JGR Solid Earth

RESEARCH ARTICLE

10.1029/2021JB021925

Special Section:

Machine learning for Solid Earth observation, modeling and understanding

Key Points:

- Multidimensional approach of evaluating quartz trace element data using a machine-learning tool
- Accuracy comparisons between low-dimensional and high-dimensional approaches in quartz classification
- A developed web-based app for quartz classification

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

K.-F. Qiu,
kunfengqiu@qq.com

Citation:

Wang, Y., Qiu, K.-F., Müller, A., Hou, Z.-L., Zhu, Z.-H., & Yu, H.-C. (2021). Machine learning prediction of quartz forming-environments. *Journal of Geophysical Research: Solid Earth*, 126, e2021JB021925. <https://doi.org/10.1029/2021JB021925>

Received 23 FEB 2021

Accepted 11 JUL 2021

Machine Learning Prediction of Quartz Forming-Environments

Yu Wang¹ , Kun-Feng Qiu¹ , Axel Müller^{2,3}, Zhao-Liang Hou⁴ , Zhi-Hai Zhu⁵, and Hao-Cheng Yu¹

¹State Key Laboratory of Geological Processes and Mineral Resources, School of Earth Science and Resources, China University of Geosciences, Beijing, China, ²Natural History Museum, University of Oslo, Oslo, Norway, ³Natural History Museum, London, UK, ⁴Department of Geology, University of Vienna, Vienna, Austria, ⁵School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, China

Abstract Trace elements of quartz document the physical-chemical evolutions of quartz growth, which has been a great and most applied tool in the study of geological settings in quartz-forming environments. A classic method is using graphic diagram plots visualizing the quartz trace element discriminations and trends, examples including the Al-Ti diagram (Rusk, 2012, https://doi.org/10.1007/978-3-642-22161-3_14) and the Ti-Al-Ge diagram (Schrön et al., 1988, https://www.researchgate.net/publication/236149159_Geochemische_Untersuchungen_an_Pegmatitquarzen). However, those diagrams are limited to two dimensions and cannot show the information in a higher dimension. In the study, we thus used a machine learning-based approach to evaluate quartz trace elements, and visualized them for the first time in the high-dimensional diagrams. We revisited 1,626 quartz samples from nine geological environments from previous studies, and applied a support vector machine to characterize values of the contained trace elements, including Al, Ti, Li, Ge, and Sr. We demonstrate that support vector machines can identify the crystallization environment of quartz with a significantly higher accuracy than the traditional plotting methods. Our work can massively improve the confidence on distinguishing quartz origin from different geological environments with a high efficiency. The method may also be applicable for other minerals, and we anticipate our research is a starting point for investigating mineral trace elements with machine learning techniques. Our quartz classifier can be accessed via <https://quartz-classifier.herokuapp.com>.

Plain Language Summary Quartz is a widespread mineral in the Earth's crust in various environments. Previous studies have made an effort in discriminating different geological settings by plotting quartz trace elements on two-dimensional (2D) diagrams. However, subsequent studies found these discrimination diagrams have some issues mainly because 2D diagrams have essential limitations to higher dimensions of information. In the present study, we apply a multidimensional approach of evaluating quartz trace element data using a machine learning tool. We use quartz trace element data from nine geological environments obtained by previous studies. Our study shows machine learning provides a better result than 2D diagrams and allows the more precise identification of deposit types using quartz chemistry.

1. Introduction

Quartz is an abundant and important mineral in the crust that occurs in a wide range of geological environments. Crystalline quartz records the geo-environmental evolutions, and because of its abundance, is a perfect pathfinder mineral to reveal changes of the physical-chemical conditions. As the main gangue mineral in magmatic, transitional magmatic-hydrothermal, and hydrothermal deposits, its trace element abundance and distribution has been increasingly utilized for genetic classifications and to reconstruct the physicochemical conditions during quartz formation (Götte & Ramseier, 2012; Götte et al., 2011; Kempe et al., 2012; Rusk, 2012).

Using trace elements to classify source rocks has been studied over the last decades. A conventional approach to visualize quartz trace element relationship is reinforced by diagrams such as Al versus Ti biplot and Ti-Al-Ge ternary plot (Breiter et al., 2020; Götze, 2009; Müller et al., 2000, 2002, 2010; Rusk, 2012; Schrön et al., 1988).

Figure 1. Locations of samples used in this study. (a) World map. (b) A larger view of the box in Figure 2a showing locations in Czech Republic. IRG is the abbreviation for Intrusion-Related Gold deposits. “Complex” refers to localities that comprise more than one deposit type.

However, the subsequent studies show that there is an increasing demand to more precisely classify quartz populations, and potentially use more than three element concentrations based on much larger data sets (Deng et al., 2020, 2021; Feng et al., 2020; Fu et al., 2020; Li et al., 2020; Müller et al., 2018; Pacák et al., 2019; Peterková & Dolejš, 2019; Rottier & Casanova, 2020; Tanner et al., 2013; Zhang et al., 2019). These studies also revealed that many data do not fit the existing classification ranges using the two types of discrimination plots. Another issue is, when data scatter close to a classification boundary, any interpretation would be ambiguous.

