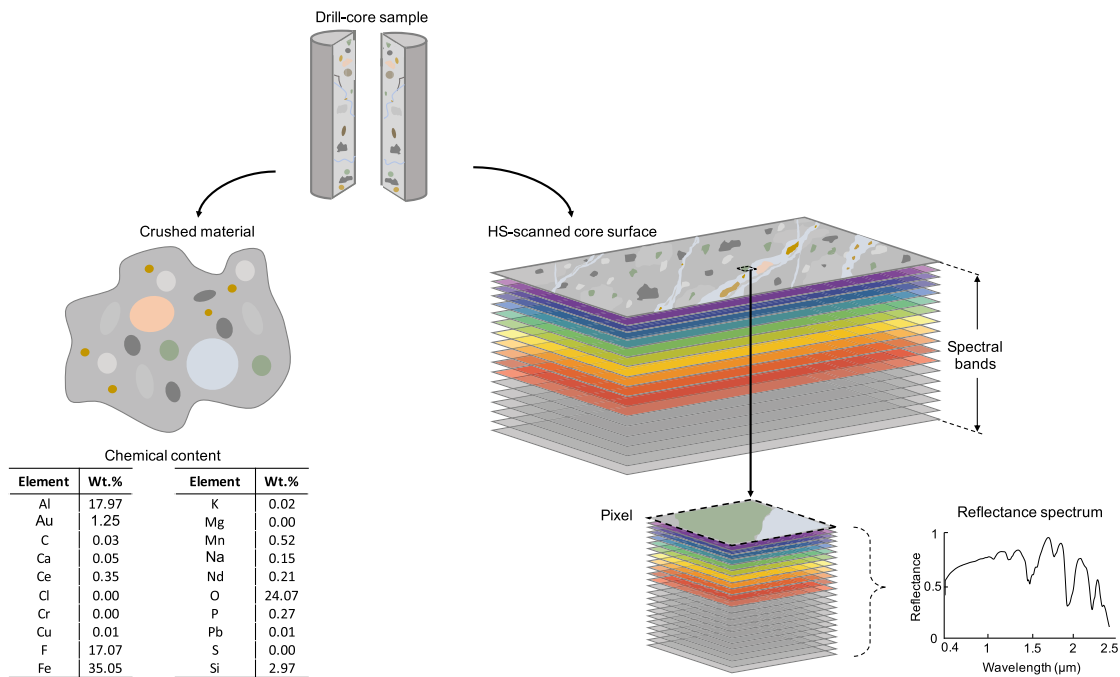| Chemical content			
Element	Wt.%	Element	Wt.%
Al	17.97	K	0.02
Au	1.25	Mg	0.00
C	0.03	Mn	0.52
Ca	0.05	Na	0.15
Ce	0.35	Nd	0.21
Cl	0.00	O	24.07
Cr	0.00	P	0.27
Cu	0.01	Pb	0.01
F	17.07	S	0.00
Fe	35.05	Si	2.97

Fig. 1. Illustration of the drill-core hyperspectral and geochemical data. Hyperspectral data are commonly acquired from the surface of half of the drill-core sample kept for the archive, and geochemical data are obtained from either point measurements performed along the core or crushed material.

with available reference spectral libraries (e.g., USGS Spectral library [28]) [29].

Recently, machine learning (ML) based methods have been suggested for the analysis of drill-core hyperspectral data to improve accuracy, speed, and robustness [30]–[32]. The ML techniques offer automatic means to unravel underlying relationships in large amounts of data. However, defining meaningful mineral classes and obtaining representative samples to train an ML model is not straightforward in drill-core samples. In our previous attempts [33]–[35], we proposed the use of high-resolution mineralogical data, obtained with SEM-based Mineral Liberation Analysis software, to generate reference data and train a supervised ML algorithm. This strategy allows upscaling the very detailed mineralogical information from small and relevant sections of the core to entire drill-core sections. Although SEM-based analytical techniques provide valuable submicrometer mineralogical information, their usage is still limited in the mining industry due to their costs and availability; one single thin section takes up to 6 h for its acquisition and analysis in a specialized laboratory.

Analytical techniques that provide geochemical information (e.g., element concentrations) are by far the most common approach in operational exploration. One reason for this is that geochemical data are used for whole-rock composition analyses, which aid to the delineation of lithological boundaries and units [36], [37]. Moreover, drill-core geochemical data are useful for targeting important concentrations of elements that reveal specific and relevant minerals of the geological systems and, therefore, serve as proxies for economically viable accumulations [38]–[40]. Although the geochemical data provide valuable information, they suffer from a scale limitation. Geochemical

data are frequently obtained either from point measurements or the so-called assays, performed on crushed material from specific drill-core sections. Usually, geostatistical interpolation techniques are used to estimate missing information across the drill-cores, which introduces uncertainties. On the contrary, hyperspectral data provide high-resolution and large-scale information but lacks the ability to provide quantitative concentrations of critical elements.

Fig. 1 demonstrates the links between drill-core geochemical and hyperspectral data. Typically, the drill-core samples are cut in halves, and one half is crushed and used for laboratory analyses. The other half that is often kept for the archive can be scanned to acquire the hyperspectral data. Geochemical data are usually represented as a table including the analytical values of chemical elements determined, for example, by laboratory assays of the crushed materials. Hyperspectral data are represented in the form of a 3-D data cube, in which a reflectance spectrum is associated with each pixel. Theoretically, the integration of hyperspectral and geochemical data is beneficial to estimate the detailed spatial distribution of elements concentrations across the whole drill core. Details on the fundamental links between hyperspectral and geochemical data are presented in Section II.

Previous studies have suggested the combined use of geochemical and hyperspectral data for the interpretation and validation of the results [25], [39], [41]–[44]. For example, in [42], the geochemical and hyperspectral datasets are processed and analyzed separately, and their results are compared and used for a complete understanding and analysis of the correlation between stratigraphic control as well as mineralogical alterations. In [45], the authors proposed a combination strategy and linked hyperspectral and geochemical data to successfully infer

geochemical and mineralogical characteristics of a specific rock type from hyperspectral data. From their results, it can be seen that the predicted geochemistry showed greater detail than the geochemical data. This is because of the better spatial resolution of hyperspectral data. Although drill-core geochemical data have been applied in the previously mentioned studies to interpret and validate hyperspectral data, to the best of our knowledge, a systematic link of hyperspectral and geochemical data has not been addressed in the literature.

In this article, we develop a new ML-based technique for the integration of drill-core geochemical and hyperspectral data. We propose to exploit both datasets to upscale the element abundance information obtained from the geochemical assays to the entire borehole by means of the hyperspectral data. The hyperspectral data have a higher spatial resolution (i.e., 1.6 mm/pixel) than the geochemical data (i.e., assays cover around 1 m long). To tackle the difference in spatial resolution, we perform a superpixel segmentation of the hyperspectral data. Each superpixel corresponds to an object that contains spatially connected and spectrally similar pixels [46]. Our proposed method integrates geochemical data as a complementary source of information within an ML classification system. It fills the gap between the scales of the measurements in these two data sources by applying a superpixel-based approach.

In our proposed method, we extract labels from the geochemical data obtained from the assay measurements along an entire drill-core, and for each of the extracted labels, considering the superpixels, a representative spectrum is assigned. Then, the geochemical labels with the representative spectra are used as input to train an ML model. Once the training is performed and the model parameters are estimated, the model is used to automatically classify the superpixels derived from the complete drill-core hyperspectral data and, thus, spread the chemistry. This approach requires minimum human interaction, and therefore, the results are not impacted by different expert interpretations.

Another important contribution of this article is the integration of hyperspectral data covering different regions of the electromagnetic spectrum, i.e., visible to near-infrared (VNIR), short-wave infrared (SWIR), and the long-wave infrared (LWIR). This integration allows a more comprehensive and complete analysis including not only the elements that influence the diagnostic spectrum absorptions in the VNIR-SWIR but also in the LWIR. There are several studies in the literature confirming the advantages of such integration [45], [47]–[51]. In this article, we suggest to use the composite kernel support vector machine (ckSVM) algorithm [52], [53] to balance the spectral information contained in the hyperspectral data covering the VNIR-SWIR and LWIR regions of the electromagnetic spectrum. We chose ckSVM because it has been proven to improve the performance of the classification by simultaneously exploiting information from two datasets [52]–[54]. Moreover, it allows considering *a priori* knowledge, which can be beneficial if there is a tendency in the mineralogy of the samples towards minerals with diagnostic spectrum absorptions in specific regions of the electromagnetic spectrum.

In our experiments, we considered geochemical and hyperspectral data of a core of about 330 m long drilled from the Rajapalot gold–cobalt prospect located at the Ylitornio and Rovaniemi municipalities in northern Finland. The hyperspectral data cover the VNIR-SWIR region of the electromagnetic spectrum with 340 bands and the LWIR region with 94 bands. Regarding the geochemical data, 190 measurements assays comprising the concentrations of 49 chemical elements in parts per million (ppm) are available.

The rest of this article is organized as follows. Section II explains the fundamental links between geochemical and hyperspectral data. Section III presents the proposed framework. Section IV describes the datasets used in this study and the experimental results achieved. Section V presents the discussion. Finally, Section VI concludes this article.

II. BASIS FOR THE INTEGRATION OF GEOCHEMICAL AND HYPERSPECTRAL DATA

The acquisition of geochemical assays in drill-cores is routinely performed over sections covering areas ranging from few centimeters to couple of meters, depending on the purpose of the study [55]. Geochemical data refer to the abundance of different chemical elements (e.g., major, minor, and trace elements) in the form of proportions or percentages of the whole sample [e.g., weight percent or parts per million (ppm)] subject to a constant sum [56]. Although geochemical data represents a key factor in characterizing geological deposits and calculating grade of the ore resource, the precision and accuracy of the ore body valuation are related to the amount of data available [55]. This is an issue since hundreds of meters of cores are extracted during exploration campaigns, and geochemical data acquisition is generally destructive, time-consuming, and costly.

Hyperspectral data are represented by reflectance spectrum profiles, where minima, called absorptions features, occur due to different physical and chemical processes when incoming electromagnetic radiations (light) interact with the elements and molecular bonds of the samples under observation. These absorption features illustrate the relation between geochemical and hyperspectral data. For example, in the SWIR region of the electromagnetic spectrum, molecular bonds in CO_3 generate an absorption feature between 2306 and 2365 nm, and in this region, the main absorption feature of calcite ($CaCO_3$) is located [57] (see example in Fig. 2). In the LWIR, the most intense spectrum absorptions are due to stretching and bending processes within molecules of e.g., Si–O, Si–O–Si, Si–O–Al, CO_3, and SO_4 [58]. For more details on the minerals with diagnostic responses in the VNIR-SWIR and LWIR regions of the electromagnetic spectrum, we refer readers to [33].

