Revision 3

Identifying serpentine minerals by their chemical compositions with machine learning

Word Count: 8454

Shichao Ji[1,2], Fang Huang[3*], Shaoze Wang[4], Priyantan Gupta[5], William Seyfried, Jr[6], Hejia Zhang[7], Xu Chu[8], Wentao Cao[9], J ZhangZhou[10]

[1] CAS Key Laboratory of Mineralogy and Metallogeny/Guangdong provincial Key Laboratory of Mineral Physics and Materials, Guangzhou Institute of Geochemistry, Chinese Academy of Sciences (CAS), Guangzhou 510640, China

2 CAS Center for Excellence in Deep Earth Science, Guangzhou 510640, China

[3] CSIRO Mineral Resources, Kensington, WA 6151, Australia

[4] Surface Science Lab, Physics and Astronomy Department, Ohio University, Athens, OH 74501, USA

[5] Indian Institute of Technology, Kharagpur, West Bengal 721302, India

[6] Department of Earth and Environmental Sciences, University of Minnesota, Minneapolis, MN 55455, USA

[7] School of the Environment, Yale University, New Haven, CT 06511, USA

[8] Department of Earth Sciences, University of Toronto, Toronto, ON M5S 3B1, Canada

[9] Department of Geology and Environmental Sciences, State University of New York at Fredonia, Fredonia, NY 14063, USA

22 [10] School of Earth Sciences, Zhejiang University, Hangzhou 310058, China

23 * **E-mail:** *f.huang@csiro.au*

24

This is the peer-reviewed, final accepted version for American Mineralogist, published by the Mineralogical Society of America.
The published version is subject to change. Cite as Authors (Year) Title. American Mineralogist, in press.
DOI: https://doi.org/10.2138/am-2022-8688. http://www.minsocam.org/

ABSTRACT

The three main serpentine minerals, chrysotile, lizardite, and antigorite, form in various geological settings, and have different chemical compositions and rheological properties. The accurate identification of serpentine minerals is thus of fundamental importance to understanding global geochemical cycles and the tectonic evolution of serpentine-bearing rocks. However, it is challenging to distinguish specific serpentine species solely based on geochemical data obtained by traditional analytical techniques. Here, we apply machine learning approaches to classify serpentine minerals based on their chemical compositions alone. Using the Extreme Gradient Boosting (XGBoost) algorithm, we trained a classifier model (overall accuracy of 87.2%) that is capable of distinguishing between low-temperature (chrysotile and lizardite) and high-temperature (antigorite) serpentines mainly based on their SiO_2, NiO, and Al_2O_3 contents. We also utilized a *k*-means model to demonstrate that the tectonic environment in which serpentine minerals form correlates with their chemical compositions. Our results obtained by combining these classification and clustering models imply the increase of Al_2O_3 and SiO_2 contents and the decrease of NiO content during the transformation from low- to high-temperature serpentine (i.e., lizardite and chrysotile to antigorite) under greenschist–blueschist conditions. These correlations can be used to constrain mass transfer and the surrounding environments during the subduction of hydrated oceanic crust.

Keywords: Serpentine, machine learning, XGBoost, classifications, *k*-means, clustering

INTRODUCTION

The hydration of upper mantle rocks and metamorphism are the two predominate processes producing serpentine minerals (John et al. 2011; Plümper et al. 2017; Tamblyn et al. 2019; Zhang et al. 2019; Ulrich et al. 2020), resulting in the wide occurrence of serpentines in various geological settings (Rüpke et al. 2004; Reynard 2013; Canales et al. 2017). Lizardite and chrysotile are produced during near-surface metasomatism or alteration, corresponding to greenschist facies or lower-grade metamorphism (Auzende et al. 2006; Koutsovitis 2017; Boskabadi et al. 2020), whereas antigorite dominates in blueschist- and eclogite-facies ultramafic rocks, which usually represent subduction contexts at depths exceeding 75 km (Hattori and Guillot 2007; Zhang et al. 2009, 2019; Tamblyn et al. 2019). Antigorite is also less hydrous and contains more minor fluid-mobile elements than lizardite and chrysotile (Deschamps et al. 2010; Debret et al. 2019). Antigorite can also endure greater shear strain than lizardite or chrysotile, indicating the different rheological properties of the serpentine minerals (Hirauchi and Katayama 2013; Plissart et al. 2019). Therefore, distinguishing serpentine minerals from one another is of fundamental importance for constraining geothermal gradients and mass transfers, as well as for understanding the mechanism of strain localization, which may induce exhumation, deformation, and seismicity in subduction zones (Boudier et al. 2010; Deschamps et al. 2010; Reynard 2013; Schwartz et al. 2013; Rouméjon et al. 2019; Wheat et al. 2020).

Serpentine minerals are hydrous, Mg-rich, 1:1 (i.e., one tetrahedral silicate sheet, mainly Si^{4+} occupying the tetrahedral sites, and one octahedral sheet, mainly Mg^{2+} occupying the octahedral sites, stack alternately) trioctahedral clay minerals with the ideal composition $Mg_3Si_2O_5(OH)_4$. Because of their chemical compositional differences, the tetrahedral sheets are slightly smaller than the octahedral sheets (Whittaker and Wicks 1970), and different sheet

arrangements accommodate this mismatch, producing the different structures of the serpentine minerals, with lizardite, chrysotile, and antigorite being the three most abundant serpentine minerals (Auzende et al. 2002; Boudier et al. 2010; Deschamps et al. 2013). Common substitutions occurring in natural samples are Fe^{3+}, Ti^{4+}, Mn^{2+}, or Cr^{3+} for Mg^{2+} in the octahedral sites and Al^{3+} for Si^{4+} in the tetrahedral sites (Evans 2004).

Modern techniques used to discriminate between serpentine minerals present specific advantages and disadvantages. Serpentine minerals can be distinguished based on structural differences (Mellini 1986; Banfield et al. 1995; Schwartz et al. 2013; Rouméjon et al. 2019), generally by transmission electron microscopy (TEM), requiring important sample preparation. Serpentine minerals can also be identified from their Raman spectra (Rinaudo et al. 2003; Petriglieri et al. 2015). Auzende et al. (2004) and Groppo et al. (2006) found that antigorite can be identified from the frequencies at which Si-O-Si and the inner and outer OH stretchings occur, and the translation modes of OH-Mg-OH linkages (vibrating at 500–550 cm^{-1}) can be used to distinguish between chrysotile and lizardite. To identify serpentine based on these vibrational differences, the serpentine species being analyzed should be well-crystallized and sufficiently compositionally distinct. However, serpentine minerals may share the same chemical compositions via substitutions (e.g., the substitution of Ni, Fe, and Al for Mg in octahedral sheets and Al for Si in tetrahedral sheets; O'Hanley and Dyar 1998), and poorly crystallized serpentine minerals are common. The associated shifts of the vibrational frequencies of these bonds can result in different serpentine minerals sharing similar Raman spectra (Wang et al. 2015). Therefore, it may be difficult to reliably distinguish poorly crystallized serpentine minerals based only on their chemical compositions and spectroscopic features using traditional geological methodologies.

Recent studies suggest that machine learning (ML) approaches offer new opportunities for classifying minerals based on their chemical compositions in high-dimensional space (Petrelli and Perugini 2016; O'Sullivan et al. 2020; Valetich et al. 2021; Wang et al. 2021). Because they are driven purely by data, ML models can unravel complexities in large datasets through a learning process unimpeded by *a priori* defined conceptual models (e.g., the definition of an equilibrium state; Hazen et al. 2019; Caricchi et al. 2020; Petrelli et al. 2020).

Here, we tested the potential of ML approaches as an alternative tool for classifying serpentine minerals. We first compiled and filtered a dataset comprising eight elemental concentrations in serpentine minerals. Then, we trained the Extreme Gradient Boosting (XGBoost) classification and *k*-means clustering models, employing strategies to avoid overfitting. XGBoost was applied to classify the serpentine minerals at first. Based on the results of these models, we discussed the classification performance and feature importance, which can be used to explore the features dominating the chemical differences between serpentines crystallized in natural systems. Then, the *k*-means algorithm was used to cluster each class. We tested the clustering results to identify the tectonic environments in which serpentine minerals form and to constrain the typical variations in SiO_2, NiO, and Al_2O_3 contents during the replacement of chrysotile/lizardite by antigorite in subduction zones.

DATA AND METHODS

Data Compilation

The dataset used in this study contains compositional data for serpentine-group minerals from 50 publications (see Supplementary Reference List), containing 1,375 chemical analyses of 566 serpentine phases. To explore the relationship between the serpentine minerals and their formation

conditions, we grouped the minerals as low-temperature, (L_Srp, 534 analyses of 264 lizardite and chrysotile samples) *vs*, high-temperature serpentines (H_Srp, 841 analyses of 302 antigorite samples; Supplementary Table S2). Lizardite has flat, stretched tetrahedral sheets with interlayer H-bonds, favoring the coupled substitution of Fe, Ni, and Al for Mg in the octahedral sheets and Al for Si in the tetrahedral sheets (Mellini 1986; Viti and Mellini 1997). Chrysotile forms multi-walled nanotubes or nanoscrolls (Wicks and Whittaker 1975). Antigorite displays curved, wavy layers similar to Roman roof tiles (e.g., Banfield et al. 1995; Ji et al. 2018): the octahedral sheets are continuous and wavy, whereas the tetrahedral sheets undergo periodic reversals along the *a*-axis to connect to the concave half-waves of adjacent octahedral sheets (Capitani and Mellini 2004). These reversals bind the antigorite layers through strong, mainly covalent Si–O bonds (Evans et al. 2013) (Fig. S1).

The samples in our dataset are mainly from three geological settings: ophiolites, modern seafloor hydrothermal systems, and subduction zones (e.g., Sachan et al. 2007; Zhang et al. 2019; Majumdar et al. 2020). The serpentine minerals within these samples have been accurately characterized by X-ray diffraction, Raman spectroscopy, scanning electron microscopy, and/or TEM (e.g., O'Hanley and Wicks 1995; Rouméjon et al. 2019; Shen et al. 2020), and their compositions determined by electron microprobe analysis, X-ray fluorescence, and/or energy dispersive spectrometry (e.g., Hirose et al. 2006; Wu et al. 2018). The concentrations of 12 oxides (SiO_2, TiO_2, Al_2O_3, Cr_2O_3, FeO_{tot} [total iron expressed as FeO], MnO, NiO, MgO, CaO, Na_2O, K_2O, and H_2O) are commonly reported for the major chemical compositions of serpentine minerals. However, CaO, Na_2O, and K_2O contents are negligible because the large ionic radii of Ca^{2+}, Na^{+}, and K^{+} make their substitutions for smaller cations (Mg^{2+} and Si^{4+}) almost impossible. Thus, the occurrence of CaO, Na_2O, and K_2O contents in these publications is most likely due to impure

materials (e.g., the presence of pyroxene), and we excluded these three oxides from our processed dataset (Supplementary Table S2). We also excluded H_2O content because most studies did not report loss on ignition for serpentine minerals.