Driven by data availability and computational scale, machine learning offers new opportunities for developments and applications in geosciences. For example, machine learning helps geologists with choosing the intervals of drill core for assay sampling (Caté et al., 2017), classifying lithology from remote sensing data (Cracknell & Reading, 2014; Yu et al., 2012), and detrital provenance studies from mineral trace element composition data (O’Sullivan, 2020). Machine learning is effective in high-dimensional spaces.

The quartz study challenges stated above and the availability of the machine learning technique motivated this work. First, we present a compilation of previously published quartz trace-element compositional data comprising 5,341 quartz analyses from nine different geological environments. The comprehensive data set can also be used for further studies. Then, we compared four machine learning classification algorithms on the classifying source rock of quartz problem and chose the superior one to tune and yield the final classifier. The classifier was trained from the compiled data set and can be used to predict the genetic classes of new quartz based on its trace element analyses. The classifier enables the determination of source rock types from quartz trace element information, which is a powerful tool in quartz study and ore prospecting.

2. Compiled Data and Genetic Quartz Classes

The quartz trace element data were collected from 38 publications and comprise 5,341 analyses from 48 locations worldwide (Figure 1). The applied analytical methods, bedrock types, and references of each location are provided in Table S1. The sample set includes the following rock and mineralization types: granites (Breiter et al., 2012, 2013, 2017, 2019; Jacamon & Larsen, 2009; Monnier et al., 2018; Pacák et al., 2019;

Figure 2. Trace element concentrations of different genetic types of quartz. The height of the colored bars represents the interquartile range. The horizontal black lines within the colored bars are the median and the white dots with black edges represent the mean value. “Whiskers” of each box illustrate the maximum values lying within 1.5 times the interquartile range beyond the edges of the bars. The black dots represent the outliers deviating by more than $\pm 1.5\sigma$. IRG-Intrusion-related Au deposits.

Peterková & Dolejš, 2019), pegmatites (Beurlen et al., 2011; Breiter et al., 2019; Larsen et al., 2000, 2004; Müller et al., 2008; Pacák et al., 2019; Peterková & Dolejš, 2019; Rottier & Casanova, 2020), greisen (Breiter et al., 2017, 2019; Monnier et al., 2018; Müller et al., 2018; Peterková & Dolejš, 2019), porphyry Cu-Au (Bennett, 2014; Landtwing & Pettke, 2005; Mao et al., 2017, 2018; Maydagán et al., 2015; Monnier et al., 2018; Müller et al., 2010; Rottier & Casanova, 2020; Rusk et al., 2006, 2008; Tanner et al., 2013), skarn Au-Cu (Zhang et al., 2019), epithermal Au-Ag (Maydagán et al., 2015; Rottier & Casanova, 2020; Rusk et al., 2006, 2008; Tanner et al., 2013), intrusion-related Au (IRG; Pacák et al., 2019; Wertich et al., 2018), Carlin-type Au (Li et al., 2020; Lubben, 2000; Rusk et al., 2008; Yan et al., 2020), and orogenic Au and Au-Sb (Chen et al., 2019; Feng, 2020; Feng et al., 2020; Hu, 2020; Oliver et al., 2015; Pacák et al., 2019; Wolff, 2012). Figure 2 provides the compilation of all quartz trace element data discriminated based on deposit types. As shown in Figure 2, different types of quartz can be distinguished to some certain degree by each element. For example, porphyry type can be separated from most of the other types by the Ti content; Skarn type quartz has little overlap with granite type, pegmatite type, greisen type, porphyry type, and orogenic type by using Sr concentration. Combining multi-elements, the different types of quartz can be discriminated more distinctively.

3. Approach

3.1. Machine Learning Framework

The indicator used to evaluate classifiers is macro F1-score since the class distribution of our data set is skewed. F1-score is the harmonic mean between precision and recall, where precision and recall calculations are respectively given by Equations 1 and 2. A macro F1-score computes the metric independently for each class and then takes the average (hence treating all classes equally).

$$\text{precision} = \frac{\text{ture positives}}{\text{no. of predicted positives}} \quad (1)$$

$$\text{recall} = \frac{\text{true positives}}{\text{no. of actual positives}} \quad (2)$$

We tested four classic machine learning classification algorithms applying the fivefold cross-validation method (An et al., 2007; Breiman, 2001; Chang & Lin, 2011; Ng et al., 2011). The macro F1-scores of the four algorithms in descending ranking are random forest (0.83 ± 0.03), support vector machine with radial basis function (RBF) kernel (0.82 ± 0.02), artificial neuron network with one 100-neuron hidden layer (0.79 ± 0.02) and support vector machine with linear kernel (0.69 ± 0.03). Although random forest got the highest macro F1-score, tuning the number of trees in the forest did not improve the performance. Support vector machine with RBF kernel turned out to be the best choice.

Support vector machines are a set of supervised learning methods for classification, regression, and outliers detection (Chang & Lin, 2011). Work flows for the support vector machines see Hsu et al. (2003) and <https://scikit-learn.org/stable/modules/svm.html>. For our data set and problem definition, the following approach was applied: data pre-processing → feature selection → data splitting → tuning hyperparameters → retrain the estimator and testing the estimator → model persistence.

