Relationships between intensity of deformation induced Cr-Al chemical zoning and geometrical properties of spinel: An approach applying machine learning analysis

Tae-Hoon Uhmb*[1] and Katsuyoshi Michibayashi[1, 2]

1. Department of Earth and Planetary Sciences, Graduate School of Environmental Studies, Nagoya University, Nagoya 464-8601, Japan

2. Volcanoes and Earth's Interior Research Center, Research Institute for Marine Geodynamics, JAMSTEC, Yokosuka 237-0061, Japan

* Corresponding author

Abstract

Chemical zoning of minerals, which is commonly caused by incomplete chemical reaction, is often utilized to study magma cooling processes, and metamorphic rock reaction paths. Cr-Al chemical zoning of spinel has been reported as induced by deformation (lattice diffusion). However, there are no studies that address the correlates of chemical distribution (intensity) with deformation, and application methods of the Cr-Al chemical zoning. In this study, we observed differences for intensities of the Cr-Al chemical zoning with various geometrical properties of spinel grains within a dunite sample in the Transition Zone in the Horoman Peridotite Complex in Hokkaido, Japan. Using machine learning analysis, we present relationships between the intensities of the Cr-Al zoning and geometrical properties of spinel grains. We examine connections between the relationships and deformation mechanisms of spinel and estimate deformation temperature based on the results. As a consequence, the spinel grains are clustered into three groups based on the chemical zoning intensity. The intensity is more importantly affected by grain size than aspect ratio and is much greater with increasing grain size. These results suggest that lattice diffusion actively contributes more to total diffusion creep of spinel than grain boundary diffusion with increasing grain size. The deformation temperature of spinel is estimated as 1250°C–1100 °C by comparing diffusion flux ratio (R_{diff}) and the spinel grains.

Plain Language Summary

Chemical zoning is a feature characterized by optical changes in color or extinction from core to rim in many species of minerals. The features have been used to estimate magmatic or metamorphic conditions they are generated, because chemical zoning is commonly developed by chemical reactions during magma cooling process and metamorphism. Previous studies reported the Cr-Al chemical zoning within spinel is derived by not cooling process or metamorphism, but deformation. But there are no studies investigating characteristics and application methods of the Cr-Al chemical zoning. In this study, we observed that spinel grains show different characteristics of the Cr-Al chemical zoning according to their size and shape within a dunite in Hokkaido, Japan. Machine learning analysis is a good tool to help us understand hidden relationships between factors which we do not know. By using the utility of the tool, we analyzed the relationships between the characteristics of the Cr-Al chemical zoning and their size and shape. Given the analyzed results, we interpreted connections between the relationships and deformation mechanisms of spinel. And we estimated deformation temperature of the spinel grains by using the connections.

Keywords: machine learning, Cr-Al chemical zoning, spinel, diffusion creep, lattice diffusion, grain boundary diffusion, deformation temperature.

Key points

1. Intensities of Cr-Al chemical zoning of spinel are different according to geometrical properties of each spinel.
2. Using machine learning analysis, we present relationships between the intensity of chemical zoning and geometrical properties of spinel.
3. The relationships are explained by diffusion flux ratio of the spinel with diffusion mechanisms, implying new application method of the Cr-Al chemical zoning.

1. Introduction

Chemical zoning is a chemical disequilibrium feature that is observed in igneous and metamorphic rocks from various mineral species (Hollister, 1970; Loomis, 1976; Tracy and McClellan, 1985; Hickmott and Shimizu, 1990; Allaby, 2013). Chemical zoning is commonly caused by incomplete chemical reactions attempting to maintain chemical equilibrium with magma or slow diffusion kinetics of cation inhibiting chemical homogenization during metamorphism or change in the chemical reaction environment (Vance, 1965; Nakamura, 1973; Loomis, 1983; Hickmott et al., 1987; Chakraborty and Ganguly, 1991; Hoskin and Schaltegger, 2003). Given the characteristics of chemical zoning, chemical zoning is commonly used as evidence for interpreting the reaction history of metamorphic rocks and cooling processes of magma and so on (Vance, 1965; Nakamura, 1973; Loomis, 1983; Hickmott et al., 1987; Chakraborty and Ganguly, 1991; Hoskin and Schaltegger, 2003).

Ozawa (1989) firstly reported that the Cr-Al chemical zoning of elongated spinel grains in naturally deformed peridotites can be derived by deformation (lattice diffusion), suggesting a model in which the chemical zoning induced by diffusion creep is derived by a difference of diffusivity between Cr and Al during deformation. Suzuki et al. (2008) demonstrated that detailed lattice diffusion processes induced the Cr-Al chemical zoning of spinel grains by measuring Cr and Al self-diffusion coefficients from the Cr-Al interdiffusion experiment. Ozawa (1983, 1989) also reported that the chemical distributions of Cr from spinel grains showing the chemical zoning are different in accordance with each spinel grain, and the differences in element concentration in the spinel grains can be used as a tool to discriminate deformation conditions. However, no studies have been conducted to analyze and interpret the chemical properties of the Cr-Al chemical zoning and distinguished application methods for the chemical properties to estimate deformation conditions.

In this study, we observed that each spinel grain showing the Cr-Al chemical zoning exhibits each different distributions of Cr and Al depending on their geometrical properties including grain size and aspect ratio from a dunite sample in Horoman Peridotite Complex. Grain size is well known for one of important factor controlling chemical distributions of chemical zoning in various minerals by providing information on growth rates of minerals during metamorphism, it has been broadly discussed

as a tool estimating metamorphic grade (Jones and Galwey, 1966; Kretz, 1973; Cashman and Ferry, 1988; Carlson, 1989; Carlson and Denison, 1992; Chernoff and Carlson, 1997; Denison and Carlson, 1997; Gaidies et al., 2008; Prenzel et al., 2009). The model for Ozawa (1989) and Suzuki et al. (2008) is demonstrating that the Cr-Al chemical zoning of spinel is induced by deformation, and aspect ratio of elongated mineral is widely used to first order factor describing deformation (Bell, 1978; George, 1978; Thayer, 1980; Misseri & Boudier, 1985; Platt and Behrmann, 1986; Goodge et al., 1993; Grujic and Mancktelow, 1995; Michibayashi et al., 2013). The observations in this study and the importance for the geometrical properties discussed from previous studies represent that there might be important relationships between the geometrical properties and the Cr-Al chemical zoning of spinel, but no study for the relationships has ever been attempted.

Machine learning is an algorithm making a model presenting relationships between features of data from a data set, well known as training data. Due to this specialty of machine learning, machine learning is often utilized in various fields to perceive hidden patterns or connections between data that have relationships we do not know (Wang et al., 2009; Angra and Ahuja, 2017; Shinde and Shah, 2018; Dhall et al., 2020; Shaukat et al., 2020; Dahiya et al., 2022). Considering the usage of machine learning, we applied the machine learning to find out the relationships that have not been studied.

Using EPMA analysis, we firstly investigate various chemical features of the Cr-Al chemical zoning observed from spinel grains. We then used machine learning algorithms to investigate the relationships between Cr-Al chemical zoning and geometrical properties of spinel grains. We distinguish the deformation temperature of the dunite sample based on the analyzed data and interpret connections between the analyzed relationships and diffusion mechanisms of spinel grains.

2. Characteristics of Cr-Al chemical zoning in spinel grains

2.1 Analytical Method

At Nagoya University in Nagoya, Japan, a scanning electron microscope (SEM) equipped with electron backscatter diffraction (EBSD) equipment (HITACHI S-3400N Type II with HKL Channel5) is used to analyze the thin section. The accelerating voltage is 20 kV, the working distance is 28 mm,

105 the sample tilt is 70°, and the low-vacuum mode is 30 Pa. EBSD data are collected using large-area
106 mapping with a step size of 10 μm which is approximately 10 times smaller than the average grain size.
107 Line scan analysis with an accelerating voltage of 20 kV, a working distance of 28 mm, and a low-
108 vacuum mode of 30 Pa is used to collect energy dispersive X-ray spectroscopy (EDS) data.

109 The chemical distribution of Cr and Al in spinel grains is investigated using both area mapping
110 and point analysis of an electron-probe microanalyzer (EPMA, JEOL JXA-8800R) at the Nagoya
111 University with 100 nA and 12 nA of beam current and 20 kV and 15 kV of accelerating voltage,
112 respectively. Area mapping is operated with a 1 μm of step size for spinel grains.

113 Image analysis software ImageJ is used to analyze the geometrical properties (grain size and
114 aspect ratio) of spinel grains. Grain boundaries of spinel grains are traced from optical
115 photomicrographs. The geometrical properties are estimated from best-fit ellipses of the tracing grain
116 boundaries computed using ImageJ software.

117

118 **2.2 Sample description**

119 The analyzed spinel grains were collected from a dunite sample located in the dunite layer
120 within the Transition Zone between the Upper and Lower Zone in the Horoman Peridotite Complex in
121 Hokkaido, northern Japan (Niida, 1974; Ozawa and Takahashi, 1995; Takazawa et al., 1999; Ozawa,
122 2004; Takahashi, 2004; Sawaguchi, 2004; Malaviarachchi et al., 2008; Malaviarachchi et al., 2010;
123 Yoshikawa et al., 2019) (Fig. 1a-b). A lineation characterized by a linear arrangement of elongated
124 spinel grains on a foliation defined by grain shape preferred orientation of olivine and diopside was
125 observed in the dunite sample. Because the chemical zoning is expected to develop along with the
126 elongated orientation of spinel grains, microstructures were analyzed from a thin section cut
127 perpendicular to the foliation and parallel to the lineation. The sample consists mainly of olivine with
128 minor spinel and diopside (Fig. 1c-d). The grain size of olivine is coarse in a range of about 100–2000
129 μm. Grain boundaries are interlobate and locally straight. Olivine grains show grain shape preferred
130 orientation with sweeping undulose extinction and deformation band. Spinel grains exhibit various
131 grain sizes ranging from 20 to 2200 μm. Spinel grains are elongated parallel to the preferred orientation

of olivine grains with various aspect ratios. Figure 2 presents the CPOs of olivine grains from the thin section. The lineation (X) and the foliation normal (Z) characterize the structural framework of CPOs. The CPOs are plotted by using one point per grain onto lower-hemisphere equal-area projection. The CPO patterns show that the [100] axes are parallel to the X, and the [010] axes are normal to the foliation, suggesting an A-type of olivine crystal fabric. These microstructures indicate that olivine grains were deformed by plastic deformation with spinel grain.

2.3 Cr-Al chemical zoning

Distributions of Cr are identified from area mapping of EPMA for spinel grains showing Cr-Al chemical zoning by greatest concentration at both tip areas of the short axis and lowest concentration at both tip areas of the long axis of spinel (Fig. 3a). The distributions of Al are characterized by a reverse relationship with the distributions of Cr. Distributions of Cr and Al are distinguished as intermediate concentrations in the center area. These characteristics are identical to those reported by Ozawa (1989) for multipolar Cr-Al chemical zoning (Fig. 4). Although most of the spinel grains exhibit an ellipsoidal shape with a preferred orientation of long axes, each spinel grain has different concentrations of Cr and Al between at center area and both tip areas (Fig. 3). In this study, we defined the intensity of Cr-Al chemical zoning as the degree of difference between Cr and Al concentrations. Figure 3 displays three representative spinel samples exhibiting different intensities of Cr-Al chemical zoning. Table 1 exhibits chemical compositions measured by point analysis of EPMA for typical points (P1 to P9 in Figure 3) within the representative spinel samples. S1 is a spinel grain with strong intensity of Cr-Al zoning and the gaps of compositions between P1 and P2 are 4.038 wt% for Cr and 3.773 wt% for Al. S2 shows relatively weak intensity of Cr-Al zoning and the differences in compositions between P4 and P5 are smaller than the gaps of S1 as 1.16 wt% for Cr and 1.308 wt% for Al. Some spinel grains, such as S3, even have homogeneous chemical distribution. The differences in Cr and Al compositions between P7 and P8 of S3 are 0.263 wt% and 0.064 wt%, respectively. Spinel grains showing the high intensity of Cr-Al zoning exhibit high differences of wt% for Cr and Al between the center area and both tip areas. There are no typical differences of wt% for those for other grains having homogenous chemical

159 distribution. These chemical distributions of Cr and Al suggest that the gap of Cr and Al distributions
160 between the center area and both tip areas reflect the intensity of Cr-Al chemical zoning, and there are
161 differences in the intensity of Cr-Al chemical zoning for each spinel grain.
162
163 **2.4 Intensity of Cr-Al chemical zoning**
164 We estimated chemical distributions of Cr and Al for each spinel grain by using EDS line scan
165 analysis to assess a quantification value for the intensity of Cr-Al chemical zoning. Given the inverse
166 chemical distribution relationship between Cr and Al, we measured EDS data for only Cr to simplify
167 data analysis. The line scan analysis is conducted along the long axis of each spinel grain from its
168 margin to the other margin (Fig. 5). We analyzed 87 spinel grains showing various geometrical
169 properties to determine the intensity of Cr-Al zoning as numerical values. The analyzed data are
170 exported by files in CSV format for each spinel grain. Each data consists of 200 values that represent
171 the chemical distribution of Cr. The values are listed in the order from the starting point to the endpoint
172 of a line scan with outliers. To correct the values by sorting outliers, the statical outlier identification
173 method is applied by using the Interquartile range (IQR) (Yang et al., 2019). IQR is defined by the
174 difference between the values of 75% (Q_3) and 25% (Q_1) of data.
175 $$IQR = Q_3 - Q_1$$
176 Outliers are the values above $Q_3 + 1.5 \times IQR$ (maximum) or under $Q_1 - 1.5 \times IQR$ (minimum).
177 The method is applied to the values of each spinel grain and the sorted outliers are replaced by the
178 average for each spinel data. The corrected values are divided into five equal parts, A through E and
179 average values are calculated for each part. The average values for A, E, and C parts represent average
180 values of Cr at the margin (starting point) area, the other margin (endpoint) area, and the center area,
181 respectively. The intensity of Cr-Al zoning is expressed by calculating differences in the average values
182 between C and A (ΔCA), and C and E parts (ΔCE) (Fig. 5). The intensity is much higher as the
183 differences are higher.
184
185 **3. Machine learning analysis**

3.1 Machine learning framework

Machine learning methods are classified into two types based on how they use data: unsupervised machine learning methods that learn patterns in feature data without label data, and supervised machine learning methods that learn functions or relationships between feature and label data considering examples of feature and label data (Bergen et al., 2019). Mean-shift Clustering, that is unsupervised machine learning method, and Decision Tree and Random Forest, which are supervised machine learning methods were used to estimate the relationship between the intensity of Cr-Al chemical zoning and geometrical properties of spinel grains. The intensity of Cr-Al zoning ($\Delta CA,\ \Delta CE$) were converted to standardized values. The standardized values for each spinel grain are projected as data points in a scatterplot. The Mean-shift Clustering is used to define the clustering patterns of the data points. The data points are set as featuring data for applying to the machine learning method and the defined types of clusters are established as label data for the feature data. The Decision Tree and Random Forest are used to recognize the relationship based on the label and feature data. The following is the detailed workflow for running machine learning: data pre-processing → data clustering → data splitting → classification and node analysis → estimating feature importance → constructing probability map.

