Classification of lithostratigraphic and alteration units from drillhole lithogeochemical data using machine learning: a case study from the Lalor volcanogenic massive sulphide deposit, Snow Lake, Manitoba, Canada

Antoine Caté[1], Ernst Schetselaar[2], Patrick Mercier-Langevin[3], and Pierre-Simon Ross[1]

[1] Institut national de la recherche scientifique, Centre Eau Terre Environnement, 490 rue de la Couronne, Québec (QC), G1K 9A9, Canada

[2] Geological Survey of Canada, 601 Booth Street, Ottawa (ON), K1A 0E8, Canada

[3] Geological Survey of Canada, 490 rue de la Couronne, Québec (QC), G1K 9A9, Canada

Abstract

Classification of rock types using geochemical variables is widely used in geosciences, but most standard classification methods are restricted to the simultaneous use of two or three variables at a time. Machine learning-based methods allow for a multivariate approach to classification problems, potentially increasing classification success rates. Here a series of multivariate machine learning classification algorithms, together with different sets of lithogeochemistry-derived variables, are tested on samples collected at the Lalor Zn-Cu-Au volcanogenic massive sulphide deposit, to discriminate volcanic units and alteration types. Support Vector Machine and Ensemble method algorithms give the best performance on both classification exercises. Untransformed chemical element concentrations with high classification power are the best-performing variables. Classification success rates are equal or better than those obtained using standard classification methods and are satisfactory enough for the use of the resulting predictions for 2D and 3D modelling of geological units.

Highlights

- Machine learning algorithms are used for multivariate geochemical classification.
- Volcanic units and alteration types are discriminated using untransformed chemical element concentrations.
- Support Vector Machine and Ensemble methods yield the highest classification success scores.

Keywords

Lalor; Snow Lake; mining exploration; lithogeochemistry; multivariate classification; machine learning

1. Introduction

Machine learning is increasingly being used to aid interpretation of geological data (e.g., O'Brien et al., 2015; Rodriguez-Galiano et al., 2015; Sadeghi and Carranza, 2015; Kirkwood et al., 2016). Contrary to traditional geochemical classification diagrams, which are generally limited to two or three variables at a time (e.g., Pearce and Norry, 1979; De La Roche et al., 1980; Wood, 1980; Verma and Agrawal, 2011), machine learning algorithms such as neural networks and support vector machines allow for the simultaneous use of multiple variables. These approaches reduce interpretation bias and can outperform the traditional graphical or statistical classification methods (Friedman et al., 2001). The application of these algorithms however requires iterative and empirical tuning of weights and parameters for approximating an optimal classification function. As a result, their application can be considered as a ‘black box’ approach by some, mainly due to the lack of a simple link between the weights and estimated parameters, and the classification function being approximated.

This study illustrates the power of multivariate classification methods applied on drillhole geochemical data from altered volcanic rocks hosting the volcanogenic massive sulphide (VMS) Lalor deposit in the Snow Lake area in Manitoba, Canada. VMS deposits generally consist of stratiform to stratabound ore lenses underlain by discordant sulphide stringer (feeder) zones. These deposits are closely associated with volcanic rocks (Franklin et al., 2005; Galley et al., 2007a), and recognizing specific volcanic units, or volcanic horizons, is key in defining vectors towards favourable host rocks (e.g., Gibson et al., 1999). In addition, the formation of VMS deposits is associated with extensive, up to regional scale, hydrothermal alteration of the host rocks (e.g., Galley, 1993; Galley et al., 1993), and variations in alteration styles and mineral assemblages in space are critical exploration vectors toward ore at the regional and deposit scales. The Lalor deposit is an excellent area for testing the multivariate classification methodology. Lithologies, alteration, mineralization, and the metamorphic and tectonic contexts are well studied (e.g., Tinkham, 2013; Caté et al., 2015; Schetselaar et al., 2017), and an extensive set of data has been collected by the company exploiting Lalor (Hudbay) and several scientific teams (Geological Survey of Canada, Laurentian University and Manitoba Geological Survey). Moreover, the local geology is complex, with a wide variety of volcanic lithologies overprinted by complex hydrothermal, deformation and metamorphic events (Caté et al., 2015; Caté, 2016).

Supervised multivariate classification can help categorizing and mapping volcanic rocks and alteration types that have been identified and discriminated on a well-studied subset (the training set) of geochemical drillhole data (e.g., Abbaszadeh et al., 2015). One significant challenge in such environments is to differentiate between the protolith signature (e.g., Ross et al., 2014) and the signal specific to the overprinting hydrothermal alteration (i.e. post-depositional geochemical modifications to the protolith signature) (e.g., Ross et al., 2016). Protoliths are finite, spatially and statistically coherent features for the most part, whereas alteration 'units' are gradational and irregular in nature. The performance of a series of classifiers and multivariate geochemical datasets, including variable transformations, are specifically tested for the classification of volcanic units and alteration types in this paper. The classification results are plotted in 3D space and on conventional classification diagrams to validate their geological significance and determine their success rates. Our results indicate that machine learning models based on lithogeochemical data can be efficient classifiers for lithostratigraphic units and alteration types. Both of these applications, however, necessitate to carefully select discriminative variables and algorithms to obtain high classification success rates.

2. Geological setting

Lalor is a Zn-Cu-Au VMS deposit located in the Snow Lake arc assemblage of the Paleoproterozoic Flin Flon greenstone belt (Galley et al., 2007b). The deposit is currently being mined by HudBay Minerals Inc. (Hudbay) and has been studied in detail (Bailes et al., 2013; Tinkham, 2013; Caté et al., 2014a; Caté et al., 2014b; Lam et al., 2014; Mercier-Langevin et al., 2014; Bellefleur et al., 2015; Caté et al., 2015; Duff et al., 2015; Schetselaar and Shamsipour, 2015; Caté, 2016; Duff, 2016; Schetselaar et al., 2017).