3.1.1. Data Pre-Processing

Standardization of data sets is a common requirement for many machine learning estimators to avoid features in greater numeric ranges dominating those in smaller numeric ranges meanwhile reducing the calculation expense (Hsu et al., 2003). Since the median values are generally much lower than mean values for each element in the data set (Figure 2), the data are first transformed using a log-ratio transformation to get a Gaussian distribution. Then a “StandardScaler” (centering of data by removing the mean value of each feature, then scale it by dividing non-constant features by their standard deviation) is implemented to get zero mean and unit variance distributions.

3.1.2. Feature Selection

Any valid information could be used as a feature to train the classifier, such as grain diameters, isotope compositions, and trace elements concentrations of quartz. In our data set, only trace element concentrations of quartz are compiled. Since the data set is a compilation from various sources of individual data set obtained by different analytical methods and research teams, it contains many missing values. Imputing a large number of missing values may lead to false positives (Seijo-Pardo et al., 2019). Reasons for the value missing are complex: Some data are under detection limits, while others are not provided by the original publications. To avoid any misleadingness by missing values, missing values of required features were not compensated, and samples missing required values were not selected to train the classifier.

The frequency of each element’s analysis occurring in the data set was summed up and sorted in descending order and regarded as the plotting order of the diagrams in Figure 2. Then, n most frequent elements were selected as features to train the classifier. Note that the order of Li and Ge were swapped when plotting the learning curves in Figure 3 to make the $n = 3$ case comparable to the ternary diagram which was introduced by Schrön et al. (1988) and modified by Götze (2009). Since there are not many choices for selecting features, the learning curves of each $n \in (2, 7)$ were plotted (Figure 3) to get an intuition of the relationship between the classifier’s performance and the number of features. It is apparent that as n increases, the performance improves.

Figure 3. Learning curve for $n \in (2, 7)$ where n represents numbers of the most frequent elements selected as feature of support vector machine algorithm. Note that positions of Li and Ge were exchanged, in order to make the case $n = 3$ comparable to the Ti-Al-Ge ternary diagram shown in Figure 5b. The score represents macro F1-score, which is the indicator to evaluate the performance of the classifier.

When $n > 5$, the numbers of samples decrease dramatically. Three of the nine classes have no more than 50 analyses available. With such small number of training examples, the model is more prone to overfitting. When $n > 6$, some classes do not have samples anymore. The final choice $n = 5$ (Al, Ti, Li, Ge, and Sr) was made by valuing and comparing better performance gained by introducing more features and the risk of

Figure 4. The cross-validation macro F1-score across the γ and C grid search.

overfitting due to only much smaller data sets for each class are available when n is large. This also ensures the classifier to be more applicable when only a small subset of elements analysis is available as input. The selected data set contains 1,626 samples (orogenic Au: 339, greisen: 325, Cu-Au porphyry: 272, granite: 237, Au-Cu skarn: 203, Carlin-type Au: 127, IRG: 96, pegmatite: 65, and epithermal: 26).

Training on imbalanced data set are risked in neglecting minority classes. However, support vector machines perform well with moderately imbalanced data without modifications (Akbani et al., 2004), and the test result shows the unaffected by an imbalance with the selected data set. The application of oversampling (He & Garcia, 2009) did not improve the performance of the classifier. Thus, no extra implements were applied to handle the imbalance.

3.1.3. Data Splitting

The whole selected data set was randomly split into training sets (80%) and testing sets (20%) by hold-out method while keeping the proportions of each class. The training set was used to train the classifier while the testing set was utilized to evaluate the classifier performance.

3.1.4. Tuning Hyperparameters

Hyperparameters are parameters that are not directly learned within estimators. For the support vector machine with RBF kernel used for this study, there are two hyperparameters: C and γ . They were determined by grid search techniques with the fivefold cross-validation method (Hsu et al., 2003) combined with the highest macro F1-score (Figure 4).

3.1.5. Retrain the Estimator and Test the Estimator

The best combination suggested by grid search was chosen as the final hyperparameter. Then the whole training set was retrained to produce a final classifier. This classifier was then evaluated on the testing set.

3.1.6. Model Persistence

The classifier model was saved for future use in the file named “quartz_clf.joblib.” The file to run the classifier is “Quartz_Classifier.py.” Code and data necessary to reproduce the classifier are also available at https://github.com/sleepypepperhead/quartz_classifier_2 and <http://doi.org/10.5281/zenodo.4077298>.

3.2. Library

The work was done in the “python” environment. The code used to make the diagrams and machine learning model used in this study, represents a combination of libraries (including pandas and numpy for data analysis and manipulation; matplotlib, ternary (Marc et al., 2019) and seaborn for plotting the diagrams; adjusttext (Flyamer et al., 2020) and geopandas for making the map in Figure 1; scikit-learn for machine learning; mlxtend (Raschka, 2018) for plotting decision regions).

