American Mineralogist, Volume 107, pages 1886–1900, 2022

Estimating ferric iron content in clinopyroxene using machine learning models

Wei-hua Huang[1,2], Yang Lyu[3], Ming-hao Du[4], Can He[5], Shang-de Gao[6], Ren-jun Xu[6], Qun-ke Xia[1,2,†], and J ZhangZhou[1,2,*]

[1]School of Earth Sciences, Zhejiang University, Hangzhou 310027, China
[2]Key Laboratory of Big Data and Deep Earth Resources of Zhejiang Province, Hangzhou 310027, China
[3]Gemological Institute, China University of Geosciences, Wuhan 430074, China
[4]School of Geoscience and Info-Physics, Central South University, Changsha 410083, China
[5]School of Earth Sciences and Engineering, Sun Yat-Sen University, Guangzhou, 510275, China
[6]Center for Data Science, Zhejiang University, Hangzhou 310058, China

Abstract

Clinopyroxene ferric iron content is an important consideration for garnet-clinopyroxene geothermometry and estimations of water storage in the Earth's interior but remains difficult and expensive to measure. Here, we develop seven classic algorithms and machine learning methods to estimate $Fe^{3+}/\Sigma Fe$ in clinopyroxene using major element data from electron microprobe analyses. The models were first trained using a large data set of clinopyroxene $Fe^{3+}/\Sigma Fe$ values determined by Mössbauer spectroscopy and spanning a wide compositional range, with major uncertainties ranging from 0.25 to 0.3 and root-mean-square errors on the test data set ranging from 0.071 to 0.089. After dividing the entire data set into three compositional sub-data sets, the machine learning models were trained and compared for each sub-data set. Our results suggest that ensemble learning algorithms (random forest and Extra-Trees) perform better than principal component analysis-based elastic net polynomial, artificial neural network, artificial neural network ensemble, decision trees, and linear regressions. Using a sub-data set excluding clinopyroxene in spinel peridotite and omphacite in eclogite, the new models achieved uncertainties of 0.15 to 0.2 and root-mean-square errors on the test data set ranging from 0.051 to 0.078, decreasing prediction errors by 30–40%. By incorporating compositional data on coexisting spinel, new models for clinopyroxene in spinel peridotite show improved performance, indicating the interaction between spinel and clinopyroxene in spinel peridotite. Feature importance analysis shows Na^+, Ca^{2+}, and Mg^{2+} to be the most important for predicting Fe^{3+} content, supporting the coupled substitution between Ca^{2+}-M^{2+} and Na^+-M^{3+} in natural clinopyroxenes. The application of our models to garnet-clinopyroxene geothermometry greatly improves temperature estimates, achieving uncertainties of ±50 °C, compared with uncertainties of ±250 °C using previous models assuming all Fe as Fe^{2+} or calculating Fe^{3+} by charge conservation. Differences in the ferric iron contents, as calculated using the machine learning models, of clinopyroxenes that did or did not experience hydrogen diffusion during their crystallization from basaltic magma support a redox-driven mechanism for hydrogen diffusion in clinopyroxene.

Keywords: Clinopyroxene, $Fe^{3+}/\Sigma Fe$, machine learning, geothermometer, redox