Quantitative geochemical fingerprinting: Machine learning to trace tephra to source

MS Bolton[1], BJL Jensen[1], KL Wallace[2], N Praet[3], D Fortin[4], DS Kaufman[5], and M De Batist[3]

[1] *Earth and Atmospheric Sciences, University of Alberta, AB, Canada*
[2] *Alaska Volcano Observatory, United States Geological Survey, Anchorage, AK, USA*
[3] *Department of Geology, Ghent University, Gent, Belgium*
[4] *Department of Geography and Planning, University of Saskatchewan, Saskatoon, SK, Canada*
[5] *Departments of Geology and Environmental Sciences, Northern Arizona University, Flagstaff, AZ, USA*

Volcanic glass, the primary component of ash (tephra), can be used to identify volcanic eruptions and sources based on their "geochemical fingerprints". However, cross-correlating unknown geochemistries to reference material traditionally requires extensive interpretation of chemical plots. This process can be time-consuming, even when conducted by experts with the aid of digital visualization tools. However, using quantitative methods, particularly machine learning classifiers, to discriminate and identify tephras has the potential to greatly expedite this process while removing an element of subjectivity from geochemical correlation.

The purpose of this work was to evaluate the applicability of machine learning algorithms in tracing tephras to their volcanic sources. We explored this using Alaska's volcanically active and geochemically complex Aleutian Arc-Alaska Peninsula and Wrangell volcanic field. Multiple supervised classification algorithms, including neural networks, support vector machines, decision trees, and random forests among others, were tested to predict the origin of tephras from 10 sources. The models were trained on nearly 2,000 electron microprobe analysis (EMPA) glass measurements from the University of Alberta Earth and Atmospheric Science's Tephra Collection and Geochemical Database and the Alaska Volcano Observatory. Performance was evaluated internally (via cross-validation on training data) and externally (using held-out data from the source dataset and literature).

Collectively, the algorithms accurately predicted tephra sources. Random forests and artificial neural networks and their aggregation as an ensemble were consistently accurate (> 0.96), even in class-imbalanced scenarios. In order to evaluate the practicality of the computational methods, we also trialed the best-performing methods on a suite of glass geochemical data from sediment cores from Eklutna Lake, south-central Alaska. The model predictions agreed with traditional, plot-based correlations, and provided useful correlation pseudo-probabilities, even when mixed geochemical populations were present. We propose machine learning is a valuable tool to aid tephrochronologists in rapidly evaluating large datasets, which in turn facilitates expert assessment of correlations, especially when integrated with contextual data (e.g. stratigraphy and chronology).

Corresponding author: bolton1@ualberta.ca