The Lalor deposit consists of stratigraphically and structurally stacked ore lenses (Bellefleur et al., 2015; Caté et al., 2015) hosted in volcanic and subvolcanic rocks informally categorized into units and groups of units (Figure 1 and Table 1; Caté, 2016).

The host rocks of the ore lenses are known informally as the Lalor volcanic succession (Caté, 2016). This succession comprises the Footwall volcaniclastic unit, the Moore volcanics (composed of the Moore basalt and the stratigraphically younger Upper Moore mafic unit), the Lalor rhyolite, and the ‘Lalor’ Powderhouse dacite (Figure 1). These units dip ~30° to the east-northeast and face upward. Below the Lalor volcanic succession and to the West of it, the Western volcanic succession is composed of the ‘Western’ Powderhouse dacite, which is interpreted as a structurally-distinct sliver of the Powderhouse dacite present in the Lalor volcanic succession (Caté, 2016). The Balloch volcanic succession structurally overlies the Lalor volcanic succession. It is composed of steeply dipping WSW-facing and overturned volcanic units (Bailes et al., 2013). These units are the North Balloch rhyodacite, the Balloch basalt, the Ghost Lake rhyodacite, the Threehouse volcanics (North Balloch mafic intrusive, Threehouse diorite, Threehouse mafic unit and Upper Threehouse mafic unit), and the North Chisel dacite. Mafic, intermediate and felsic dykes are present within all units. The Moore and Threehouse volcanic assemblages are two groups of volcanic and intrusive units sharing similar geochemistry and magmatic origin (Caté, 2016) but present at distinct stratigraphic positions.

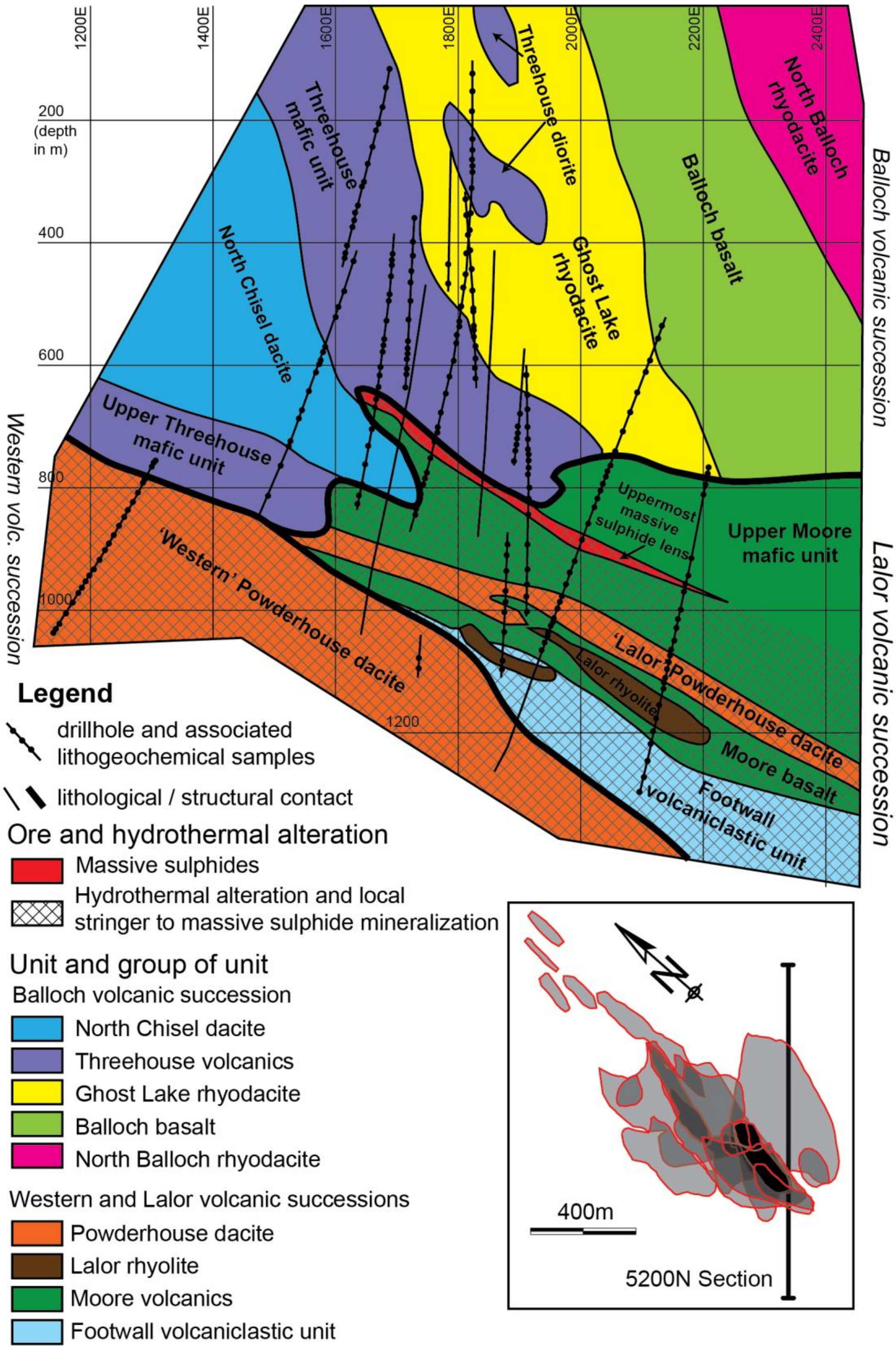

Figure 1: Section 5200N of the Lalor deposit, after Bailes et al. (2013) and Caté (2016). The North Balloch mafic intrusive does not appear in this section, but it is present within the North Balloch rhyodacite elsewhere

in the study area. The location map of the section (view from above) is presented with the simplified traces of ore lenses.