4. Result and Discussion

4.1. Limitation of 2D Space

Schrön et al. (1988), Götze (2009), Müller et al. (2010), Rusk (2012), Breiter et al. (2014), and Pacák et al. (2019) did significant work on the visualization of quartz trace elements. In this study, we also tried to adapt and enhance the Al versus Ti and Ti-Al-Ge plots. Using the same elements of the Al versus Ti and Ti-Al-Ge plot, we trained the corresponding two-feature support vector machines from the compiled data

Figure 5. Decision boundaries predicted by two-feature support vector machines (a) The decision boundaries predicted by a support vector machine for Al versus Ti diagram calculated from 4,547 analyses (porphyry: 1,078; IRG: 670; granite: 661; orogenic: 607; skarn: 203; greisen: 377; epithermal: 308; Carlin: 181; pegmatite: 462). The cross-validation macro F1-score of this classifier is 0.365. The three bold line regions are according to Rusk (2012). (b) The decision boundaries predicted by a support vector machine for the Ti-Al-Ge ternary diagram, calculated from 3,257 samples (porphyry: 455; IRG: 291; granite: 654; orogenic: 499; skarn: 203; greisen: 374; epithermal: 153; Carlin: 177; pegmatite: 451). The cross-validation macro F1-score of this classifier is 0.330. (Adapted after Schrön et al., 1988). IRG, Intrusion-related Au deposits.

set and plot the decision regions to visualize them (Figure 5). However, these methods of utilizing 2D space for classification have some intrinsic shortcomings and limitations.

The Al versus Ti binary plot at its essential is a decision boundary for classification, and when $n = 2$ (Figure 3), the support vector machine essentially learned a decision boundary in the same space since they both use the same two features (Ti and Al). Generally, the cross-validation score is supposed to be lower than the training score. In the case $n = 2$, the training score curve is as low as around 0.45 at sample numbers $>3,000$, which means that the cross-validation score is not supposed to be higher than 0.45. The corresponding two-feature SVM classifier has the macro F1-score of 0.365 (Figure 5a). At such high sample numbers, the trend of the cross-validation score curve flattens out, which means that it is also subjected to high bias and low feature count. Adding more samples may help a little but will not exceed its corresponding training score. The case $n = 3$ is similar and only slightly better. Although the $n = 3$ case has the same three elements as the Ti-Al-Ge ternary diagram, the $n = 3$ case is 3D space while the Ti-Al-Ge ternary diagram is still 2D

Figure 6. Confusion matrix of the testing set to evaluate the accuracy of a classification. IRG, Intrusion-related Au deposits.

Figure 3 demonstrates that the classifier performance improves with the increase of the number of features n (number of elements). Choosing $n = 5$ is a compromise between classifier performance, and keeping enough samples for each class based on the current data set. Note that the score in the learning curve diagram (Figure 3) is slightly different from the score in the classification report (Table 1). Because when plotting the learning curve, the support vector machine was not tuned to use the optimal hyperparameters, and the evaluating samples were selected in different random ways. In all the cases, the high variance between the training scores and cross-validation scores and the trend of the learning curves indicate that the classifier will still benefit from the increase of training set, particularly when $n \geq 4$. So, if a training

space. They are essentially different. The two-feature SVM classifier corresponding to the Ti-Al-Ge ternary diagram has the macro F1-score of 0.330 (Figure 5b), which is obviously lower than the $n = 3$ case (Figure 3). This explains why the 2D plots do not work well for quartz classification using trace element contents.

4.2. Classifying in High Dimensional Space

The confusion matrix of the testing set is shown in Figure 6, and the classification report calculated from 50 executions is provided in Table 1. The classification report is a summary of the precision, recall, F1-score for each class. The confusion matrix shows the detailed classification result for each class of the testing set. This method has an accuracy of $86 \pm 2\%$. The nine genetic quartz types are included in the classification. The results presented here demonstrate a high performance of the support vector machine technique for the genetic quartz classification based on the trace element contents of quartz. Figure 7 shows a series of 2D planes of the classifier and some scatters around the planes, which gives an intuitive understanding of the classification model (Migut et al., 2015).

Table 1
Classification Report

	Precision ^a	Recall ^b	F1-score ^c	Support ^d
IRG	0.81 ± 0.09	0.82 ± 0.09	0.81 ± 0.07	18.72 ± 4.01
Carlin	0.95 ± 0.05	0.94 ± 0.05	0.94 ± 0.04	25.16 ± 4.38
Epithermal	0.98 ± 0.07	0.86 ± 0.16	0.91 ± 0.11	5.06 ± 1.57
Granite	0.88 ± 0.04	0.94 ± 0.04	0.91 ± 0.03	48.20 ± 5.17
Greisen	0.84 ± 0.05	0.83 ± 0.05	0.83 ± 0.03	64.04 ± 5.53
Orogenic	0.86 ± 0.04	0.85 ± 0.04	0.85 ± 0.03	69.40 ± 5.27
Pegmatite	0.77 ± 0.13	0.76 ± 0.12	0.76 ± 0.10	12.58 ± 2.78
Porphyry	0.89 ± 0.04	0.91 ± 0.04	0.90 ± 0.03	54.24 ± 5.72
Skarn	0.79 ± 0.07	0.75 ± 0.06	0.76 ± 0.04	40.60 ± 5.50
Accuracy			0.86 ± 0.02	338
Macro avg ^e	0.86 ± 0.02	0.85 ± 0.03	0.85 ± 0.02	338
Weighted avg ^f	0.86 ± 0.02	0.86 ± 0.02	0.86 ± 0.02	338

Note. Classification report for 50 executions.