The analysis was done in Python 3.7.0 and Jupyter Notebook 6.3.0 environment. The codes in this study are utilized to construct machine learning models and make diagrams by collaborating various libraries including scikit-learn, pandas, numpy, matplotlib, and seaborn.

3.2 Pre-processing

Data standardization is required as a stage of pre-processing before machine learning analysis to not only accelerate the calculation of machine learning but also improve model accuracy (Hsu et al., 2003). The analyzed raw data show diversity in scale. The data standardization is used to transform the analyzed raw data into a format showing a particular scale by normalizing the raw data. In this study, the data standardization is applied by using a “StandardScaler” algorithm (from Scikit-learn in a Python library). The algorithm converts original data showing variable scales into normalized data showing

that the mean value is zero and the standard deviation is one. Normalized data are obtained by removing the mean value from each original value and dividing by their standard deviation.

$$z = \frac{x - u}{s}$$

Where x, u, and s are the value, the mean, and standard deviation of original data, and z is the value of the normalized data.

3.3 Mean-shift clustering

The Mean-shift Clustering is a method that clusters data points considering locations of a maximum density of data points (Cheng, 1995; Tuzel, 2009; Zhang et al., 2018). The clustering method is built on the principle of kernel density estimation (KDE) which is a method to calculate the probability density function of data. The KDE can be written as

$$f(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$$

where $f(x)$ is density function for a given point, x_i is the point for input data, K is Kernel function (Gaussian kernel), h is bandwidth. Since h is the only variable parameter in the equation, the density function significantly depends on h.

The Mean-shift Clustering works by shifting a point from a random location to a much denser point as a starting point to find out a center point of the densest cluster of data points upon the density function, repeatedly calculating the mean coordinate of data points. The data points are clustered based on the calculated center points. Because the clustering method works on the density function, the number of clusters for the data points can be changed with a predefined value of h. The Mean-shift Clustering has the advantage of calculating the most optimal number of clusters by estimating the most suitable bandwidth for the data points, as opposed to other clustering algorithms that require a subjective decision by the user to cluster data points or set up the number of clusters. As a result, we used a "MeanShift" algorithm (from Scikit-learn in a Python library) to cluster points of the data and an "estimate bandwidth" to estimate the most optimal number of clusters.

To purify the data from outliers that may damage clustering result, outlier data are detected

and filtered by using three kinds of outlier detection methods utilizing Mean-shift, DBSCAN, and K-means algorithms, respectively, as preprocessing step before the clustering (Hautamäki et al., 2005; Çelik et al., 2011; Thang and Kim, 2011; Gan and Ng, 2017; Yang et al., 2021). We considered data commonly detected from the three detecting methods as outliers. A value of 'k' in Mean-shift outlier detection is set up on the basis of logarithmic relationship between the value of 'k' and data size referred from Yang et al. (2021). In DBSCAN outlier detection, Epsilon is estimated from K-distance graph, and min sample is set as minimum for analyzing data. The number of clusters (k) in K-means outlier detection is specified from values of Within Cluster Sum of Squared Errors (WSS) and Silhouette scores.

3.4 Data splitting

Data splitting is one of the stages that randomly divides data into train data sets and test data sets to estimate the performance of the model. The train data sets are used to train a model and the performance of the model is estimated by the test data sets. By using a "train_test_split" algorithm (from Scikit-learn in a Python library), all of the data in this study are split into train data sets and test data sets as 90% and 10% of volume size, respectively. Proportions of the label data are maintained while splitting the data sets.

3.5 Model assessment

To assess the performances of the models, we evaluated accuracy, recall, precision, f1 score, and ROC-AUC score for each model. The accuracy is the ratio of data sets correctly predicted from the trained model to test data sets. The accuracy is expressed by

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

where TP and FP are cases predicted positive and it is true and false, respectively. TN and FN are cases predicted negative and it is true and false, respectively. The accuracy is maximum as 1.0 when several correctly predicted data sets are equal to the number of the test data sets. The recall and precision are defined by

$$recall = \frac{TP}{(FN + TP)}$$

$$precision = \frac{TP}{(FP + TP)}$$

$(FN + TP)$ means numbers of actual positive cases, and $(FP + TP)$ means numbers of cases predicted positive regardless of both true and false. The recall and precision show a trade-off relationship. If the score of the recall increase, the score of the precision decrease. To define the trade-off relationship, we also calculated $f1$ score. The $f1$ score, which indicates how much the recall and precision are skewed, is a harmonic mean combining the recall and precision and expressed by

$$f1 = 2 \times \frac{precision \times recall}{precision + recall}$$

The $f1$ score has a relatively high value when the recall and precision are not biased to either side. Finally, the ROC-AUC score is calculated to estimate the prediction performance of models. The ROC curve depicts the trade-off relationships between true positive rate and false-positive rate with changing threshold. AUC stands for the area under the ROC curve. So, the ROC-AUC score represents the area under the ROC curve, and the closer to 1, the more reliable model is.

The recall, precision, $f1$ score, and ROC-AUC score are not appropriate for assessing the performances of multiclass models with more than two species of label data but are suitable for assessing those of the binary models with only two species of label data. Then, to assess the performances of the models in this study, we utilized the accuracy, recall, precision, $f1$ score, and ROC-AUC score in cases of binary models and utilized only the accuracy in cases of multiclass models. The assessments are evaluated using the "cross_val_score" algorithm (from Scikit-learn in a Python library), which calculates the average of each score from cross-validation (cv = 10), which means resampling 10 times and calculating the performances for each resample.

3.6 Criteria analysis using Decision tree

A Decision tree is a machine learning algorithm that constructs tree structures to find a model predicting the value of label data based on training features and label data (Myles et al., 2004; Cho and

290 Kurup, 2011). The tree structures consist of nodes and branches (Fig. 6). The nodes are divided into the
291 root node which is starting node, decision nodes that can be divided into nodes, and the terminal node
292 that cannot be split anymore. Each node contains information about the node's criterion, gini score
293 (gini), the number of data contained in the node (samples), the number of data belonging to each label
294 data (value), and the type of label data (class). The gini score is a score assessing the purity of the node.
295 The gini score increases with increasing species of label data included in the node. The gini score is
296 zero when the node only contains a single label data. As a result, the gini scores of all terminal nodes
297 are zero. The nodes are classified by the branches. The branch on the left and right sides represents the
298 true and false criterion of each node, respectively. The algorithm of the Decision tree works downward
299 from the top (root node) to the bottom (decision node or terminal node) by partitioning off nodes based
300 on the decision rules of each node. The classification is achieved by repeating this process and growing
301 the tree structure. In this study, we used a "graphviz" algorithm (from Scikit-learn in a Python library)
302 to visualize the tree structure as a figure and a "DecisionTreeClassifier" algorithm (from Scikit-learn in
303 a Python library) to mathematically interpret the criteria of the classification for the data sets by
304 analyzing the criterion of each node in the visualized tree structure.
305
306 **3.7 Importance assessment using Random Forest**
307 Importance assessment in machine learning is a method of calculating which feature data has
308 the greatest influence on label data as scores for each feature by using a specific machine learning
309 algorithm (Sung and Mukkamala, 2003; Hu et al., 2009; Park and Kim, 2019). The scores are assigned
310 based on the importance of the features. We can interpret relationships between feature and label data
311 using the method by referring to the feature importance. Although there are several ways to estimate
312 the feature importance according to machine learning algorithms, we used a method based on gini
313 importance and the Random Forest algorithm. Random Forest is an ensemble machine learning
314 algorithm operated by assembling a great number of specific Decision trees. A prediction of Random
315 Forest is achieved by a voting majority for the prediction results of each specific Decision tree. Gini
316 importance, which is defined by a function describing homogeneity of label data (impurity function) in

the Random Forest algorithm, is utilized to calculate node importance. Assuming a Decision tree with two terminal nodes split from one decision node, the node importance is expressed by

$$ni_j = w_j C_j - w_{left(j)} C_{left(j)} - w_{right(j)} C_{right(j)}$$

where ni_j is the importance of node j, w_j is the weighted number of samples on node j, and C_j is Gini importance of node j. $left(j)$ and $right(j)$ are left and right terminal nodes of node j, respectively. The importance of node j is increased with decreasing impurity of node j. The feature importance of feature i (fi_i) is estimated by dividing the sum of the importance of nodes split by feature i by the sum of the importance of all nodes.

$$fi_i = \frac{\sum_{j=node\ j\ split\ by\ feature\ i} ni_j}{\sum_{k \in all\ nodes} ni_k}$$

To express fi_i as value suited to Random Forest level, the feature importance is transformed to normalized value by dividing fi_i by the sum of all importance of features and the normalized feature importance on each tree is divided by the number of all Decision trees.

$$normfi_i = \frac{fi_i}{\sum_{l \in all\ features} fi_l}$$

$$RFfi_i = \frac{\sum_{t \in all\ trees} normfi_{it}}{T}$$

The $normfi_i$ and $RFfi_i$ are normalized feature importance and final feature importance fitted to Random Forest for feature i, respectively. To estimate the importance of features classified from the Random Forest model, we used a "feature importances" algorithm with a "RandomForestClassifier" algorithm from Scikit-learn in a Python library.

3.8 Probability map constructed by logistic regression

Logistic regression is another popular machine learning algorithm that conducts classification by calculating the probability for each event using a logistic or sigmoid function via a non-linear transformation of ordinary least squares for linear regression. The logistic function distinguishes between the logistic regression and linear regression models. The logistic function is expressed by

$$y = \frac{1}{1 + e^{-x}}$$

y is the probability and x is a weighted linear combination of feature data. The probability is assumed to limit between 0 and 1, as a binary system. x is defined as

$$x = b_0 + b_1 z_1 + b_2 z_2 + \cdots + b_n z_n$$

where b_0 is the intercept, n is the number of variables, b_i (i =1, 2, 3, …, n) is the corresponding coefficients, and z_i is the independent variable for feature data. Considering the model, we can identify the probability of each feature data for label data showing the binary system as a statistical value. Relationships between feature and label data can be expressed as a probability map by projecting the probabilities against feature parameters. In this study, we used a "LogisticRegression" algorithm (from Scikit-learn in a Python library) to find out appropriate models representing relationships between feature data and label data. The verified model is visualized by constructing a probability map for each feature data to interpret the importance of features classified from the Random Forest model.

4. Result

4. 1 Scatterplot for the intensity of Cr-Al chemical zoning

Figure 7 shows a scatterplot for the ΔCA and ΔCE, which represent the intensity of Cr-Al zoning. Each data point is standardized from the pre-processing stage. The closer the ΔCA and ΔCE to 0, the more evenly distributed Cr is within the spinel grains. The intensity of Cr-Al zoning is intense with increasing the ΔCA and ΔCE to a positive value. The data points are distributed from −2 to 3 for the ΔCA and from −3 to 3 from the ΔCE. Although the ΔCA is diversely distributed with the ΔCE, the distribution of the data points indicates proportional relationship between the ΔCA and ΔCE. Data points that ΔCA is less than 0 present ΔCE less than 1. Data points that ΔCE is more than 1 show ΔCE more than 0. This relationship is originally derived by characteristics of Cr-Al chemical zoning. Considering the features of Cr-Al chemical zoning, Cr is symmetrically distributed along the long axis and minimally concentrated at both tip areas. The ΔCA and ΔCE proportionally change in accordance with the intensity of Cr-Al zoning.

4.2 Criteria for clustering

Fig. 8 displays results for outlier detection utilizing the three detecting methods as mentioned in section 3.3. We considered that final outliers are data commonly detected from each of the methods, and filtered the final outliers. The purified data points are clustered using the Mean-shift Clustering with 0.938 of bandwidth calculated by the 'estimate bandwidth' algorithm (Fig. 9). The data points are divided into 3 clusters as cluster 1 to cluster 3. The numbers of data points of the 1, 2, and 3 cluster are 54, 17, and 11, respectively. To interpret clustering criteria, the ΔCA and ΔCE data (feature data) and types of the clusters (label data) for each spinel grain are split into train data set including 73 spinel data and test data set including 9 spinel data, and the data are classified using the Decision Tree algorithm. We estimated only accuracy to assess the performance of this model because this is a multiclass model as mentioned in section 3.5. The estimated average accuracy is 0.95. The visualized tree structure trained by the train data set is shown in Figure 10. There are 11 nodes containing a root node, 4 decision nodes, and 6 terminal nodes. Each cluster is classified by the 6 terminal nodes. Important nodes are root or decision nodes containing criterion that generates decision or terminal nodes consisting of a large number of samples and low gini score. Important nodes for cluster 1 are the root node 1 and decision node 2 deriving the terminal node 4. The root node 1 and decision nodes 3 and 6 are important nodes for cluster 2 characterized by the terminal node 11. Important nodes for cluster 3 are the root node 1 and decision node 3 defining the terminal node 7. The criteria of the important nodes for each of the cluster are expressed by

$$cluster\ 1 = \{\Delta CA \leq 0.176, \Delta CE \leq 0.311\}$$

$$cluster\ 2 = \{\Delta CA > 0.176,\ -0.527 < \Delta CE \leq 0.826\}$$

$$cluster\ 3 = \{\Delta CA > 0.176, \Delta CE > 0.826\}$$

Since both the ΔCA and ΔCE represent the difference of Cr and Al distributions between the tip and center area, the criteria can be simplified by considering the intersection of the criteria for each cluster.

$$cluster\ 1 = \{\Delta \leq 0.176\}$$

$$cluster\ 2 = \{0.176 < \Delta \leq 0.826\}$$

$$cluster\ 3 = \{\Delta > 0.826\}$$

Δ is the intersection of the criteria of the ΔCA and ΔCE. The simplified criteria numerically explain that spinel grains are systemically clustered into three clusters based on the difference of Cr and Al distributions between the tip and center area and the difference are high in an order of cluster 3, cluster 2, and cluster 1. Because the difference represents the intensity of the Cr-Al chemical zoning, each cluster can be interpreted as spinel grains having homogeneous chemical distribution for cluster 1, the weak intensity of Cr-Al chemical zoning for cluster 2, and the strong intensity of Cr-Al chemical zoning for cluster 3.