Hydrothermal alteration overprints the volcanic rocks in the deposit vicinity (Figure 1) and these altered volcanic rocks were subsequently affected by regional deformation and metamorphism, which makes it very difficult to reliably discriminate units and alteration types solely based on visual inspection. In these situations, lithogeochemical analyses provide additional, and often critical, insights on the nature of the protolith of altered rocks (e.g., Barrett and MacLean, 1994). A series of diagrams from the literature (Winchester and Floyd, 1977; Pearce, 1996; Ross and Bédard, 2009) have been used in Caté et al., 2014a (Figure 2) to determine the geochemical signature of volcanic units in the Lalor area. The Zr/TiO_2 versus SiO_2 diagram (Figure 2A) gives insight on the magmatic differentiation and the alkalinity of rocks. However, SiO_2 concentrations are affected by alteration, causing a noticeable spread in the data. The Nb/Y versus Zr/Ti diagram (Figure 2B) gives similar information and is not significantly affected by alteration at Lalor, hence providing better clustering for discriminating volcanic rocks. The log Zr/Y versus log Th/Yb diagram (Figure 2C) classifies the magmatic affinity of volcanic units. The combined use of these diagrams allows naming and discriminating each volcanic unit despite some partial overlap. Despite being relatively widely used, these classification diagrams still use only a few major oxides and trace elements, which leads to partly subjective class definitions and potentially limits classification performance.

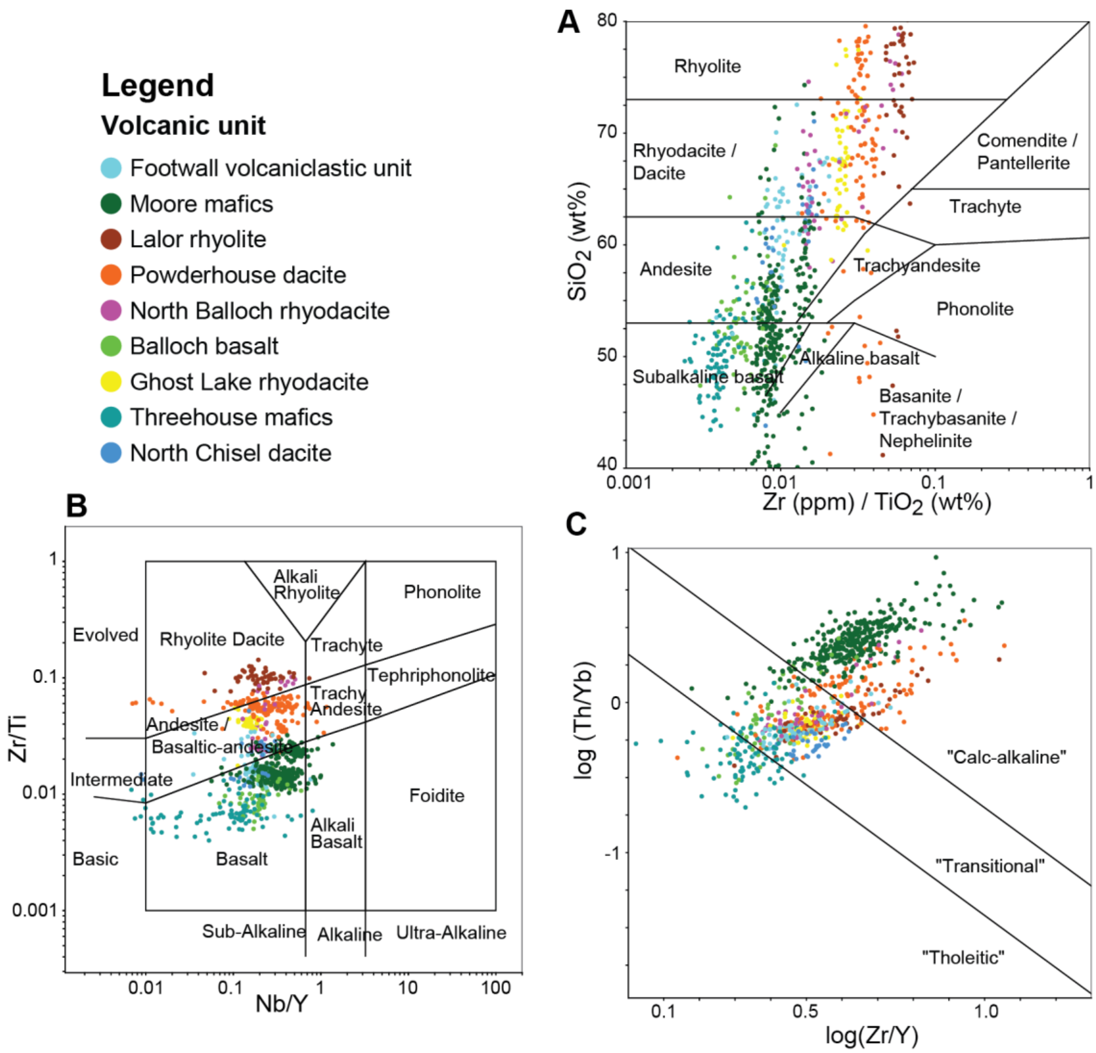

Figure 2: Discriminant geochemical diagrams for the volcanic and intrusive units and groups of units of the Lalor area with samples from the training dataset (data from the Geological Survey of Canada; Caté et al., 2017). A: Winchester and Floyd (1977) classification diagram; B: Pearce (1996) classification diagram modified from Winchester and Floyd (1977); C: Magmatic affinity diagram from Ross and Bédard (2009). Note that some lithostratigraphic units defined in the Lalor VMS camp plot in single fields whereas others straddle field boundaries in the diagrams.