III. PROPOSED APPROACH

The proposed system to integrate geochemical and hyperspectral data for the analysis of drill-core samples is summarized in Fig. 3. The algorithm is divided into two main steps: the training and the prediction phase. In the training phase, first, the Gaussian mixture model (GMM) is applied to label
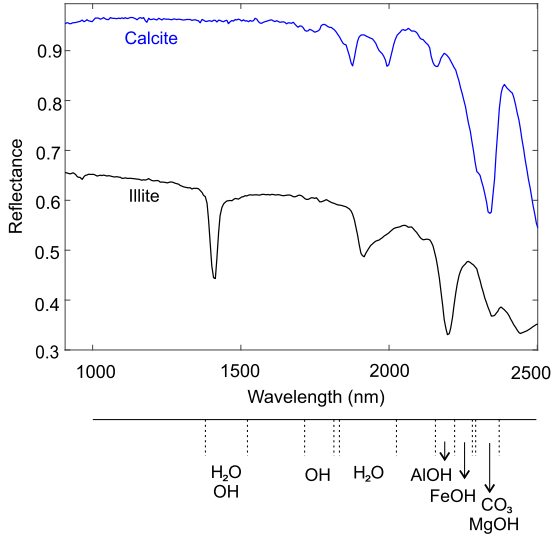
Fig. 2. Example spectra from the USGS spectral library and major spectral absorption bands in the SWIR [28], [57].

the transformed geochemical data. In a parallel process, the hyperspectral VNIR-SWIR and LWIR data are concatenated and segmented into superpixels. For the regions where the geochemical assays were performed, the weighted mean spectra of the extracted superpixels are calculated and linked with the computed geochemical labels. To finalize the training phase, the labels and their respective spectra are used to train a ckSVM model. In the prediction phase, the trained model is used to classify the superpixels and derive a classification map. Details of the approach are presented in the following subsections.

A. Geochemical Data Labeling

The transformation of the geochemical data consists of collinearity removal, logarithmic transformation, and dimensionality reduction. In general, variables are considered collinear if the data vectors representing them lie on a subspace of lower dimension. For two variables, this subspace is linear. A broader definition considers two variables collinear if they lie almost on the same line and, thus, if the angle between the data vectors is small [59]. The Belsley diagnostics [59] are implemented to assess the strength and source of collinearity among the variables. The Belsley method identifies the linearly correlated variables by setting a threshold for the variance-decomposition proportions known as the tolerance threshold. The variance-decomposition proportions refer to the group of variables affected by dependencies and their influence [60]. A transformation is then needed before applying ML algorithms to transfer the geochemical data from a constrained sample space into an unconstrained space. In such a way, the constant-sum constraint embedded in any compositional data is removed, while the true covariance structure of the data remains the same [61]. The data are adapted using the centered log-ratio (CLR) algorithm [62]. CLR is a logarithm transformation of the data scaled by its geometric mean [63]. Finally, the dimension of the data is reduced using

principal component analysis (PCA) [64] to transform the set of interrelated variables to a new set of uncorrelated variables. The first principal components retain most of the variance present in the original variables.

Once the geochemical data are transformed, the number of classes that better summarizes the geochemical data, k, is estimated using the Elbow method. This method is based on the within-cluster sum of square errors metric [65]. This metric corresponds to the squared average Euclidean distance of all the points within each class to the centroid of the class to which they are assigned. After estimation of the appropriate number of classes, the geochemical data are finally clustered using the GMM [66], [67]. We chose the GMM due to its well-known flexibility and advantages of finding the probability of belonging to different components within the dataset.

The GMM has gained significant attention due to its capability of capturing non-Gaussian statistics of multivariate data. This method is a parametric model that considers the data as originated from a weighted sum of several multivariate Gaussian sources that allows computing posterior probability density functions for each observation [66]. A GMM with k number of components can be expressed as

$$f_{\mathbf{x}}(\mathbf{x}) = \sum_{k=1}^{K} \pi_k N\left(\mathbf{x}; \mu_k, \mathbf{T}_k^{-1}\right)$$

where $\mathbf{x}$ is a multidimensional random vector and the multivariate normal probability density function is $N(\mathbf{x}; \mu_k, \mathbf{T}_k^{-1})$. The probability density function is characterized by the mean vector μ_k, the inverse covariance matrix $\mathbf{T}_k$, and a mixing proportion weight $\pi_k \geq 0$, which is subject to the constraint $\sum_{k=1}^{K} \pi_k = 1$ [68]. The learning of the GMM algorithm is performed by maximizing the likelihood function with the expectation–maximization approach, which estimates the mixture parameters from the data once k is specified [67]. The algorithm will iterate over these steps until convergence and each geochemical assay will be then assigned to their specific class. To estimate the composition of each class, the average of the chemical composition of the assays belonging to each class is considered.

B. Superpixel Segmentation

To account for the differences in spatial resolution between hyperspectral and geochemical data, the hyperspectral datasets are segmented into superpixels. These superpixels are used in both training and prediction phases. In the training phase, the superpixels that cover the regions of geochemical assays are used to derive the training spectra. For each geochemical region, a representative spectrum is calculated by the mean of the corresponding superpixels weighted by the size of the superpixel. The usage of the mean tends to average and smooth the variability in the spectra; however, in this work, this is negligible because each geochemical measurement summarizes information extracted from a region of around 1 m long. In the prediction phase, the classification is performed at the object level; it means that the superpixels are being classified instead of the pixels. For each superpixel, the average of all the pixels
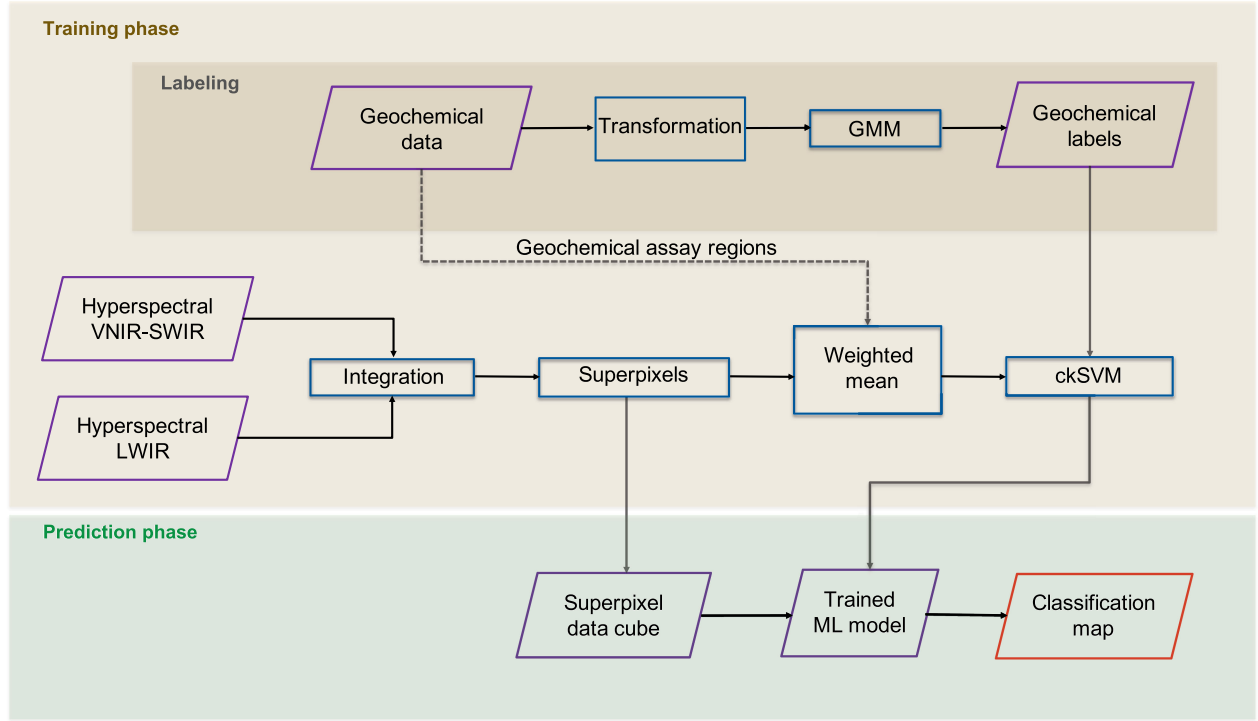
Fig. 3.   Flowchart of the proposed approach to integrate geochemical and hyperspectral data for the analysis of drill-core samples.

belonging to that superpixel is calculated. Thus, a superpixel data cube is generated for each hyperspectral dataset.

The segmentation is performed with the well-established entropy rate superpixel (ERS) segmentation [69]. ERS is known for considering boundary adherence, meaning that it is able to capture small details. A good boundary adherence influences compactness, regularity, and smoothness. ERS is considered a stable algorithm since it has been shown that the standard deviation in the generated superpixels is low [70]. The ERS belongs to the family of graph-based algorithms. These methods treat the images as indirect graphs and perform the segmentation based on edge weights [70]. Each graph is partitioned into a connected subgraph by choosing a subset of edges $\mathbf{A} \subseteq \mathbf{E}$ [71].

The objective function in the ERS algorithm consists of two terms. The first is an entropy rate term $H(\mathbf{A})$ of a random walk on a graph, which favors the formation of compact and homogeneous superpixels. The second, $B(\mathbf{A})$, is a balancing term that fosters superpixels with similar sizes and, thus, reduces the number of unbalanced superpixels. Both terms optimize the superpixel segmentation process

$$\mathbf{A}^* = \arg\max \operatorname{Tr}\{H(\mathbf{A}) + \alpha B(\mathbf{A})\}, \quad \text{s.t. } \mathbf{A} \subseteq \mathbf{E}.$$

α is used to balance the contributions of the entropy rate and the balancing term. $\mathbf{E}$ corresponds to the edge set [69], [71].

C. Classification

Once the labels and the representative samples are selected, they are given as input to a supervised ML algorithm. For this,

the ckSVM algorithm [52] was chosen. This algorithm is an extension of the well-known support vector machine (SVM) [72], [73], and it allows a more advanced integration of the VNIR-SWIR and LWIR hyperspectral data since different kernels are built for each dataset. We suggested to use SVM because it has been proven to give good results for mineral mapping [33], and it has a good generalization ability. Moreover, SVM is an effective classification method that performs well with limited number of training samples, which tend to be the case in drill-cores where usually detailed *a priori* information is only available in small sections of the core. The good performance of SVM with a small set of training samples is because the training samples close to the class boundary, so-called support vectors, are the ones that influence the most the location of the hyperplane [73]–[75].