4.3 Feature importance

To determine which geometrical factor controls the intensity of Cr-Al zoning, more importantly, the feature importance is calculated using the feature data that includes information for the ΔCA and ΔCE, grain sizes (R), and aspect ratios (L) of spinel grains and label data composed by types of clusters. Figure 11 depicts the significance of each feature data. Because this model is also multiclass, we calculated only accuracy to assess the performance of this model. The calculated average accuracy is 0.9139. The ΔCA and ΔCE are the feature data displaying the first and second highest scores of feature importance, indicating that the ΔCA and ΔCE have the greatest impact on determining types of the clusters. Grain size is the third most important, and at the same time, the most important among the geometrical properties. The grain size score indicates that grain size is closely related to the types of clusters. Aspect ratio is the feature data showing the lowest score of feature importance. The score of aspect ratio reflects that there are relatively low relationships between aspect ratio and the intensity of Cr-Al chemical zoning.

4.4 Probability map

We inspect probability maps by using the logistic regression method to examine the relative relationships between each feature and the Cr-Al chemical zoning in greater detail. Figure 12 depicts the probability maps for features including the ΔCA and ΔCE, grain size (R), and aspect ratio (L). To

represent the ΔCA and ΔCE data as a unified value, the ΔCA and ΔCE is simplified by calculating average. The average of the ΔCA and ΔCE is $aver\Delta$. The probability is expressed as values between 1 to 0 and the closer it is to 1, the more likely it is to be included in clusters 2 and 3. In contrast, the closer the value is to 0, the higher the possibility of being included in cluster 1. Given the criteria for clusters 1, 2, and 3 in section 4.2, high values of the probability imply that possibility that the Cr-Al chemical zoning can be observed from the spinel grains is high. The probability maps are constructed by considering species of features as three cases; the $aver\Delta$ and grain size, the $aver\Delta$ and aspect ratio, and aspect ratio and grain size. The performances of these models are estimated by evaluating scores for accuracy, recall, precision, $f1$, and ROC-AUC, because this is binary model as mentioned in section 3.5. Scores representing performances for each probability map are described in Table 2

The probability map for the $aver\Delta$ and grain size shows that the probability increases with increasing $aver\Delta$ and grain size. The variances in probability with changing grain size are less than those with changing the $aver\Delta$. Intervals between each of contour line are relatively narrow, suggesting that the criteria for distinction are relatively clear. The scatter plot for the $aver\Delta$ and grain size also represent the meaning of the narrow intervals of the contour lines. In the probability map for the $aver\Delta$ and aspect ratio, although changes in the probability for the $aver\Delta$ display a similar trend with those in the probability map for the $aver\Delta$ and grain size, there are no relationship for aspect ratio. The gaps between the contour lines are also similar to those for the probability map for the $aver\Delta$ and grain size. In the probability map for the aspect ratio and grain size, the probability is increase with increasing aspect ratio and grain size. Variances of the probability with changing grain size are bigger than those with changing aspect ratio. The intervals are much wider than the intervals of the probability maps for the $aver\Delta$ and grain size, and the $aver\Delta$ and aspect ratio.

5. Discussion

5.1 Interpretation for the analyzed data

In sections 4.3 and 4.4, the machine learning analysis exhibits the feature importance representing many important factors for the intensity of Cr-Al zoning and the probability maps showing

448 relative relationships between the factors and the intensity of Cr-Al zoning. We interpret the analyzed
449 data and the relative relationships in this section.
450 In Figure 11, the ΔCA and ΔCE show the first and second highest scores for feature
451 importance representing the greatest influence on identifying the types of the clusters. Because the data
452 points are originally clustered based on the ΔCA and ΔCE, it is obvious that they are the most
453 important factors to the intensity of Cr-Al zoning. The connection is also represented by the probability
454 maps for the ΔCA and ΔCE (Fig. 12a-b). Feature importance indicates grain size is the third most
455 important factor to the intensity of Cr-Al zoning. The importance score of grain size is smaller than
456 those of the ΔCA and ΔCE and greater than that of the aspect ratio. This importance of grain size can
457 be also certified from the probability maps (Fig. 12a, c). Although the variances of the probability with
458 varying grain sizes are much smaller than those with varying $aver\Delta$ in the probability map for the
459 $aver\Delta$ and grain size (Fig. 12a), the variances of the probability with altering grain size are much
460 bigger that those with altering aspect ratio in the probability map for the aspect ratio and grain size (Fig.
461 12c). Both probability maps for grain size show that the probability increases with increasing grain size,
462 implying that the Cr-Al chemical zoning is more easily observed with increasing grain size. Figure 13a
463 depicts the grain size distribution for each cluster type. The grain size trend in clusters 1, 2, and 3 reflects
464 the relationships between grain size and the Cr-Al zoning. The aspect ratio has the lowest importance
465 score, about 10 times lower than ΔCA and ΔCE and 5 times lower than that of grain size. The
466 probability maps for aspect ratio also reflect the gaps of importance scores for aspect ratio (Fig. 12b-c).
467 In the probability map for the $aver\Delta$ and aspect ratio (Fig. 12b), even if variances of the probability
468 is relatively large with changing the $aver\Delta$, there are no variances for the probability with changing
469 aspect ratio because the $aver\Delta$ much affect to the Cr-Al zoning compared to aspect ratio as about 10
470 times enormously. The probability map for aspect ratio and grain size, on the other hand, displays
471 effects for aspect ratio due to relatively small differences in importance scores (Fig. 12c). The
472 probability increases with increasing aspect ratio, implying that possibility of observing the Cr-Al
473 chemical zoning increases with increasing aspect ratio. The intervals of contour lines are wider than
474 those of the other maps because the criteria for aspect ratio constructing the probability map is relatively

475 unclearness. These relationships are displayed in the scatter plot for aspect ratio and grain size, and box
476 plots showing aspect ratios of spinel grains for each type of cluster (Figure 13b). Although the aspect
477 ratio for each type of cluster exhibits variation and the aspect ratio for cluster 1 is the lowest, there is
478 no critical difference in aspect ratio between clusters 1, 2, and 3.
479
480 **5.2 Interpretation for relationships between grain size and the Cr-Al chemical zoning**
481 As stated in section 5.1, grain size is proportional to the intensity of Cr-Al chemical zoning.
482 Based on Ozawa's model, we assumed that spinel grains with Cr-Al chemical zoning are deformed by
483 lattice diffusion to interpret the relationship. Considering the assumption, the relationship between grain
484 size and the intensity of Cr-Al chemical zoning can be interpreted as lattice diffusion becoming
485 significant with increasing grain size of spinel. This interpretation can be explained by the ratio between
486 grain boundary diffusion flux and lattice diffusion flux. The ratio (r_{diff}) is expressed by
487 $$r_{diff} = \frac{D_b \delta_b}{D_l R}$$
488 D_b and D_l is grain boundary diffusion coefficient and lattice diffusion coefficient, respectively. δ_b is
489 grain boundary width and R is grain size. The r_{diff}, which is inverse proportional to grain size, and
490 is related to dominant types of diffusion creep mechanisms. Lattice diffusion becomes important with
491 decreasing the r_{diff}, and grain boundary diffusion become dominant with increasing the r_{diff}.
492 Considering the ratio, D_b, D_l, and δ_b are relatively constant between minerals of the same species
493 within the same rock sample and the only variable for the r_{diff} becomes R. Based on the relations,
494 lattice diffusion is more active with increasing grain size. Considering the Ozawa's model that the Cr-
495 Al chemical zoning in spinel grain is kinetic demixing caused by the difference of diffusivity between
496 Cr and Al during cation transport through the lattice, lattice diffusion in spinel grain is associated with
497 flux of Cr and Al. This means that the more lattice diffusion is active, the more cation is transported
498 through the lattice and the high flux of cation derives the high intensity of Cr-Al zoning. Then, the
499 relationship between grain size and the intensity of Cr-Al chemical zoning suggests that the intensity is
500 increase with increasing grain size because lattice diffusion is much more active in coarser spinel grains.

For types of clusters, lattice diffusion is most intensely activated in cluster 3, intermediately in cluster 2, and relatively not activated in cluster1.

5.3 The R_{diff} and deformation temperature

The ratio for diffusion flux (r_{diff}) is controlled by grain boundary diffusion coefficient (D_b) and lattice diffusion coefficient (D_l), grain boundary width (δ_b), and grain size (R). And the D_b and D_l are changed with varying temperatures. Thus, we can calculate variations of the r_{diff} following changing of temperature and the R. There were studies investigating specific values of the r_{diff} representing the transition of diffusion creep mechanisms from grain boundary diffusion to lattice diffusion. Considering the specific values of the r_{diff}, critical grain size (R_c), which is grain size deriving the transition of diffusion creep mechanisms, can be determined according to each temperature condition. By comparing the calculated R_c and grain size distribution for each type of the clusters, we estimated the deformation temperature of spinel grain.

The D_b is grain boundary diffusion coefficient for the slowest atom on grain boundary of spinel grain and the D_l is lattice diffusion coefficient for the slowest atom within the lattice of spinel grain, because the slowest atoms control diffusion creep of mineral. To calculate the r_{diff}, we supposed that O and Cr are the slowest atom in grain boundary and lattice of spinel grain, respectively (Joesten, 1991; Suzuki et al., 2008; Nakakoji and Hiraga, 2018). The δ_b is assumed as 1 nm. Because there has been no direct estimation of the O grain boundary diffusion coefficient for spinel grain, the O grain boundary diffusion coefficient is derived by referring to results of previous studies (Oishi and Ando, 1975; Reddy and Cooper, 1981; Ando and Oishi, 1983), in which O self-diffusion coefficient of polycrystalline spinel is approximately 4 orders faster than that of single spinel crystal due to diffusion of O improved along grain boundaries. Given the result, we calculated the approximate O grain boundary diffusion coefficient by multiplying 4 orders by the O self-diffusion coefficient estimated by the result. According to Suzuki et al. (2008), the approximately calculated O grain boundary diffusion coefficient is about 5 orders greater than the O self-diffusion coefficient estimated by Reddy and Cooper

527 (1981) and Ando and Oishi (1983). The Cr lattice diffusivity is computed by extrapolating Cr self-
528 diffusion coefficients data from Suzuki et al. (2008). We evaluated the r_{diff} for various temperatures
529 and grain sizes by applying the calculated D_b and D_l to the r_{diff} (Table3).

530 Swaroop et al. (2005) reported that the R_c deriving a transition from grain boundary diffusion
531 to lattice diffusion is determined by

532 $$R_c \approx \frac{1.2 D_b \delta}{D_l}$$

533 The specific value of the r_{diff} can be calculated by modifying the equation.

534 $$0.83 \approx \frac{D_b \delta}{D_l R} = r_{diff}$$

535 Shibutani et al. (1998) reviewed that there are three types of diffusion processes controlling cavity
536 growth derived by diffusion, depending on the r_{diff}. The r_{diff} for the types of processes are
537 introduced as

538 (a) $r_{diff} > 1$: Grain boundary diffusion dominant

539 (b) $0.1 < r_{diff} < 1$: Grain boundary diffusion + Lattice diffusion

540 (c) $r_{diff} < 0.1$: Lattice diffusion dominant

541 Since this model indicates that the transition occurs from the (b) range of the r_{diff}, the transition may
542 have occurred from a grain size achieving $r_{diff} = 1$.

543 We estimated the r_{diff} (Table 3) and the R_c (Table 4) with various temperature and grain
544 sizes based on the two models. Both the model represents that the r_{diff} is decrease with increasing
545 grain size and temperature, and the R_c decrease with increasing temperature, suggesting that the
546 relative significance of lattice diffusion to total diffusion creep increases with increasing temperature
547 (Fig. 14).

548 The activity of lattice diffusion for types of the clusters discussed in section 4.3 show that
549 lattice diffusion begins to be active from cluster 2. Deformation temperature can be estimated by
550 comparing the grain sizes of Q1 (73.5 μm) and minimum (22 μm) of cluster 2, and the critical grain
551 sizes in Table 4. By considering the difference between the Q1 and minimum grain sizes, deformation

552 temperature is expressed as ranges. For the equation of Swaroop et al. (2005), the temperature
553 representing the critical grain size closest to the grain size of Q1 is 1100°C and the temperature for the
554 grain size of minimum is 1250°C. Then, the deformation temperature range is estimated as 1250°C–
555 1100°C. The estimated deformation temperature calculated by supposing the equations of Shibutani et
556 al. (1998) is also 1250°C–1100°C.

