Supplemental Information of Article

Geochemical Profile Characterization of Mine Tailings by Exploited Element as Input for Receptor Models: Case of Chilean Tailings (Cu-Au-Ag-Mo-Fe-Zn-Pb-Kaolin-$CaCO_3$)

Felipe André Reyes Reyes [1,*], Sebastián Pérez Cortés [1] and Ernesto Gramsch Labra [2]

[1] Department of Mining Engineering, Faculty of Engineering, Universidad de Santiago de Chile, Santiago 9170124, Chile; sebastian.perez@usach.cl
[2] Department of Physics, Faculty of Sciences, Universidad de Santiago de Chile, Santiago 9170124, Chile; egramsch@gmail.com
* Correspondence: felipe.reyesr@usach.cl

Workflow analysis for the Characterization of Chilean Mine Tailings

Accordingly, the proposed workflow can be directly applied to other national or regional tailings datasets by following the same sequence of data preprocessing, standardization, hierarchical clustering, dendrogram inspection, and geochemical interpretation.

Step 1. Data acquisition and initial screening: The analysis is based on the national mine tailings geochemical database provided by SERNAGEOMIN. The initial dataset consists of 2,012 geochemical records corresponding to 628 tailings deposits distributed across 17 exploited resource categories. Samples were classified according to the main commodity associated with each mining operation. The dataset was screened to retain only solid tailings samples with complete geochemical characterization, exclude duplicated records and samples with missing or inconsistent identifiers, and select chemical elements with sufficient analytical coverage across samples to ensure statistical robustness.

Step 2. Data cleaning and preprocessing: Concentration data were subjected to a systematic preprocessing protocol prior to multivariate analysis. This consisted of a representativeness analysis of non-detects and below-detection-limit values being treated using a consistent substitution criterion (e.g., half the detection limit or minimum reported value), ensuring numerical stability without artificially inflating variability.

Step 3. Data log transformation and z-score standardization: Given the wide dynamic range and compositional nature of geochemical data, concentrations were log-transformed to reduce skewness and mitigate the influence of extreme values and standardized using Z-score normalization (mean = 0, standard deviation = 1). This step ensures that all chemical

variables contribute equally to distance calculations, preventing dominance by high-abundance elements.

Step 4. Multivariate analysis implementation: In this study, hierarchical cluster analysis constitutes the primary classification tool used to define geochemical profiles of mine tailings. The clusters derived from the dendrogram represent the basis for the subsequent grouping and interpretation of tailings types. Hierarchical cluster analysis was conducted on a standardized dataset. It starts by calculating the Euclidean distance as a measure of dissimilarity and by applying Ward's linkage method (Ward.D2) to minimize within-cluster variance. Results are graphically presented using dendrograms, a graphical representation used in hierarchical cluster analysis to illustrate how a set of samples is progressively grouped based on their multivariate similarity.

In the dendrogram, each leaf corresponds to an exploited resource category associated with the tailings. The branches represent successive mergers of these resource-based profiles according to their geochemical similarity. The vertical axis reflects the level of dissimilarity at which resource categories are joined: low-height merges indicate resource types with highly similar geochemical signatures, whereas higher merges indicate increasingly distinct geochemical compositions. Dendrogram height was used as a metric to select the final number of clusters based on similarity structure and geochemical interpretability. In this study, the 17 exploited resource categories were grouped into four geochemically coherent clusters.

All multivariate analyses were conducted using the R statistical software environment (version 4.5.0). Data preprocessing, manipulation, and aggregation were performed using the tidyverse package, while additional data reshaping utilities were supported by reshape2. Hierarchical cluster analysis and data standardization were implemented using base R functions, and the visualization and interpretation of multivariate results were supported by the pheatmap and factoextra packages.

Step 5. Definition of geochemical profiles: The selection of the final number of clusters was based on a combined evaluation of dendrogram structure and its geochemical interpretability, rather than on an arbitrary numerical criterion. Specifically, the dendrogram was examined to identify major fusion levels characterized by pronounced increases in linkage distance, which indicate the merging of geochemically dissimilar groups. The chosen cut level was located immediately below these large jumps in dendrogram height, thereby preserving internally coherent groups while avoiding excessive aggregation.

This initial selection was further refined through geochemical interpretability by assessing whether the resulting clusters exhibited distinct and meaningful elemental signatures

consistent with known mineralogical associations, ore types, and tailings characteristics reported in the literature. Cluster solutions that produced geochemically heterogeneous or poorly interpretable groupings were discarded.

The clusters identified through hierarchical analysis were subsequently presented and examined using boxplots, which were employed to illustrate the distribution of elemental and compound concentrations within each cluster. This graphical approach allowed the predominance or absence of key geochemical variables to be clearly visualized, as well as the comparison of concentration ranges, central tendencies, and variability among clusters.

For each cluster, the concentration distributions of individual elements and oxides were analyzed and systematically compared across the different clusters, enabling the identification of characteristic geochemical signatures. Particular attention was given to elements and compounds that are diagnostic of specific mineralogical assemblages, beneficiation processes, or tailings types.

The interpretation of these distributions was further supported by an exhaustive review of the relevant literature, which provided an independent geochemical and mineralogical context for the observed patterns. This combined statistical and literature-based approach ensured robustness in the interpretation of each cluster and minimized the risk of attributing spurious meaning to purely statistical groupings. As a result, the cluster characterization reflects geochemically consistent and interpretable groupings that can be reliably used to generalize the behavior of extensive geochemical tailings datasets such as the one analyzed in this study.



Figure S1: Workflow for the Geochemical Characterization and Multivariate Classification of Chilean Mine Tailings.