SVM exploits the training data and searches for a separating class boundary, known as hyperplane, with the largest margin [72]. It was originally developed as a linear classifier to solve binary classification problems. However, the decision boundaries in classification problems are nonlinear. SVM makes use of kernel methods to tackle this issue. The data are projected into a high-dimensional feature space, where the samples are linearly separable [76]. The final hyperplane decision function in a binary SVM classifier is defined as

$$f(\mathbf{x}) = \left(\sum_{i=1}^{n} \alpha_i y_i \mathbf{K}(\mathbf{x}_i, \mathbf{x}) + b \right)$$

where $\mathbf{x}_i \in \mathbb{R}^d$, $i = 1, \ldots, n$, is a set of n training samples with their corresponding class labels $y_i \in \{1, +1\}$, α_i denotes the Lagrange multipliers, and b refers to the bias. $\mathbf{K}$ denotes the

kernel function which can be defined, for a mapping function ϕ, as $\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$. SVM can be converted to a multiclass classification problem with the well-known one-against-one or the one-against-all strategies. For a detailed review on the SVM technique, we refer readers to [76].

Composite kernels were originally developed to enhance classification accuracies by incorporating spatial content in addition to the spectral information [52]. In this work, the datasets to be combined are the VNIR-SWIR and LWIR hyperspectral data. For this, the two generated kernels are $\mathbf{K}^s(\mathbf{x}_i^s, \mathbf{x}_j^s)$ that refers to the VNIR-SWIR kernel and $\mathbf{K}^l(\mathbf{x}_i^l, \mathbf{x}_j^l)$ that refers to the LWIR kernel. For the kernel combination, different approaches have been developed. In this study, we implement the *weighted summation kernel* because it has the advantage of introducing *a priori* knowledge in the classifier by specifying different weights (ω) for each source of information [52]

$$\phi(\mathbf{x}_i) = \{\varphi_1(\mathbf{x}_i^s), \varphi_2(\mathbf{x}_i^l)\}$$

where φ_1 and φ_2 are the two nonlinear transformations of $\mathbf{x}_i^s$ and $\mathbf{x}_i^l$, respectively. Thus, the sum of the VNIR-SWIR and LWIR dedicated kernels matrices can be performed. Moreover, ω, a positive real-valued free parameter $(0 < \omega < 1)$, is considered to introduce a weight for each kernel. From this, the kernel function can be computed as follows:

$$\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) = \omega \mathbf{K}_s(\mathbf{x}_i^s, \mathbf{x}_j^s) + (1 - \omega)\mathbf{K}_l(\mathbf{x}_i^l, \mathbf{x}_j^l).$$

ω represents the tradeoff between both datasets, and it is tuned in the training process.

IV. EXPERIMENTAL RESULTS

A. Data Description

Geochemical and hyperspectral data of a complete drill core of about 330 m long provided by Mawson Resources Limited company were used to showcase the proposed approach. The drill core was cut every 1 m and stored in boxes containing around 5 m of core.

The borehole was drilled in the Rajapalot gold-cobalt prospect located at the northern edge of the Peräpohja Belt, specifically in the Ylitornio and Rovaniemi municipalities of northern Finland. The Rajapalot prospect lays within an approximately 10 $\times$ 10 km gold-enriched area, known as Rompas-Rajapalot, and comprises several gold rich occurrences. The rock types in the area are mostly altered quartzites, carbonate-bearing rocks, and amphibolites [77]. In general, the host rocks consist of an isoclinally folded package of amphibolite facies. This package has been divided into two sequences separated by an interpreted unconformity. A series of siliciclastic, dolomitic carbonate, and albite-altered metasedimentary rocks characterize the first sequence. The second sequence is metasedimentary consisting of pelitic turbidites, arkosic sands, carbonates, various quarzitic sandstones, and sulphidic bituminous rocks [78]. The unconformity between the two sequences is a boundary between largely oxidized rocks from the first sequence and reduced rocks from the second sequence [79]. Two main mineralization styles can be

	FENIX	OWL
Spectral range	463 - 956 nm VNIR 1004 - 2478 nm SWIR	7500 - 11976 nm LWIR
Spectral resolution	3.5 nm 12 nm	100 nm
Spectral band width	3.4 nm 6 nm	48 nm
Number of bands	73 267	94
Pixel size	1.6 mm/pixel	
Scanning speed	170 mm/s @ 1mm pixel size	
Field of view	32.3°	

described: a sulphidic Fe–Mg formation, which includes chlorite, Fe–Mg amphiboles, tourmaline, and pyrrhotite (often associated with quartz veins) in the upper part [80], and a sulphidic K–Fe alteration including muscovite, biotite, chlorite, quartz, albite, Mg–Fe amphiboles, tourmaline, pyrrhotite, scheelite, pyrite, chalcopyrite, bismuth tellurides, gold, and cobaltite [78], [79].

B. Data Acquisition

1) Geochemical Data: The geochemical data available from the drill-core comprise the ppm concentrations of 49 chemical elements: Ag, As, Au, Ba, Be, Bi, Cd, Ce, Co, Cr, Cs, Cu, Ga, Ge, Hf, In, La, Li, Mn, Mo, Nb, Ni, P, Pb, Rb, Re, S, Sb, Sc, Se, Sn, Sr, Ta, Te, Th, Tl, U, V, W, Y, Zn, Zr, Al, Ca, Fe, K, Mg, Na, and Ti. A total of 190 assays were performed covering regions of approximately 1 m long in the suspected Au grade intersections and adjacent regions. The wall rock sections with inferred mineralizations were sampled at lengths between 2 and 3 m.

The geochemical data were provided by Mawson Resources Limited company and were acquired using several analytical methods to cover a variety of elements: atomic absorption spectroscopy, inductively coupled plasma–atomic emission spectroscopy, and fire assay were used to measure gold; inductively coupled plasma–mass spectrometry was used for a multielement work. For more details on these methods, the readers are referred to [81].

2) Hyperspectral Data: For the acquisition of the hyperspectral data, three cameras were used on the SPECIM SisuRock scanner: a high-resolution RGB, a SPECIM FENIX, and a SPECIM OWL cameras. The scanner is an automatic imaging workstation equipped with a mobile tray that carries the drill-core boxes under the field of view of the cameras. The FENIX camera provides coregistered VNIR-SWIR hyperspectral data. It contains two sensors to cover both regions of the electromagnetic spectrum. The OWL is a cooled LWIR camera. Both the FENIX and OWL cameras are push-broom sensors. The pixel size in both the VNIR-SWIR and LWIR hyperspectral data is around 1.6 mm/pixel. More details on the specifications used for the acquisition of the data are shown in Table I.

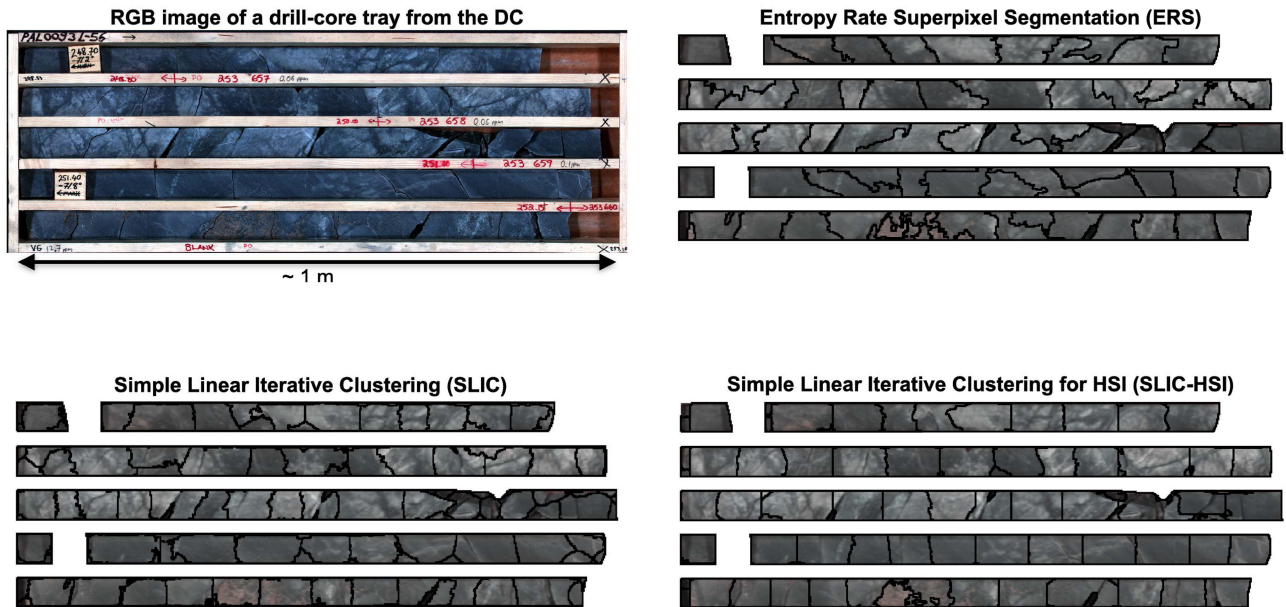**Superpixel segmentation algorithms using 125 superpixels**

Fig. 4. Comparison of superpixel segmentation algorithms: ERS segmentation, SLIC, and SLIC-HSI.

The entire system used for the acquisition of the hyperspectral data is capable of acquiring data from up to 200 boxes per day. For this study, the surface of the drill-core halved with a diamond saw was scanned. The drill-core covers around 325.80 m and is stored in 74 trays.

The preprocessed hyperspectral data were provided by the mining company at a processing level that includes coregistration, radiometric corrections, masking of boxes, and corrections for any possible wavelength shift and dead pixels. Data were available for further processing after a final quality control check.

C. Experimental Setup

To derive the labels from the geochemical data, a collinearity test with tolerance threshold of 0.5 was implemented. We chose this value because it is commonly considered that for variables highly associated, a decisive factor for collinearity relying on high variance-decomposition proportions sets the threshold greater than 0.5 [60]. It means that all the variables with a variance-decomposition proportion exceeding 0.5 are assumed to exhibit collinearity and are discarded. At this stage, ten elements were removed: As, Ce, Ga, Hf, La, Sc, Ta, V, Zr, and Ti. From the PCA, a total of seven principal components were chosen because they contain 99% of the total variance of the data. Finally, the GMM method was implemented to label the geochemical data. For this, the Elbow method estimated the appropriate number of classes, k, as 5.