557 The CPO patterns of the olivine grains and the P-T trajectory of the Horoman Peridotite
558 Complex based on chemical compositional zoning of pyroxenes and whole-rock compositions (Ozawa
559 and Takahashi, 1995 Takahashi, 2004) support the proposal that the estimated deformation temperature
560 range is 1250°C–1100°C. The A-type CPO of the olivine grains represents that the olivine grains are
561 deformed in high-temperature conditions (approximately > 1100°C). Since we supposed the
562 deformation of the olivine grains forming A-type CPO coincides with the deformation of the spinel
563 grains in section 2.1 based on their microstructures, the range of the estimated temperature is consistent.
564 Based on the P-T trajectory of the Horoman Peridotite Complex, the equilibrium temperature of spinel
565 peridotite for the Transition Zone including the dunite sample is estimated to be the intermediate
566 temperature between the equilibrium temperature of spinel peridotite of the Upper (1150°C–1100°C)
567 and Lower Zone (950°C–900°C). According to Takahashi (2004), the temperature difference between
568 the Upper and the Lower Zones is gradual. Given the sampling area that is located very near to the
569 Upper Zone, the equilibrium temperature for the dunite sample may be closer to that of the Upper Zone
570 than the Lower Zone, implying that the estimated temperature range is consistent.

571

572 **5.4 Interpretation for relationships between aspect ratio and the Cr-Al chemical zoning**

573 Although aspect ratio is less correlated with the intensity of Cr-Al chemical zoning than grain
574 size, aspect ratio also represents a relationship with the intensity (Fig. 12c). The probability map and
575 scatter plot in Fig. 12c describe two of characteristics for aspect ratio, which are that the probability
576 decrease with decreasing aspect ratio and aspect ratio for spinel grains smaller than about 100 μm is
577 relatively low, representing that spinel grains having homogeneous chemical distribution show
578 relatively fine grain size and low aspect ratio. According to the characteristics, the relationship between

aspect ratio and the chemical zoning is correlated with grain size. The effect of the rounding process introduced by Toriumi (1987), Okamoto and Michibayashi (2005), and Uhmb and Michibayashi (2022) was thought to be responsible for the relationship between chemical zoning, aspect ratio, and grain size. Isolated mineral inclusions tend to round out to a spherical shape to minimize their surface energy, which is referred to as the rounding process, during the post-deformational annealing stage, and the effects of the rounding process are related to grain size and temperature because the rounding process is controlled by diffusion creep. The rounding process has greater effects as grain size decreases, increasing deformation temperature, and increasing annealing time. Because the deformation temperature and annealing time are identical along with the dunite sample, the effect of the rounding process on the sample is only activated by the following variable of grain size. Given the effect of the rounding process, we thought that the fine spinel grains became more round in shape than coarse grains during the annealing stage and the trend of decreasing the probability with decreasing aspect ratio is derived from the relationship between grain size and the chemical zoning, that the chemical zoning is less observable in finer grains.

6. Summary and conclusion

We investigated the relationship between the intensity of Cr-Al chemical zoning (ΔCA and ΔCE) and the geometrical properties of spinel grains collected from a dunite sample within the dunite layer in the Transition Zone of the Horoman Peridotite Complex, northern Japan using unsupervised (Mean-shift Clustering) and supervised (Decision tree, Random Forest, and Logistic regression) machine learning analyses.

- Using Mean-shift Clustering, the ΔCA and ΔCE data were clustered into three kinds of clusters (cluster 1, 2, and 3).
- The Decision tree classification revealed that the spinel grains are systematically clustered based on degree of the intensity of Cr-Al chemical zoning.
- The feature importance analyzed by Random Forest method feature importance analysis suggested that grain size is more important factor closely related to the intensity of Cr-Al

chemical zoning than aspect ratio.

- The probability maps constructed by the Logistic regression displayed the relationships between the trends of the geometrical properties and the intensity of Cr-Al chemical zoning.

The relationship is interpreted as lattice diffusion is most strongly activated in cluster 1, intermediately in cluster 2, and not activated in cluster 1, implying that lattice diffusion is much more active in coarser spinel grains and there is a critical grain size switching diffusion mechanism. Given the relationships, we estimated the critical grain size (d_c) for the spinel grains with various temperature conditions (Fig. 14). The comparison between the estimated d_c and observed grain size from the dunite sample suggests that deformation temperature of the spinel grains within the sample is 1250°C–1100°C.

Our newly suggested model in this study, which is constructed by machine learning analyses, describes the relationships between the intensity of the Cr-Al chemical zoning and the geometrical properties. The model allows us to understand motive of various intensities of the Cr-Al chemical zoning and diffusion mechanisms of spinel grains according to their geometrical properties, and further derive a new application method of the chemical zoning for estimating deformation temperature of spinel grains. It will be achievable to not only find out new model for the chemical zoning of spinel grains but improve the model, as more other data of spinel are applied to the machine learning analysis.

Acknowledgments

To compute all of data in this study, we utilized Python 3.7.0 and Jupyter Notebook 6.3.0 with libraries including pandas, numpy, sklearn, scikit, and matplotlib. This study was supported by a student grant awarded to T.U. by the Graduate School of Environmental Studies (Nagoya University), the Fujii International Scholarship, the THERS Interdisciplinary Frontier Next Generation Researcher fund, and a research grant awarded to K.M. by the Japan Society for the Promotion of Science (Kiban-S 16H06347).

Data Availability Statement

All data for machine learning analysis are available as the format using in this study from

633 Figshare (https://doi.org/10.6084/m9.figshare.20131157.v1). The all data include not only modeling
634 codes described by python languages, but also analyzing data for spinel grains.
635
636 **References**

637 Allaby, M. (Ed.). (2013). A dictionary of geology and earth sciences. Oxford University Press.
638 Ando, K., & Oishi, Y. (1983). Effect of Ratio of Surface Area to Volume on Oxygen Self-Diffusion
639 Coefficients Determined for Crushed MgO-Al2O3 Spinels. *Journal of the American Ceramic*
640 *Society*, *66*(8), C-131.
641 Angra, S., & Ahuja, S. (2017, March). Machine learning and its applications: A review. In 2017
642 international conference on big data analytics and computational intelligence (ICBDAC) (pp. 57-
643 60). IEEE.
644 Bell, T. (1978). Progressive deformation and reorientation of fold axes in a ductile mylonite zone: the
645 Woodroffe thrust. Tectonophysics, 44(1-4), 285-320.
646 Bergen, K. J., Johnson, P. A., de Hoop, M. V., & Beroza, G. C. (2019). Machine learning for data-driven
647 discovery in solid Earth geoscience. *Science*, *363*(6433), eaau0323.
648 Carlson, W. D. (1989). The significance of intergranular diffusion to the mechanisms and kinetics of
649 porphyroblast crystallization. Contributions to Mineralogy and Petrology, 103(1), 1-24.
650 Carlson, W. D., & Denison, C. (1992). Mechanisms of porphyroblast crystallization: results from high-
651 resolution computed X-ray tomography. Science, 257(5074), 1236-1239.
652 Cashman, K. V., & Ferry, J. M. (1988). Crystal size distribution (CSD) in rocks and the kinetics and
653 dynamics of crystallization. Contributions to Mineralogy and Petrology, 99(4), 401-415.
654 Çelik, M., Dadaşer-Çelik, F., & Dokuz, A. Ş. (2011, June). Anomaly detection in temperature data using
655 DBSCAN algorithm. In 2011 international symposium on innovations in intelligent systems and
656 applications (pp. 91-95). IEEE.
657 Chakraborty, S., & Ganguly, J. (1991). Compositional zoning and cation diffusion in garnets. In
658 Diffusion, atomic ordering, and mass transport (pp. 120-175). Springer, New York, NY.
659 Cheng, Y. (1995). Mean shift, mode seeking, and clustering. *IEEE transactions on pattern analysis and*

machine intelligence, *17*(8), 790-799.

Chernoff, C. B., & Carlson, W. D. (1997). Disequilibrium for Ca during growth of pelitic garnet. Journal of metamorphic Geology, 15(4), 421-438.

Cho, J. H., & Kurup, P. U. (2011). Decision tree approach for classification and dimensionality reduction of electronic nose data. *Sensors and Actuators B: Chemical*, *160*(1), 542-548.

Dahiya, N., Gupta, S., & Singh, S. (2022). A Review Paper on Machine Learning Applications, Advantages, and Techniques. ECS Transactions, 107(1), 6137.

DENISON*, C., Carlson, W. D., & Ketcham, R. A. (1997). Three-dimensional quantitative textural analysis of metamorphic rocks using high-resolution computed X-ray tomography: Part I. Methods and techniques. Journal of Metamorphic Geology, 15(1), 29-44.

Dhall, D., Kaur, R., & Juneja, M. (2020). Machine learning: a review of the algorithms and its applications. Proceedings of ICRIC 2019, 47-63.

Gaidies, F., De Capitani, C., & Abart, R. (2008). THERIA_G: a software program to numerically model prograde garnet growth. Contributions to Mineralogy and Petrology, 155(5), 657-671.

Gan, G., & Ng, M. K. P. (2017). K-means clustering with outlier removal. Pattern Recognition Letters, 90, 8-14.

GEORGE JR, R. P. (1978). Structural petrology of the Olympus ultramafic complex in the Troodos ophiolite, Cyprus. Geological Society of America Bulletin, 89(6), 845-865.

Goodge, J. W., Hansen, V. L., Peacock, S. M., Smith, B. K., & Walker, N. W. (1993). Kinematic evolution of the Miller Range shear zone, central Transantarctic Mountains, Antarctica, and implications for Neoproterozoic to early Paleozoic tectonics of the East Antarctic margin of Gondwana. Tectonics, 12(6), 1460-1478.

Grujic, D., & Mancktelow, N. S. (1995). Folds with axes parallel to the extension direction: an experimental study. Journal of Structural Geology, 17(2), 279-291.

Hautamäki, V., Cherednichenko, S., Kärkkäinen, I., Kinnunen, T., & Fränti, P. (2005, June). Improving k-means by outlier removal. In Scandinavian conference on image analysis (pp. 978-987). Springer, Berlin, Heidelberg.

687 Hickmott, D. D., Shimizu, N., Spear, F. S., & Selverstone, J. (1987). Trace-element zoning in a
688 metamorphic garnet. Geology, 15(6), 573-576.
689 Hickmott, D. D., & Shimizu, N. (1990). Trace element zoning in garnet from the Kwoiek Area, British
690 Columbia: disequilibrium partitioning during garnet growth?. Contributions to Mineralogy and
691 Petrology, 104(6), 619-630.
692 Hollister, L. S. (1970). Origin, mechanism, and consequences of compositional sector-zoning in
693 staurolite. American Mineralogist: Journal of Earth and Planetary Materials, 55(5-6), 742-766.
694 Hoskin, P. W., & Schaltegger, U. (2003). The composition of zircon and igneous and metamorphic
695 petrogenesis. Reviews in mineralogy and geochemistry, 53(1), 27-62.
696 Hsu, C. W., Chang, C. C., & Lin, C. J. (2003). A practical guide to support vector classification.
697 Hu, H. Y., Lee, Y. C., Yen, T. M., & Tsai, C. H. (2009). Using BPNN and DEMATEL to modify
698 importance–performance analysis model–A study of the computer industry. *Expert systems with*
699 *applications*, *36*(6), 9969-9979.
700 Joesten, R. (1991). Grain-boundary diffusion kinetics in silicate and oxide minerals. In *Diffusion,*
701 *atomic ordering, and mass transport* (pp. 345-395). Springer, New York, NY.
702 JONES, K. A., & GALWEY, A. K. (1966). Size distribution, composition, and growth kinetics of garnet
703 crystals in some metamorphic rocks from the west of Ireland. Quarterly Journal of the Geological
704 Society, 122(1-4), 29-44.
705 Kretz, R. (1974). Some models for the rate of crystallization of garnet in metamorphic
706 rocks. Lithos, 7(3), 123-131.
707 Loomis, T. P. (1976). Irreversible reactions in high-grade metapelitic rocks. Journal of Petrology, 17(4),
708 559-588.
709 Loomis, T. P. (1983). Compositional zoning of crystals: a record of growth and reaction history. In
710 Kinetics and equilibrium in mineral reactions (pp. 1-60). Springer, New York, NY.
711 Malaviarachchi, S. P., Makishima, A., & Nakamura, E. (2010). Melt–peridotite reactions and fluid
712 metasomatism in the upper mantle, revealed from the geochemistry of peridotite and gabbro from
713 the Horoman peridotite massif, Japan. Journal of Petrology, 51(7), 1417-1445.