Abbreviation: IRG, Intrusion-related Au.

^a“How many are correctly classified among that class,” calculated by Equation 1. ^b“How many of this class you find over the whole number of elements of this class,” calculated by Equation 2. ^cThe harmonic mean between precision and recall. ^dThe number of occurrences of the given class in the test data set. ^eAveraging the unweighted mean per label. ^fAveraging the support-weighted mean per label.

set of better quality is acquired (both in numbers of samples and valid features), the genetic type of quartz can be predicted with increased certainty. Despite the small number of input features, this study shows that support vector machine classification is useful for the classification of genetic quartz types based just on the trace element contents of quartz.

The support vector machine algorithm with a grid search for tuning hyperparameters eliminates most of the artificial factors in both trainings the model and predicting new samples, which eliminates the bias issues caused by both lack of the number of features and visual evaluation of 2D diagrams. The overlapping issue is solved by increasing dimensions (more trace elements to constrain the types). Moreover, this method is highly reproducible and universal. The method can also be implemented on other minerals whose features, such as trace elements or isotopes, reflect the type of source rock. The classifier can be used easily by inputting the path of the data file to the program. And it will output the predicted type and probabilities of possible outcomes for valid samples in the data file.

5. Conclusions

Our new suggested discrimination of genetic quartz applies a multidimensional approach using support vector machines. This classifier allows us to discriminate nine genetic types of quartz. The results of classification presented above, demonstrate a high performance of the support vector machine technique for classifications of quartz from variable environments based upon knowledge of the trace element compositions of quartz. As more

Figure 7. Six two-dimensional planes of the classifier showing some slices of the decision boundaries and some scatters around the planes.

trace elements data of quartz are published in the future, it will be possible to classify additional categories of quartz, and further improve the performance of the classifier. The statistical methods used in the present study may also be applied to or combined with other minerals. To make the technique more accessible, we deploy the quartz classifier at <https://quartz-classifier.herokuapp.com>.

Data Availability Statement

All data are available at <http://doi.org/10.5281/zenodo.4077298>.

Acknowledgments

This study was financially supported by the National Key Research Program (2019YFA0708603), the National Natural Science Foundation (91962106, 42072087, 41702069), the Beijing Nova Program (Z201100006820097), the 111 Project of Ministry of Science and Technology of China (BP0719021), the State Key Laboratory of Ore Deposit Geochemistry (201704), and the European Union's Horizon 2020 Innovation Program (Grant agreement no 869274, project GREENPEG: New Exploration Tools for European Pegmatite Green-Tech Resources).

References

- Akbani, R., Kwek, S., & Japkowicz, N. (2004). Applying support vector machines to imbalanced datasets. In *Paper presented at the European Conference on Machine Learning* (pp. 39–50). Springer. https://doi.org/10.1007/978-3-540-30115-8_7
- An, S., Liu, W., & Venkatesh, S. (2007). Fast cross-validation algorithms for least squares support vector machine and kernel ridge regression. *Pattern Recognition*, 40, 2154–2162. <https://doi.org/10.1016/j.patcog.2006.12.015>
- Bennett, M. M. (2014). *Cathodoluminescence and fluid inclusion characteristics of hydrothermal quartz from porphyry deposits*. Colorado School of Mines, Arthur Lakes Library.
- Beurlen, H., Muller, A., Silva, D., & Da Silva, M. R. R. (2011). Petrogenetic significance of LA-ICP-MS trace-element data on quartz from the Borborema Pegmatite Province, northeast Brazil. *Mineralogical Magazine*, 75, 2703–2719. <https://doi.org/10.1180/minmag.2011.075.5.2703>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32. <https://doi.org/10.1023/a:1010933404324>
- Breiter, K., Ackerman, L., Ďurišová, J., Svojtka, M., & Novák, M. (2014). Trace element composition of quartz from different types of pegmatites: A case study from the Moldanubian Zone of the Bohemian Massif (Czech Republic). *Mineralogical Magazine*, 78, 703–722. <https://doi.org/10.1180/minmag.2014.078.3.17>
- Breiter, K., Ackerman, L., Svojtka, M., & Müller, A. (2013). Behavior of trace elements in quartz from plutons of different geochemical signature: A case study from the Bohemian Massif, Czech Republic. *Lithos*, 175, 54–67. <https://doi.org/10.1016/j.lithos.2013.04.023>
- Breiter, K., Badanina, E., Ďurišová, J., Dosbaba, M., & Syritso, L. (2019). Chemistry of quartz—A new insight into the origin of the Orlovka Ta-Li deposit, Eastern Transbaikalia, Russia. *Lithos*, 348–349, 105206. <https://doi.org/10.1016/j.lithos.2019.105206>
- Breiter, K., Ďurišová, J., & Dosbaba, M. (2017). Quartz chemistry—A step to understanding magmatic-hydrothermal processes in ore-bearing granites: Čínovec/Zinnwald Sn-W-Li deposit, Central Europe. *Ore Geology Reviews*, 90, 25–35. <https://doi.org/10.1016/j.oregeorev.2017.10.013>
- Breiter, K., Ďurišová, J., & Dosbaba, M. (2020). Chemical signature of quartz from S- and A-type rare-metal granites—A summary. *Ore Geology Reviews*, 125, 103674. <https://doi.org/10.1016/j.oregeorev.2020.103674>