The superpixel segmentation was performed in two different setups that were chosen after testing different settings and taking special care at the differences in resolution between the hyperspectral and geochemical data and the homogeneity of the superpixels. These two setups consider different spatial resolutions to assess the performance of the segmentation in compensation of the difference between the hyperspectral and geochemical spatial resolution. The first setup, referred from now on as *SP-765*, derives superpixels with higher spatial resolution that are intended to cover almost half of the diameter. Considering that the diameter of the cores used in this study is of approximately 50 mm, the size of the superpixels in this first setup was planned to be about 18×18 mm to also ensure a buffer for possible variations. In the hyperspectral data, the diameter of the cores encompass around 30 pixels, therefore, to achieve the desired superpixel segmentation the ERS algorithm was needed to be set to derive a maximum of 765 superpixels. In the second scenario, *SP-125*, with coarser resolution, the superpixels are intended to cover almost the entire diameter of the core. Therefore, the size of the superpixels was planned to be around 45×45 mm to keep the mentioned buffer. To be able to obtain this, the number of superpixels was set to 125.

To demonstrate the efficiency of the ERS algorithm, Fig. 4 shows a comparison with two versions of the simple linear iterative clustering (SLIC) segmentation algorithm: the traditional SLIC algorithm [82], [83] and a SLIC modification for hyperspectral images (SLIC-HSI) [84]. The comparison is performed using 125 superpixels because this represents a low number of superpixels, but it is still enough to highlight the main patterns in the data. In general, results in Fig. 4 show that SLIC superpixels tend to have sharper edges than the ERS superpixels. Structures such as vein and fractures are better segmented with the ERS algorithm. Moreover, the processing time of the ERS algorithm is quite acceptable, i.e., slightly more than SLIC and significantly less than SLIC-HSI algorithms. To complement the analysis of our proposed method and evaluate the advantages of

TABLE II
TOTAL NUMBER OF SAMPLES AVAILABLE PER CLASS IN THE
TRAINING AND TEST SETS

Class	Total	Training	Test
1	105	10	95
2	49	10	39
3	18	10	8
4	6	5	1
5	12	10	2

the superpixel segmentation, a pixel-based (PB) approach was also considered. For this PB procedure, the mean of the regions from where the geochemical data were available was selected as the representative spectra for the labels.

For the ckSVM algorithm, radial basis function (RBF) kernels were implemented due to the simplicity to calibrate their parameters. An equal weight, $\omega = 0.5$, was given to both kernels, the VNIR-SWIR and the LWIR kernels, to avoid influencing the results according to the mineralogical knowledge known from the samples. Moreover, a traditional SVM and the Random Forest (RF) algorithms were also implemented as a comparison to complement the analysis of our proposed method. For the implementation of SVM and RF, the VNIR-SWIR and LWIR hyperspectral data were integrated at the feature level. Thus, a total of 434 bands were available in the integrated data cube. In the case of SVM, only one kernel was required and an RBF kernel was used. The optimal hyperplane parameters for both SVM and ckSVM were determined using fivefold cross-validation. When tuning the parameters in SVM, cross-validation is commonly chosen over arbitrarily initializing the hyperplane parameters, because it automatizes the algorithm, results are more reliable, and it also overcomes the sensitivity to the regularization parameters characteristic of SVM. The typical number for the cross-validation when tuning an SVM classifier is 5 because with a higher number, the probability of finding the best parameter increases, and this leads to higher computational time, and with a lower number, suboptimal parameters might be selected [85], [86]. Parameters for the RF algorithm were set as in [33] because these parameters showed stable results in our previous experiments when mapping minerals in drill-core hyperspectral data.

As shown in Table II, an equal number of samples per class were considered for the training of the ML models. For all the classes, except for *class 4*, ten samples were used. For *class 4*, only six samples were available in total. Thus, five samples were used for the training of *class 4* and one for testing. The sampling was performed randomly, and the experiment was repeated 15 times to avoid any bias and to obtain statistically sound results.

D. Assessment of the Geochemical Data Labeling

Training labels derived from the processing of the geochemical data described in Section III-A are shown in the left plot of Fig. 5 (*DC*). The white areas represent regions where geochemical data were not available. Seven chemical elements (i.e., Al, Fe, Mg, Na, K, S, and Ca) are used for the visualization of the main composition of each class shown in the bar plots on the

right side of Fig. 5. These are the chemical elements, available in the geochemical data, forming molecules that have the main diagnostic responses in the VNIR-SWIR and LWIR regions of the electromagnetic spectrum. In general, from these seven elements, Al is the most abundant element in all the classes. After Al, Fe tends to be the most dominant except for *class 2*, where K is the second most abundant element. In *classes 1* and *5*, main differences are with the concentration of Na, K, and S. *Classes 3* and *4* have similar patterns for the mineral concentrations, and their occurrence is highly interrelated at around 240–260 m. However, *classes 3* and *4* are dominant in Na and S, respectively. This similar pattern could suggest that *class 4* is located closer to the mineralized region, whereas *class 3* maps the transition toward the more altered regions. From the spider plot in Fig. 5, the relative abundance of each class can be seen, being *class 1* the most abundant (see Table II for more details on the number of samples available per class).

For each of the geochemical measurements and assigned labels, a representative spectrum was obtained from the corresponding hyperspectral data. As an example, Fig. 6 shows the segmentation of one drill-core tray using a maximum of 125 superpixels; the area enclosed in the red boxes is the region where one geochemical assay was performed. The spectra in dashed lines correspond to the mean of each superpixel found in this region. The spectrum in blue is the mean of the region, and the spectrum in red is the weighted mean of the region. For both superpixel segmentation setups, the weighted mean is considered as the representative spectrum. The region enclosed in the red boxes covers from 251.10 to 252.15 m, and the assigned label for the entire region corresponds to *class 1*. For each of the superpixels derived in this region, a class label is then predicted during the classification step.

E. Quantitative and Qualitative Assessment

The models trained using the geochemical and hyperspectral data of the drill-core training samples were used to classify all the remaining hyperspectral data of the drill-core boxes. The classification was performed at the object level (i.e., classifying the superpixels data cube). The drill-core test samples were used to quantitatively assess the different approaches. Accuracies were calculated at the pixel level, to ensure the availability of enough test samples, and are reported in Table III. From these results, it can be seen that, in general, the accuracies per class increase with the use of superpixels, and as it is expected, lowest accuracies are obtained for *class 4* due to the limited number of test samples (see Table II). The overall accuracies also show that the superpixel-based approach outperforms the classification at the PB level. Between the two superpixel segmentation setups, the segmentation using 125 superpixels outperforms the scenario with 765 superpixels. The improvement in the accuracies when using superpixels is because superpixels better compensate the difference in the spatial resolution between the hyperspectral and geochemical data. Accuracies improved especially when using 125 superpixels because the *SP-125* setup is the one that better resembles the resolution of the geochemical data.
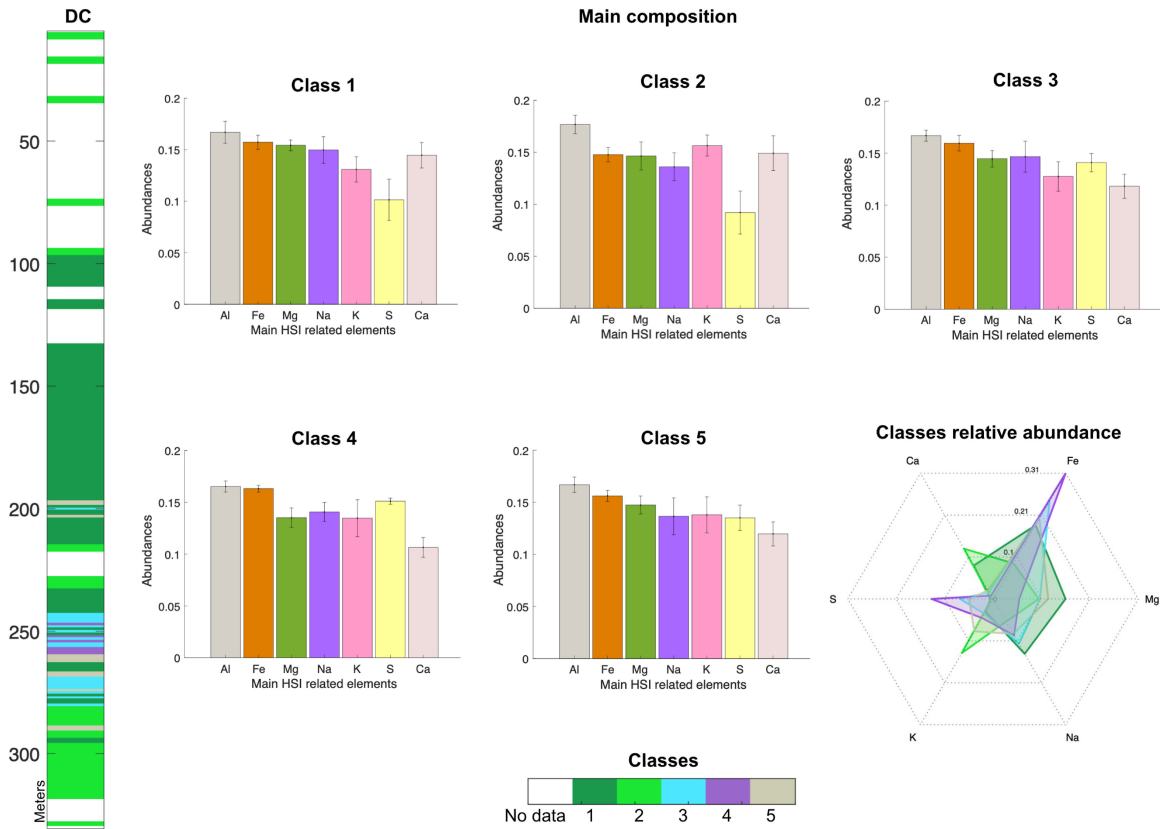
Fig. 5. Labels derived from the available geochemical data of the drill-core (DC). Main composition of each class and their relative abundances are shown in the bar plots and the spider plot. Abundances of the main chemical elements linkable with VNIR-SWIR and LWIR hyperspectral data (HSI) are used for this purpose.
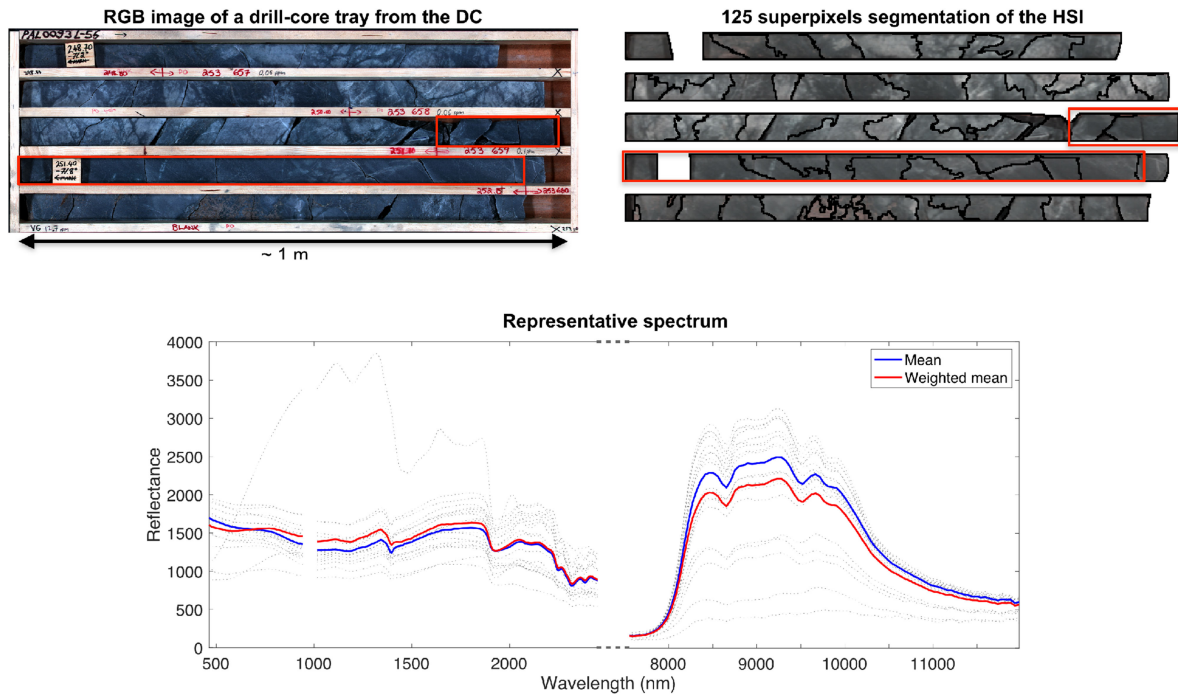