Data Processing

To train the classification algorithm, we applied the XGBoost algorithm directly to the unprocessed, labeled (L_Srp *vs*. H_Srp) dataset because it can handle missing data. However, we did some processing to the original dataset before applying the *k*-means clustering (Fig. 1). (1) Missing data imputation: missing data from the mineralogically labeled (L_Srp *vs*. H_Srp) unprocessed dataset were imputed by the *k*-nearest neighbor (KNN) algorithm (Gheyas and Smith 2010). (2) Centered Log-Ratio (CLR) transformation: like all other compositional data, our dataset suffers from data closure problems due to the multivariate nature of geochemical data (Aitchison 1982), which we chose to address by applying the CLR transformation (Aitchison 1982). Because this transformation requires strictly positive values, we replaced any zero values with 0.0001. We also reduced the complexity of the 1,375 × 8 data matrix by employing principal component analysis (PCA). PCA measures interrelationships among multiple variables in numerous dimensions using a covariance matrix (Ringnér 2008). We used the first three principal components (explaining over 85% of the variance in the L_Srp and H_Srp datasets) to run the *k*-means algorithm for visualization purposes. After applying the *k*-means algorithm, the resulting datasets with predicted labels were named *L_Srp_km* and *H_Srp_km* (Supplementary Tables S3 and S4, respectively) (Fig. 1).

This is the peer-reviewed, final accepted version for American Mineralogist, published by the Mineralogical Society of America.
The published version is subject to change. Cite as Authors (Year) Title. American Mineralogist, in press.
DOI: https://doi.org/10.2138/am-2022-8688. http://www.minsocam.org/

Machine learning (ML) models

Given a set of observations (here, serpentine chemical compositions), ML models classify the dataset. We trained the classifier using the XGBoost, support vector machine (SVM), and random forest (RF) algorithms and their results were similar (not shown). We chose the XGBoost algorithm for our classification because it does not require imputation of missing values. Next, we used both the *k*-means and Gaussian mixture algorithms to cluster the data; again, both algorithms showed consistent results (not shown). Here, we showed and discussed only the clustering results obtained using the well-known unsupervised learning approach of the *k*-means algorithm. As mentioned above, we also applied PCA to the dataset before the *k*-means clustering to visualize the results in three dimensions.

Extreme Gradient Boosting (XGBoost). XGBoost is a highly efficient and flexible end-to-end boosting tree system (Chen and Guestrin 2016), improved by the gradient-boosting decision tree. It uses a parallel computing strategy in the training process to maintain optimal calculation speeds while adding a common term to the objective function. This prevents overfitting and optimizes computing resources during the adjustment of the objective function value. Unlike using feature vectors to calculate similarity, boosted trees are constructed to intelligently obtain feature scores, revealing the importance of each feature to the training model. When a feature is used with higher frequency to make critical decisions in boosted trees, that feature is given a higher score. The principle of the algorithm is to continuously add trees to fit the residuals of the previous prediction, summing the prediction scores of each tree to obtain the final score. More specific information on the XGBoost algorithm is available in Chen and Guestrin (2016). We used the open-source Python package ("xgboost") to implement the algorithm.

Model optimization with Scikit-learn. Scikit-learn is a Python module integrating a wide range of state-of-the-art machine learning algorithms for medium-scale supervised and unsupervised problems (Pedregosa et al., 2011). Before training the model, the dataset was randomly split into training (70%) and testing subsets (30%). The XGBoost algorithm uses numerous hyperparameters that affect the performance of the final model. Here, we selected the optimal hyperparameter combination by grid search, a typical parameter-tuning approach that methodically builds and evaluates a model in a specific grid (Hsu et al. 2003). We also performed tenfold cross-validation by building the prediction model with 90% of the training subset during the grid search and subsequently using the model to predict the withheld 10%. This was performed iteratively until all observations had been withheld and subsequently predicted. Tenfold cross-validation is important for avoiding overfitting and evaluating the model's predictive capability (Fig. S2). We performed the grid search over a 1,500-parameter space to identify the optimal parameter combination and define the final classifier. Then, the classifier was evaluated based on its performance when applied to the testing subset.

Model evaluation with confusion matrix. A confusion matrix (Supplementary Table S5) contains information about the actual and predicted classifications, which is used to evaluate the model performance. Based on the confusion matrix, four classification performance measures are typically used, defined based on the total numbers of true positive (TP), false positive (FP), true negative (TN), and false negative (FN) predictions: precision, recall, accuracy, and the traditional F-measure or balanced F-score (F_1 score).

This is the peer-reviewed, final accepted version for American Mineralogist, published by the Mineralogical Society of America.
The published version is subject to change. Cite as Authors (Year) Title. American Mineralogist, in press.
DOI: https://doi.org/10.2138/am-2022-8688. http://www.minsocam.org/

In this context, precision is a measure of accuracy, provided that a specific class has been predicted:

$$\text{Precision} = TP/(TP + FP) \quad \text{(Eq. 1)}$$

Recall measures the ability of the model to select instances of a specific class from a dataset:

$$\text{Recall} = TP/(TP + FN) \quad \text{(Eq.2)}$$

Accuracy is a straightforward measurement of the proportion of the total number of predictions that were correct:

$$\text{Accuracy} = (TP + TN)/(TP + TN + FP + FN) \quad \text{(Eq. 3)}$$

The F_1 score is the harmonic mean of precision and recall:

$$F_1 \text{ score} = 2*(\text{Precision} - \text{Recall})/(\text{Precision} + \text{Recall}) = 2TP/(2TP + TN + FP) \quad \text{(Eq. 4)}$$

The F_1 score is high when both precision and recall are high.

***k*-means clustering.** The *k*-means clustering algorithm aims to partition a given dataset into *k* clusters based on the distances between data points and cluster centroids (Zhou et al. 2018; Ghezelbash et al. 2020). Because the naive *k*-means algorithm can be highly computationally expensive and very sensitive to the initialization of *k* centroids, multiple variations have been proposed to improve the algorithm's efficiency and accuracy. Here, we used the *k*-means++ algorithm, which adopts a smarter centroid initialization (Arthur and Vassilvitskii 2007) than the naive *k*-means algorithm, which randomly chooses the initial *k* centroids. Generally, the procedure is as follows:

1. The algorithm chooses the first centroid randomly from a given dataset.
2. The distance between each data point and the nearest centroid is calculated.

This is the peer-reviewed, final accepted version for American Mineralogist, published by the Mineralogical Society of America.
The published version is subject to change. Cite as Authors (Year) Title. American Mineralogist, in press.
DOI: https://doi.org/10.2138/am-2022-8688. http://www.minsocam.org/

3. The data point farthest from the previously chosen centroid(s) is selected as the next new centroid.
4. Steps 2 and 3 are iterated until *k* centroids are found.
5. Each data point is assigned to a cluster corresponding to the nearest centroid.
6. The *k* cluster centers are taken as *k* new centroids (cluster centers) when Step 5 is completed.
7. Steps 5 and 6 are iterated until the positions of the centroids remain fixed between subsequent iterations.

RESULTS

Chemical compositions of serpentine minerals

The two classes of serpentine-group minerals have different compositional patterns (Fig. 2). H_Srp have systematically higher SiO_2, whereas L_Srp spans a wide range of SiO_2 contents (36.0–46.0 wt%). Based on their higher Al concentrations, Al substitution into crystal sites is slightly more frequent in H_Srp than in L_Srp. Furthermore, although both H_Srp and L_Srp span similar ranges of FeO_{tot} contents (0.0–12.5 wt%), >75% of H_Srp contain ≤5.0 wt% FeO_{tot}. H_Srp tend to contain less MnO and NiO than L_Srp. H_Srp MgO contents are concentrated in the range 37.0–39.0 wt%, whereas those of most L_Srp are more broadly distributed over 37.0–41.0 wt%. The distributions of TiO_2 (0–0.2 wt%) and Cr_2O_3 contents (0.0–1.3 wt%) are almost identical in H_Srp and L_Srp.

XGBoost model

Confusion matrix and performance evaluation. The raw confusion matrix of true (known) and predicted L_Srp and H_Srp labels after applying the XGBoost algorithm to the test dataset (i.e., 30% of the unprocessed dataset) is shown in Figure 3a and normalized to the known total number of analyses with each label in Figure 3b (i.e., the rows of the matrix sum to 1). The classification results indicate an overall accuracy of 87.2% (Table 1), with 86.2% of L_Srp and 87.4% of H_Srp correctly classified (Fig. 3b). Of the 160 L_Srp examples in the test dataset, 22 were falsely predicted as H_Srp, and 31 of the 253 H_Srp examples were falsely classified as L_Srp (Fig. 3a). These results indicate that the ML model provides reasonably accurate classifications of serpentine minerals.

Feature importance. Here, 'feature' refers to an elemental concentration, and feature importances reflect the relative influences of constituent elements on the model predictions. To identify the main features controlling the performance of the XGBoost algorithm in classifying serpentines, we calculated feature importances for the eight elemental concentrations considered. SiO_2 content had the highest feature importance score (0.29), implying its importance in classifying serpentine minerals (Fig. 3c). The classification was also fairly sensitive to NiO content, with a feature importance score of 0.15. Therefore, the classification of serpentine minerals performs well when SiO_2 and NiO are used to evaluate the models. Al_2O_3, MgO, FeO_{tot}, MnO, and TiO_2 contents had similar but weaker influences on the classification than NiO content (Fig. 3c).

k-means model

In our dataset, fewer than 10 analyses were available for some locations, leading to non-representative clustering. Therefore, we excluded these analyses before applying the *k*-means

clustering model. We trained separate *k*-means models for the L_Srp and H_Srp samples and used their silhouette scores (measuring the mean distance between each sample to its cluster centroid) to evaluate their performances. As shown in Figure S3, the clustering results for L_Srp and H_Srp samples have maximum respective silhouette scores of 0.57 and 0.54 when $k = 4$, making this the optimum number of clusters. In addition, four well-separated groups are observed in the PCA plots of the first three PCs (Fig. 4). These results indicate that the datasets were well separated into four clusters.

