Innovation in Characterization-Proyecto Touro – Galicia, Spain

Tom Meuzelaar[1], Pablo Núñez Fernández[2], Morgan Warren[1]

[1]*Life Cycle Geo, LLC., 729 Main Street, Longmont CO 80501, tom@lifecyclegeo.com, ORCID 0000-0003-3119-8390*

[2]*Cobre San Rafael SL, B70491121, La Mina, S/N 15822, Touro (San Xoan), La Coruña, Spain, pablo.nunez@atalayamining.com, ORCID 0000-0002-6218-0479*

Abstract

Mine waste characterization tests conducted on metamorphosed mine materials can substantially underpredict acid generation potential and overpredict acid neutralization potential if improper laboratory methods are used. Supervised machine learning methods were used to correct 60,000 such erroneous sulfur assay data points for the Proyecto Touro exploration assay dataset. A trained machine learning model showed excellent results, being able to predict sulfur concentrations with up to 93% accuracy. The innovative methods used in this study indicate that erroneous data generated from use of improper laboratory characterization tests does not necessarily need to be discarded, as machine learning algorithms can sometimes be used to correct them.

Keywords: Waste characterization, environmental geochemistry, unsupervised machine learning, supervised machine learning, digital transformation, artificial intelligence, data science

Introduction

Environmental characterization testing must be conducted on future mine waste material to determine its potential acid rock drainage (ARD) and metals leaching characteristics. When carrying out such tests on metamorphosed mine materials it is important to select the correct laboratory methods as tests that do not fully digest metamorphosed sulfides (e.g., pyrrhotite) can lead to substantially underpredicted acid generation potential (AGP) while tests that don't account for metalliferous carbonates (e.g., rhodochrosite, siderite) or graphite can lead to overpredicted acid neutralization potential (ANP); this can lead to considerable underprediction of overall material ARD potential (Meuzelaar *et al*., 2021a).

When Cobre San Rafael obtained the Touro copper property[1], it also acquired multiple legacy assay datasets. The legacy waste and ore sulfur assay data from Proyecto Touro in Galicia, Northwestern Spain were frequently obtained using weak acid methods that did not fully digest metamorphosed sulfide minerals (principally pyrrhotite). Additionally, the presence of both graphite and manganese–iron carbonates in Touro added considerable risk of ANP overestimation. While environmental characterization tests conducted on newer ore and waste samples collected by Cobre San Rafael for Proyecto Touro aimed to properly address these risks, the older legacy dataset contained considerable inaccuracy, especially in sulfur measurement. Accordingly, there was a need to correct the underestimated and erroneous legacy sulfur data. The cost to re-assay 60,000 erroneous sulfur data points was deemed prohibitive, therefore innovative machine learning algorithms were employed to attempt to correct this data. The innovative approach relied on re-analysis of a small subset of samples, and training a machine learning model to 1) predict sulfur concentrations based on concentrations of other chemical components in the data subset, and 2) predict corrected sulfur values for the entire exploration assay database.

[1]Atalaya Mining currently owns 10% but has a phased earn-in agreement for up to 80% ownership

Background

Proyecto Touro is located in the Galicia province in northwest Spain. It is a brownfield copper project, having been historically mined from 1973 to 1986. The mine footprint comprises six separate deposits (Arinteiro, Vieiro, Bama, Brandelos, Monte de las Minas and Arca). Among them, Monte de las Minas and Arca have not been mined yet. Regionally, the deposits are located in the Órdones Complex, an extensively metamorphosed allochthnous unit within the larger northwest Iberian Massif. Member units are predominantly meta-sedimentary (paragneiss) and meta-volcanic (amphibolite) lithologies. Copper mineralization occurs as chalcopyrite mostly within the meta-volcanics in a Besshi-type volcanogenic massive sulfide configuration. Both paragneiss and amphibolite contain considerable metamorphosed pyrite, as pyrrhotite, and this represents the primary mineral that gives future Touro waste rock ARD potential. Cobre San Rafael aims to properly characterize and understand waste rock ARD potential as part of its long-term water and materials management plan, and to assist with permitting.

Methods

For this study, Cobre San Rafael provided assay data for 5,880 samples comprising eight different ore and waste lithologies (Table 1, below) with measurements of 49 different chemical elements. Some samples contained very little sulfur (<0.01 wt. %), while others such as the Massive Sulfide lithology samples contained sulfur concentrations as high as 11.8 wt. %. The original legacy data sulfur analyses were obtained using a three-acid aqua regia digestion that was insufficient to fully digest the metamorphosed pyrite and pyrrhotite in the samples.