714 Malaviarachchi, S. P., Makishima, A., Tanimoto, M., Kuritani, T., & Nakamura, E. (2008). Highly
715 unradiogenic lead isotope ratios from the Horoman peridotite in Japan. Nature Geoscience, 1(12),
716 859-863.
717 Michibayashi, K., Suzuki, M., & Komori, N. (2013). Progressive deformation partitioning and
718 recrystallization of olivine in the lithospheric mantle. Tectonophysics, 587, 79-88.
719 Misseri, M., & Boudier, F. (1985). Structures in the Canyon Mountain ophiolite indicate an island-arc
720 intrusion. Tectonophysics, 120(3-4), 191-209.
721 Myles, A. J., Feudale, R. N., Liu, Y., Woody, N. A., & Brown, S. D. (2004). An introduction to decision
722 tree modeling. *Journal of Chemometrics: A Journal of the Chemometrics Society*, *18*(6), 275-285.
723 Nakakoji, T., & Hiraga, T. (2018). Diffusion creep and grain growth in forsterite+ 20 vol% enstatite
724 aggregates: 2. Their common diffusional mechanism and its consequence for weak-temperature-
725 dependent viscosity. *Journal of Geophysical Research: Solid Earth*, *123*(11), 9513-9527.
726 Nakamura, Y. (1973). Origin of sector-zoning of igneous clinopyroxenes. American Mineralogist:
727 Journal of Earth and Planetary Materials, 58(11-12), 986-990.
728 Niida, K. I. Y. O. A. K. I. (1974). Structure of the Horoman ultramafic massif of the Hidaka metamorphic
729 belt in Hokkaido, Japan. J. Geol. Soc. Japan, 80, 31-44.
730 Okamoto, A., & Michibayashi, K. (2005). Progressive shape evolution of a mineral inclusion under
731 differential stress at high temperature: Example of garnet inclusions within a granulite-facies
732 quartzite from the Lützow-Holm Complex, East Antarctica. *Journal of Geophysical Research:*
733 *Solid Earth*, *110*(B11).
734 Ozawa, K. (1989). Stress-induced Al–Cr zoning of spinel in deformed peridotites. *Nature*, *338*(6211),
735 141-144.
736 Ozawa, K., & Takahashi, N. (1995). PT history of a mantle diapir: the Horoman peridotite complex,
737 Hokkaido, northern Japan. *Contributions to Mineralogy and Petrology*, *120*(3), 223-248.
738 Ozawa, K. (2004). Thermal history of the Horoman peridotite complex: a record of thermal perturbation
739 in the lithospheric mantle. *Journal of Petrology*, *45*(2), 253-273.
740 Park, S., & Kim, J. (2019). Landslide susceptibility mapping based on random forest and boosted

741 regression tree models, and a comparison of their performance. *Applied Sciences*, *9*(5), 942.
742 Platt, J. P., & Behrmann, J. H. (1986). Structures and fabrics in a crustal-scale shear zone, Betic
743 Cordillera, SE Spain. Journal of Structural Geology, 8(1), 15-33.
744 Prenzel, J., Abart, R., & Keller, L. (2009). Complex chemical zoning in eclogite facies garnet reaction
745 rims: the role of grain boundary diffusion. Mineralogy and Petrology, 95(3), 303-313.
746 Shaukat, K., Luo, S., Varadharajan, V., Hameed, I. A., Chen, S., Liu, D., & Li, J. (2020). Performance
747 comparison and current challenges of using machine learning techniques in
748 cybersecurity. Energies, 13(10), 2509.
749 Shibutani, T., Kitamura, T., & Ohtani, R. (1998). Creep cavity growth under interaction between lattice
750 diffusion and grain-boundary diffusion. *Metallurgical and Materials Transactions A*, *29*(10),
751 2533-2542.
752 Shinde, P. P., & Shah, S. (2018, August). A review of machine learning and deep learning applications.
753 In 2018 Fourth international conference on computing communication control and automation
754 (ICCUBEA) (pp. 1-6). IEEE.
755 Sung, A. H., & Mukkamala, S. (2003). Identifying important features for intrusion detection using
756 support vector machines and neural networks. In *2003 Symposium on Applications and the Internet,*
757 *2003. Proceedings.* (pp. 209-216). IEEE.
758 Suzuki, A. M., Yasuda, A., & Ozawa, K. (2008). Cr and Al diffusion in chromite spinel: experimental
759 determination and its implication for diffusion creep. *Physics and Chemistry of Minerals*, *35*(8),
760 433-445.
761 Swaroop, S., Kilo, M., Argirusis, C., Borchardt, G., & Chokshi, A. H. (2005). Lattice and grain
762 boundary diffusion of cations in 3YTZ analyzed using SIMS. *Acta materialia*, *53*(19), 4975-4985.
763 Sawaguchi, T. (2004). Deformation history and exhumation process of the Horoman Peridotite Complex,
764 Hokkaido, Japan. *Tectonophysics*, *379*(1-4), 109-126.
765 Takazawa, E., Frey, F. A., Shimizu, N., Saal, A., & Obata, M. (1999). Polybaric petrogenesis of mafic
766 layers in the Horoman peridotite complex, Japan. *Journal of Petrology*, *40*(12), 1827-1851.
767 Thang, T. M., & Kim, J. (2011, April). The anomaly detection by using dbscan clustering with multiple

parameters. In 2011 International Conference on Information Science and Applications (pp. 1-5). IEEE.

Thayer, T. P. (1963). Flow-layering in alpine peridotite-gabbro complexes. Spec Pap Mineral Soc Am, 1, 55-62.

Toriumi, M. (1987). Progressive deformation and annealing of quartz inclusion in porphyroblastic feldspar during synmetamorphic non-coaxial deformation. *The Journal of the Japanese Association of Mineralogists, Petrologists and Economic Geologists*, *82*(4), 123-131.

Tracy, R. J., & McLellan, E. L. (1985). A natural example of the kinetic controls of compositional and textural equilibration. In Metamorphic reactions (pp. 118-137). Springer, New York, NY.

Tuzel, O., Porikli, F., & Meer, P. (2009). Kernel methods for weakly supervised mean shift clustering. In *2009 IEEE 12th International Conference on Computer Vision* (pp. 48-55). IEEE.

Uhmb, T. H., & Michibayashi, K. (2022). A shape-change model for isolated K-feldspar inclusions within a shear zone developed in the Teshima granite, Ryoke metamorphic belt, Japan: Estimation of the duration of deformation in a natural shear zone. Tectonophysics, 229229.

Vance, J. A. (1965). Zoning in igneous plagioclase: patchy zoning. The Journal of Geology, 73(4), 636-651.

Wang, H., Ma, C., & Zhou, L. (2009, December). A brief review of machine learning and its application. In 2009 international conference on information engineering and computer science (pp. 1-4). IEEE.

Yang, J., Rahardja, S., & Fränti, P. (2019). Outlier detection: how to threshold outlier scores?. In *Proceedings of the international conference on artificial intelligence, information processing and cloud computing* (pp. 1-6).

Yang, J., Rahardja, S., & Fränti, P. (2021). Mean-shift outlier detection and filtering. Pattern Recognition, 115, 107874.

Yoshikawa, M., Niida, K., & Green, D. H. (2019). Dunite channels within a harzburgite layer from the Horoman peridotite complex, Japan: Possible pathway for magmas. Island Arc, 28(1), e12279.

Zhang, P., Du, K., Tannant, D. D., Zhu, H., & Zheng, W. (2018). Automated method for extracting and analysing the rock discontinuities from point clouds based on digital surface model of rock mass.

795 Engineering Geology, 239, 109-118.

Figure 1. (a) Modified Ozawa and Takahashi (1995) geological map of the Horoman Peridotite Complex in Hokkaido, Japan (1995). The sampling area is the dunite layer near the Upper Zone within the Transition Zone. (b) Photograph of the dunite sample outcrop. The foliation is marked by dash lines. (c) Representative phase maps for the dunite sample were obtained by SEM-EBSD. (d) Optical photomicrographs (PPL) of the spinel grains within the phase map of (c). The elongated spinel grains show various grain sizes and aspect ratios.

Figure 2. CPOs of olivine grains in the dunite sample. Equal-area lower-hemisphere projection. Right projections are point plots for all olivine data onto each axis and left projections are contour plots for the same data.

Figure 3. EPMA-WDS mapping data for three representative spinel samples showing each different intensity of Cr-Al chemical zoning. Spinel samples exhibited relatively (a) strong (S1), (b) intermediate (S2), and (c) weak (S3) intensity of Cr-Al chemical zoning. White circles nominated from P1 to P9 within each spinel sample are the location conducting point analysis. Detailed point analysis data are set out in Table 1.

Figure 4. Schematic figure of multipolar Cr-Al chemical zoning in two dimensions reported by Ozawa (1989), modified from Suzuki et al. (2008).

Figure 5. Representative EDS line scan data for Cr. The EDS data observed along the long axis of the elongated spinel from tip to opposite tip is divided into five equal parts from A to E. $\Delta \boldsymbol{CA}$ and $\Delta \boldsymbol{CE}$ are differences between the average intensity of E part (Aver. C part) and the average intensity of each tip part (Aver. C part and Aver. E part).

Figure 6. Schematic illustration explaining the tree structure of the Decision tree.

Figure 7. Scatterplot of the ΔCA versus the ΔCE for the spinel samples.

Figure 8. Results of outlier detection for the spinel data. Each of detection method is (a) Mean-shift, (b) DBSCAN, and (c) K-means, respectively. Final filtered data are outliers commonly detected from the each of detection mothods.

Figure 9. Clustering result for the data points. The data points are divided into 3 clusters. The marks for each cluster are displayed.

Figure 10. The visualized tree structure classifies the spinel data. Each node numbered from 1 to 11 contains information including gini score (gini), number of data (sample), numbers of data for each label data (value), and type of label data (class). A scatterplot located on the upper left side shows the region of clusters classified by this tree structure.

Figure 11. A bar chart describing the feature importance of the feature data. Values beside each bar are the scores for feature importance.

Figure 12. Probability maps for (a) the averΔ and grain size, (b) the averΔ and aspect ratio, and (c) aspect ratio and grain size. The probability is expressed in brighter color as it increased, and in a darker color as it decreased. Red lines are contour lines for 0.8, 0.6, 0.4, 0.2 of probabilities. Clustered spinel data points are overlaid on the probability maps.

Figure 13. Box plots for (a) grain size and (b) aspect ratio in different types of clusters. Values in each box plot are (a) grain size and (b) aspect ratio of maximum, Q3 (75 %), median, Q1 (25 %), and minimum. Bar charts within (a) show logarithmic grain size distributions for each cluster.

Figure 14. The relative importance of the diffusion creeps mechanism to total diffusion creep of

chromite spinel with various grain size and temperature conditions were calculated using Shibutani et al model. (1998). Grain boundary and lattice diffusion, are denoted by GB and L, respectively. The black dash lines represent boundaries changing dominant diffusion mechanisms for the model of Shibutani et al. (1998). A white dash line is a transition boundary of diffusion mechanisms from grain boundary diffusion to lattice diffusion based on the model of Swaroop et al. (2005). Characters colored red represent the intensity of Cr-Al chemical zoning.

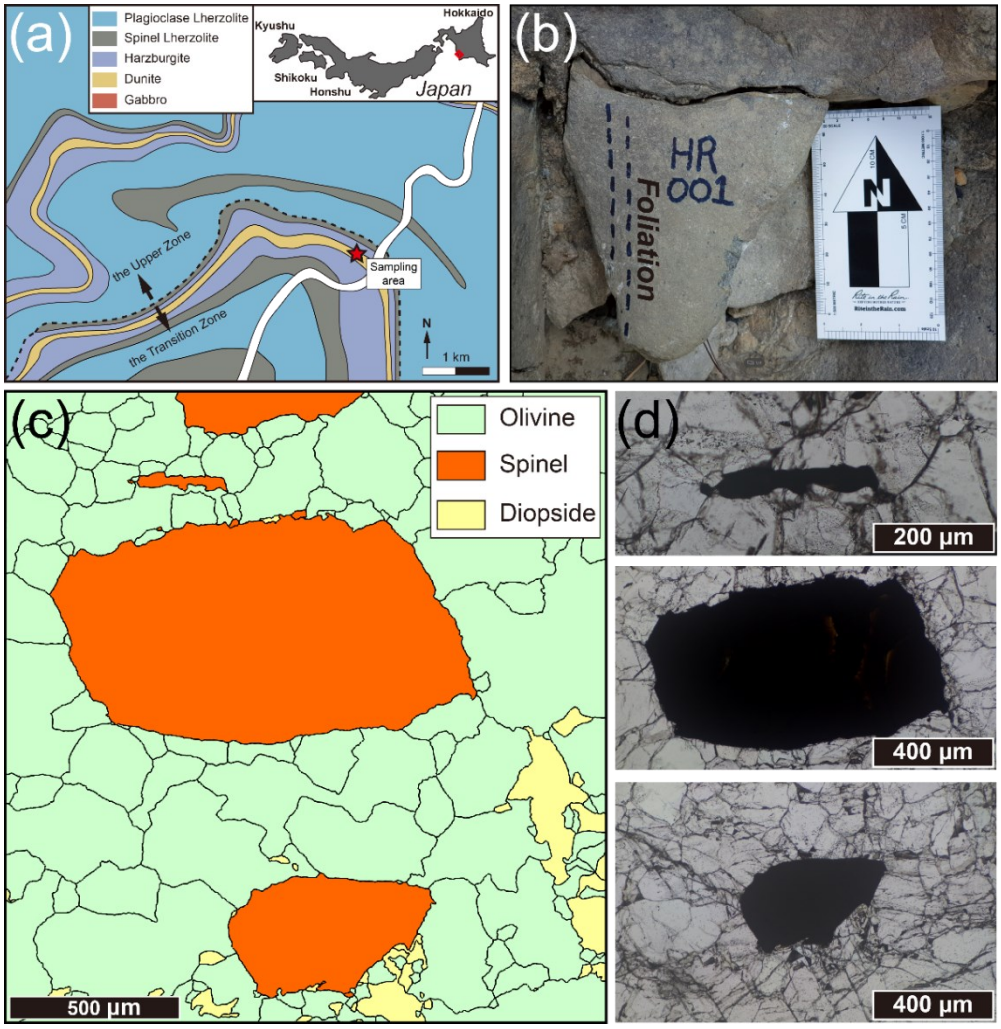

Figure 1. (a) Modified Ozawa and Takahashi (1995) geological map of the Horoman Peridotite Complex in Hokkaido, Japan (1995). The sampling area is the dunite layer near the Upper Zone within the Transition Zone. (b) Photograph of the dunite sample outcrop. The foliation is marked by dash lines. (c) Representative phase maps for the dunite sample were obtained by SEM-EBSD. (d) Optical photomicrographs (PPL) of the spinel grains within the phase map of (c). The elongated spinel grains show various grain sizes and aspect ratios.

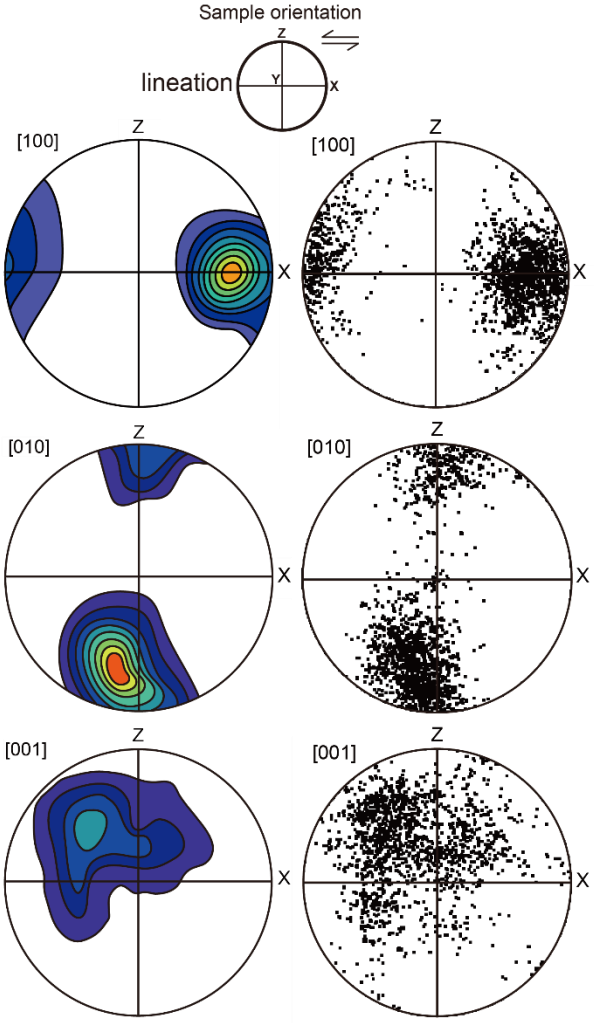

Figure 2. CPOs of olivine grains in the dunite sample. Equal-area lower-hemisphere projection. Right projections are point plots for all olivine data onto each axis and left projections are contour plots for the same data.