- Breiter, K., Svojtka, M., Ackerman, L., & Švecová, K. (2012). Trace element composition of quartz from the Variscan Altenberg-Teplce caldera (Krušné hory/Erzgebirge Mts, Czech Republic/Germany): Insights into the volcano-plutonic complex evolution. *Chemical Geology*, 326, 36–50. <https://doi.org/10.1016/j.chemgeo.2012.07.028>
- Caté, A., Perozzi, L., Gloaguen, E., & Blouin, M. (2017). Machine learning as a tool for geologists. *The Leading Edge*, 36, 215–219. <https://doi.org/10.1190/tle36030215.1>
- Chang, C., & Lin, C. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2, 1–27. <https://doi.org/10.1145/1961189.1961199>
- Chen, B., Deng, J., Wei, H., & Ji, X. (2019). Trace element geochemistry in quartz in the Jinqingding Gold Deposit, Jiaodong Peninsula, China: Implications for the gold precipitation mechanism. *Minerals*, 9, 326. <https://doi.org/10.3390/min9050326>
- Cracknell, M. J., & Reading, A. M. (2014). Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. *Computers & Geosciences*, 63, 22–33. <https://doi.org/10.1016/j.cageo.2013.10.008>
- Deng, J., Qiu, K., Wang, Q., Goldfarb, R., Yang, L., Zi, J., et al. (2020). In situ dating of hydrothermal monazite and implications for the geodynamic controls on ore formation in the Jiaodong gold province, eastern China. *Economic Geology*, 115, 671–685. <https://doi.org/10.5382/econgeo.4711>
- Deng, J., Wang, Q., Gao, L., He, W., Yang, Z., Zhang, S., et al. (2021). Differential crustal rotation and its control on giant ore clusters along the eastern margin of Tibet. *Geology*, 49, 428–432. <https://doi.org/10.1130/G47855.1>
- Feng, K. (2020). *Refined metallogenic processes and fluid evolutions of gold deposits in the Penglai-Qixia belt, Jiaodong (PhD Thesis)*. University of Chinese Academy of Sciences.
- Feng, Y., Zhang, Y., Xie, Y., Shao, Y., Tan, H., Li, H., & Lai, C. (2020). Ore-forming mechanism and physicochemical evolution of Gutaishan Au deposit, South China: Perspective from quartz geochemistry and fluid inclusions. *Ore Geology Reviews*, 119, 103382. <https://doi.org/10.1016/j.oregeorev.2020.103382>
- Flyamer, I., Weber/GwendalD, S., Xue, Z., Colin, Li, A., Neste, C. V., et al. (2020). *Phlya/adjustText: 0.8 beta*. Zenodo. <https://doi.org/10.5281/ZENODO.3924114>
- Fu, S., Lan, Q., & Yan, J. (2020). Trace element chemistry of hydrothermal quartz and its genetic significance: A case study from the Xikuangshan and Woxi giant Sb deposits in southern China. *Ore Geology Reviews*, 126, 103732. <https://doi.org/10.1016/j.oregeorev.2020.103732>
- Götze, T., Pettke, T., Ramseier, K., Koch-Müller, M., & Mullis, J. (2011). Cathodoluminescence properties and trace element signature of hydrothermal quartz: A fingerprint of growth dynamics. *American Mineralogist*, 96, 802–813. <https://doi.org/10.2138/am.2011.3639>
- Götze, T., & Ramseier, K. (2012). Trace element characteristics, luminescence properties and real structure of quartz. In J. Götze, & R. Möckel (Eds.), *Quartz: Deposits, mineralogy and analytics* (pp. 265–285). Springer. https://doi.org/10.1007/978-3-642-22161-3_12
- Götze, J. (2009). Chemistry, textures and physical properties of quartz—Geological interpretation and technical application. *Mineralogical Magazine*, 73, 645–671. <https://doi.org/10.1180/minmag.2009.073.4.645>
- He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21, 1263–1284. <https://doi.org/10.1109/tkde.2008.239>
- Hsu, C., Chang, C., & Lin, C. (2003). *A practical guide to support vector classification* (p. 16).
- Hu, H. (2020). *Hydrothermal alteration and mineralization processes of the Jiaojia-type gold deposits, Jiaodong Peninsula, China (PhD Thesis)*. University of Chinese Academy of Sciences.