Fig. 6. Example of the selection of the representative spectrum for one of the geochemical assays available. The spectra shown in the plot belong to the area where a geochemical measurement was performed, which is enclosed in the red boxes. The dashed lines correspond to the mean of each superpixel within the red boxes.

TABLE III
CLASSES AND OVERALL ACCURACIES (%)

Classes	RF			SVM			ckSVM		
	PB	SP-765	SP-125	PB	SP-765	SP-125	PB	SP-765	SP-125
1	56.07	57.56	60.08	46.74	45.55	56.56	71.32	**73.98**	72.42
2	56.20	54.31	**58.34**	49.42	41.60	42.48	51.42	49.27	53.60
3	37.26	43.01	45.63	37.91	45.59	46.59	39.10	49.43	62.22
4	24.72	25.09	**32.82**	10.89	11.21	20.95	4.54	9.71	8.52
5	56.26	53.14	62.23	47.96	52.58	61.87	49.36	54.59	**62.47**
Overall	53.98	54.67	58.11	45.81	44.32	51.64	60.99	63.27	**64.92**

Concerning the performance of the algorithms, although the highest accuracies for the class with limited number of test samples (*class 4*) were obtained by RF, in general, ckSVM outperforms RF and SVM in all the tested scenarios, being RF the one performing better after ckSVM. To complement the quantitative analysis of the proposed method retrieving the highest values (i.e., *SP-125*), we report the root-mean-squared error (RMSE) metric. To calculate the RMSE, the average of the element concentrations of each class was weighted by the percentage of each class obtained in the resultant classified maps. These values were compared against the concentrations obtained from the geochemical data of the drill-core for each assay (for more details on the element concentrations of each class, see the *main composition* in Fig. 5). The RMSEs obtained by RF, SVM, and ckSVM are 0.0457, 0.0493, and 0.0399, respectively. These values also confirm the good performance of the proposed method.

To qualitatively evaluate the proposed approach, the resulting maps from the classification of the hyperspectral data were resampled to the resolution of the geochemical data. The most dominant label in the classification maps for each of the sections where geochemical data were available was considered (see an example of the *SP-125* classification in Fig. 7). In general, the reconstructed labels (*HSI max*) show a comparable distribution of the classes, being, for example, *class 2* and *class 1* highly abundant at the bottom and middle of the core, respectively. Between 250 and 300 m, a more structure can be seen in the *HSI max* labels (e.g., *class 5* is alternating with *classes 1* and *2*). However, the relative abundance for each class (*HSI classes 1–5*) shows that the patterns of the derived labels from geochemical data (*DC*) remain visible. To illustrate this, at around 290 m the geochemical data (*DC*) shows *class 2*, whereas the hyperspectral data (*HSI max*) shows *class 1*. However, the *HSI class* plots show that *HSI class 2* is also highly abundant in this region.

The classified maps illustrate that this method provides a good estimation of the chemical elements distribution and relative abundance. As an example, *classes 3–5*, which are the ones with the higher concentration of S, were mapped at the bottom of the drill core (see *HSI max* in Fig. 7), and based on the description of the geological deposit presented in Section IV-A, the mineralization expected to have the higher concentration of S is the second one that is located at the bottom part of the deposit. This second mineralization includes not only pyrrhotite but also pyrite, chalcopyrite, cobaltite, amongst others, which are minerals containing S.

Finally, the *HSI* plot in Fig. 8 reveals the significant performance in the estimation of chemistry in the regions where geochemical data are not available. For example, at around 50 m depth, geochemical assays were not available; however, from the results of the proposed approach, it can be seen that this region is dominated by *classes 1* and *2*. Going more into details, Fig. 8 zooms into examples of the obtained classification maps. The first map (covering from 44.17 to 48.53 m) shows that not only *classes 1* and *2* are present, but also *class 3* has been mapped. Similarly, the second map shows that although from the labeling of the geochemical data (*DC*), this region (around 260 m) belongs to *classes 4* and *5*, the region is not that homogeneous and *classes 1–3* are also present.

V. DISCUSSION

Our proposed superpixel-based approach considers hyperspectral data covering the VNIR-SWIR and LWIR regions of the electromagnetic spectrum; this brings the advantage of identifying a large range of elements and, thus, minerals. For example, *classes 1* and *5* have their main differences in composition with Na and K (see classes *main composition* in Fig. 5). This would show a slight shift in the spectrum absorptions in the SWIR. Nonetheless, these two classes have been mapped individually (see *HSI max* between 150 and 210 m in Fig. 7), which can be attributed to the use of the LWIR data since S has an effect in these wavelengths.

To select the representative hyperspectral sample that corresponds to each geochemical assay, the mean of the regions covering the assays, weighted by the superpixels sizes, was selected. The advantage of this weighted mean relies on the fact that the composition of small grains, which can be reflected in the assays, will not be smoothed in the spectra, and by giving a weight to the mean according to the superpixels, its influence is also perceived. Moreover, in this approach, the classification of the hyperspectral data is also performed at the object level, meaning that the superpixels are classified instead of the pixels. In general, and as it is expected, the superpixel-based approach outperforms the PB approach (see accuracies in Table III), and this is because the geochemical data encompass rather big sections in comparison to the resolution of the pixels in the hyperspectral data. Thus, the superpixels compensate for this difference in the spatial resolution and better approximate the resolution of the geochemical data. More specifically, the *SP-125* setup is generating the most accurate results because the superpixels in this setup are large enough to balance the difference in resolution between the hyperspectral and the geochemical data. Although
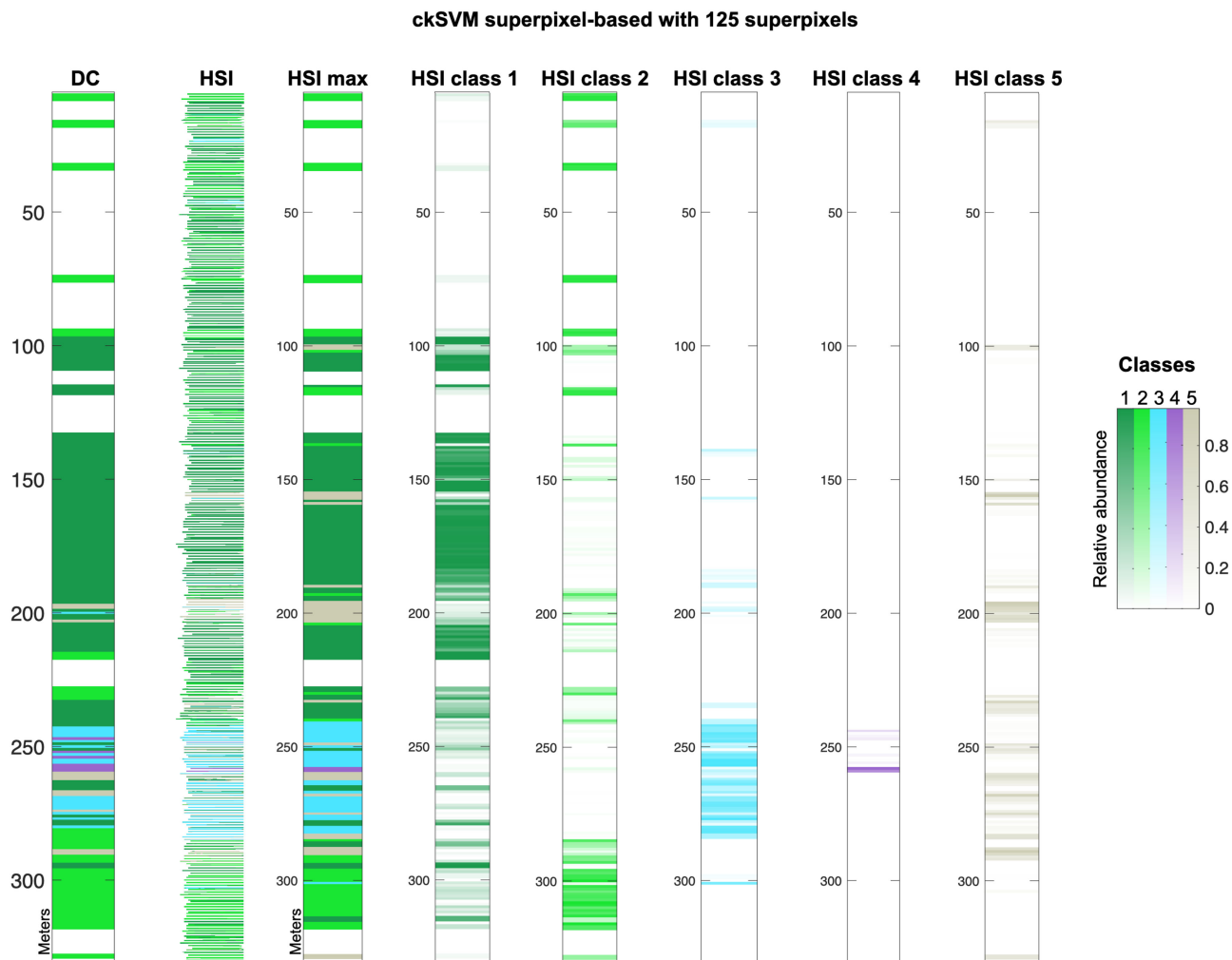
Fig. 7. Classification results of the hyperspectral data by ckSVM using labels derived from the geochemical data and a maximum of 125 superpixels for the segmentation. From left to right: labels derived from the available geochemical data (*DC*), hyperspectral classified map (*HSI*), most dominant labels (*HSI max*) and relative abundances of the five classes (*HSI classes 1–5*) obtained per geochemical assay from the resultant hyperspectral classified maps.

the obtained values for the accuracies are relatively low, these values, in general, represent a good performance considering the low spatial resolution of the geochemical measurements. To validate this performance, we applied the proposed approach on a second bore-hole drilled during the same campaign in the same geological deposit and several meters away from the drill-core presented in this study. Results confirmed that the *SP-125* setup performs most accurately. To classify drill-cores with a different diameter, we suggest performing the superpixel segmentation considering almost the entire diameter of the cores but leaving a buffer in the case of variations in the segmentation, as the *SP-125* setup considers.