The Lalor volcanic succession is affected by extensive syn- and post-VMS alteration that has partly obliterated the primary textures, mineralogy and geochemistry of the volcanic rocks (Figure 1; Caté et al., 2015). Alteration styles have been grouped by their chemical affinity and intensity (Table 1; Caté et al., 2015). The intense K, K-Mg-Fe, Mg-Fe and Mg-

Ca alterations are present in the Lalor volcanic succession as haloes around the ore lenses and in the footwall. Zones of moderate-intensity alteration with variable chemical signatures are also present in the footwall of the deposit (moderate footwall alteration) and in the Western volcanic succession (distal alteration; Caté, 2016). They are grouped here as ‘moderate alteration’ for simplicity. Post-VMS Ca metasomatism is present in all volcanic successions and overprints syn-VMS alteration (Caté et al., 2015). The Snow Lake area has been affected by middle-amphibolite grade metamorphism (Froese and Gasparrini, 1975; Menard and Gordon, 1997) resulting in unusual metamorphic mineral assemblages in altered rocks comprising chlorite, amphiboles, muscovite, aluminosilicates, quartz, staurolite, garnet, cordierite, carbonates, talc and diopside (Zaleski et al., 1991; Galley et al., 1993; Caté et al., 2015). The geochemical signature of alteration in VMS deposits can be represented in a box-plot diagram modified from Large et al. (2001) (Figure 3). Least altered rocks mostly plot in the fields of unaltered basalt, andesite, dacite, and rhyolite. Moderately altered rocks plot in the least altered fields or at higher Alteration Index (AI) values. Most intensely altered rocks do not display AI and chlorite-carbonate-pyrite index (CCPI) values matching that of least altered rocks, and have extremely high AI (>80) and/or CCPI (>95) values. The high AI and CCPI values for altered rocks are in agreement with the mineralogical assemblages (Caté et al., 2015) and $\delta^{18}O$ variations at deposit scale (Mercier-Langevin et al., 2014). Significant overlaps exist in the distribution of alteration types within the diagram, especially between least altered and moderately altered rocks.

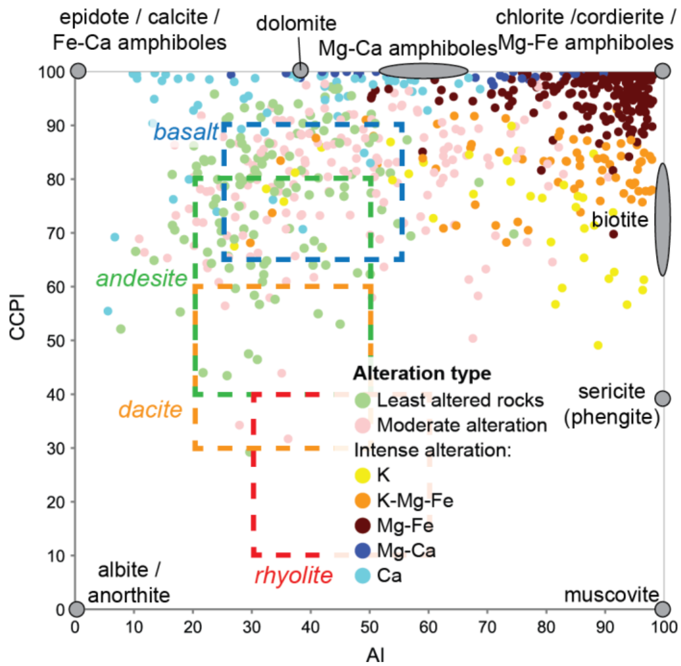

Figure 3: Box-plot diagram (modified from Large et al., 2001) showing the geochemical signature of samples from the training dataset affected by the different alteration types. The main alteration-related minerals present at Lalor are indicated. Fields representing the general distribution of least altered volcanic rocks (basalt, andesite, dacite and rhyolite) are from Gifkins et al. (2005). AI = 100(K_2O+MgO) / (K_2O+MgO+Na_2O+CaO); CCPI = 100(MgO+FeO) / (MgO+FeO+Na_2O+K_2O).

The Lalor deposit and its host rocks have been affected by polyphase deformation during the Trans-Hudson Orogen (Lucas et al., 1996; Kraus and Williams, 1999; Caté et al., 2014b) dominated by the D_2 event, which is characterized by a SSW verging fold and thrust tectonics with associated S_2 foliation and L_2 stretching lineation, the first being axial planar to F_2 folds.

Table 1: Volcanic and intrusive units and alteration types present at Lalor (compilation from Caté, 2016).

Volcanic and intrusive units
Lalor and Western volcanic successions:
Footwall volcaniclastic unit, Moore volcanics, Lalor rhyolite, Powderhouse dacite
Balloch volcanic succession:
North Balloch rhyodacite, Balloch basalt, Ghost Lake rhyodacite, Threehouse volcanics, North Chisel dacite
Alteration types
Unaltered: *Least altered*
Syn-VMS hydrothermal alteration:
Moderate alteration, K, K-Mg-Fe, Mg-Fe, Mg-Ca
Post-VMS metasomatism: *Ca*

3. Materials and methods

3.1. Lithogeochemical database

The geochemical data used for the classification of lithostratigraphic and alteration units consist of major oxide and trace element analyses of 7335 drillcore samples acquired by Hudbay and the Geological Survey of Canada (Caté et al., 2017). A total of 54 elements were analyzed on most of the samples. The geochemical dataset contains a very small proportion (<0.3%) of analyses under the detection limit, which have been arbitrarily set to half the detection limit to avoid ‘zero’ values in the database. A total of 44 samples with missing data were discarded.

Samples collected by the Geological Survey of Canada were individually described in detail and well constrained in terms of stratigraphic position, lithology, volcanic unit and alteration type (Caté, 2016). The samples consist of 20 cm-long full-core or half-core sections. They were analyzed by Activation Laboratories Inc., Ancaster, Ontario using a combination of methods that provide precise and accurate results for each element (see Caté, 2016 p. 27 for details on analytical procedure). Precision, accuracy and blanks were monitored by the authors. These analyses provide a training dataset for the classification. Two distinct training sets have been defined for the two series of classes (lithostratigraphic units and alteration types). For each training set, four series of predictor variables derived from elemental analyses were selected.