- Jacamon, F., & Larsen, R. B. (2009). Trace element evolution of quartz in the charnockitic Kleivan granite, SW-Norway: The Ge/Ti ratio of quartz as an index of igneous differentiation. *Lithos*, 107, 281–291. <https://doi.org/10.1016/j.lithos.2008.10.016>
- Kempe, U., Götze, J., Dombon, E., Monecke, T., & Poutivtsev, M. (2012). Quartz regeneration and its use as a repository of genetic information. In J. Götze, & R. Möckel (Eds.), *Quartz: Deposits, mineralogy and analytics* (pp. 331–355). Springer. https://doi.org/10.1007/978-3-642-22161-3_15
- Landtving, M. R., & Pettke, T. (2005). Relationships between SEM-cathodoluminescence response and trace-element composition of hydrothermal vein quartz. *American Mineralogist*, 90, 122–131. <https://doi.org/10.2138/am.2005.1548>
- Larsen, R. B., Henderson, I., Ihlen, P. M., & Jacamon, F. (2004). Distribution and petrogenetic behaviour of trace elements in granitic pegmatite quartz from South Norway. *Contributions to Mineralogy and Petrology*, 147, 615–628. <https://doi.org/10.1007/s00410-004-0580-4>
- Larsen, R. B., Polve, M., & Juve, G. (2000). Granite pegmatite quartz from Evje-Iveland: Trace element chemistry and implications for the formation of high-purity quartz. *Norges Geologiske Undersøkelse*, 436, 57–66.
- Li, J., Hu, R., Xiao, J., Zhuo, Y., Yan, J., & Oyebamiji, A. (2020). Genesis of gold and antimony deposits in the Youjiang metallogenic province, SW China: Evidence from in situ oxygen isotopic and trace element compositions of quartz. *Ore Geology Reviews*, 116, 103257. <https://doi.org/10.1016/j.oregeorev.2019.103257>
- Lubben, J. D. (2000). *Silicification across the Betze-Post Carlin-type AU deposit: Clues to ore fluid properties and sources, northern Carlin trend, Nevada*. UNLV.
- Mao, W., Rusk, B., Yang, F., & Zhang, M. (2017). Physical and chemical evolution of the Dabaoshan Porphyry Mo deposit, South China: Insights from fluid inclusions, cathodoluminescence, and trace elements in quartz. *Economic Geology*, 112, 889–918. <https://doi.org/10.2113/econgeo.112.4.889>
- Mao, W., Zhong, H., Zhu, W.-G., Lin, X.-G., & Zhao, X.-Y. (2018). Magmatic-hydrothermal evolution of the Yuanzhuding porphyry Cu-Mo deposit, South China: Insights from mica and quartz geochemistry. *Ore Geology Reviews*, 101, 765–784. <https://doi.org/10.1016/j.oregeorev.2018.08.016>
- Marc, H., Weinstein, B., Woodcock, T. G., Simon, C., chebeeTi, Morgan, W., et al. (2019). *marcharper/python-ternary: Version 1.0.6*. Zenodo. <https://doi.org/10.5281/ZENODO.2628066>
- Maydagán, L., Franchini, M., Rusk, B., Lentz, D. R., McFarlane, C., Impiccini, A., et al. (2015). Porphyry to epithermal transition in the Altar Cu-(Au-Mo) deposit, Argentina, studied by cathodoluminescence, LA-ICP-MS, and fluid inclusion analysis. *Economic Geology*, 110, 889–923. <https://doi.org/10.2113/econgeo.110.4.889>
- Migut, M. A., Worring, M., & Veenman, C. J. (2015). Visualizing multi-dimensional decision boundaries in 2D. *Data Mining and Knowledge Discovery*, 29, 273–295. <https://doi.org/10.1007/s10618-013-0342-x>
- Monnier, L., Lach, P., Salvi, S., Melleton, J., Bailly, L., Beziat, D., et al. (2018). Quartz trace-element composition by LA-ICP-MS as proxy for granite differentiation, hydrothermal episodes, and related mineralization: The Beauvoir Granite (Echassières district), France. *Lithos*, 320, 355–377. <https://doi.org/10.1016/j.lithos.2018.09.024>
- Müller, A., Herklotz, G., & Giegling, H. (2018). Chemistry of quartz related to the Zinnwald/Cinovec Sn-W-Li greisen-type deposit, Eastern Erzgebirge, Germany. *Journal of Geochemical Exploration*, 190, 357–373. <https://doi.org/10.1016/j.jgexplo.2018.04.009>