Regarding the sampling strategy, the number of test samples remaining in *class 4* affected the class accuracy, having *class 4* the lowest accuracy of all the classes. In general, having ten training samples per class is quite low, which indicates the method capabilities in extrapolating the geochemical data from only a few measurements to the entire drill-core. The drill-core training set was used to train three supervised ML algorithms: RF, SVM, and ckSVM. From the accuracies obtained, it can be seen that

ckSVM outperforms RF and SVM in all the tested scenarios and RF outperforms SVM. A similar pattern has been observed in the results from the second bore-hole, where ckSVM outperforms RF and SVM classification with about 15% more accurate results. Another advantage of the ckSVM is that it permits to adjust the impact of each data source (i.e., VNIR-SWIR or LWIR) in the classification by tuning the weight, ω, of each kernel. This allows considering specific weights for each dataset in case the drill-core can be better evaluated with a specific dataset according to the mineralogical composition of the geological system. For example, in the case of an epithermal system where most of the mineralogy has the diagnostic absorption features in the SWIR region of the electromagnetic spectrum, giving a higher weight to the VNIR-SWIR kernel could highlight relevant patterns.

From the results, it can be seen that performing the geochemical assays covering regions of about 1 m can lead to a biased generalization of the composition of the core. To illustrate this, the second map shown at the right side of Fig. 8 corresponds to an area where geochemical measurements were taken. Here, the
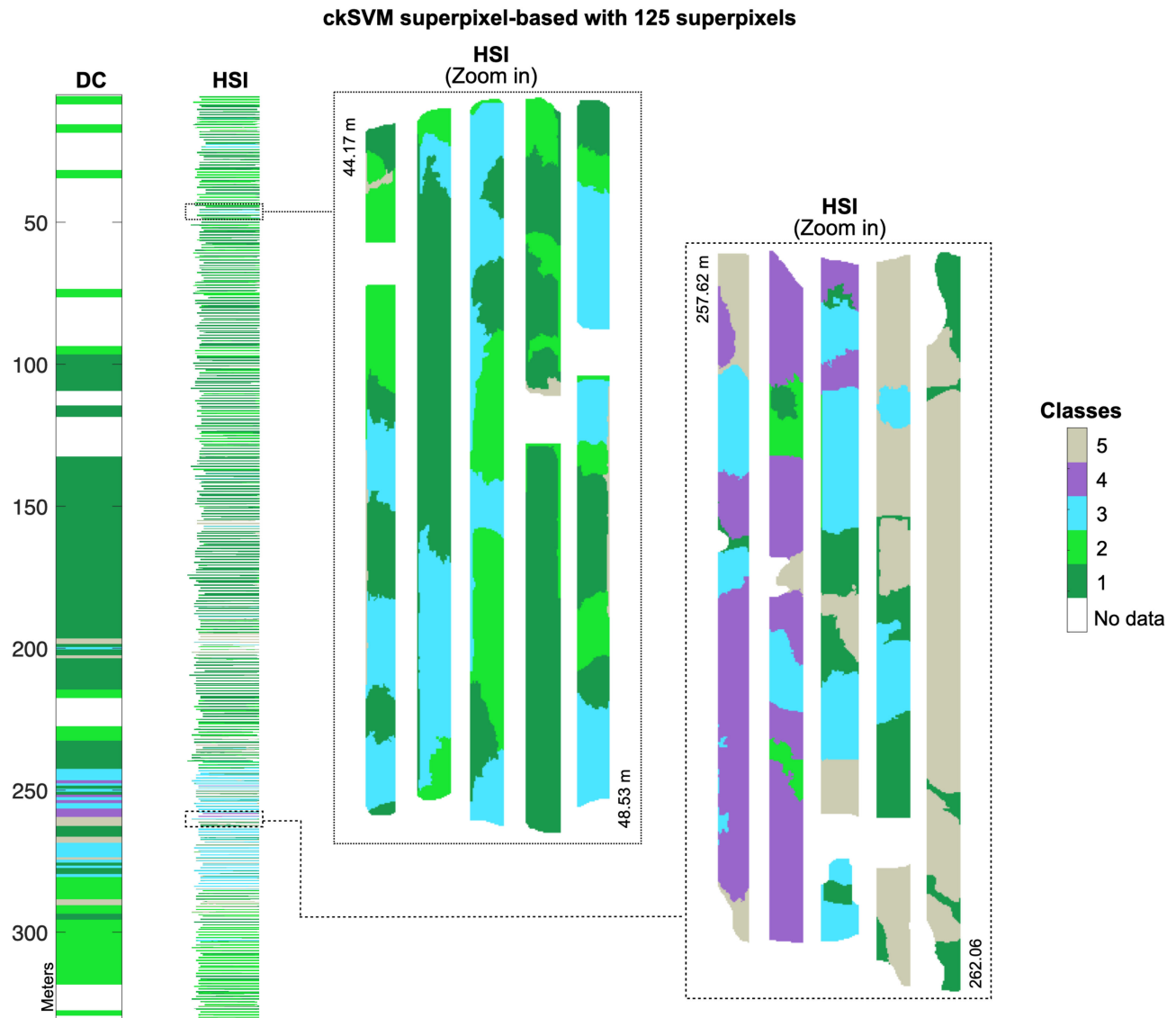
Fig. 8. Zoom to the classification results of the hyperspectral data by ckSVM using labels derived from the geochemical data and a maximum of 125 superpixels for the segmentation. From left to right: labels derived from the available geochemical data (*DC*), hyperspectral classified maps (*HSI*), zoom to maps between 44.17 and 48.53 m and 257.62 and 262.06 m, respectively.

measurements covering the two first drill-core samples (from 257.62 to 259.62 m) were assigned *class 4* in the geochemical data (see *DC* in Fig. 8); however, as can be seen from the hyperspectral classified map, this is not a homogeneous region and *class 4* is not the only class present here. Thus, reducing the size and number of the geochemical assays and having a more selective strategy to perform the measurements can be beneficial. In addition, this highlights the capabilities of hyperspectral data to support the analyses of drill-cores, especially because the classification of one drill-core hyperspectral tray of about 143 640 pixels (i.e., as the data used in this article) takes between 1 and 3 min allowing a faster characterization than the analytical tests. Finally, the proposed approach to integrate drill-core geochemical and hyperspectral data can be implemented on a different set of drill-cores with only taking into consideration the diameter size of the cores and tuning the parameters of the

classification algorithms. However, it is important to keep in mind that a representative amount of samples for each class should be available to avoid a low performance in the classes accuracies as seeing in the obtained classification accuracies of *class 4*.

VI. CONCLUSION

In this article, we developed a new ML-based technique to integrate geochemical and hyperspectral data for the analysis of drill-cores. Our proposed methodology follows a superpixel-based approach where superpixels are considered to tackle the difference between the spatial resolution of the geochemical and hyperspectral data. Labels containing elements abundance composition are derived from the geochemical assays, and a representative sample is obtained from the hyperspectral data

for each geochemical assay. These representative samples and the extracted labels are then used as a training set to build a supervised ML model using ckSVM. This allows upscaling of the detailed information available in geochemical assays to entire bore-holes. Moreover, the developed approach also considers the integration of hyperspectral data covering the VNIR-SWIR and the LWIR regions of the electromagnetic spectrum.

From the experimental results, we conclude that our developed method allows an effective integration of drill-core geochemical and hyperspectral data. With this approach, the composition of the chemical elements available in the geochemical assays can be successfully extrapolated to entire drill cores. Moreover, our proposed method provides an accurate mapping tool, where, due to the integration of VNIR-SWIR and LWIR hyperspectral data, a larger set of chemical elements and, therefore minerals, is considered. This integration enables a more comprehensive and complete analysis since a larger set of elements is being mapped. Moreover, the availability of two kernels, one for the VNIR-SWIR and one for the LWIR, and the possibility of setting different weights for each of them could be beneficial when relevant mineralizations with absorption features in specific parts of the electromagnetic spectrum are known and want to be emphasized. Experimental results show that the upscaling capabilities of our proposed method could be beneficial to support the selection of relevant samples and the area where the more detailed qualitative and quantitative analyses are appropriate. Moreover, this method allows reducing the number of samples needed for such analyses and still supports the geologist and experts with the analysis of the cores. To fully validate the proposed approach, its implementation on a dataset from a different site is suggested.

For our future developments, we will evaluate the possibility and performance of the integration of drill-core geochemical and hyperspectral data using spectral–spatial kernels. As part of this, different possibilities for the extraction of the geochemical labels and the extrapolation of the data will be considered.

REFERENCES

[1] *PDAC Special Edition: World Exploration Trends 2018*, S&P Global Market Intelligence, New York, NY, USA, 2019.

[2] S. Gandhi and B. Sarkar, "Drilling," in *Essentials of Mineral Exploration and Evaluation*, B. S. S. M. Gandhi, Ed. Amsterdam, The Netherlands: Elsevier, 2016, ch. 8, pp. 199–234.

[3] G. Krahenbuhl, P. Hapugoda, K. Warren, and G. O'Brien, "A new method for obtaining detailed mineral information on individual coal particles at the size that they are used in coke making," in *Proc. Bowen Basin Symp.*, 2015, pp. 35–40.

[4] S. Haldar, "Exploration geochemistry," in *Mineral Exploration: Principles and Applications*. Amsterdam, The Netherlands: Elsevier, 2013, ch. 4, pp. 55–71.

[5] W. Nikonow and D. Rammlmair, "Automated mineralogy based on micro-energy-dispersive X-ray fluorescence microscopy (μ-EDXRF) applied to plutonic rock thin sections in comparison to a mineral liberation analyzer," *Geosci. Instrum., Methods Data Syst.*, vol. 6, no. 2, pp. 429–437, 2017.