Of the L_Srp analyses, ~68.0% were classified into Cluster L_1 and are globally distributed (Fig. 5a). Furthermore, analyses from most locations fall into multiple clusters, with the exception of the landward slope of the Middle America Trench off Guatemala (Fig. 5a), for which all analyses were classified as Cluster L_1. Cluster L_2 of L_Srp only spread in four locations, including New Caledonia, NW Iran, the Eastern Desert of Egypt, and NE Corsica (Fig. 5a). These *k*-means results imply that analyses from British Columbia Canada, Elba Island, New Caledonia, and the North Qaidam orogen (Tibet) are characterized by only two meaningful classes, whereas analyses from the other eight areas (i.e., excluding Guatemala) are characterized by three or four clusters (Fig. 5a). We note that the classification of most of these analyses as Cluster L_1 indicates that the major chemical compositions of chrysotile and lizardite do not strongly reflect their sampling locations.

Of the H_Srp analyses, 87.9% were classified as either Cluster H_1 or H_2 (Fig. 5b; note that Clusters H_1–4 are distinct from Clusters L_1–4). NW Iran was the only location without any samples classified as Cluster H_2. Analyses from Cuba, the Hess Deep, and Japan were uniquely classified as Cluster H_1. Only analyses from New Caledonia and Qinling China were categorized into just two clusters: Clusters H_2 and H_4 and Clusters H_2 and H_3, respectively (Fig. 5b).

This is the peer-reviewed, final accepted version for American Mineralogist, published by the Mineralogical Society of America.
The published version is subject to change. Cite as Authors (Year) Title. American Mineralogist, in press.
DOI: https://doi.org/10.2138/am-2022-8688. http://www.minsocam.org/

The remaining nine locations showed broadly similar clustering distributions (Fig. 5b), implying that the compositional clustering of H_Srp (antigorite) is not largely due to the chemical environment in which is formed.

DISCUSSION

Classification of serpentine minerals

Raman analyses are currently the most widely used traditional analytical method for discriminating between serpentine minerals (Rinaudo et al. 2003). The presence of peaks at ~1050 cm^{-1} and 3680-3700 cm^{-1}, representing antisymmetric Si-O_b-Si and OH stretching, respectively, usually indicates that a polymorph is an antigorite (Auzende et al. 2004; Petriglieri et al. 2015), whereas chrysotile and lizardite are identified based on the presence of a peak attributed to antisymmetric stretching modes of Si-O_{nb} groups at ~1100 cm^{-1} (Groppo et al. 2006). However, these peaks may be absent in poorly crystallized serpentines. Valid peaks may also not occur if the sample was first analyzed by electron microprobe analysis which could destroy its structure by the electron beam. Another problem in most applications of Raman spectroscopy is the strong fluorescence background (Wei et al., 2015). Therefore, the high quality of the sample and appropriate methods to suppress the fluorescence background are required to collect useful Raman results. In contrast, our quantitative ML model can distinguish between serpentine minerals based only on their major chemical compositions. Our model can distinguish between L_Srp and H_Srp with 87.2% accuracy (Table 1), which is far more reliable than traditional geochemical methods.

Our binary classification model sometimes misclassified serpentines (Figs. 3a, 3b), possibly due to mineral replacements among the serpentine minerals. Antigorite is more stable than chrysotile and lizardite at high temperatures (~250 to >500 °C; O'Hanley and Wicks 1995; Ghaderi et al. 2015), and usually forms through prograde and retrograde metamorphic reactions during

317 subduction and obduction, respectively (Guillot et al. 2001; Boudier et al. 2010; Wu et al. 2018;
318 Zhang et al. 2019). Field observations and laboratory evidence suggest that chrysotile and lizardite
319 can precede replacement by antigorite during prograde metamorphism (Koutsovitis 2017; Plissart
320 et al. 2019), which is associated with the formation of brucite or a modest influx of SiO_2 (e.g.,
321 O'Hanley and Wicks 1995; Evans 2010; Rouméjon et al. 2019). These two mineral replacement
322 processes likely explain why antigorite contains more Si than lizardite and chrysotile (Fig. 2a),
323 and are probably why the classification between H_Srp (antigorite) with L_Srp (chrysotile and
324 lizardite) performed so well.

325 These same transformations also imply that the composition of antigorite could be inherited
326 from previous chrysotile and/or lizardite, which may explain the rare misclassification of H_Srp
327 as L_Srp in our model. Furthermore, Evans (2004) suggested that, although serpentinization may
328 force the crystallization of chrysotile, a considerable amount of chrysotile recrystallizes in
329 antigorite-bearing serpentinites surrounded by abundant Si- and Mg-oversaturated fluids.
330 Therefore, chrysotile may also replace antigorite without any associated compositional change
331 (Evans 2004; Ulrich et al. 2020), which may explain the rare misclassification of L_Srp as H_Srp
332 in our model (Figs. 3a, 3b).

333 Our feature importance results indicate that the chemical differences between serpentine
334 minerals mainly arise from their SiO_2 contents (Fig. 3c). Consistent with previous results (e.g.,
335 Deschamps et al. 2013) and the SiO_2 distribution plot (Fig. 2), antigorite samples contain more
336 SiO_2 than lizardite and chrysotile. Indeed, the tetrahedral sheets with periodic reversals along the
337 *a*-axis in antigorite predestine the formation of antigorite to consume more SiO_2 than the formation
338 of chrysotile and lizardite. NiO contents were also important to the performance of our
339 classification (Fig. 3c): chrysotile and lizardite span a fairly wide range of NiO concentrations (0–

0.6 wt%), whereas the NiO contents of antigorite are more restricted to the range 0–0.2 wt%. This result is consistent with the conclusions of McCollom and Bach (2009), who identified the formation of Fe-Ni alloy during the replacement of chrysotile and lizardite by antigorite, indicating that Ni migrates from serpentine to the surrounding environment. Al_2O_3 was the third-most important feature for classifying serpentine minerals (Fig. 3c). The distribution plot reveals that, although serpentines have a wide range of Al_2O_3 contents (0.0–6.7 wt%), those of H_Srp are mainly restricted to the range 0.4–3.0 wt%, slightly above the range for L_Srp (Fig. 2c). Because Al generally substitutes for Si in the tetrahedral sheets of serpentine, the higher Al_2O_3 contents of H_Srp are the consequence of the higher SiO_2 contents of H_Srp. Unexpectedly, the substitutions of Fe, Mn, Ti, and Cr into the structures of serpentine minerals had only minimal contributions to the classification (feature importances of 0.09, 0.09, 0.09, and 0.07, respectively; Fig. 3c). The different impacts of these substitutions suggest that Ni and Al equilibria between serpentine minerals and their surrounding environments depend on the minerals' structures and formation conditions, whereas those involving Fe, Mn, Ti, and Cr do not.

Application of the *k*-means algorithm to determine the environmental conditions of crystallization

Our clustering model can be applied to determine the original geological environments in which serpentine crystallized. The derived silhouette coefficients of the *k*-means model performances for L_Srp (chrysotile and lizardite) indicate that four clusters best classified the data (Fig. 5a). Kodolányi and Pettke (2011) confirmed that the alteration of oceanic crust is dominant in Guatemala, and no high temperature (>300–400 °C) alteration phases have been detected there, indicating that Guatemala chrysotile and lizardite formed at low temperatures (<300 °C).

Therefore, we consider that the temperature during the formation of L_Srp in Cluster L_1 is constrained to <300 °C, consistent with previous field observations and experimental results (Evans 2004; Schwartz et al. 2013). Cluster L_1 may also represent L_Srp that were not modified by subduction zones because L_Srp tends to crystallize during water-rock interactions on the seafloor, in the absence of intense tectonic activity and major shear zones (Federico et al. 2005; Putlitz et al. 2005; Obeid et al. 2016; Imayama et al. 2020). Because chrysotile and lizardite tend to crystallize in such geological environments, the classification of ~68.0% of L_Srp into Cluster L_1 is consistent with these observations (Fig. 5a). L_Srp samples in Cluster L_1 occur in all locations listed in our L_Srp dataset, indicating that it is not practical to attempt to distinguish which L_Srp phase is produced upon alteration of oceanic crust on the seafloor at a given sampling location based on chemical composition alone. Furthermore, the hydration of ultramafic or mafic rocks by meteoric water at Earth's surface (Zhang et al. 2019; Ulrich et al. 2020), the equilibration of fluids with shallow subducting sediments (Wu et al. 2018), or upwelling fluids sourced from subducted slabs (Martin et al. 2020) dehydrated under sub-greenschist conditions can also produce both chrysotile and lizardite. Unfortunately, we failed to assign any of this information to Clusters L_2–4. The incorporation of more data, including isotopic results (e.g., $\delta^{11}B$, $\delta^{18}O$) should improve our models, allowing us to constrain specific *P*-*T* conditions and explore the evolution of geological processes and elemental cyclings (Zhang et al. 2019; Martin et al. 2020).

We also obtained four clusters by applying the *k*-means algorithm to the H_Srp chemical compositions (Fig. 5b). Ribeiro da Costa et al. (2008) suggested that antigorite at the Mid-Atlantic ridge may result from extensive deformation at low temperature rather than high-temperature prograde metamorphism. Rouméjon et al. (2019) further proposed that the presence of a Si-rich fluid is the only requirement for the crystallization of antigorite. Indeed, antigorite formed under

This is the peer-reviewed, final accepted version for American Mineralogist, published by the Mineralogical Society of America.
The published version is subject to change. Cite as Authors (Year) Title. American Mineralogist, in press.
DOI: https://doi.org/10.2138/am-2022-8688. http://www.minsocam.org/

conditions unrelated to subduction was classified into Cluster H_2 in our study (Fig. 5b). In contrast, antigorite produced by blueschist to eclogite facies metamorphism (i.e., 390–650 °C) was classified into Clusters H_1, H_3, and H_4 (Lafay et al. 2013; Vitale Brovarone and Herwartz 2013; Martin et al. 2020; Ulrich et al. 2020). Antigorite from Tibet, China, and Iran may have undergone metamorphism at ultra-high pressure (21–35 kbar), but at temperatures not exceeding 700 °C (Zhang et al. 2019; Moghadam and Stern 2015), indicating that the stability of antigorite is more sensitive to temperature than pressure. Nonetheless, the prograde recrystallization of chrysotile and/or lizardite to form some of the subduction-related H_Srp (Rouméjon et al. 2019; Martin et al. 2020) may result in misclassifications between L_Srp and H_Srp, consistent with our XGBoost results (Fig. 3).

For both L_Srp and H_Srp, samples from different locations can be classified into the same cluster, suggesting that the specific sampling location has a negligible effect on compositional clustering. However, our results confirm that the geological environment determines the serpentine species and chemical composition. We note that our dataset only contains 1,375 analyses, limiting our efforts to assess the *P-T* conditions required to form serpentine of specific chemical compositions. Nevertheless, subduction-related serpentines seem to contain less TiO_2, Al_2O_3, and Cr_2O_3 (Figs. S4 and S5).