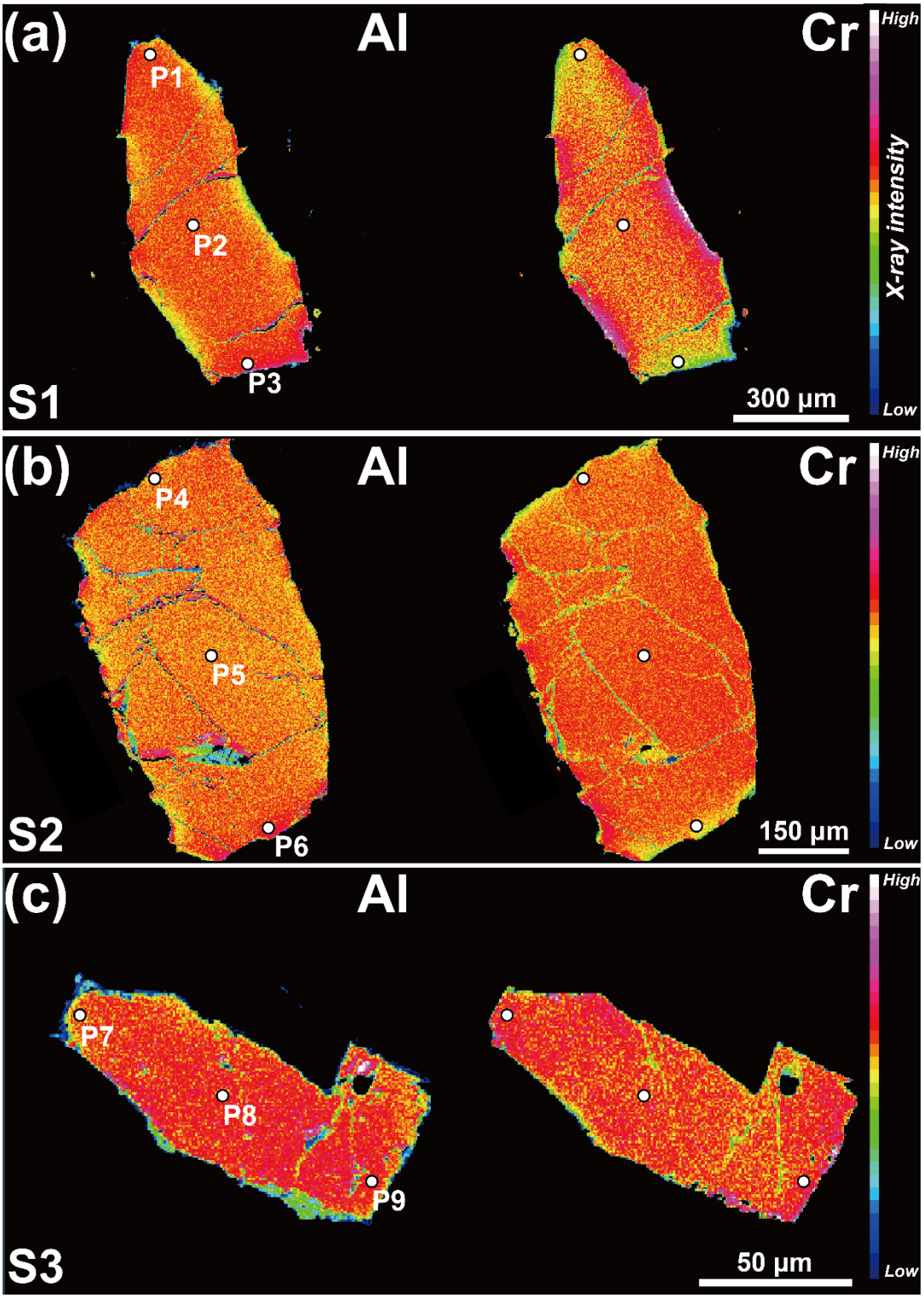

Figure 3. EPMA-WDS mapping data for three representative spinel samples showing each different intensity of Cr-Al chemical zoning. Spinel samples exhibited relatively (a) strong (S1), (b) intermediate (S2), and (c) weak (S3) intensity of Cr-Al chemical zoning. White circles nominated from P1 to P9 within each spinel sample are the location conducting point analysis. Detailed point analysis data are set out in Table 1.

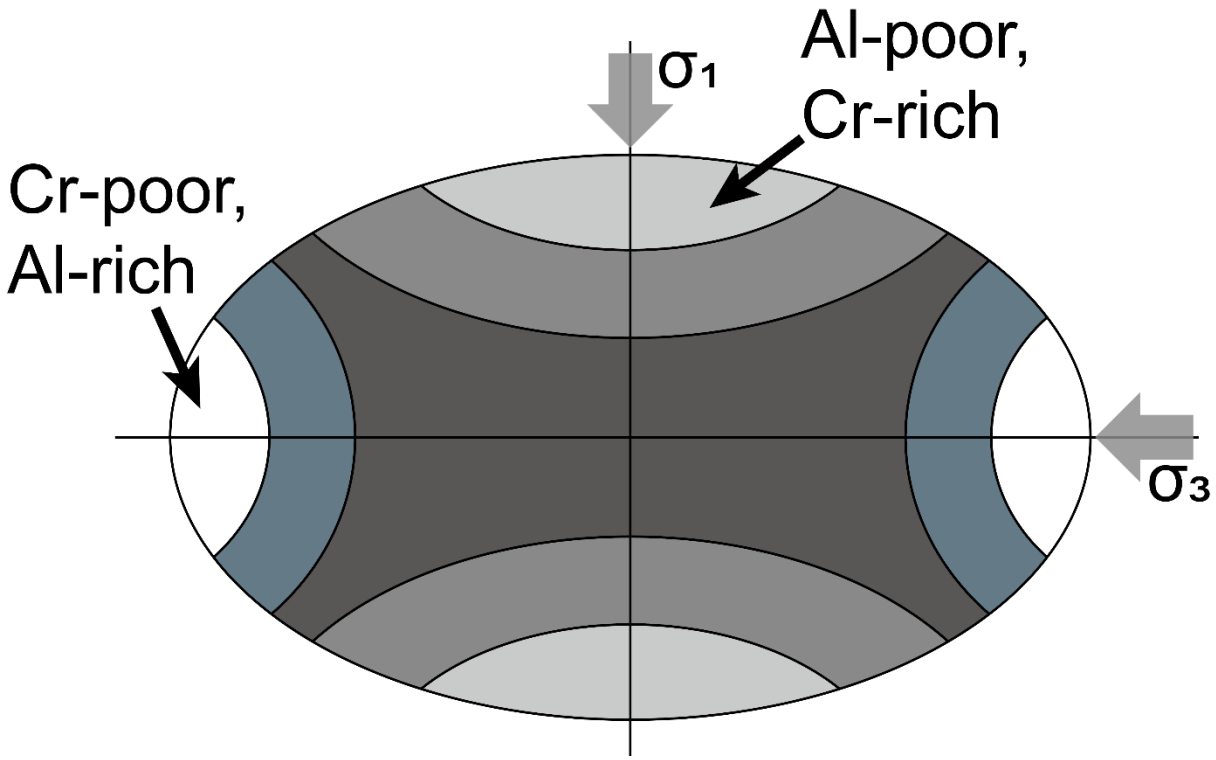

Figure 4. Schematic figure of multipolar Cr-Al chemical zoning in two dimensions reported by Ozawa (1989), modified from Suzuki et al. (2008).

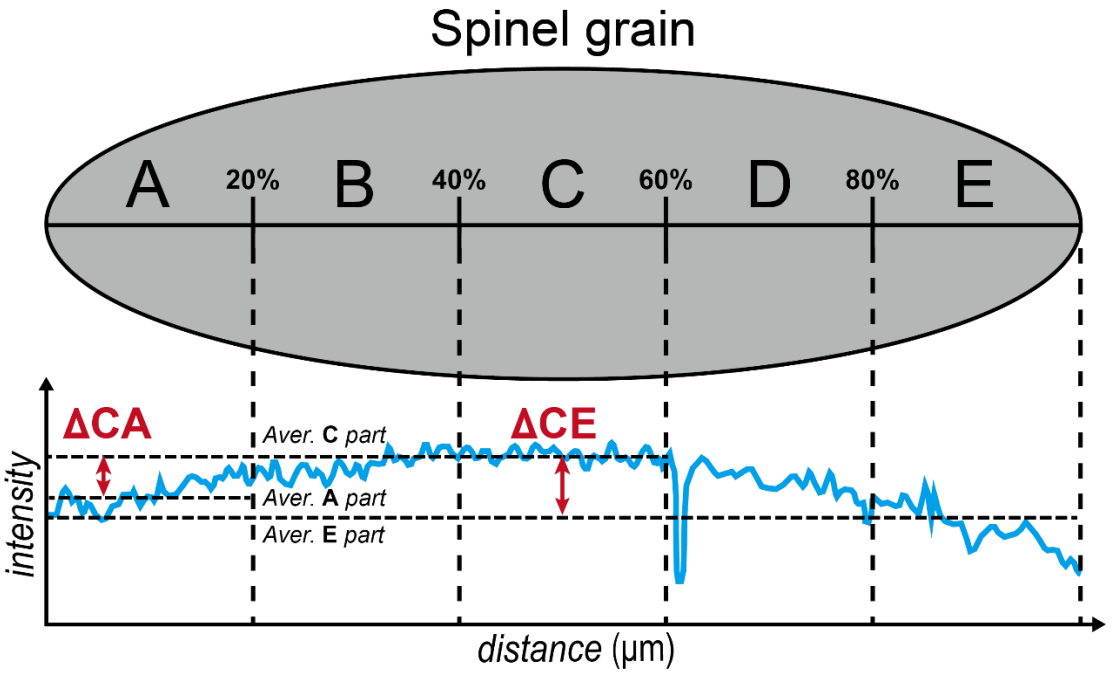

Figure 5. Representative EDS line scan data for Cr. The EDS data observed along the long axis of the elongated spinel from tip to opposite tip is divided into five equal parts from A to E. ΔCA and ΔCE are differences between the average intensity of E part (Aver. C part) and the average intensity of each tip part (Aver. C part and Aver. E part).

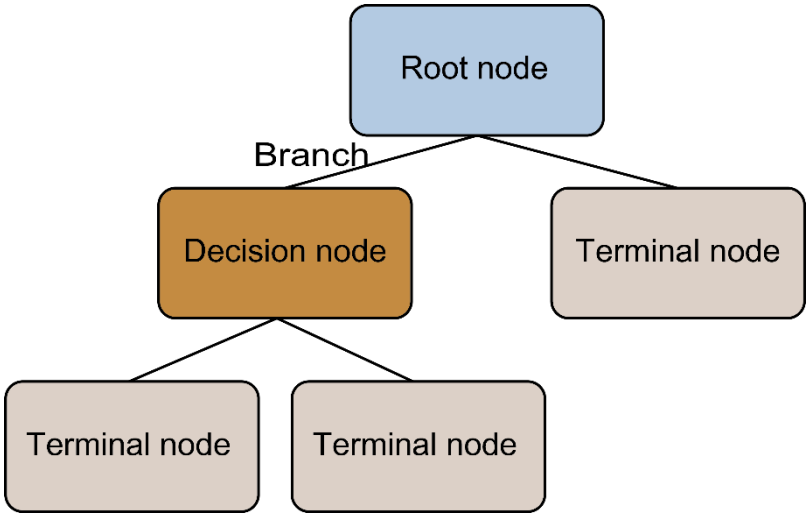

Figure 6. Schematic illustration explaining the tree structure of the Decision tree.

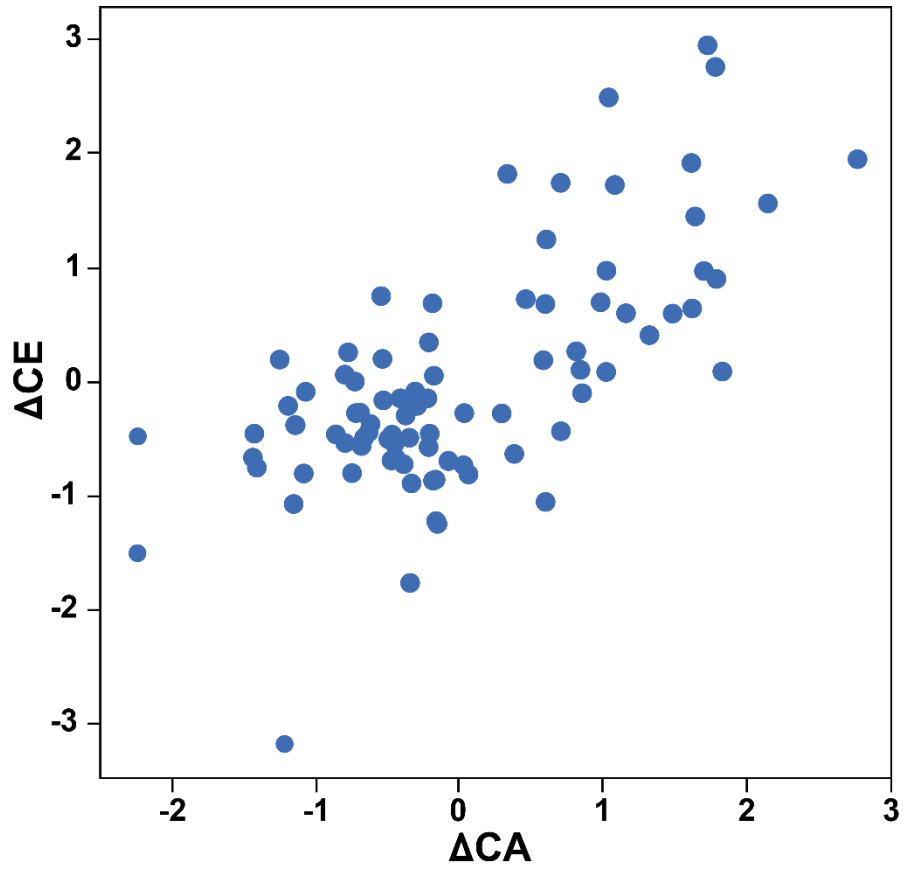

Figure 7. Scatterplot of the ΔCA versus the ΔCE for the spinel samples.