Samples collected by Hudbay were analyzed by Activation Laboratories Inc., Ancaster, Ontario. Sample length varies but each sample had to be uniform in texture and composition. Major elements were determined using metaborate-tetraborate fusion followed by inductively coupled plasma atomic emission spectrometry. Minor and trace elements were determined by a combination of metaborate-tetraborate fusion, four-acids digestion and two-acids digestion followed by inductively coupled plasma atomic emission spectrometry mass spectrometry or inductively coupled plasma atomic emission spectrometry. Duplicates, standards and blanks were analyzed, but monitoring was not performed by the authors.

3.2. Labelling training sets

A total of 922 samples from drillholes investigated by the Geological Survey of Canada were considered for the training of predictive models (Caté, 2016). These samples are well constrained and were acquired from carefully logged drillholes making them ideal candidates for training models. Analyses of veins and other heterogeneities potentially affecting results were removed from the database. The training set of lithostratigraphic units contains 837 samples from drillholes investigated by the Geological Survey of Canada. A total of 85 samples with an uncertain lithostratigraphic assignation were not taken into account. A unit (or group of units) name as presented in the legend of Figure 1 is attributed to each sample. Classes were attributed using a combination of: 1) geochemical signatures (i.e., Figure 2 and several other diagrams shown by Caté, 2016, and listed in Table 2); 2) volcanic textures and mineralogy preserved from the alteration and indicative of the physical and compositional nature of the units when present (Table 3); and 3) the spatial distribution of volcanic units as presented in Figure 1 (see Caté, 2016, chapters 3 and 4 for more details).

Table 2: List of diagrams used to determine the geochemical signature of volcanic units and groups of volcanic units at Lalor.

Name	Elements	Reference
La/Yb vs. Zr/Ti	La, Ti, Yb, Zr	
Th/Yb vs. Zr/Ti	Th, Ti, Yb, Zr	
TAS	K_2O, Na_2O, SiO_2	Le Maître, 1989
Zr/Ti vs. Nb/Y	Nb, Ti, Y, Zr	Pearce, 1996
Th-Co Discrimination Diagram	Co, Th	Hastie et al., 2007
Th/Yb vs. Zr/Y	Th, Y, Yb, Zr	Ross and Bédard, 2009
AFM	FeO, K_2O, MgO, Na_2O	Kuno, 1968 and Irvine and Baragar, 1971
Spider diagram	Ce, Dy, Er, Eu, Gd, Hf, La, Lu, Nb, Nd, Pr, Sm, Ta, Tb, Th, Ti, Y, Yb, Zr	

Table 3: Typical mineralogical composition, volcanic textures and lithofacies for each volcanic unit and group of volcanic units at Lalor, for the least altered rocks. These features can be partially or totally obliterated in rocks affected by hydrothermal alteration.

Unit	Composition	Textures and lithofacies
Footwall volcaniclastic unit	Intermediate	Volcaniclastic
Moore volcanics	Mafic to intermediate	Coherent with feldspar phenocrysts or volcaniclastic
Lalor rhyolite	Felsic	Coherent to breccia
Powderhouse dacite	Felsic	Coherent to volcaniclastic - feldspar phenocrysts
North Balloch rhyodacite	Felsic to intermediate	Coherent to volcaniclastic
Balloch basalt	Mafic	Volcaniclastic to coherent
Ghost Lake rhyodacite	Felsic	Volcaniclastic to coherent
Threehouse volcanics	Mafic	Volcaniclastic, intrusive or coherent - feldspar (rarely amphibole) phenocrysts

North Chisel dacite	Intermediate	Volcaniclastic

Alteration type is labelled on 680 training samples out of the 922. A total of 242 samples with unclear or undefined alteration type were not taken into account. The alteration type was attributed based solely on the mineralogical composition (based on a visual inspection) of samples using key minerals indicator of the geochemical signature of the alteration as discriminants (Table 4), as detailed in Caté et al. (2015) and Caté (2016). A subsequent verification of the validity of these types based on geochemical diagrams (see below) was completed.

Table 4: Summary of the discriminative mineralogy of alteration types

Alteration type	Discriminant mineralogy
Least altered	Absence or trace amounts of metamorphosed alteration-associated minerals (e.g., muscovite, Mg-Fe amphiboles, chlorite, cordierite, staurolite)
Moderately altered	Presence of metamorphosed alteration-associated minerals, >5% feldspar, preserved volcanic textures
K	>5% muscovite, <5% feldspar
K-Mg-Fe	>5% biotite, <5% muscovite, Mg-Fe amphiboles, cordierite, chlorite and/or Ca amphiboles, <5% feldspar
Mg-Fe	>5% chlorite, Mg-Fe amphibole or cordierite , <5% feldspar
Mg-Ca	>20% chlorite with >5% carbonate and/or Ca-amphiboles
Ca	Ca-amphibole and/or epidote assemblages overprinting other mineral assemblages

3.3. Classifier variables

The success rate of multivariate classification is strongly influenced by the input data and how it has been preprocessed (e.g., Domingos, 2012). For each classification exercise, a total of four distinct sets of predictor variables were built to test their effect on classification success.

Magmatic rocks can be discriminated using a restricted set of elements that are dependent on the formation and evolution of magmas and less susceptible to hydrothermal alteration and metasomatism (Winchester and Floyd, 1977; Pearce et al., 1984; Pearce, 1996). These are known as immobile elements (Winchester and Floyd, 1977) and in a VMS setting, they typically include Al, Zr, Ti, Nb, Y, Hf, Ta, Th and heavy Rare Earth Elements (Gifkins et al., 2005). Ratios of immobile elements remain constant regardless of the hydrothermal alteration intensity. For the classification of volcanic units, a total of four sets of variables were created (Table 5). The first set corresponds to the element ratios (plus SiO_2) used in binary classification diagrams used to determine the geochemical signature of volcanic rocks at Lalor (Figure 2). The second set (restricted set of elements) corresponds to the concentrations in elements used to derive the previous ratios in addition to the concentrations in elements utilized in extended spider diagrams in Caté et al. (2014a) for volcanic rocks classification. The third set of variables (extended set of elements) corresponds to an extended selection of 26 elements that were shown to have an important classification power in altered volcanic rocks (Pearce, 1996). Most of the elements and oxides in these three sets are immobile in most VMS settings, except SiO_2 and sometimes the light REE (e.g., MacLean and Kranidiotis, 1987).