IMPLICATIONS

Serpentinization occurs ubiquitously in seafloor ultramafic rocks and subduction zones, exerting a strong influence on the geochemistry of those systems (Schmidt et al. 2011; Schroeder et al. 2015; Seyfried et al. 2015; Scott et al. 2017; Debret et al. 2019). Serpentine-group minerals can structurally accommodate water and elements, and minerals thus play a significant role in the

409 alteration process (Evans 2008; Hacker 2008; Iyer et al. 2008; Jöns et al. 2010; Schrenk et al. 2013;
410 Canales et al. 2017; Mayhew et al. 2018). Therefore, the transformation from L_Srp to H_Srp
411 promotes elemental cycling (Malvoisin 2015; Mayhew et al. 2018). Our feature importance data
412 and the distributions of serpentine chemical compositions (Figs. 2a, 3c) indicate that the
413 enrichment of Si in serpentine is associated with subduction-generated Si-rich fluids. This sheds
414 light on the factors controlling the solubility of hydrous melts and dilute aqueous solutions in
415 which the dissolved silicate components also play significant roles (Manning 2004). Our feature
416 importance data indicate that Ni migrates from serpentine to the surrounding environment during
417 mineral replacement (Figs. 2g, 3c) and may be consumed by Fe-Ni alloy when dominated by
418 ferrous iron (McCollom and Bach 2009), redistributing Fe among the solids. In contrast, Al
419 diffuses into H_Srp during replacement (Figs. 2c, 3c) in association with the increase in SiO_2
420 content (Figs. 2a, 3c). This indicates that Al tends to substitute for Si in the tetrahedral sheets
421 (Evans et al. 2013), which may stabilize antigorite at higher temperatures (Padrón-Navarta et al.
422 2013). Our XGBoost model thus documents the importance of serpentine mineral replacements
423 during mass transfers in subduction zones and relates serpentine chemical compositions to their
424 geological environments, whereas traditional geochemical methods have limited capabilities to
425 provide such information on alteration chemistry (e.g., Barbier et al. 2020).

426 The dataset used to train these ML algorithms only contains 1,375 analyses for which the
427 concentrations of the eight analyzed elements were reported. Due to the lack of *P*-*T* data, our
428 discussion on serpentine crystallization and geological environments is qualitative. Model
429 performance will improve when more data becomes available. Similar ML algorithms could be
430 applied to track the sources of fluids participating in the formation of serpentine minerals by
431 incorporating $^{87}Sr/^{86}Sr$ data to further constrain the evolution of metasomatic processes (Consuma

et al. 2020). Additional data on Fe^{2+}/Fe^{3+} ratios and B isotopes will be useful for classifying the redox state of serpentine minerals and revealing the influence of mineral replacements on the redox budget during subduction (Zhang et al. 2019, 2021). Furthermore, applying such ML algorithms to a dataset incorporating the *P-T* conditions of serpentine crystallization, the age of the lithosphere, and convergence rates will improve our understanding of Earth's tectonic history (Blanco-Quintero et al. 2011; Wu et al. 2018; Cannaò et al. 2020).

DATA AVAILABILITY STATEMENT

Python scripts for our ML classification and clustering models are available on Zenodo at https://doi.org/ 10.5281/zenodo.7551437. After setting up a python environment with the essential packages installed, the models can be run on Windows, Mac OS X, or Linux.

ACKNOWLEDGMENTS

This work was supported by Fundamental Research Funds for the Central Universities (Grant No. 226-2022-00176) and startup funding to ZhangZhou J. We acknowledge Robert Dennen for polishing the language of the paper.

REFERENCES CITED

Aitchison, J. (1982) The Statistical Analysis of Compositional Data. Journal of the Royal Statistical Society: Series B (Methodological), 44(2), 139–160.

Arthur, D., Vassilvitskii, S. (2007) k means++: The advantages of careful seeding. Proceedings of the Eighteenth Annual ACM-SIL Symposium on Discrete Algorithms. Philadelphia, USA: SIAM, 1027-1035.

Auzende, A.-L., Devouard, B., Guillot, S., Daniel, I., Baronnet, A., and Lardeaux, J.-M. (2002) Serpentinites from Central Cuba: petrology and HRTEM study. European Journal of Mineralogy, 14(5), 905–914.

Auzende, A.-L, Daniel, I., Reynard, B., Lemaire, C., and Guyot, F. (2004) High-pressure behaviour of serpentine minerals: a Raman spectroscopic study. Physics and Chemistry of Minerals, 31, 269-277.

Auzende, A.-L., Guillot, S., Devouard, B., and Baronnet, A. (2006) Serpentinites in an Alpine convergent setting: Effects of metamorphic grade and deformation on microstructures. European Journal of Mineralogy, 18(1), 21–33.

Banfield, J.F., Bailey, S.W., Barker, W.W., and Smith, R.C. (1995) Complex polytypism; relationships between serpentine structural characteristics and deformation. American Mineralogist, 80(11–12), 1116–1131.

Barbier, S., Huang, F., Andreani, M., Tao, R., Hao, J., Eleish, A., Prabhu, A., Minhas, O., Fontaine, K., Fox, P., and Minhas, O. (2020) A Review of H_2, CH_4, and Hydrocarbon Formation in Experimental Serpentinization Using Network Analysis. Frontiers in Earth Science, 8, 209.

Blanco-Quintero, I.F., Rojas-Agramonte, Y., García-Casco, A., Kröner, A., Mertz, D.F., Lázaro, C., Blanco-Moreno, J., and Renne, P.R. (2011) Timing of subduction and exhumation in a subduction channel: Evidence from slab melts from La Corea Mélange (eastern Cuba). Lithos, 127, 86–100.

Boskabadi, A., Pitcairn, I.K., Leybourne, M.I., Teagle, D.A.H., Cooper, M.J., Hadizadeh, H., Bezenjani, R.N., and Bagherzadeh, R.M. (2020) Carbonation of ophiolitic ultramafic rocks: Listvenite formation in the Late Cretaceous ophiolites of eastern Iran. Lithos, 352–353, 105307.

Boudier, F., Baronnet, A., and Mainprice, D. (2010) Serpentine Mineral Replacements of Natural Olivine and their Seismic Implications: Oceanic Lizardite versus Subduction-Related Antigorite. Journal of Petrology, 51(1–2), 495–512.

Canales, J.P., Carbotte, S.M., Nedimović, M.R., and Carton, H. (2017) Dry Juan de Fuca slab revealed by quantification of water entering Cascadia subduction zone. Nature Geoscience, 10(11), 864–870.

Cannaò, E., Scambelluri, M., Bebout, G.E., Agostini, S., Pettke, T., Godard, M., and Crispini, L. (2020) Ophicarbonate evolution from seafloor to subduction and implications for deep-Earth C cycling. Chemical Geology, 546, 119626.

Capitani, G., and Mellini, M. (2004) The modulated crystal structure of antigorite: The m = 17 polysome. American Mineralogist, 89(1), 147-158.

Caricchi, L., Petrelli, M., Bali, E., Sheldrake, T., Pioli, L., and Simpson, G. (2020) A Data Driven Approach to Investigate the Chemical Variability of Clinopyroxenes From the 2014–2015 Holuhraun–Bárdarbunga Eruption (Iceland). Frontiers in Earth Science, 8, 18.

Chen, T., and Guestrin, C. (2016) XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). San Francisco California USA: ACM.

Consuma, G., Braga, R., Giovanardi, T., Bersani, D., Konzett, J., Lugli, F., Mazzucchelli, M., and Tropper, P. (2020) In situ Sr isotope analysis of mantle carbonates: Constraints on the evolution and sources of metasomatic carbon-bearing fluids in a paleo-collisional setting. Lithos, 354–355, 105334.

Debret, B., Albers, E., Walter, B., Price, R., Barnes, J.D., Beunon, H., Facq, S., Gillikin, D.P., Mattielli, N., and Williams, H. (2019) Shallow forearc mantle dynamics and geochemistry: New insights from IODP Expedition 366. Lithos, 326–327, 230–245.

Deschamps, F., Guillot, S., Godard, M., Chauvel, C., Andreani, M., and Hattori, K. (2010) In situ characterization of serpentinites from forearc mantle wedges: Timing of serpentinization and behavior of fluid-mobile elements in subduction zones. Chemical Geology, 269(3), 262–277.

Deschamps, F., Godard, M., Guillot, S., and Hattori, K. (2013) Geochemistry of subduction zone serpentinites: A review. Lithos, 178, 96–127.

Evans, B.W. (2004) The Serpentinite Multisystem Revisited: Chrysotile Is Metastable. International Geology Review, 46, 479–506.

——— (2008) Control of the Products of Serpentinization by the $Fe^{2+}Mg_{-1}$ Exchange Potential of Olivine and Orthopyroxene. Journal of Petrology, 49, 1873–1887.

——— (2010) Lizardite versus antigorite serpentinite: Magnetite, hydrogen, and life(?). Geology, 38, 879–882.

Evans, B.W., Hattori, K., and Baronnet, A. (2013) Serpentinite: What, Why, What? Element, 9, 99-106.

Federico, L., Capponi, G., Crispini, L., Scambelluri, M., and Villa, I. (2005) $^{39}Ar/^{40}Ar$ dating of high-pressure rocks from the Ligurian Alps: Evidence for a continuous subduction–exhumation cycle. Earth and Planetary Science Letters, 240(3–4), 668–680.

Ghaderi, N., Zhang, H., and Sun, T. (2015) Relative stability and contrasting elastic properties of serpentine minerals from first-principles calculations. Journal of Geophysical Research: Solid Earth, 120(7), 4831–4842.

Gheyas, I.A., and Smith, L.S. (2010) A neural network-based framework for the reconstruction of incomplete data sets. Neurocomputing, 73(16–18), 3039–3065.

Ghezelbash, R., Maghsoudi, A., and Carranza, E.J.M. (2020) Optimization of geochemical anomaly detection using a novel genetic K-means clustering (GKMC) algorithm. Computers and Geosciences, 134, 104335.

Groppo, C., Rinaudo, C., Cairo, S., Gastaldi, D., and Compagnoni, R. (2006) Micro-Raman spectroscopy for a quick and reliable identification of serpentine minerals from ultramafics. European Journal of Mineralogy, 18(3), 319–329.

Guillot, S., Hattori, K.H., de Sigoyer, J., Nägler, T., and Auzende, A.-L. (2001) Evidence of hydration of the mantle wedge and its role in the exhumation of eclogites. Earth and Planetary Science Letters, 193(1–2), 115–127.

Hacker, B.R. (2008) H_2O subduction beyond arcs. Geochemistry, Geophysics, Geosystems, 9(3), Q03001.

Hattori, K.H., and Guillot, S. (2007) Geochemical character of serpentinites associated with high- to ultrahigh-pressure metamorphic rocks in the Alps, Cuba, and the Himalayas: Recycling of elements in subduction zones: Geochemistry of serpentinites. Geochemistry, Geophysics, Geosystems, 8(9), Q09010.