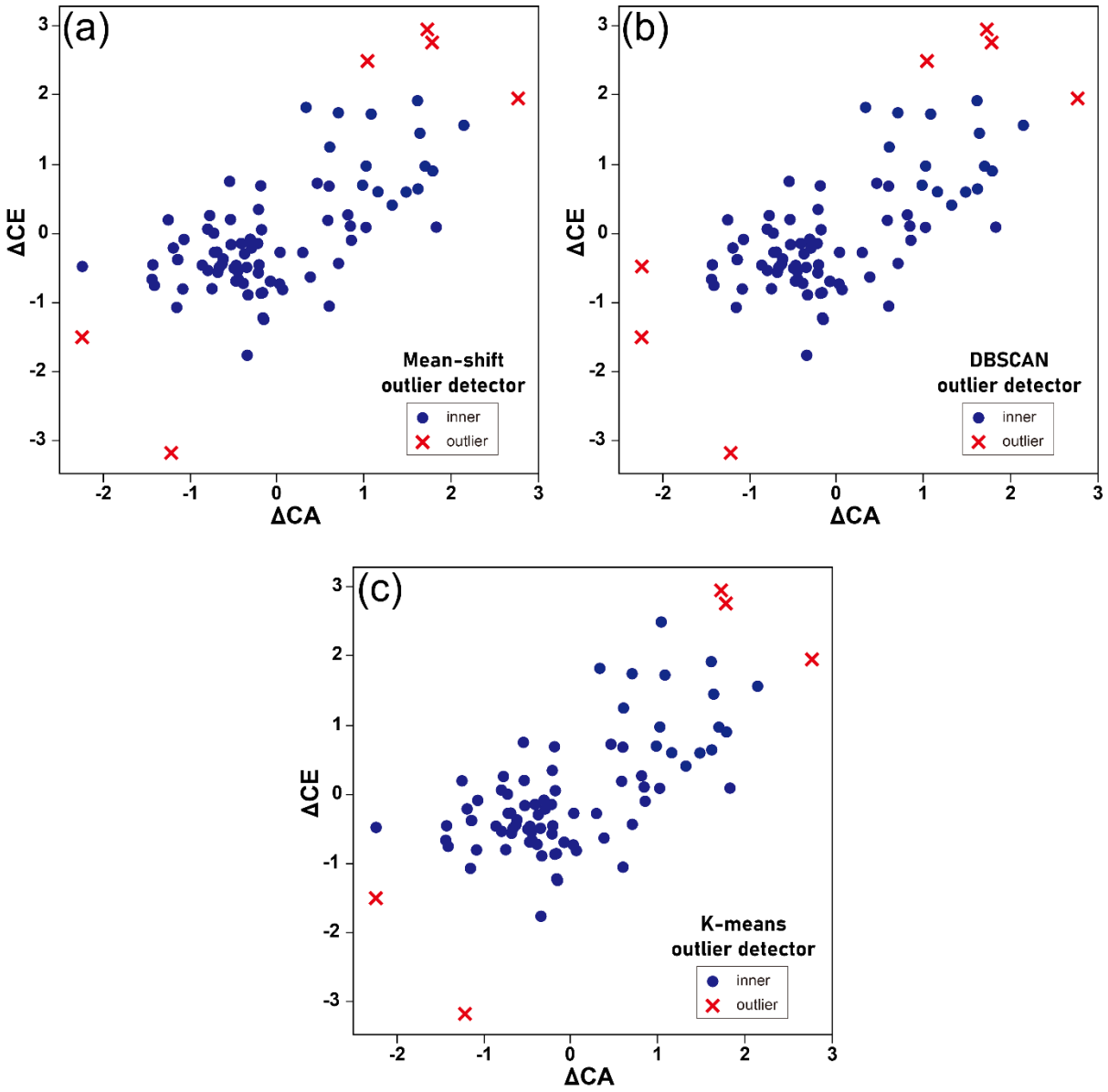

Figure 8. Results of outlier detection for the spinel data. Each of detection method is (a) Mean-shift, (b) DBSCAN, and (c) K-means, respectively. Final filtered data are outliers commonly detected from the each of detection mothods.

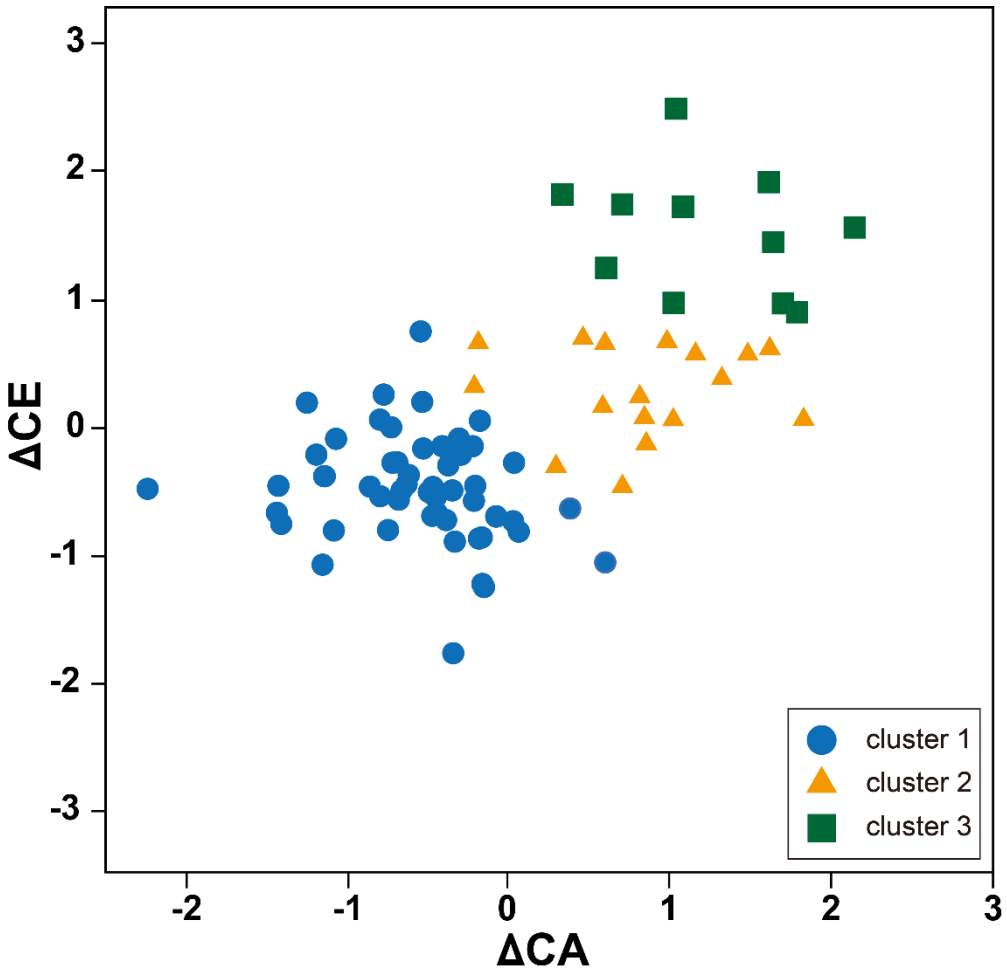

Figure 9. Clustering result for the data points. The data points are divided into 3 clusters. The marks for each cluster are displayed.

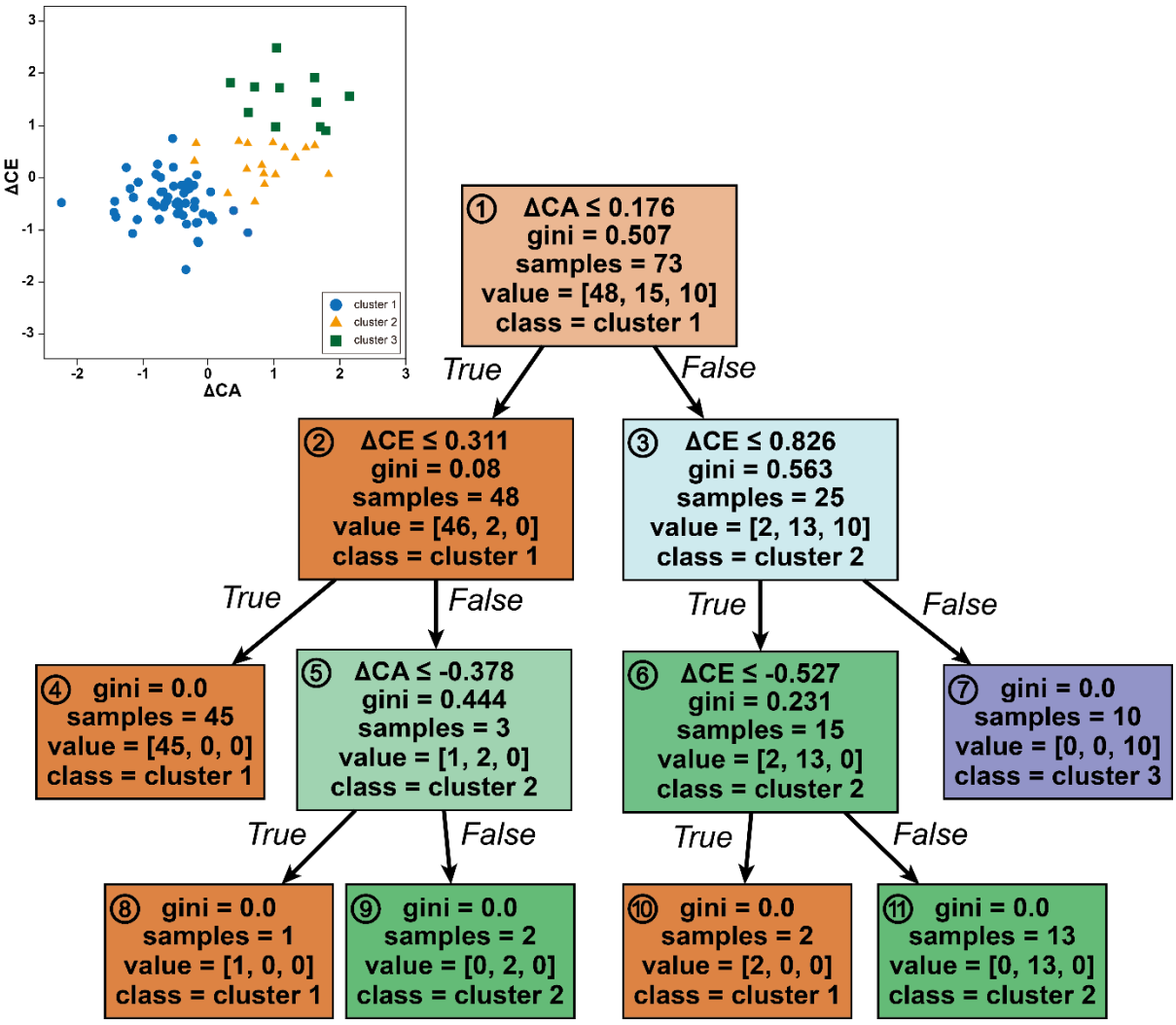

Figure 10. The visualized tree structure classifies the spinel data. Each node numbered from 1 to 11 contains information including gini score (gini), number of data (sample), numbers of data for each label data (value), and type of label data (class). A scatterplot located on the upper left side shows the region of clusters classified by this tree structure.

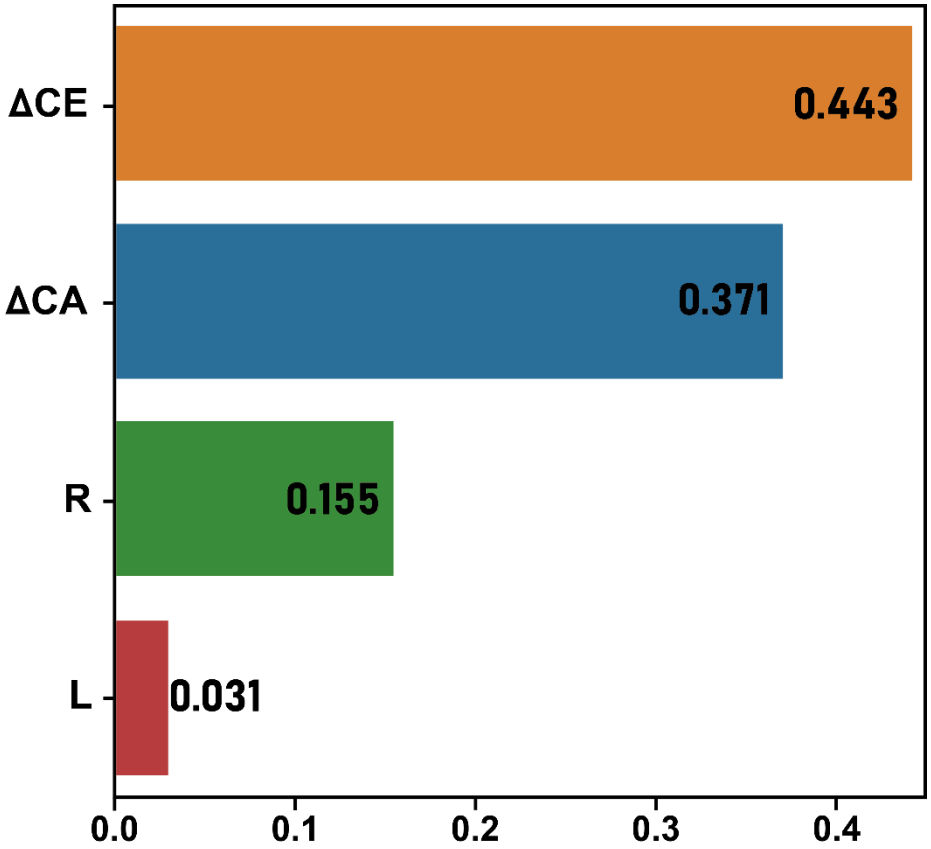

Figure 11. A bar chart describing the feature importance of the feature data. Values beside each bar are the scores for feature importance.