Because geochemical analyses are compositional data, they are affected by the closure problem, and element concentrations do not vary independently (Aitchison, 1982; Pawlowsky-Glahn and Egozcue, 2006). To test the effect of data closure on classification, the extended set of elements was converted in centered-log-ratios (CLR; Aitchison, 1982) in the fourth set of variables. This transformation opens the data and thus removes spurious correlations between elements related to the closure effect.

Table 5: Sets of variables used for multivariate classification. Alteration indexes are from Ishikawa et al., 1976 (AI), Large et al., 2001 (CCPI), Kishida and Kerrich, 1987 (Muscovite Saturation Index (MSI) and Carbonate Saturation Index (CSI)) and Gemmell, 2006 (Sodium-Sulphide Index (SSI)).

Classification of volcanic units		
1	Element ratios	SiO_2 (ppm), Zr/TiO2, Nb/Y, Th/Yb, Zr/Y
2	Elements (restricted)	SiO_2, TiO2, Nb, Zr, Y, Th, La, Ce, Pr, Nd, Sm, Gd, Tb, Dy, Ho, Er, Yb, Lu (in ppm)
3	Elements (extended)	SiO_2, Al_2O_3, TiO_2, P_2O_5, Nb, Zr, Y, Th, Cr, Ni, Sc, V, La, Ce, Pr, Nd, Sm, Gd, Tb, Dy, Ho, Er, Tm, Yb, Lu, Co (in ppm)
4	CLR-transformed elements	Log of the elements (extended set) divided by their geometric mean
Classification of alteration types		
1	Alteration indices	AI [100*(K_2O+MgO)/(K_2O+MgO+CaO+Na_2O)], CCPI [100*(FeO+MgO)/ (FeO+MgO+Na_2O+K_2O)], MSI [(3*2*K_2O/94.196)/(2*Al_2O_3/101.961276)], SSI [100*(S/32.066)/(S/32.066+2*Na_2O/94.196)], CSI [(CO_2/44.0095)/(CaO/56.0774+MgO/40.3044+FeO/71.8444)]
2	Elements (restricted)	SiO_2, Al_2O3, MgO, Fe_2O_3, CaO, Na_2O, K_2O, CO_2, S (in ppm)
3	Elements (extended)	SiO_2, Al_2O_3, MgO, Fe_2O_3, CaO, Na_2O, K_2O, MnO, CO_2, S, Ba, Sr, Rb, Ag, As, Bi, Cd, Cu, Pb, Sb, Zn, Ni (in ppm)
4	CLR-transformed elements	Log of the elements (extended set) divided by their geometric mean

Geochemical discrimination of alteration types is mainly based on mobile major elements, volatiles and sulphur (e.g. MacLean and Kranidiotis, 1987; Barrett and MacLean, 1994; Piché and Jébrak, 2004). For the classification of alteration type, a total of four sets of variables were tested (Table 5). The first set corresponds to a series of alteration indices combining several elements used in Caté (2016) to illustrate the different alteration types at Lalor. The second set is composed of all the elements and oxides forming the alteration indices. The third set corresponds to an extended set of elements with major oxides, CO_2, S, alkaline and alkaline-earth elements and trace metals. The trace elements added in the third set are typically mobile in VMS environments and/or related to mineralization (Gifkins et al., 2005). The last set of variables corresponds to the CLR-transformed third set of variables.

3.4. Multivariate classification

Multivariate classification is widely and successfully used in science (e.g., Haaland et al., 1997), and has many applications in geosciences and mineral exploration (Schetselaar et al., 2000; Cracknell et al., 2014; Abbaszadeh et al., 2015; Carranza and Laborte, 2015; O'Brien et al., 2015) including lithological discrimination in VMS environments (e.g., Fresia et al., 2017). Multivariate classification resorts to using several variables (X_1, X_2,…, X_{n-1}, X_n) that describe a set of samples, and that will allow to discriminate between classes among these samples. In supervised classification, an algorithm will divide the n dimensional space into volumes attributed to each class using a labelled training set for which the class of each sample is already attributed. The rest of the dataset is then classified by subjecting all the remaining (or unlabelled) samples to the classification model based on the location of each sample in the n-dimensional space. A total of five classification algorithms have been tested using the Python Scikit-learn module (Pedregosa et al., 2011).

K-nearest neighbour

The supervised K-nearest neighbor (KNN) classification method is based on the selection of a number (K) of training samples closest in the Euclidean space from the sample that has to be classified. The classification criterion is the predominant class within the K samples (Peterson, 2009). The K variable is the main adjustable parameter of the method. A weighting function of the Euclidean distance between the classified sample and the training samples can be introduced.

Gaussian naïve Bayesian

The naïve Bayesian classifier (e.g. Androutsopoulos et al., 2000; Flach and Lachiche, 2004; Zhang, 2004) is based on the Bayes theorem, which describes the probability of an event using one or several attributes with the equation

$$P(A|B) = \frac{\mathrm{P(B|A)}\ P(A)}{\mathrm{P(B)}}$$

where P(A) and P(B) are the probability of respectively A and B to occur. P(A|B) is the probability of A to occur if B is true. P(B|A) is the probability of B to occur if A is true. In this study, A is the discriminant class and B is the set of variables attributed to each sample. The Gaussian naïve Bayesian (GNB) classification is based on the 'naïve' assumption of independence between input variables, and of a normal distribution of these variables for each class, which is generally not true in geochemistry.