Hazen, R.M., Downs, R.T., Eleish, A., Fox, P., Gagné, O.C., Golden, J.J., Grew, E.S., Hummer, D.R., Hystad, G., Krivovichev, S.V., and others. (2019) Data-Driven Discovery in Mineralogy: Recent Advances in Data Resources, Analysis, and Visualization. Engineering, 5(3), 397–405.

Hirauchi, K., and Katayama, I. (2013) Rheological contrast between serpentine species and implications for slab-mantle wedge decoupling. Tectonophysics, 608, 545-551.

Hirose, T., Bystricky, M., Kunze, K., and Stünitz, H. (2006) Semi-brittle flow during dehydration of lizardite–chrysotile serpentinite deformed in torsion: Implications for the rheology of oceanic lithosphere. Earth and Planetary Science Letters, 249(3–4), 484–493.

Hsu, C., Chang, C., and Lin, C. (2003) A practical guide to support vector classification. Technical report, Department of Computer Science and Information Engineering (p. 16), National Taiwan University, Taipei.

Imayama, T., Ueda, H., Usuki, T., Minami, M., Asahara, Y., and Nagahashi, T. (2020) Variability of protoliths and pressure-temperature conditions of amphibolites from the Ohmachi Seamount (Izu-Bonin-Mariana arc): evidence of a fossil subduction channel in a modern intra-oceanic arc. Mineralogy and Petrology, 114, 305-318.

Iyer, K., Austrheim, H., John, T., and Jamtveit, B. (2008) Serpentinization of the oceanic lithosphere and some geochemical consequences: Constraints from the Leka Ophiolite Complex, Norway. Chemical Geology, 249(1–2), 66–90.

Ji, S., Zhu, J., He, H., Tao, Q., Zhu, R., Ma, L., Chen, M., Li, S., and Zhou, J. (2018) Conversion of serpentine to smectite under hydrothermal condition: Implication for solid-state transformation. American Mineralogist, 103(2), 241–251.

John, T., Scambelluri, M., Frische, M., Barnes, J.D., and Bach, W. (2011) Dehydration of subducting serpentinite: Implications for halogen mobility in subduction zones and the deep halogen cycle. Earth and Planetary Science Letters, 308(1), 65–76.

Jöns, N., Bach, W., and Klein, F. (2010) Magmatic influence on reaction paths and element transport during serpentinization. Chemical Geology, 274(3), 196–211.

Kodolányi, J., and Pettke, T. (2011) Loss of trace elements from serpentinites during fluid-assisted transformation of chrysotile to antigorite — An example from Guatemala. Chemical Geology, 284(3–4), 351–362.

Koutsovitis, P. (2017) High-pressure subduction-related serpentinites and metarodingites from East Thessaly (Greece): Implications for their metamorphic, geochemical and geodynamic evolution in the Hellenic–Dinaric ophiolite context. Lithos, 276, 122–145.

Lafay, R., Deschamps, F., Schwartz, S., Guillot, S., Godard, M., Debret, B., and Nicollet, C. (2013) High-pressure serpentinites, a trap-and-release system controlled by metamorphic conditions: Example from the Piedmont zone of the western Alps. Chemical Geology, 343, 38–54.

Majumdar, A.S., Ray, D., and Shukla, A.D. (2020) Serpentinization of olivine–gabbro in Central Indian ridge: Insights into H_2 production during alteration in lower oceanic crust and sustenance of life at slow–spreading ridges. Lithos, 374–375, 105730.

Malvoisin, B. (2015) Mass transfer in the oceanic lithosphere: Serpentinization is not isochemical. Earth and Planetary Science Letters, 430, 75-85.

Manning, C.E. (2004) The chemistry of subduction-zone fluids. Earth and Planetary Science Letters, 224, 1-14.

Martin, C., Flores, K.E., Vitale-Brovarone, A., Angiboust, S., and Harlow, G.E. (2020) Deep mantle serpentinization in subduction zones: Insight from in situ B isotopes in slab and mantle wedge serpentinites. Chemical Geology, 545, 119637.

Mayhew, L.E., Ellison, E.T., Miller, H.M., Kelemen, P.B., and Templeton, A.S. (2018) Iron transformations during low temperature alteration of variably serpentinized rocks from the Samail ophiolite, Oman. Geochimica et Cosmochimica Acta, 222, 704–728.

McCollom, T.M., and Bach, W. (2009) Thermodynamic constraints on hydrogen generation during serpentinization of ultramafic rocks. Geochimica et Cosmochimica Acta, 73(3), 856–875.

Mellini, M. (1986) Chrysotile and polygonal serpentine from the Balangero serpentinite. Mineralogical Magazine, 50(356), 301–305.

Moghadam, H.S., and Stern, R.J. (2015) Ophiolites of Iran: Keys to understanding the tectonic evolution of SW Asia: (II) Mesozoic ophiolites. Journal of Asian Earth Sciences, 100, 31–59.

Obeid, M.A., Khalil, A.E.S., and Azer, M.K. (2016) Mineralogy, geochemistry, and geotectonic significance of the Neoproterozoic ophiolite of Wadi Arais area, south Eastern Desert, Egypt. International Geology Review, 58(6), 687–702.

O'Hanley, D.S., and Dyar, M.D. (1998) The composition of chrysotile and its relationship with lizardite. The Canadian Mineralogist, 36, 727–739.

O'Hanley, D.S., and Wicks, F.J. (1995) Conditions of formation of lizardite, chrysotile and antigorite, Cassiar, British Columbia. The Canadian Mineralogist, 33, 753-773.

O'Sullivan, G., Chew, D., Kenny, G., Henrichs, I., and Mulligan, D. (2020) The trace element composition of apatite and its application to detrital provenance studies. Earth-Science Reviews, 201, 103044.

Padregosa, E., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderpls, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Ducchesnay, É. (2011) Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825-2830.

Padrón-Navarta, J.A., Sánchez-Vizcaíno, V.L., Hermann, J., Connolly, J.A.D., Garrido, C.J., Gómez-Pugnaire, M.T., and Marchesi, C. (2013) Tschermak's substitution in antigorite and consequences for phase relations and water liberation in high-grade serpentinites. Lithos, 178, 186–196.

Petrelli, M., and Perugini, D. (2016) Solving petrological problems through machine learning: the study case of tectonic discrimination using geochemical and isotopic data. Contributions to Mineralogy and Petrology, 171(10), 81.

Petrelli, M., Caricchi, L., and Perugini, D. (2020) Machine Learning Thermo-Barometry: Application to Clinopyroxene-Bearing Magmas. Journal of Geophysical Research: Solid Earth, 125, e2020JB020130.

Petriglieri, J.R., Salvioli-Mariani, E., Mantovani, L., Tribaudino, M., Lottici, P.P., Laporte-Magoni, C., and Bersani, D. (2015) Micro-Raman mapping of the minerals of serpentine: Micro-Raman mapping of the minerals of serpentine. Journal of Raman Spectroscopy, 46(10), 953–958.

Plissart, G., González-Jiménez, J.M., Garrido, L.N.F., Colás, V., Berger, J., Monnier, C., Diot, H., and Padrón-Navarta, J.A. (2019) Tectono-metamorphic evolution of subduction channel serpentinites from South-Central Chile. Lithos, 336–337, 221–241.

Plümper, O., John, T., Podladchikov, Y.Y., Vrijmoed, J.C., and Scambelluri, M. (2017) Fluid escape from subduction zones controlled by channel-forming reactive porosity. Nature Geoscience, 10(2), 150–156.

Putlitz, B., Cosca, M.A., and Schumacher, J.C. (2005) Prograde mica 40Ar/39Ar growth ages recorded in high pressure rocks (Syros, Cyclades, Greece). Chemical Geology, 214(1–2), 79–98.

Reynard, B. (2013). Serpentine in active subduction zones. Lithos, 178, 171–185.

Ribeiro da Costa, I., Barriga, F.J.A.S., Viti, C., Mellini, M., and Wicks, F.J. (2008) Antigorite in deformed serpentinites from the Mid-Atlantic Ridge. European Journal of Mineralogy, 20, 563-572.

Rinaudo, C., Gastaldi, D., and Belluso, E. (2003) Characterization of chrysotile, antigorite and lizardite by FT-Raman spectroscopy. The Canadian Mineralogist, 41(4), 883–890.

Ringnér, M. (2008) What is principal component analysis? Nature Biotechnology, 26(3), 303–304.

Rouméjon, S., Andreani, M., and Früh-Green, G.L. (2019) Antigorite crystallization during oceanic retrograde serpentinization of abyssal peridotites. Contributions to Mineralogy and Petrology, 174(7), 60.

Rüpke, L.H., Morgan, J.P., Hort, M., and Connolly, J.A.D. (2004) Serpentine and the subduction zone water cycle. Earth and Planetary Science Letters, 223(1), 17–34.

Sachan, H.K., Mukherjee, B.K., and Bodnar, R.J. (2007) Preservation of methane generated during serpentinization of upper mantle rocks: Evidence from fluid inclusions in the Nidar ophiolite, Indus Suture Zone, Ladakh (India). Earth and Planetary Science Letters, 257(1–2), 47–59.

Schmidt, K., Garbe-Schönberg, D., Koschinsky, A., Strauss, H., Jost, C.L., Klevenz, V., and Königer, P. (2011) Fluid elemental and stable isotope composition of the Nibelungen hydrothermal field (8°18′S, Mid-Atlantic Ridge): Constraints on fluid–rock interaction in heterogeneous lithosphere. Chemical Geology, 280(1), 1–18.

Schrenk, M.O., Brazelton, W.J., and Lang, S.Q. (2013) Serpentinization, Carbon, and Deep Life. Reviews in Mineralogy and Geochemistry, 75(1), 575–606.

Schroeder, T., Bach, W., Jöns, N., Jöns, S., Monien, P., and Klügel, A. (2015) Fluid circulation and carbonate vein precipitation in the footwall of an oceanic core complex, Ocean Drilling Program Site 175, Mid-Atlantic Ridge. Geochemistry, Geophysics, Geosystems, 16(10), 3716–3732.

Schwartz, S., Guillot, S., Reynard, B., Lafay, R., Debret, B., Nicollet, C., Lanari, P., and Auzende, A.L. (2013) Pressure–temperature estimates of the lizardite/antigorite transition in high pressure serpentinites. Lithos, 178, 197–210.

Scott, S.R., Sims, K.W.W., Frost, B.R., Kelemen, P.B., Evans, K.A., and Swapp, S.M. (2017) On the hydration of olivine in ultramafic rocks: Implications from Fe isotopes in serpentinites. Geochimica et Cosmochimica Acta, 215, 105–121.