[6] R. Fandrich, Y. Gu, D. Burrows, and K. Moeller, "Modern SEM-based mineral liberation analysis," *Int. J. Mineral Process.*, vol. 84, nos. 1–4, pp. 310–320, 2007.

[7] D. Layton-Matthews, C. Hamilton, and M. B. Mcclenaghan, "Mineral chemistry: Modern techniques and applications to exploration," in *Application of Indicator Mineral Methods to Mineral Exploration*. Ottawa, ON, Canada: Geological Survey of Canada, 2014.

[8] R. N. Clark, "Spectroscopy of rocks and minerals, and principles of spectroscopy," in *Remote Sensing for the Earth Science: Manual of Remote Sensing*, A. N. Rencz, Ed, vol. 3. New York, NY, USA: Wiley, 1999, ch. 1, pp. 3–58.

[9] F. van der Meer, "Analysis of spectral absorption features in hyperspectral imagery," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 5, no. 1, pp. 55–68, 2004.

[10] E. Littlefield, W. Calvin, P. Stelling, and T. Kent, "Reflectance spectroscopy as a drill core logging technique: An example using core from the Akutan geothermal exploration project," *Geothermal Resour. Council Trans.*, vol. 36, pp. 1283–1291, 2012.

[11] W. M. Calvin and E. L. Pace, "Mapping alteration in geothermal drill core using a field portable spectroradiometer," *Geothermics*, vol. 61, pp. 12–23, 2016.

[12] F. A. Kruse, "Identification and mapping of minerals in drill core using hyperspectral image analysis of infrared reflectance spectra," *Int. J. Remote Sens.*, vol. 17, no. 9, pp. 1623–1632, 1996.

[13] G. S. Taylor, "Mineral and lithology mapping of drill core pulps using visible and infrared spectrometry," *Natural Resour. Res.*, vol. 9, no. 4, pp. 257–268, 2000.

[14] G. Bonifazi, N. Picone, and S. Serranti, "Ore minerals textural characterization by hyperspectral imaging," in *Proc. SPIE*, vol. 8655, 2013, Art. no. 865510.

[15] A. J. Mauger, J. L. Keeling, and J. F. Huntington, "Alteration mapping of the Tarcoola Goldfield (South Australia) using a suite of hyperspectral methods," *Appl. Earth Sci.*, vol. 116, no. 1, pp. 2–12, 2007.

[16] E. F. Littlefield and W. M. Calvin, "Geothermal exploration using imaging spectrometer data over Fish Lake Valley, Nevada," *Remote Sens. Environ.*, vol. 140, pp. 509–518, 2014.

[17] F. A. Kruse, R. L. Bedell, J. V. Taranik, W. A. Peppin, O. Weatherbee, and W. M. Calvin, "Mapping alteration minerals at prospect, outcrop and drill core scales using imaging spectrometry," *Int. J. Remote Sens.*, vol. 33, no. 6, pp. 1780–1798, 2012.

[18] N. Xu, Y.-X. Hu, B. Lei, Y.-T. Hong, and F.-X. Dang, "Mineral information extraction for hyperspectral image based on modified spectral feature fitting algorithm," *Spectroscopy Spectral Anal.*, vol. 31, no. 6, pp. 1639–1643, Jun. 2011.

[19] E. Bedini, F. van der Meer, and F. van Ruitenbeek, "Use of HyMap imaging spectrometer data to map mineralogy in the Rodalquilar caldera, Southeast Spain," *Int. J. Remote Sens.*, vol. 30, no. 2, pp. 327–348, 2009.

[20] C. Kratt, W. M. Calvin, and M. F. Coolbaugh, "Mineral mapping in the Pyramid Lake basin: Hydrothermal alteration, chemical precipitates and geothermal energy potential," *Remote Sens. Environ.*, vol. 114, no. 10, pp. 2297–2304, 2010.

[21] F. J. A. Van Ruitenbeek *et al.*, "Mapping the wavelength position of deepest absorption features to explore mineral diversity in hyperspectral images," *Planetary Space Sci.*, vol. 101, pp. 108–117, 2014.

[22] T. J. Roache *et al.*, "Epidote-clinozoisite as a hyperspectral tool in exploration for Archean gold," *Australian J. Earth Sci.*, vol. 58, no. 7, pp. 813–822, 2011.

[23] M. Mathieu, R. Roy, P. Launeau, M. Cathelineau, and D. Quirt, "Alteration mapping on drill cores using HySpex SWIR-320m hyperspectral camera: Application to the exploration of an unconformity-related uranium deposit (Saskatchewan, Canada)," *J. Geochem. Exploration*, vol. 172, pp. 71–88, 2017.

[24] M. Tappert, B. Rivard, D. Giles, R. Tappert, and A. Mauger, "Automated drill core logging using visible and near-infrared reflectance spectroscopy: A case study from the Olympic Dam IOCG deposit, South Australia," *Econ. Geol.*, vol. 106, no. 2, pp. 289–296, 2011.

[25] P. Lypaczewski *et al.*, "Using hyperspectral imaging to vector towards mineralization at the Canadian Malartic gold deposit, Québec, Canada," *Ore Geol. Rev.*, vol. 111, 2019, Art. no. 102945.

[26] A. Rodger, C. Laukamp, M. Haest, and T. Cudahy, "A simple quadratic method of absorption feature wavelength estimation in continuum removed spectra," *Remote Sens. Environ.*, vol. 118, pp. 273–283, 2012.

[27] F. van der Meer, V. Kopačková, L. Koucká, H. M. van der Werff, F. J. van Ruitenbeek, and W. H. Bakker, "Wavelength feature mapping as a proxy to mineral chemistry for investigating geologic systems: An example from the Rodalquilar epithermal system," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 64, pp. 237–248, 2018.

[28] R. Kokaly *et al.*, "USGS spectral library version 7: U.S. geological survey data series 1035," 2017, p. 61, doi: 10.3133/ds1035.

[29] J. F. Huntington *et al.*, "Automated mineralogical core logging at the Emmie Bluff iron oxide- copper- gold Prospect," *Mesa J.*, vol. 41, pp. 38–44, 2006.

[30] S. Schneider, R. J. Murphy, and A. Melkumyan, "Evaluating the performance of a new classifier—The GP-OAD: A comparison with existing methods for classifying rock type and mineralogy from hyperspectral imagery," *ISPRS J. Photogrammetry Remote Sens.*, vol. 98, pp. 145–156, 2014.

[31] L. Tusa *et al.*, "Mineral mapping and vein detection in hyperspectral drill-core scans: Application to porphyry-type mineralization," *Minerals*, vol. 9, no. 2, pp. 1–23, 2019.

[32] D. Wang *et al.*, "Automated vein detection for drill core analysis by fusion of hyperspectral and visible image data," in *Proc. 23rd Int. Conf. Mechatronics Mach. Vis. Practice*, 2016, pp. 1–6.

[33] I. C. C. Acosta, M. Khodadadzadeh, L. Tusa, P. Ghamisi, and R. Gloaguen, "A machine learning framework for drill-core mineral mapping using hyperspectral and high-resolution mineralogical data fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 4829–4842, Dec. 2019.

[34] M. Khodadadzadeh and R. Gloaguen, "Upscaling high-resolution mineralogical analyses to estimate mineral abundances in drill core hyperspectral data," in *Proc. Int. Geosci. Remote Sens. Symp.*, 2019, pp. 1845–1848.

[35] B. Koirala, M. Khodadadzadeh, C. Contreras, Z. Zahiri, R. Gloaguen, and P. Scheunders, "A supervised method for nonlinear hyperspectral unmixing," *Remote Sens.*, vol. 11, no. 20, 2019, Art. no. 2458.

[36] E. J. Hill, J. Robertson, and Y. Uvarova, "Multiscale hierarchical domaining and compression of drill hole data," *Comput. Geosci.*, vol. 79, pp. 47–57, 2015.

[37] E. J. Hill and Y. Uvarova, "Identifying the nature of lithogeochemical boundaries in drill holes," *J. Geochem. Exploration*, vol. 184, pp. 167–178, 2018.

[38] L. Jackson *et al.*, "Integrating hyperspectral analysis and mineral chemistry for geoenvironmental prediction," in *Proc. 11th ICARD IMWA MWD Conf.—"Risk to Opportunity*," 2016, pp. 1075–1080.

[39] A. Petrovic, S. D. Khan, and A. K. Thurmond, "Integrated hyperspectral remote sensing, geochemical and isotopic studies for understanding hydrocarbon-induced rock alterations," *Marine Petroleum Geol.*, vol. 35, no. 1, pp. 292–308, 2012.

[40] S. Haldar, "Geology and geochemistry," in *Platinum-Nickel-Chromium Deposits*. Amsterdam, The Netherlands: Elsevier, 2017, ch. 2, pp. 37–61.

[41] M. I. Leybourne, S. Pontual, and J. M. Peter, "Integrating hyperspectral mineralogy, mineral chemistry, geochemistry and geological data at different scales in iron ore mineral exploration," in *Proc. Iron Ore Conf.*, Perth, WA, Australia, 2013, pp. 1–10.

[42] L. Sun, S. Khan, and P. Shabestari, "Integrated hyperspectral and geochemical study of sediment-hosted disseminated gold at the Goldstrike District, Utah," *Remote Sens.*, vol. 11, no. 17, 2019, Art. no. 1987.

[43] N. Fox, A. Parbhakar-Fox, J. Moltzen, S. Feig, K. Goemann, and J. Huntington, "Applications of hyperspectral mineralogy for geoenvironmental characterisation," *Minerals Eng.*, vol. 107, pp. 63–77, 2016.

[44] A. Parbhaker-Fox, B. Lottermoser, and D. J. Bradshaw, "Cost-effective means for identifying acid rock drainage risks integration of the geochemistry- mineralogy-texture approach and geometallurgical techniques," in *Proc. 2nd AusIMM Int. Geometallurgy Conf.*, 2013, pp. 143–154.

[45] B. Rivard, N. B. Harris, J. Feng, and T. Dong, "Inferring TOC and major element geochemical and mineralogical characteristics of shale core from hyperspectral imagery," *AAPG Bull.*, vol. 102, no. 10, pp. 2101–2121, 2017.

[46] T. Lu, S. Li, L. Fang, X. Jia, and J. A. Benediktsson, "From subpixel to superpixel: A novel fusion framework for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4398–4411, Aug. 2017.

[47] F. Kruse, "Combined SWIR and LWIR mineral mapping using MASTER/ASTER," in *Proc. IEEE Geosci. Remote Sens. Symp.*, 2002, pp. 2267–2269.

[48] I. C. Acosta, M. Khodadadzadeh, L. Tusa, and R. Gloaguen, "Fusion of VNIR-SWIR and LWIR for mineral mapping in a machine learning framework," *Geophys. Res. Abstr.*, vol. 21, 2019, Art. no. 9092.