- Müller, A., Ihlen, P. M., & Kronz, A. (2008). Quartz chemistry in polygeneration Sveconorwegian pegmatites, Froland, Norway. *European Journal of Mineralogy*, 20, 447–463. <https://doi.org/10.1127/0935-1221/2008/0020-1822>
- Müller, A., Kronz, A., & Breiter, K. (2002). Trace elements and growth patterns in quartz: A fingerprint of the evolution of the subvolcanic Podlesí Granite System (Krušné hory Mts., Czech Republic). *Bulletin of the Czech Geological Survey*, 77, 135–145.
- Müller, A., Seltmann, R., & Behr, H. (2000). Application of cathodoluminescence to magmatic quartz in a tin granite—Case study from the Schellerhau Granite Complex, Eastern Erzgebirge, Germany. *Mineralium Deposita*, 35, 169–189. <https://doi.org/10.1007/s001260050014>
- Müller, A., van den Kerkhof, A. M., Behr, H.-J., Kronz, A., & Koch-Müller, M. (2010). The evolution of late-Hercynian granites and rhyolites documented by quartz—A review. *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, 100, 185–204. <https://doi.org/10.1017/S1755691009016144>
- Ng, A., Ngiam, J., Foo, C. Y., Mai, Y., & Suen, C. (2011). *Backpropagation algorithm*. Coursera.
- Oliver, N. H., Thomson, B., Freitas-Silva, F. H., Holcombe, R. J., Rusk, B., Almeida, B. S., et al. (2015). Local and regional mass transfer during thrusting, veining, and boudinage in the genesis of the giant shale-hosted Paracatu gold deposit, Minas Gerais, Brazil. *Economic Geology*, 110, 1803–1834. <https://doi.org/10.2113/econgeo.110.7.1803>
- O'Sullivan, G., Chew, D., Kenny, G., Henrichs, L., & Mulligan, D. (2020). The trace element composition of apatite and its application to detrital provenance studies. *Earth-Science Reviews*, 201, 20. <https://doi.org/10.1016/j.earscirev.2019.103044>
- Pacák, K., Zachariáš, J., & Strnad, L. (2019). Trace-element chemistry of barren and ore-bearing quartz of selected Au, Au-Ag and Sb-Au deposits from the Bohemian Massif. *Journal of Geosciences*, 64, 19–35. <https://doi.org/10.3190/jgeosci.279>
- Peterková, T., & Dolejš, D. (2019). Magmatic-hydrothermal transition of Mo-W-mineralized granite-pegmatite-greisen system recorded by trace elements in quartz: Krupka district, Eastern Krušné hory/Erzgebirge. *Chemical Geology*, 523, 179–202. <https://doi.org/10.1016/j.chemgeo.2019.04.009>
- Raschka, S. (2018). MLxtend: Providing machine learning and data science utilities and extensions to Python's scientific computing stack. *Journal of Open Source Software*, 3, 638. <https://doi.org/10.21105/joss.00638>
- Rottier, B., & Casanova, V. (2020). Trace element composition of quartz from porphyry systems: A tracer of the mineralizing fluid evolution. *Mineralium Deposita*, 56, 843–862. <https://doi.org/10.1007/s00126-020-01009-0>
- Rusk, B. (2012). Cathodoluminescent textures and trace elements in hydrothermal quartz. In J. Götze, & R. Möckel (Eds.), *Quartz: Deposits, mineralogy and analytics* (pp. 307–329). Springer Berlin Heidelberg, Springer Geology. https://doi.org/10.1007/978-3-642-22161-3_14
- Rusk, B. G., Lowers, H. A., & Reed, M. H. (2008). Trace elements in hydrothermal quartz: Relationships to cathodoluminescent textures and insights into vein formation. *Geology*, 36, 547. <https://doi.org/10.1130/G24580A.1>
- Rusk, B. G., Reed, M. H., Dilles, J. H., & Kent, A. J. (2006). Intensity of quartz cathodoluminescence and trace-element content in quartz from the porphyry copper deposit at Butte, Montana. *American Mineralogist*, 91, 1300–1312. <https://doi.org/10.2138/am.2006.1984>
- Schrön, W., Schmädicke, E., Thomas, R., & Schmidt, W. (1988). Geochemische Untersuchungen an Pegmatitquarzen. *Zeitschrift für Geologische Wissenschaften*, 16, 229–244.
- Seijo-Pardo, B., Alonso-Betanzos, A., Bennett, K. P., Bolón-Canedo, V., Josse, J., Saeed, M., & Guyon, I. (2019). Biases in feature selection with missing data. *Neurocomputing*, 342, 97–112. <https://doi.org/10.1016/j.neucom.2018.10.085>
- Tanner, D., Henley, R. W., Mavrogenes, J. A., & Holden, P. (2013). Combining in situ isotopic, trace element and textural analyses of quartz from four magmatic-hydrothermal ore deposits. *Contributions to Mineralogy and Petrology*, 166, 1119–1142. <https://doi.org/10.1007/s00410-013-0912-3>
- Wertich, V., Leichmann, J., Dosbaba, M., & Götze, J. (2018). Multi-stage evolution of gold-bearing hydrothermal quartz veins at the Moksko gold deposit (Czech Republic) based on cathodoluminescence, spectroscopic, and trace elements analyses. *Minerals*, 8, 335. <https://doi.org/10.3390/min8080335>
- Wolff, W. R. (2012). *Microstructures and trace element signatures of orogenic quartz veins in the Klondike District, Yukon Territory, Canada*. The University of British Columbia.
- Yan, J., Mavrogenes, J. A., Liu, S., & Coulson, I. M. (2020). Fluid properties and origins of the Lannigou Carlin-type gold deposit, SW China: Evidence from SHRIMP oxygen isotopes and LA-ICP-MS trace element compositions of hydrothermal quartz. *Journal of Geochemical Exploration*, 215, 106546. <https://doi.org/10.1016/j.jexplo.2020.106546>
- Yu, L., Porwal, A., Holden, E.-J., & Dentith, M. C. (2012). Towards automatic lithological classification from remote sensing data using support vector machines. *Computers & Geosciences*, 45, 229–239. <https://doi.org/10.1016/j.cageo.2011.11.019>
- Zhang, Y., Cheng, J., Tian, J., Pan, J., Sun, S., Zhang, L., et al. (2019). Texture and trace element geochemistry of quartz in skarn system: Perspective from Jiguanzui Cu–Au skarn deposit, Eastern China. *Ore Geology Reviews*, 109, 535–544. <https://doi.org/10.1016/j.oregeorev.2019.05.007>