Seyfried, W.E., Pester, N.J., Tutolo, B.M., and Ding, K. (2015) The Lost City hydrothermal system: Constraints imposed by vent fluid chemistry and reaction path models on subseafloor heat and mass transfer processes. Geochimica et Cosmochimica Acta, 163, 59–79.

Shen, T., Zhang, C., Chen, J., Hermann, J., Zhang, L., Padrón-Navarta, J.A., Chen, L., Xu, J., and Yang, J. (2020) Changes in the cell parameters of antigorite close to its dehydration reaction at subduction zone conditions. American Mineralogist, 105(4), 569–582.

Tamblyn, R., Zack, T., Schmitt, A.K., Hand, M., Kelsey, D., Morrissey, L., Pabst, S., and Savov, I.P. (2019) Blueschist from the Mariana forearc records long-lived residence of material in the subduction channel. Earth and Planetary Science Letters, 519, 171–181.

Ulrich, M., Muñoz, M., Boulvais, P., Cathelineau, M., Cluzel, D., Guillot, S., and Picard, C. (2020) Serpentinization of New Caledonia peridotites: from depth to (sub-)surface. Contributions to Mineralogy and Petrology, 175(9), 91.

Valetich, M.J., Le Losq, C., Arculus, R.J., Umino, S., and Mavrogenes, J. (2021) Compositions and Classification of Fractionated Boninite Series Melts from the Izu–Bonin–Mariana Arc: A Machine Learning Approach. Journal of Petrology, 62(2), egab013.

Vitale Brovarone, A., and Herwartz, D. (2013) Timing of HP metamorphism in the Schistes Lustrés of Alpine Corsica: New Lu–Hf garnet and lawsonite ages. Lithos, 172–173, 175–191.

Viti, C., and Mellini, M. (1997) Contrasting chemical compositions in associated lizardite and chrysotile in veins from Elba, Italy. European Journal of Mineralogy, 9(3), 585–596.

Wang, A., Freeman, J.J., and Jolliff, B.L. (2015) Understanding the Raman spectral features of phyllosilicates. Journal of Raman Spectroscopy, 46(10), 829–845. https://doi.org/10.1002/jrs.4680

Wang, Y., Qiu, K., Müller, A., Hou, Z., Zhu, Z., and Yu, H. (2021) Machine Learning Prediction of Quartz Forming-Environments. Journal of Geophysical Research: Solid Earth, 126(8).

Wei, D., Chen, S., and Liu, Q. (2015) Review of fluorescence suppression techniques In Raman spectroscopy. Applied Spectroscopy Reviews, 50(5), 387-406.

Wheat, G.C., Seewald, J.S., and Takai, K. (2020) Fluid transport and reaction processes within a serpentinite mud volcano: South Chamorro Seamount. Geochimica et Cosmochimica Acta, 269, 413–428.

Whittaker, E.J.W., and Wicks, F.J. (1970) Chemical difference among the serpentine "polymorphs": A discussion. American Mineralogist, 55, 1025-1047.

Wicks, F.J., and Whittaker, E.J.W. (1975) A reappraisal of the structures of the serpentine minerals. The Canadian Mineralogist, 13, 227–243.

Wu, K., Ding, X., Ling, M.-X., Sun, W., Zhang, L.-P., Hu, Y.-B., and Huang, R.-F. (2018) Origins of two types of serpentinites from the Qinling orogenic belt, central China and associated fluid/melt-rock interactions. Lithos, 302–303, 50–64.

Zhang, G., Zhang, L., Song, S., and Niu, Y. (2009) UHP metamorphic evolution and SHRIMP geochronology of a coesite-bearing meta-ophiolitic gabbro in the North Qaidam, NW China. Journal of Asian Earth Sciences, 35(3–4), 310–322.

Zhang, L., Sun, W., and Chen, R.-X. (2019) Evolution of serpentinite from seafloor hydration to subduction zone metamorphism: Petrology and geochemistry of serpentinite from the ultrahigh pressure North Qaidam orogen in northern Tibet. Lithos, 346–347, 105158.

Zhang, Y., Gazel, E., Gaetani, G.A., and Klein, F. (2021) Serpentinite-derived slab fluids control the oxidation state of the subarc mantle. Science Advances, 7, eabj2515.

Zhou, S., Zhou, K., Wang, J., Yang, G., and Wang, S. (2018) Application of cluster analysis to geochemical compositional data for identifying ore-related geochemical anomalies. Frontiers of Earth Science, 12(3), 491–505.

698 **Figure 1.** Operational flow chart of the methods used in this study. Step 1: the original dataset was used to
699 train the XGBoost models and evaluate model performance. Step 2: the preprocessed (KNN & CLR)
700 original dataset was used to train the *k*-means clustering models.
701
702 **Figure 2.** The distributions of the eight analyzed elemental concentrations in serpentine minerals. L_Srp
703 represents chrysotile and lizardite, and H_Srp represents antigorite.
704
705 **Figure 3.** The performance of our XGBoost model presented in the form of **(a)** nonnormalized and **(b)**
706 normalized confusion matrices. **(c)** Relative feature importances (summing to 1) during the training of the
707 XGBoost algorithm to classify serpentines as either L_Srp or H_Srp.
708
709 **Figure 4.** Distribution plots of **(a,d)** PC1 *vs.* PC2, **(b,e)** PC1 *vs.* PC3, and **(c,f)** PC2 *vs.* PC3 for **(a–c)** L_Srp
710 and **(d–f)** H_Srp. Symbol colors represent clusters identified by *k*-means analysis and symbol shapes
711 indicate sampling locations. MAR, Mid-Atlantic Ridge; IOR, Indian Oceanic Ridge; IBM, Izu-Bonin
712 Mariana arc; ED, Eastern Desert; BC, British Columbia; NQO, North Qaidam orogen.
713
714 **Figure 5.** The distributions of each cluster of major chemical compositions of **(a)** L_Srp and **(b)** H_Srp
715 identified by *k*-means analysis. BC, British Columbia; NQO, North Qaidam orogen; IBM, Izu-Bonin-
716 Mariana arc; ED, Eastern Desert; MAR, Mid-Atlantic Ridge; IOR, Indian Oceanic Ridge.
717
718

This is the peer-reviewed, final accepted version for American Mineralogist, published by the Mineralogical Society of America.
The published version is subject to change. Cite as Authors (Year) Title. American Mineralogist, in press.
DOI: https://doi.org/10.2138/am-2022-8688. http://www.minsocam.org/

719 **Figure S1.** TEM images of **(a)** chrysotile, **(b)** lizardite, and **(c)** antigorite. Inset images are schematic
720 diagrams of their respective structures, with yellow tetrahedra, purple tetrahedra, light green octahedra, and
721 purple octahedra representing SiO_4, AlO_4, $Mg(O,OH)_6$, and $Al(O,OH)_6$, respectively.

722

723 **Figure S2.** The XGBoost learning curve when using the eight selected major elements to classify L_Srp
724 and H_Srp.

725

726 **Figure S3.** Silhouette scores used to select the optimum number of clusters *k* when applying the *k*-means
727 algorithm to **(a)** L_Srp and **(b)** H_Srp.

728

729 **Figure S4.** Violin plots of the density distributions of oxide concentrations in each L_Srp cluster.

730

731 **Figure S5.** Violin plots of the density distributions of oxide concentrations in each H_Srp cluster.

732

733

734 Table 1. Summary of the performance of the XGBoost model applied to the test serpentine dataset.

Class	Precision (%)	Recall (%)	F1-score	Accuracy (%)
L_Srp	81.7	86.3	0.839	87.2
H_Srp	91.0	87.7	0.872	

735

736

Figure 1

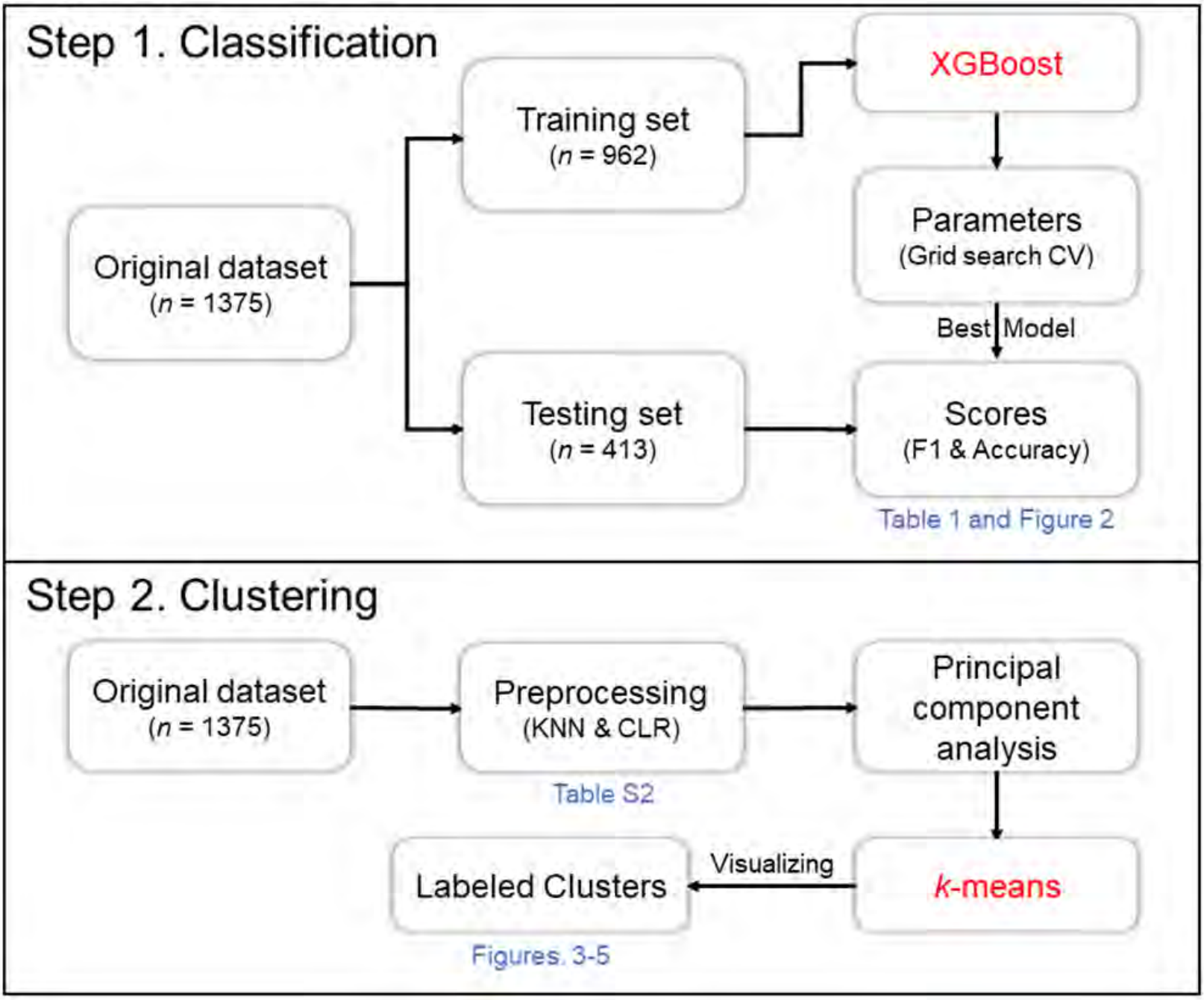

Step 1. Classification
XGBoost
Training set
(n = 962)
Parameters
(Grid search CV)
Original dataset
(n = 1375)
Best Model
Testing set
(n = 413)
Scores
(F1 & Accuracy)
Table 1 and Figure 2
Step 2. Clustering
Original dataset
(n = 1375)
Preprocessing
(KNN & CLR)
Principal component analysis
Table S2
Visualizing
Labeled Clusters
k-means
Figures. 3-5