[49] V. Kopcková and L. Koucká, "Integration of absorption feature information from visible to longwave infrared spectral ranges for mineral mapping," *Remote Sens.*, vol. 9, no. 10, pp. 8–13, 2017.

[50] G. Notesco, Y. Ogen, and E. Ben-Dor, "Integration of hyperspectral shortwave and longwave infrared remote-sensing data for mineral mapping of Makhtesh Ramon in Israel," *Remote Sens.*, vol. 8, no. 4, 2016, Art. no. 318.

[51] S. Lorenz *et al.*, "Multi-sensor spectral imaging of geological samples: A data fusion approach using spatio-spectral feature extraction," *Sensors*, vol. 19, no. 12, 2019, Art. no. 2787.

[52] G. Camps-Valls, L. Gomez-Chova, J. Muñoz-Marí, J. Vila-Francés, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 93–97, Jan. 2006.

[53] J. Li, P. R. Marpu, A. Plaza, J. M. Bioucas-Dias, and J. A. Benediktsson, "Generalized composite kernel framework for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9, pp. 4816–4829, Sep. 2013.

[54] M. Khodadadzadeh, A. Cuartero, J. Li, A. Felicisimo, and A. Plaza, "Fusion of hyperspectral and lidar data using generalized composite kernels: A case study in Extremadura, Spain," in *Proc. Int. Geosci. Remote Sens. Symp.*, 2015, pp. 61–64.

[55] D. M. Shaw and J. D. Bankier, "Statistical methods applied to geochemistry," *Geochimica et Cosmochimica Acta*, vol. 5, pp. 111–123, 1954.

[56] A. Buccianti and E. Grunsky, "Compositional data analysis in geochemestry: Are we sure to see what really occurs during natural processes?" *J. Geochem. Exploration*, vol. 141, pp. 1–5, 2014.

[57] S. Pontual, N. Merry, and P. Gamson, *Spectral Interpretation Field Manual*, AusSpec Int. Pty., Queenstown, New Zealand, vol. 1, 1997, pp. 1–169.

[58] J. Salisbury, L. S. Walter, and N. Vergo, *Mid-Infrared (2.1–25 μm) Spectral of Minerals*, 1st ed. Reston, VA, USA: U.S Geological Survey, 1987.

[59] D. A. Belsley, E. Kuh, and R. E. Welsch, *Regression Diagnostics: Identifying Influential Data and Sourcers of Collinearity*. Hoboken, NJ, USA: Wiley, 1980.

[60] D. A. Belsley, "A Guide to using the collinearity diagnostics," *Comput. Sci. Econ. Manage.*, vol. 4, no. 1, pp. 33–50, 1991.

[61] M. Kucera and B. Malmgren, "Logratio transformation of compositional data: A resolution of the constant sum constraint," *Marine Micropaleontology*, vol. 34, pp. 117–120, 1982.

[62] J. Aitchison, "The statistical analysis of compositional data," *J. Roy. Statist. Soc. Ser. B*, vol. 44, no. 2, pp. 139–177, 1982.

[63] J. J. Egozcue, V. Pawlowsky-Glahn, G. Mateu-Figueras, and C. Barceló-Vidal, "Isometric logratio for compositional data analysis," *Math. Geol.*, vol. 35, no. 3, pp. 279–300, 2003.

[64] I. T. Jolliffe, *Principal Component Analysis*, vol. 98. Berlin, Germany: Springer-Verlag, 2002.

[65] D. Ketchen and C. L. Shook, "The application of cluster analysis in strategic management research: An analysis and critique," *Strategic Manage. J.*, vol. 17, no. 6, pp. 441–458, 2002.

[66] C. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer-Verlag, 2006.

[67] G. J. McLachlan, S. X. Lee, and S. I. Rathnayake, "Finite mixture models," *Annu. Rev. Statist. Appl.*, vol. 6, no. 1, pp. 355–378, 2019.

[68] T. Veracini, S. Matteoli, M. Diani, and G. Corsini, "Fully unsupervised learning of Gaussian mixtures for anomaly detection in hyperspectral imagery," in *Proc. 9th Int. Conf. Intell. Syst. Des. Appl.*, 2009, pp. 596–601.

[69] M.-Y. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa, "Entropy rate superpixel segmentation," in *Proc. IEEE Comput. Vis. Pattern Recognit.*, 2011, pp. 2097–2104.

[70] D. Stutz, A. Hermans, and B. Leibe, "Superpixels: An evaluation of the state-of-the-art," *Comput. Vis. Image Understanding*, vol. 166, pp. 1–27, 2018.

[71] J. Jiang, J. Ma, C. Chen, Z. Wang, Z. Cai, and L. Wang, "SuperPCA: A superpixelwise PCA approach for unsupervised feature extraction of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4581–4593, Aug. 2018.

[72] B. Schölkopf and A. J. Smola, *Learning with Kernels*. Cambridge, MA, USA: MIT Press, 2002.

[73] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Lean.*, vol. 20, pp. 273–297, 1995.

[74] V. N. Vapnik, "An overview of statistical learning theory," *IEEE Trans. Neural Netw.*, vol. 10, no. 5, pp. 988–999, Sep. 1999.

[75] P. Ghamisi, J. Plaza, Y. Chen, J. Li, and A. J. Plaza, "Advanced spectral classifiers for hyperspectral images: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 1, pp. 8–32, Mar. 2017.

[76] B. Waske, J. A. Benediktsson, K. Arnason, and J. Sveinsson, "Mapping of hyperspectral AVIRIS data using machine-learning algorithms," *Can. J. Remote Sens.*, vol. 35, pp. S106–S116, 2009.

[77] J.-p. Ranta, E. Hanski, and N. Cook, "Petrography and mineral geochemistry of the paleo- proterozoic rajapalot au mineralization, Peräpohja Belt, Northern Finland," in *Proc. 13th Biennial SGA Meeting*, 2015, pp. 193–196.

[78] N. Cook and M. Hudson, "Progress report on the geology, mineralization and exploration activities on The Rompas-Rajapalot Gold-Cobalt Project, Peräpohja Belt," Mawson Resources Ltd., Lapland, Finalnd., Aug. 2018. [Online]. Available: http://www.mawsonresources.com/projects/fonland/ni43-101-technical-reports

[79] N. Taipale, "Mineralogy of the Paleoproterozoic Raja Au-Co Prospect, Northern Finland," Master's thesis, Faculty Technol., Univ. Oulu, Oulu, Finland, 2018.

[80] F. Molnár, H. O'Brien, H. Stein, and N. Cook, "Geochronology of hydrothermal processes leading to the formation of the Au–U mineralization at the Rompas Prospect, Peräpohja Belt, Northern Finland: Application of paired U–Pb dating of uraninite and ReOs dating of molybdenite to the identification of M," *Minerals*, vol. 7, no. 171, pp. 1–23, 2017.

[81] N. Petrović, D. Budelan, S. Cokić, and B. Nešić, "The determination of the content of gold and silver in geological samples," *J. Serbian Chem. Soc.*, vol. 66, no. 1, pp. 45–52, 2001.

[82] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2281, Nov. 2012.

[83] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels," EPFL, Lausanne, Switzerland, Tech. Rep. 149 300, Jun. 2010.

[84] X. Zhang, S. E. Chew, Z. Xu, and N. D. Cahill, "SLIC superpixels for efficient graph-based dimensionality reduction of hyperspectral imagery," *Proc. SPIE*, vol. 9472, 2015, Art. no. 947209.

[85] J. A. Benediktsson and P. Ghamisi, *Spectral and Spatial Classification of Hyperspectral Remote Sensing Images*. Boston, MA, USA: Artech House, 2015.

[86] P. Gurram and H. Kwon, "Sparse kernel-based ensemble learning with fully optimized kernel parameters for hyperspectral classification problems," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 787–802, Feb. 2013.

Mahdi Khodadadzadeh received the B.S. degree in electrical engineering, electronics, from the Sadjad University of Technology, Mashhad, Iran, in 2008, the M.Sc. degree in electrical engineering, communications, from Tarbiat Modares University, Tehran, Iran, in 2011, and the Ph.D. degree in information technology and communications with the Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, Cáceres, Spain, under the joint supervision of Prof. A. Plaza and Prof. J. Li, in 2015.

From 2015 to 2017, he was a Postdoctoral Researcher with the Remote Sensing Laboratory, the Department of Information Engineering and Computer Science, University of Trento, Trento, Italy. Since 2017, he has been a Machine Learning Specialist with the Exploration Technology Division, Helmholtz Institute Freiberg for Resource Technology, Freiberg, Germany, where he is leading the machine learning group activities on hyperspectral drill-core data analysis. His research interests include machine learning and image processing with application to remote sensing data analysis, and particular emphasis on hyperspectral data classification.

Raimon Tolosana-Delgado received the M.S. degree in engineering geology from the University of Barcelona, Barcelona, Spain, and the Polytechnic University of Catalonia, Barcelona, in 2001, and the Ph.D. degree in environmental engineering and physics from the University of Girona, Girona, Spain, in 2005.

He currently leads the working group on primary resources modeling with the Helmholtz Institute Freiberg for Resource Technology/Helmholtz-Zentrum Dresden-Rossendorf, Freiberg, Germany. His research interests include geostatistics, data-model merging techniques, compositional data analysis, Bayesian statistics, and R programming, in particular applied to the geosciences.

Isabel Cecilia Contreras Acosta received the B.Sc.(Hons.) degree in geological engineering from the University of Los Andes, Mérida, Venezuela, in 2013, and the M.Sc. degree in geoinformation science and earth observation for applied earth science with a specialization in natural hazards, risk, and engineering from the ITC Faculty, University of Twente, Enschede, The Netherlands, in 2017. She is currently working toward the Ph.D. degree with the Machine Learning Group, Exploration Department, Helmholtz Institute Freiberg for Resource Technology—Helmholtz-Zentrum Dresden-Rossendorf, Freiberg, Germany.

Her research interests include machine learning techniques for hyperspectral data analysis applied to mineral exploration.

Richard Gloaguen received the Ph.D. degree (*Doctor Communitatis Europae*) in marine geosciences from the University of Western Brittany, Brest, France, in collaboration with the Royal Holloway University of London, London, U.K., and University of Göttingen, Göttingen, Germany, in 2000.

He was a Marie Curie Postdoctoral Research Associate with the Royal Holloway University of London from 2000 to 2003. He led the Remote Sensing Group, University Bergakademie Freiberg, Freiberg, Germany, from 2003 to 2013. Since 2013, he has been leading the Division of Exploration Technology, Helmholtz-Institute Freiberg for Resource Technology, Freiberg. He is currently involved in unmanned-aerial-vehicle-based multisource imaging, laser-induced fluorescence, and noninvasive exploration. His research interests include multisource and multiscale remote sensing integration.