

## Genetic types of Zn-Pb deposits revealed by sphalerite geochemistry

QINGWEN ZENG<sup>1</sup>, QIHAI SHU<sup>1,\*</sup>, QINGFEI WANG<sup>1,2,†</sup>, QIAN ZHANG<sup>1</sup>, ZHONGHAI ZHAO<sup>3</sup>, XUDONG NIU<sup>1</sup>,  
FAN YU<sup>1</sup>, LITIAN ZHANG<sup>1</sup>, AND JUN DENG<sup>1,\*</sup>

<sup>1</sup>State Key Laboratory of Geological Processes and Mineral Resources, Frontiers Science Center for Deep-time Digital Earth,  
School of Earth Sciences and Resources, China University of Geosciences, Beijing 100083, China

<sup>2</sup>Jiangxi Provincial Key Laboratory of Genesis and Prospect for Strategic Minerals, East China University of Technology,  
Nanchang 330013, China

<sup>3</sup>College of Mining, Liaoning Technical University, Fuxin 123000, China

### ABSTRACT

The identification of the genetic type of a mineral deposit, based on the compositional characteristics of specific minerals, has long been a focus of interest for economic geologists and mining companies. Traditional binary plots, due to the limitation of their dimensions, fail to encompass the whole element information, potentially introducing bias to the discrimination results. This is particularly the case in classifying Zn-Pb deposits. The current study employs four widely used machine learning algorithms (random forest, extreme gradient boosting, support vector machine, and multi-layer perceptron) to train 4908 sets of element data for sphalerite compiled from five distinct Zn-Pb deposit types (VMS, SEDEX, MVT, skarn, and epithermal). The data are then visualized and interpreted through principal component analysis and t-distributed stochastic neighbor embedding, which indicate that reducing sphalerite element data to a two-dimensional projection leads to the loss of significant feature information, hindering the ability to effectively distinguish the genesis of the deposit. The machine learning results show that all four models have macro F1-scores above 0.95 on the test set, demonstrating robustness and excellent generalization ability, which reflects the reliability of using sphalerite geochemistry to distinguish Zn-Pb mineralization types. The SHAP value analysis highlights the key role of Mn, Fe, Ge, Cd, and Co concentrations in facilitating the differentiation of deposit types through machine learning algorithms. Our models show an accuracy rate of 83% in predicting the combined results on an external, independent data set. The models have also been applied to classify three Zn-Pb deposits of unknown types, and the results are consistent with geological observations. The models' parameters have been further exported and programmed into an Excel macro program and a user-friendly software application, which can be accessed via <https://sdeakii.github.io/machine-learning>.

**Keywords:** Sphalerite geochemistry, machine learning, genetic type, Zn-Pb deposit