Figure 2

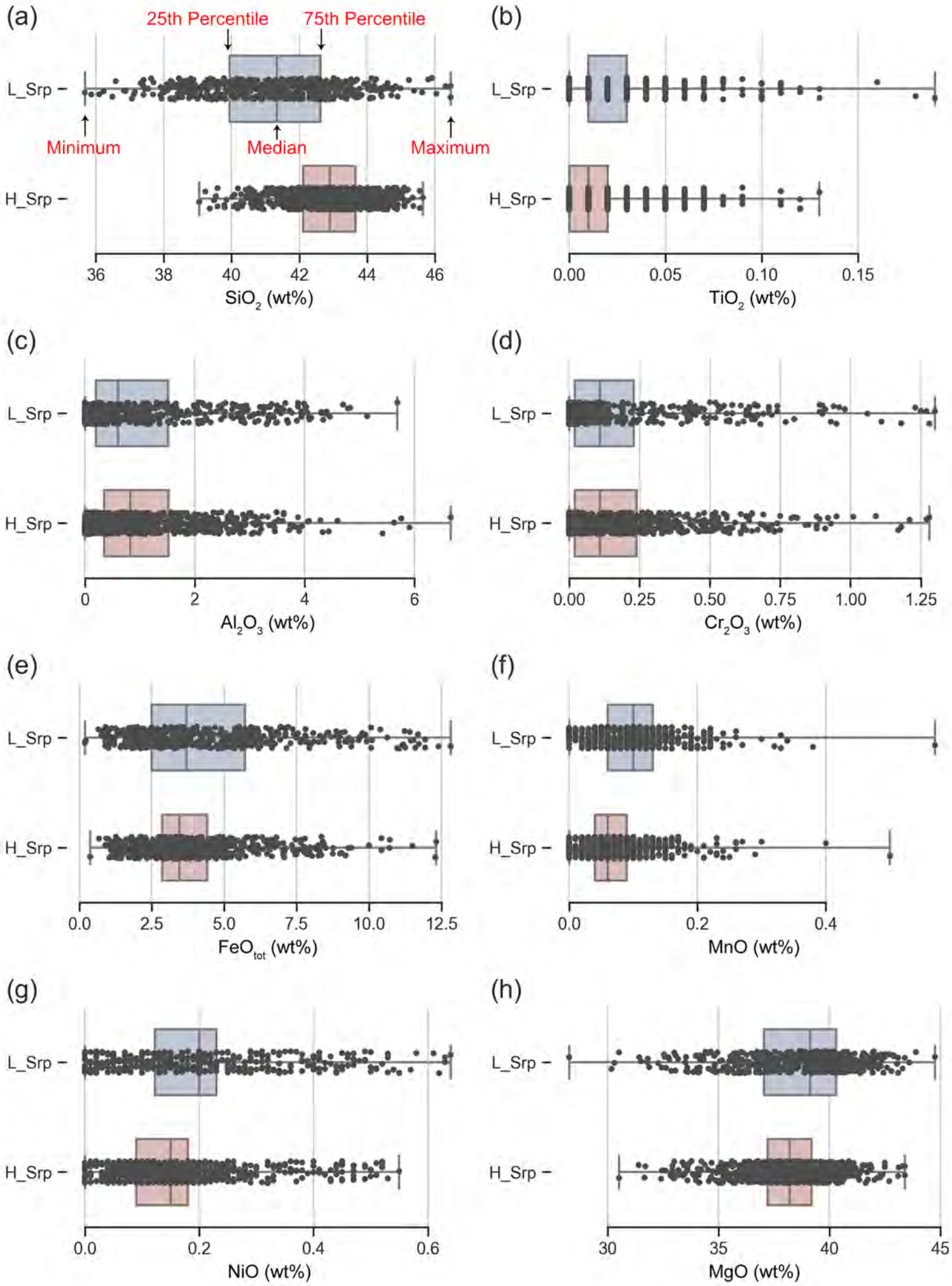
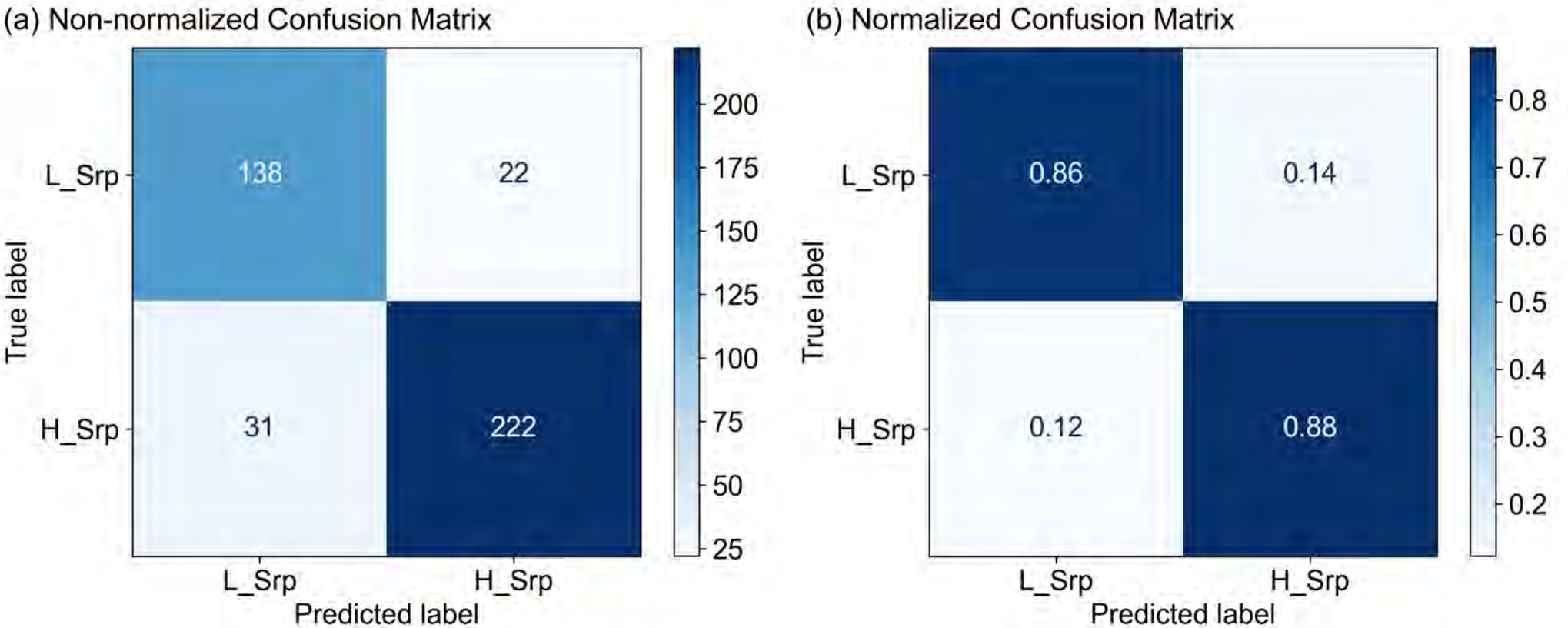

(a)
25th Percentile
75th Percentile
L_Srp
H_Srp
Minimum
Median
Maximum
36
38
40
42
44
46
SiO2 (wt%)
(b)
L_Srp
H_Srp
0.00
0.05
0.10
0.15
TiO2 (wt%)
(c)
L_Srp
H_Srp
0
2
4
6
Al2O3 (wt%)
(d)
L_Srp
H_Srp
0.00
0.25
0.50
0.75
1.00
1.25
Cr2O3 (wt%)
(e)
L_Srp
H_Srp
0.0
2.5
5.0
7.5
10.0
12.5
FeOtot (wt%)
(f)
L_Srp
H_Srp
0.0
0.2
0.4
MnO (wt%)
(g)
L_Srp
H_Srp
0.0
0.2
0.4
0.6
NiO (wt%)
(h)
L_Srp
H_Srp
30
35
40
45
MgO (wt%)

Figure 3


(a) Non-normalized Confusion Matrix
L_Srp
H_Srp
True label
138
22
31
222
L_Srp
H_Srp
Predicted label
200
175
150
125
100
75
50
25
(b) Normalized Confusion Matrix
L_Srp
H_Srp
True label
0.86
0.14
0.12
0.88
L_Srp
H_Srp
Predicted label
0.8
0.7
0.6
0.5
0.4
0.3
0.2