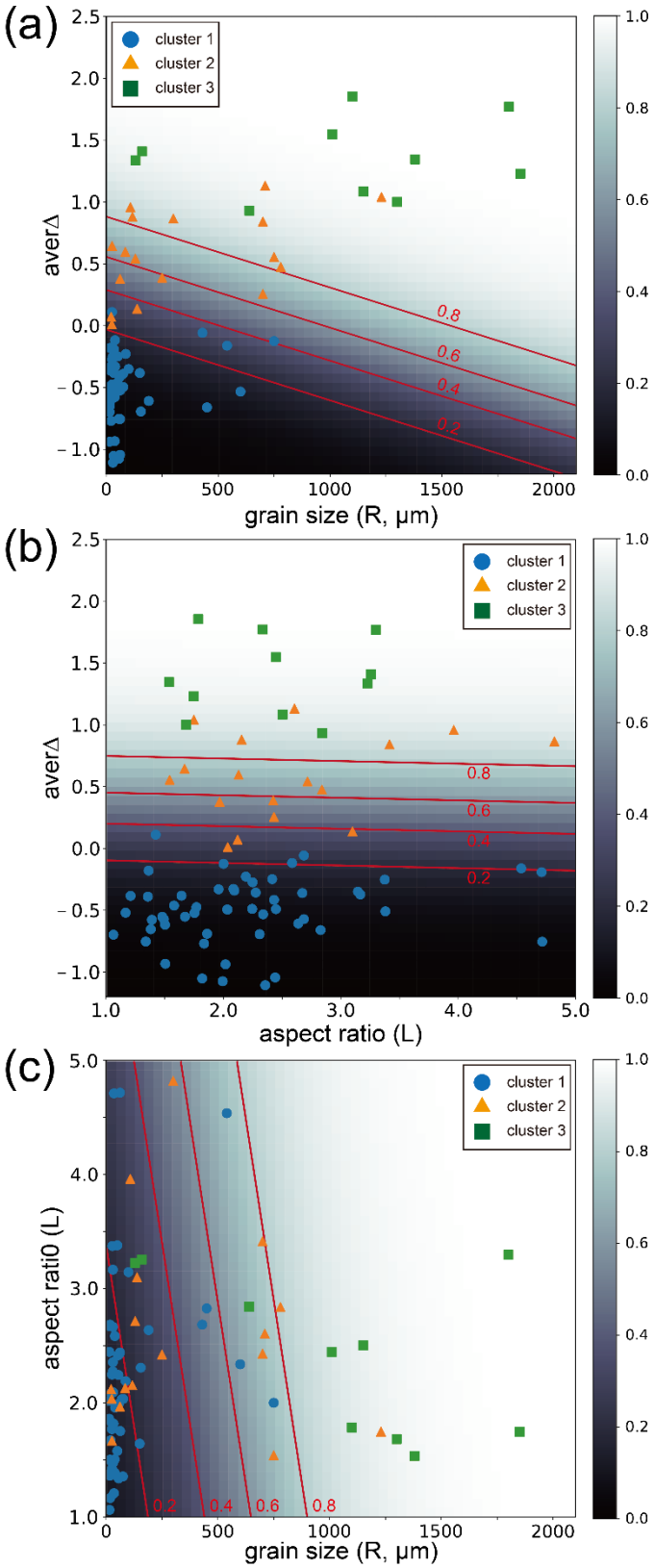

Figure 12. Probability maps for (a) the averΔ and grain size, (b) the averΔ and aspect ratio, and (c) aspect ratio and grain size. The probability is expressed in brighter color as it increased, and in a darker color as it decreased. Red lines are contour lines for 0.8, 0.6, 0.4, 0.2 of probabilities. Clustered spinel data points are overlaid on the probability maps.

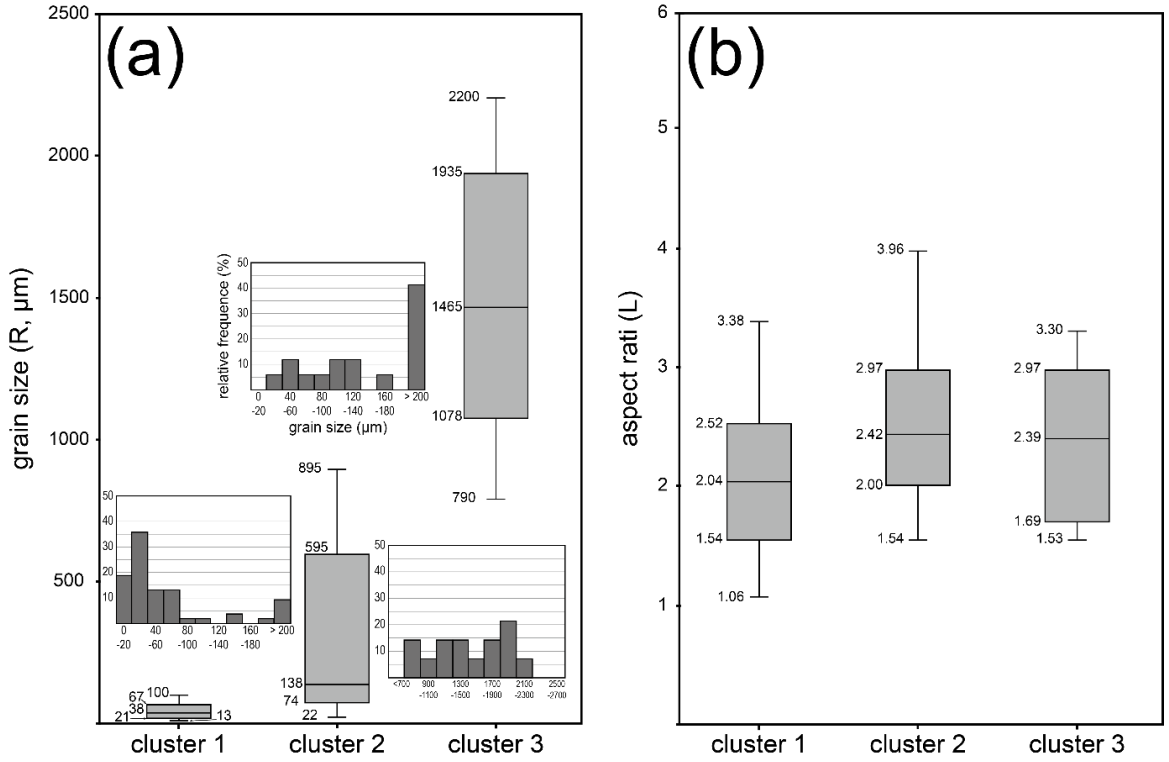

Figure 13. Box plots for (a) grain size and (b) aspect ratio in different types of clusters. Values in each box plot are (a) grain size and (b) aspect ratio of maximum, Q3 (75 %), median, Q1 (25 %), and minimum. Bar charts within (a) show logarithmic grain size distributions for each cluster.

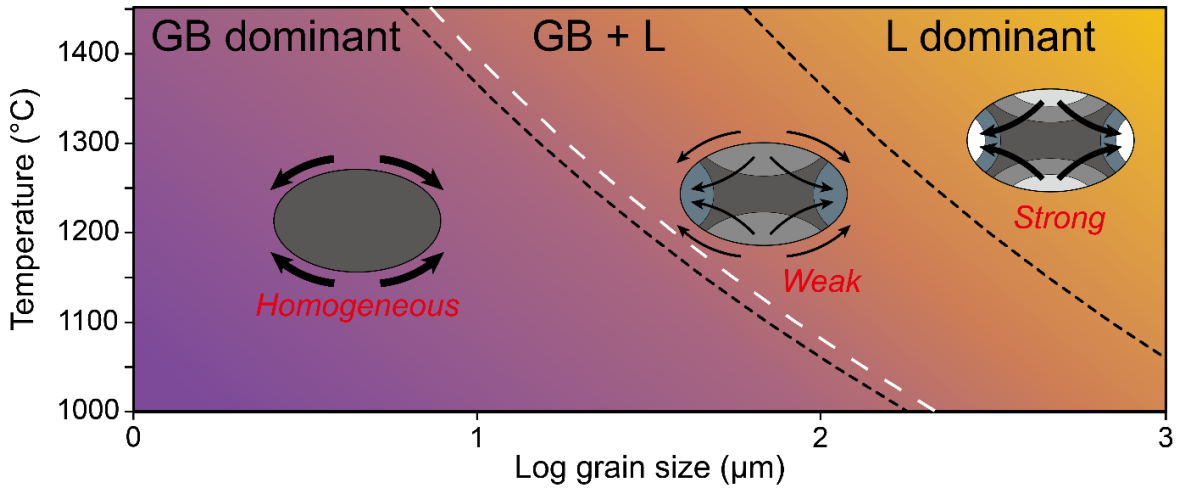

Figure 14. The relative importance of the diffusion mechanisms to total diffusion creep of chromite spinel with various grain size and temperature conditions were calculated using Shibutani et al model. (1998). Grain boundary and lattice diffusion, are denoted by GB and L, respectively. The black dash lines represent boundaries changing dominant diffusion mechanisms for the model of Shibutani et al. (1998). A white dash line is a transition boundary of diffusion mechanisms from grain boundary diffusion to lattice diffusion based on the model of Swaroop et al. (2005). Characters colored red represent the intensity of Cr-Al chemical zoning.

925 Table 1. Chemical compositions of each analysis point measured by EPMA for representative three
926 spinel grains (S1, S2 and S3).

Point No.	Al_2O_3	FeO	MnO	MgO	Cr_2O_3	NiO	TiO_2	Total (wt%)
P1	39.00	18.47	0.21	15.45	25.86	0.21	0.25	99.44
P2	35.23	18.90	0.20	14.73	29.90	0.20	0.25	99.39
P3	38.62	19.63	0.19	14.65	25.85	0.21	0.26	99.41
P4	37.99	19.63	0.19	14.65	25.85	0.21	0.26	98.78
P5	36.68	20.34	0.21	14.45	27.01	0.17	0.27	99.12
P6	37.96	19.17	0.17	15.40	26.25	0.23	0.23	99.41
P7	35.96	20.53	0.22	14.58	28.12	0.23	0.28	99.93
P8	35.90	19.53	0.20	15.28	28.39	0.21	0.23	99.73
P9	35.87	20.03	0.24	14.71	28.21	0.24	0.28	99.57

927

Table 2. Performances for each probability map. Each score is average for cross-validation (cv=10). $\boldsymbol{aver}\Delta$ is average of the $\Delta\boldsymbol{CA}$ and $\Delta\boldsymbol{CE}$. R is grain size. L is aspect ratio.

	$aver\Delta$ and R	$aver\Delta$ and L	L and R
Accuracy	0.95	0.96	0.78
Recall	0.90	0.90	0.53
Precision	0.97	1.00	0.70
f1 score	0.91	0.93	0.58
ROC-AUC	0.98	1.00	0.86

Table 3. The ratio of diffusion flux (r_{diff}) for various grain size with changing temperature condition.

T (°C) \ log grain size (μm)	0	0.25	0.5	0.75	1	1.25	1.5	1.75	2	2.25	2.5	2.75	3
1450	6.10	3.43	1.93	1.08	0.61	0.34	0.19	0.11	0.06	0.03	0.02	0.01	0.01
1400	8.12	4.57	2.57	1.44	0.81	0.46	0.26	0.14	0.08	0.05	0.03	0.01	0.01
1350	11.01	6.19	3.48	1.96	1.10	0.62	0.35	0.20	0.11	0.06	0.03	0.02	0.01
1300	15.20	8.55	4.81	2.70	1.52	0.85	0.48	0.27	0.15	0.09	0.05	0.03	0.02
1250	21.45	12.06	6.78	3.81	2.15	1.21	0.68	0.38	0.21	0.12	0.07	0.04	0.02
1200	30.98	17.42	9.80	5.51	3.10	1.74	0.98	0.55	0.31	0.17	0.10	0.06	0.03
1150	45.92	25.82	14.52	8.17	4.59	2.58	1.45	0.82	0.46	0.26	0.15	0.08	0.05
1100	70.04	39.39	22.15	12.46	7.00	3.94	2.21	1.25	0.70	0.39	0.22	0.12	0.07
1050	110.29	62.02	34.88	19.61	11.03	6.20	3.49	1.96	1.10	0.62	0.35	0.20	0.11
1000	179.98	101.21	56.92	32.01	18.00	10.12	5.69	3.20	1.80	1.01	0.57	0.32	0.18

933 Table 4. Critical grain sizes ($\boldsymbol{R_c}$) for various temperature conditions

T (°C)	[a]R_c	[b]R_c
1450	7.32	6.10
1400	9.75	8.12
1350	13.21	11.01
1300	18.24	15.20
1250	25.74	21.45
1200	37.18	30.98
1150	55.11	45.92
1100	84.05	70.04
1050	132.35	110.29
1000	215.98	179.98

934

935 Considering the models of [a]Swaroop et al. (2005) and [b]Shibutani et al. (1998)