Support vector machine

Support Vector Machine (SVM) supervised classification is based on the construction of a set of multi-dimensional hyperplanes that separate classes (Hearst et al., 1998; Bennett and Campbell, 2000). Hyperplanes are optimized by achieving the largest distance from training points. Various functions can be used to trace the hyperplanes. In this study the Gaussian radial basis function (rbf) kernel is used.

Random forest

The random forest (RF) is an ensemble method algorithm (Breiman, 2001). It consists of the combination of a series of weak learners (here decision trees) to produce a more robust prediction. Each decision tree is built from a sample of the training set (bootstrapping) and a random portion of the discriminative variables are used at each split.

Gradient tree boosting

The gradient tree boosting (GTB) algorithm is an ensemble method using a boosting procedure (Friedman, 2001). Decision trees are built in sequence with an increasing weight attributed to misclassified samples. Several parameters can be used to monitor the size of each tree and the bias versus precision trade-off. Bias represent the accuracy or the average difference between the prediction and the true value, while precision represents the reproducibility of the prediction or the standard deviation of the estimator.

Performance evaluation

Due to the relatively small number of labelled samples that can be used as training data for classification models, no independent labelled testing dataset was drawn. Instead, the success rate of each model was estimated using cross-validation, which means dividing the dataset into a training and a testing set, building a classification model based on the training set and estimating its prediction score on the testing set. Parameter tuning was completed on a wide array of parameters for each algorithm using a stratified k-fold method. This cross-validation method requires to separate the dataset in k subsets (k-folds), with the same distribution of each class in each subset (stratification). Each combination of parameters was tested k times, with training performed on k-1 subsets and testing of the prediction score on the k^{th} subset. Classification f1 scores [2*(precision*recall)/(precision+recall) with precision being true positives divided by the sum of predicted trues and recall being true positives divided by the sum of all trues] were calculated with a stratified shuffle split method. The shuffle split method randomly divides the dataset into a training and a testing set *n* times, resulting in *n* f1 scores being calculated over the *n* calculated models and their corresponding test set. It allows the calculation of an average prediction score while limiting the reduction of the number of samples in the test dataset. The standard deviation of these prediction scores is an indicator of the model variability related to the training data. Confusion matrices (tables indicating the true and predicted repartition of samples for each class, with the associated precision and recall) were calculated with a stratified k-fold method. The risk of overseeing a significant overfitting of the models due to the lack of a completely independent test set is mitigated by the use of cross-validation.

4. Results

4.1. Performance of the algorithms

Each algorithm was tested for the classification of volcanic units and alteration types. The third set of variables (Table 6) was used, since it was the most accurate (see below). For each algorithm, the classification of volcanic units is systematically more successful than that of alteration types by 9 to 16%. The success rate varies between the algorithms. The GNB yields low scores relative to the other algorithms. The KNN, SVM, RF and GTB algorithms yield success rates in a narrow range for both classification exercises, and the KNN algorithm systematically yields slightly lower scores than the SVM, RF and GTB algorithms. For the classification of volcanic units, SVM scores are significantly higher (difference higher than the standard deviation). For the classification of alteration types, SVM, RF and GTB yield similar success scores.

*Table 6: Classification success metrics for each algorithm with the average and standard deviation calculated with a shuffle split strategy (100 iterations with a random 90% of the training data used to build the model and 10% to test it). The extended set of elements variables were used as training data. The success score used here is the average f1 score = (precision * recall) / (precision + recall) of all classes weighted by the number of instances of each class. The score varies from 0 to 1, with 1 corresponding to 100% classification success.*

Classes	KNN	GNB	SVM	RF	GTB
Volcanic units	0.83 ± 0.04	0.69 ± 0.04	0.91 ± 0.03	0.85 ± 0.04	0.88 ± 0.03
Alteration types	0.69 ± 0.05	0.60 ± 0.05	0.75 ± 0.04	0.76 ± 0.05	0.76 ± 0.05

4.2. Performance of the sets of variables

All the sets of variables compiled for both labelled training sets were tested with the SVM algorithm. Both classification success score (f1 score, Table 7) and confusion matrices (Table 8 and Table 9) are used to compare performances. All sets of variables have a prediction f1 score in a close range for the classification of volcanic units (0.86-0.90) and the classification of alteration types (0.67-0.76). The range of score standard deviations varies between the classifications of volcanic units (0.03) and of alteration types (0.04-0.05). The f1 score of the set of variables composed of ratios used on the classification of volcanic units is not significantly different (i.e., the difference is lower than the standard deviation) than that of the set of elements (restricted set of elements) from which the ratios were built. Alteration indexes used for the classification of alteration types yield significantly lower scores than the set of elements from which they were built. For both classification exercises, the restricted and extended sets of elements do not show differences in f1 score higher than the standard deviation. Similarly, the use of CLR-transformed elements does not significantly increase the classification success rate.

In the case of the classification of volcanic units, the extended set of untransformed elements and the CLR-transformed set of elements yield the best results (Table 7). F1 scores are around 0.9, which is a relatively high success rate. The confusion matrix for the classification performed using the extended set of elements variables shows that misclassifications generally occur between intermediate to felsic units (Powderhouse dacite, Lalor rhyolite, Ghost Lake rhyodacite and North Balloch rhyodacite), between intermediate units (Powderhouse dacite, Footwall volcaniclastic unit and North Chisel dacite) and between mafic units (Moore mafics, Threehouse mafics and Balloch basalt). However, a significant number of misclassifications between the intermediate to felsic Powderhouse dacite and the Moore mafics occur. These two units have very distinct geochemical compositions (Figure 2) but are affected by intense alteration close to the deposit ore lenses (Figure 1; Caté, 2016). This suggests the classification of volcanic units is in part affected by alteration despite the use of chemical elements generally considered to be resistant to alteration.