As is typical for a geochemical dataset being prepared for machine learning analysis, considerable data wrangling was required to address the issue of numeric closure (Aitchison 1982) and the presence of censored data (i.e. detection limits). Major element compositional data sums to a constant of 100% which introduces artificial collinearity that need to be removed prior to statistical modeling. This collinearity, termed numeric closure, may be addressed by transforming the data using log-transforms such as the centered-log ratio (Pawlowsky-Glahn and Egozcue 2006). Censored data include those that are either above or below a laboratory detection limit. While techniques such as use of whole or half the detection limit are commonly used to address censored data, a preferred method is to impute (predict) numeric values in place of the censored constants, based on the bulk chemistry of the sample (Sanford *et al.* 1993). This imputation employs the expectation-maximization (EM) algorithm (Palarea-Albaladejo and Martin-Fernandez 2015). All data cleaning, transformation and machine learning-based analysis was performed using the R statistical computing environment (R Core Team 2017); scripts were executed with the Microsoft Azure Machine Learning Studio platform.

Machine learning algorithms are particularly useful for identifying patterns in high dimensional datasets, such as

Table 1 *Sulfur Assay Data, by Lithology.*

Logged Lithology	Number of Samples	Median Sulfur Concentration (wt. %)
Amphibolite	1742	1.9
Garnet Amphibolite	716	3.0
Ca-poor Amphibolite	1361	3.5
Breccia-Massive Sulfide	32	11.7
Biotitic Schist	297	4.3
Massive Sulfide	155	5.9
Pelitic Paragneiss	1305	0.8
Pelitic paragneiss with sulfide	272	10.4

multivariate chemical composition. The conceptual model for supervised machine learning predictions were based on the hypothesis that material sulfur concentrations are dependent on mineral type and abundance and can be predicted based on other chemical elements that are also associated with minerals that control sulfur abundance. Given that elements such as sulfur are typically found in the structure of multiple different minerals, the relationship between sulfur and other assay elements is not readily predicted by simpler methods. To achieve optimal predictive accuracy, multiple machine learning algorithms were tested including artificial neural networks (ANNs) boosted decision trees (BDTs), multiple linear regression and random forest (RF). The primary statistical criteria used to assess the accuracy of sulfur predictions were the coefficient of determination (r2) and mean squared error (MSE). To aid in interpretation, variable influence was calculated for each of the algorithms employed using a feature importance algorithm provided within the Azure environment.

Results

MSE and r^2 values for each of the algorithms are provided in Table 2. The results clearly indicate that BDTs and ANNs (both with r^2 values of 0.93 and order of magnitude lower MSE) considerably outperform multiple linear regression and the RF algorithm.

Given their high predictive accuracy, BDTs and ANNs were selected for further analysis. A comparison between machine learning-predicted sulfur and the raw training data is given in Fig. 1, which indicated that both models are least accurate in predicting low sulfur concentration data (<0.1 wt. %). Given that mine materials with sulfur concentrations below this threshold have very low ARD potential, model underperformance at these lower concentrations is considered less critical. However, it is noted that the BDT is a bit better at predicting low sulfur concentrations than the ANN algorithm.

Results of the variable influence analysis are given in Table 3 for both algorithms. For both analyses iron is, by some measure, the most critical element necessary to predict sulfur concentrations. Given that iron is one of the two primary elements comprising pyrrhotite (the other being sulfur), this result is intuitive. Beyond this, variable influence for the BDT algorithms are more intuitive as most variables are chalcophile or siderophile elements that are commonly found substituting for iron and sulfur in the crystal lattice of sulfide minerals (and also represent the primary risk for metals leaching when sulfides oxidize). The compositional variables that contribute to ANN predictive accuracy are less intuitive – they may possible have something to do with rock forming or secondary (hydrothermal) processes, but this is not immediately known.

However, the results generally validate the hypothesis that the chemical elements used to predict sulfur concentrations are, as expected, broadly controlled by mineral concentrations and crystal structure. As such, using a multivariate machine learning-based approach is a powerful tool for predicting elemental concentrations. This approach could also be used for many similar applications with environmental implications, including:

- Training a combined bulk chemistry/mineralogical composition database to predict graphite or various metal-bearing carbonates
- Predicting trace element concentrations within specific minerals (e.g., leachable

Table 2 Machine Learning Prediction Results.

Supervised Machine Learning Algorithm	Mean Squared Error MSE	Regression Coefficient r^2
Multiple Linear Regression	2.47	0.66
Random Forest	9.56	0.22
Boosted Decision Trees (BDT)	0.46	0.93
Artificial Neural Network (ANN)	0.43	0.93