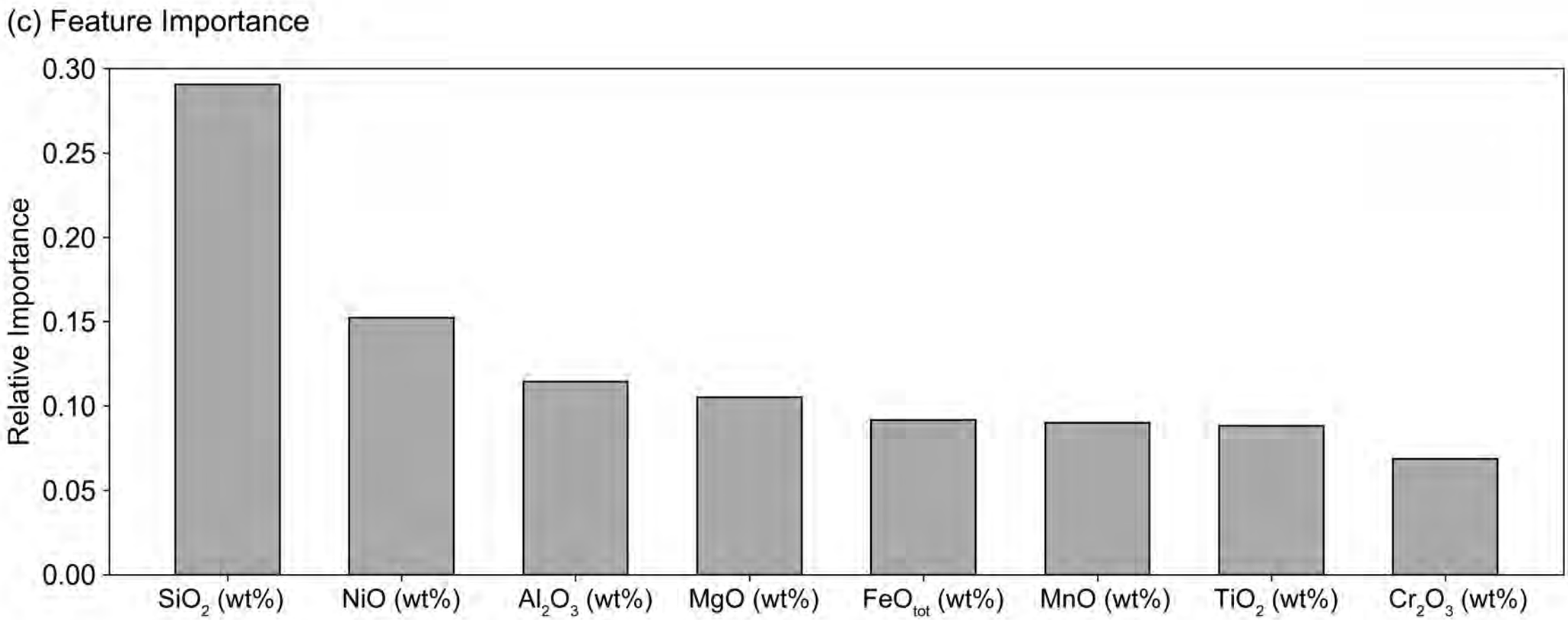

(c) Feature Importance
Relative Importance
0.30
0.25
0.20
0.15
0.10
0.05
0.00
SiO2 (wt%)
NiO (wt%)
Al2O3 (wt%)
MgO (wt%)
FeOtot (wt%)
MnO (wt%)
TiO2 (wt%)
Cr2O3 (wt%)

Figure 4

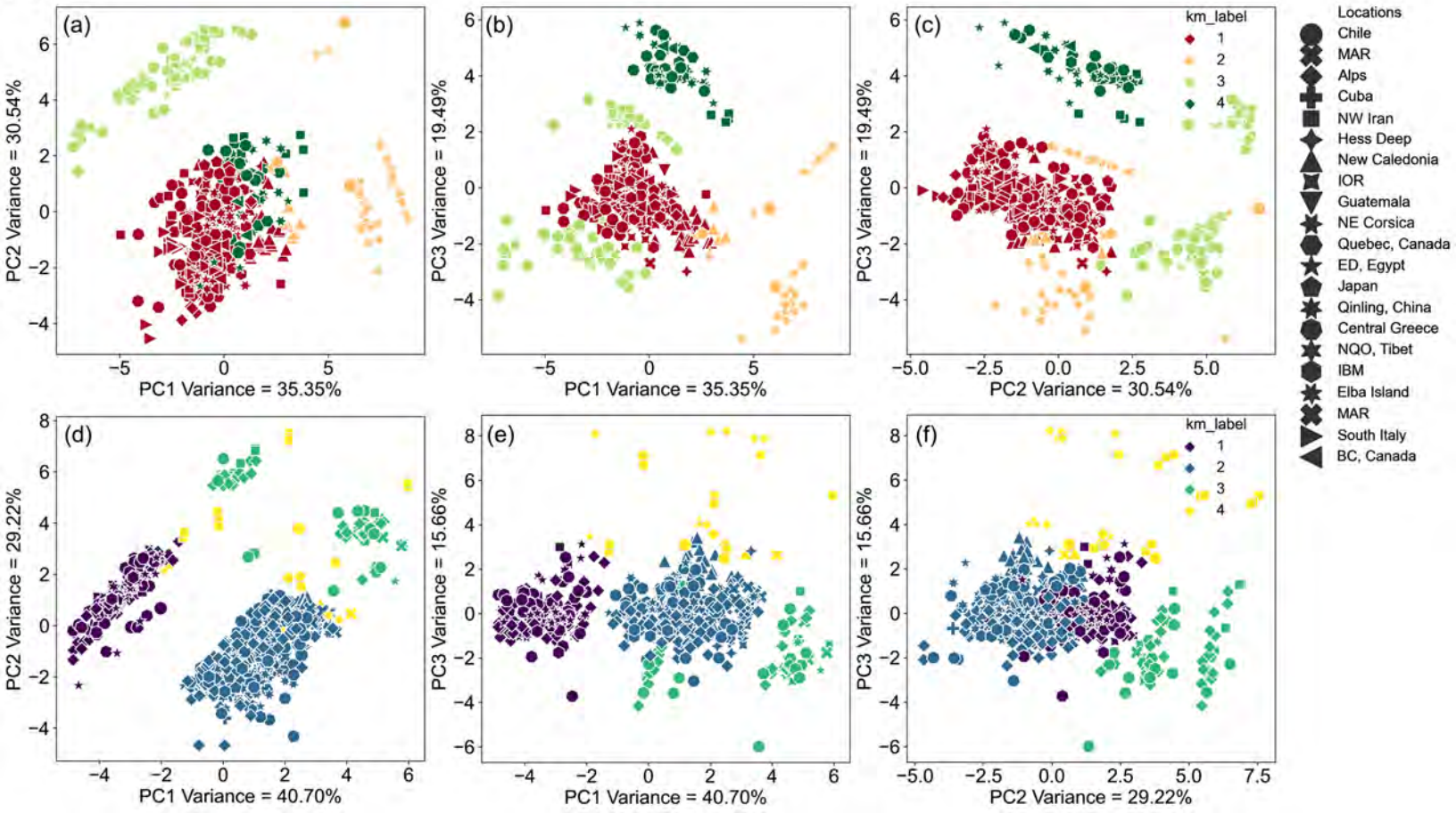
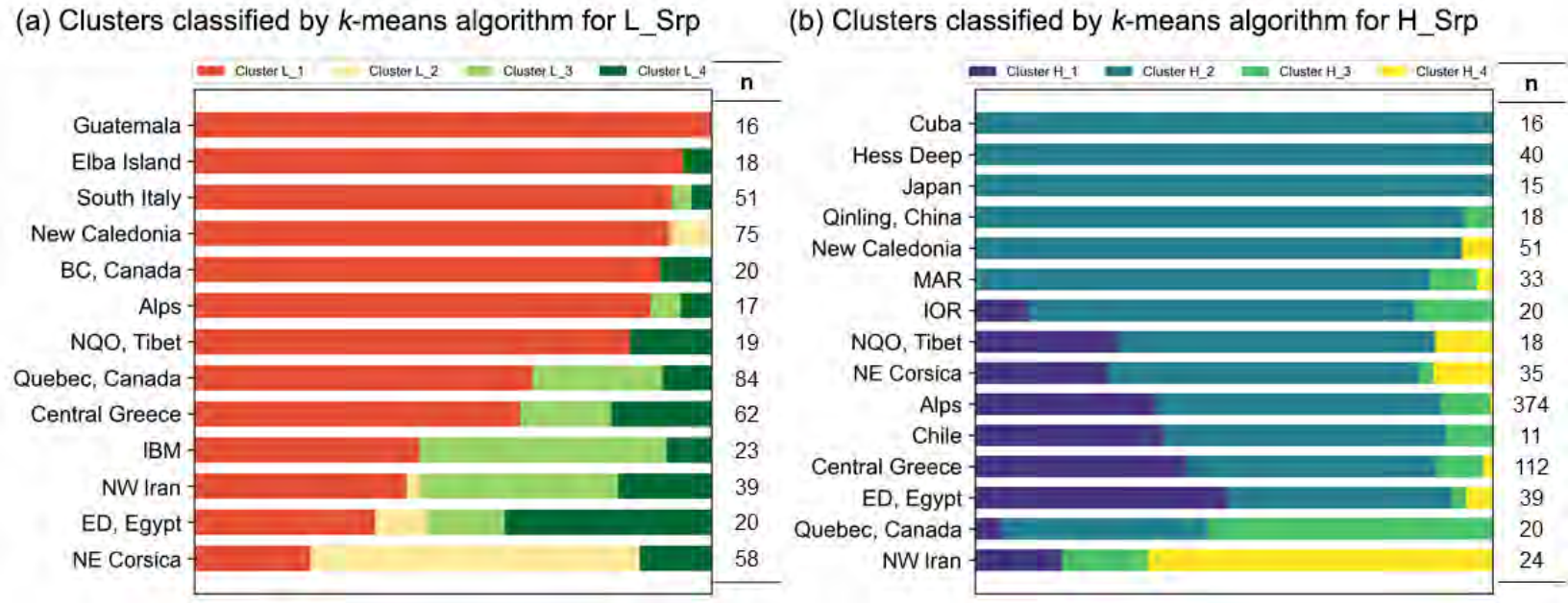

(a)
PC2 Variance = 30.54%
PC1 Variance = 35.35%
(b)
PC3 Variance = 19.49%
PC1 Variance = 35.35%
(c)
km_label
1
2
3
4
PC3 Variance = 19.49%
PC2 Variance = 30.54%
(d)
PC2 Variance = 29.22%
PC1 Variance = 40.70%
(e)
PC3 Variance = 15.66%
PC1 Variance = 40.70%
(f)
km_label
1
2
3
4
PC3 Variance = 15.66%
PC2 Variance = 29.22%
Locations
Chile
MAR
Alps
Cuba
NW Iran
Hess Deep
New Caledonia
IOR
Guatemala
NE Corsica
Quebec, Canada
ED, Egypt
Japan
Qinling, China
Central Greece
NQO, Tibet
IBM
Elba Island
MAR
South Italy
BC, Canada

Figure 5


(a) Clusters classified by k-means algorithm for L_Srp
Cluster L_1
Cluster L_2
Cluster L_3
Cluster L_4
n
Guatemala 16
Elba Island 18
South Italy 51
New Caledonia 75
BC, Canada 20
Alps 17
NQO, Tibet 19
Quebec, Canada 84
Central Greece 62
IBM 23
NW Iran 39
ED, Egypt 20
NE Corsica 58
(b) Clusters classified by k-means algorithm for H_Srp
Cluster H_1
Cluster H_2
Cluster H_3
Cluster H_4
n
Cuba 16
Hess Deep 40
Japan 15
Qinling, China 18
New Caledonia 51
MAR 33
IOR 20
NQO, Tibet 18
NE Corsica 35
Alps 374
Chile 11
Central Greece 112
ED, Egypt 39
Quebec, Canada 20
NW Iran 24