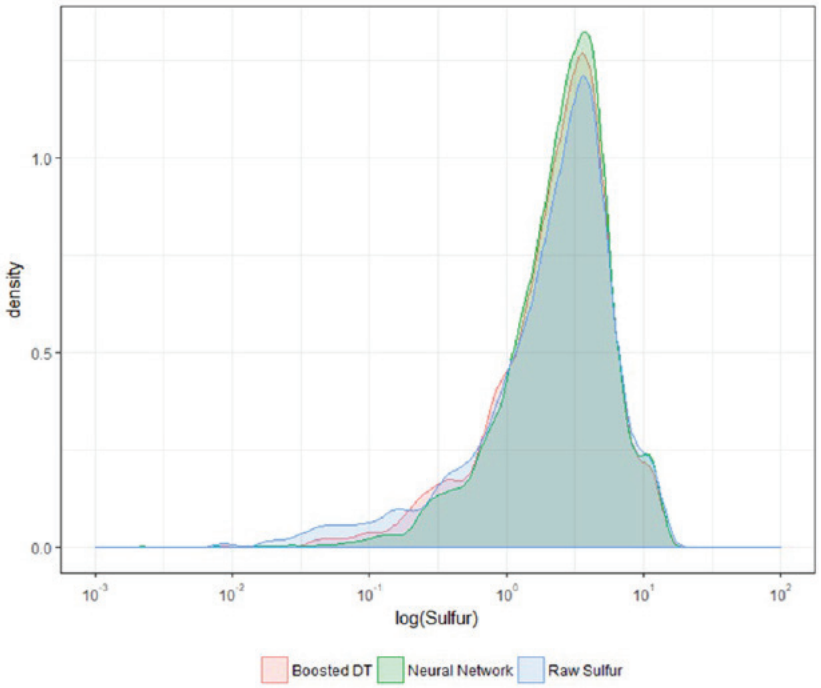

Figure 1 *Distribution of Predicted and Raw Sulfur Concentrations.*

selenium in sulfide minerals vs. "locked" selenium in silicates)

- Predicting acid-generating soluble alunite vs. non-soluble "hydrothermal" alunite vs. other acid sulfate minerals

Finally, with modern field spectroscopy-based instrumentation, an IoT (Internet of Things) configuration that includes field data acquisition, upload to a cloud-hosted database, machine learning analysis on the cloud and rapid results to support quicker decision-making, should be viable.

Conclusions

The trained machine learning model showed excellent results being able to predict sulfur concentrations with 93% accuracy. The high accuracy is the result of having sufficient assay data points and chemical parameters, as well as the fact that sulfide mineral concentrations

Table 3 *Variable Importance Analysis Results for ANN and BDT sulfur predictions*

Ranking	Boosted Decision Tree BDT		Artificial Neural Network (ANN)	
	Element	Ranking	Element	Ranking
1	Fe	0.23	Fe	0.28
2	Cd	0.13	Sr	0.09
3	Zn	0.11	Th	0.07
4	Co	0.07	Ti	0.07
5	Ag	0.06	Tl	0.05
6	Cr	0.05	Cr	0.05
7	Al	0.05	Ca	0.03
8	Ni	0.03	Ba	0.03
9	Na	0.03	U	0.03
10	Se	0.03	Rb	0.03

(largely pyrrhotite) correlate strongly with the concentrations of trace metals which substitute into the sulfide lattice.

The innovative supervised sulfur prediction, LECO digestion and Modified Sobek titration methods employed in this study indicate that erroneous data generated from use of improper laboratory tests does not necessarily need to be discarded. Rather, such methods offer a pathway to correction of erroneous data.

Acknowledgements

The authors acknowledge provision of legacy sulfur data by Cobre San Rafael. Specifically, Fernando Diaz-Riopa and Julian Sanchez of Cobre San Rafael are acknowledged, as well as Alan Nobel and Monica Barrero. Much of this work was completed as an internal Golder Associates research and development project. Microsoft graciously provided access to their Azure cloud-based analytics environment for this work.

References

Aitchison, J. (1982) The statistical analysis of compositional data (with discussion). Journal of the Royal Statistical Society, Series B (Statistical Methodology), v. 44, 139–177

Meuzelaar, T., Núñez-Fernández, P., Martín-Izard, A., Arias-Prieto, D., Díaz-Riopa, F. (2021) The Waste rock of the Touro copper deposit in Galicia, Spain: challenges for its environmental characterization. Geochemistry: Exploration, Environment, Analysis v. 21, p. 1–14 doi:10.1144

Palarea-Albaladejo J., Martin-Fernandez, J.A. (2015) zCompositions – R package for multivariate imputation of left censored data under a compositional approach. Chemometrics and Intelligent Laboratory Systems, v. 143, 85–96.

Pawlowsky-Glahn V., Egozcue J.J. (2016) Compositional data and their analysis: an introduction, in Buccianti, A., Mateu-Figueras, G. and Pawlowsky-Glahn, V. (eds) Compositional Data Analysis in the Geosciences: From Theory to Practice. Geological Society, London, Special Publications, v 264, 1–10.

R Core Team (2017) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Sanford, R.F., Pierson, C.T., Crovelli, R.A. (1993) An objective replacement method for censored geochemical data. Mathematical Geology, v 25, 59–80.