In the case of the classification of alteration types, the restricted set of elements yields the best performance, followed by the extended set of elements and the CLR-transformed elements (Table 7). The best scores are above 0.75, which is lower than for the classification of volcanic units. Most of the misclassifications occur between the least altered rocks and the moderately altered rocks (Table 9). A series of samples affected by intense syn-VMS hydrothermal alteration (K, K-Mg-Fe, Mg-Fe and Mg-Ca) are misclassified as moderate alteration. Misclassifications also occur between classes of intense hydrothermal alteration with close chemical affinity (between K and K-Mg-Fe, K-Mg-Fe and Mg-Fe, and Mg-Fe and Mg-Ca). The Ca metasomatism can be falsely predicted from, or misclassified as, least to moderately altered rocks.

Table 7: Classification success (f1 score) for each variable set using the SVM algorithm with the average and standard deviation calculated with a shuffle split strategy (100 iterations with a random 90% of the training data used to build the model and 10% to test it). The f1 score is weighted by the number of instances of each class.

Classes	**1. Ratios / indexes**	**2. Elements restricted**	**3. Elements extended**	**4. CLR**
Volcanic units	0.86 $\pm$ 0.03	0.88 $\pm$ 0.03	0.90 $\pm$ 0.03	0.90 $\pm$ 0.03
Alteration types	0.67 $\pm$ 0.05	0.76 $\pm$ 0.04	0.75 $\pm$ 0.05	0.75 $\pm$ 0.05

Table 8: Confusion matrix of the classification of volcanic units using the extended set of elements and an SVM algorithm. Columns present instances of predicted classes and rows present instances of true classes.

		PREDICTED									
	Extended elements variables	Foot. volcani. unit	Moore mafics	Lalor rhyolite	Powd. dacite	North Balloch rhyod.	Balloch basalt	Ghost Lake rhyod.	Three. mafics	North Chisel dacite	**Recall**
TRUE	Footwall volcaniclastic formation	51	2		1			1	3	2	**0.85**
	Moore mafics	1	309		12						**0.96**
	Lalor rhyolite			45	8						**0.85**
	Powderhouse dacite	2	11	4	118						**0.87**
	North Balloch rhyodacite	2		1	3	33		4			**0.77**
	Balloch basalt	1	2				48		6	1	**0.83**
	Ghost Lake rhyodacite	1			1	3		45			**0.9**
	Threehouse mafics	1	1				8		72	2	**0.86**
	North Chisel dacite	6			1				2	23	**0.72**
	Precision	**0.78**	**0.95**	**0.9**	**0.82**	**0.92**	**0.86**	**0.9**	**0.87**	**0.82**	

Table 9: Confusion matrix of the classification of alteration types using the restricted set of elements and an SVM algorithm. Columns present instances of predicted classes and rows present instances of true classes.

		PREDICTED							
	elements restricted	Least alt. rocks	Moderate alteration	K	K-Mg-Fe	Mg-Fe	Mg-Ca	Ca	**Recall**
TRUE	Least altered rocks	64	48	.	1	.	.	5	**0.54**
	Moderate alteration	54	90	.	7	6	.	1	**0.57**
	K	.	3	31	3	.	.	.	**0.84**
	K-Mg-Fe	.	5	4	57	3	.	.	**0.83**
	Mg-Fe	.	7	.	3	185	4	.	**0.93**
	Mg-Ca	.	.	.	.	6	41	.	**0.87**
	Ca	8	7	.	.	1	1	35	**0.67**
	Precision	**0.51**	**0.56**	**0.89**	**0.8**	**0.92**	**0.89**	**0.85**	

4.3. Classification of unlabelled data

The work done on labelled geochemical analyses shows that machine learning can reliably classify data, both for protoliths and alteration types. The algorithms can therefore presumably be applied to unlabelled samples, i.e. the Hudbay analyses for which the classification is not already known. All unlabelled samples were classified using SVM algorithms trained on the extended sets of elements for the classification of volcanic units and alteration types. Results have been plotted on a series of geochemical diagrams and in space, to estimate the classification success and interpret the geological significance of the results.

Geochemical diagrams with prediction results on all samples (Figure 4A, B and C) show distinct distributions of volcanic units with significant overlaps. The Moore mafics and the Powderhouse dacite have a calc-alkaline affinity, the Threehouse mafics have a tholeiitic affinity, and the other units have a dominantly transitional affinity (Figure 4A). The Threehouse and Moore mafics and the Balloch basalt plot as mafic rocks in Figure 4B. Intermediate to felsic units mainly plot in the intermediate field, with only the Lalor rhyolite being dominantly distributed in the felsic field. The distribution of each unit in the Nb/Y-Zr/Ti diagram (Figure 4B) is similar, but more widespread than that of training samples (Figure 2B). The Balloch basalt significantly overlaps with the Threehouse and the Moore mafics. Samples attributed to the North Chisel dacite and the Footwall volcaniclastic unit are distributed in the same area. Felsic units (Lalor rhyolite, Powderhouse dacite, North Balloch rhyodacite and Ghost Lake rhyodacite) plot in roughly distinct fields, and the North Balloch rhyodacite shows the same bimodal distribution observed in the training set of samples.

The distribution of predicted volcanic units in space (Figure 5) closely resembles the geological cross section (Figure 1). In the lowermost part of the model, volcanic units (Footwall volcaniclastic unit, Moore mafics, Lalor rhyolite and Powderhouse dacite) are structurally and stratigraphically imbricated, similarly to the complex distribution shown in Figure 1. All volcanic units of the Balloch volcanic succession are well delimited with few “out of place” samples, except for the Ghost Lake rhyodacite. A significant number of samples located within the Ghost Lake rhyodacite are labelled as Powderhouse dacite or North Balloch rhyodacite, which suggests mislabelling. A number of samples labelled as Threehouse mafics within the Ghost Lake rhyodacite, the Balloch basalt and the North Chisel rhyodacite correspond to the intrusive units of the Threehouse mafics (North Balloch mafic intrusive and Threehouse diorite).

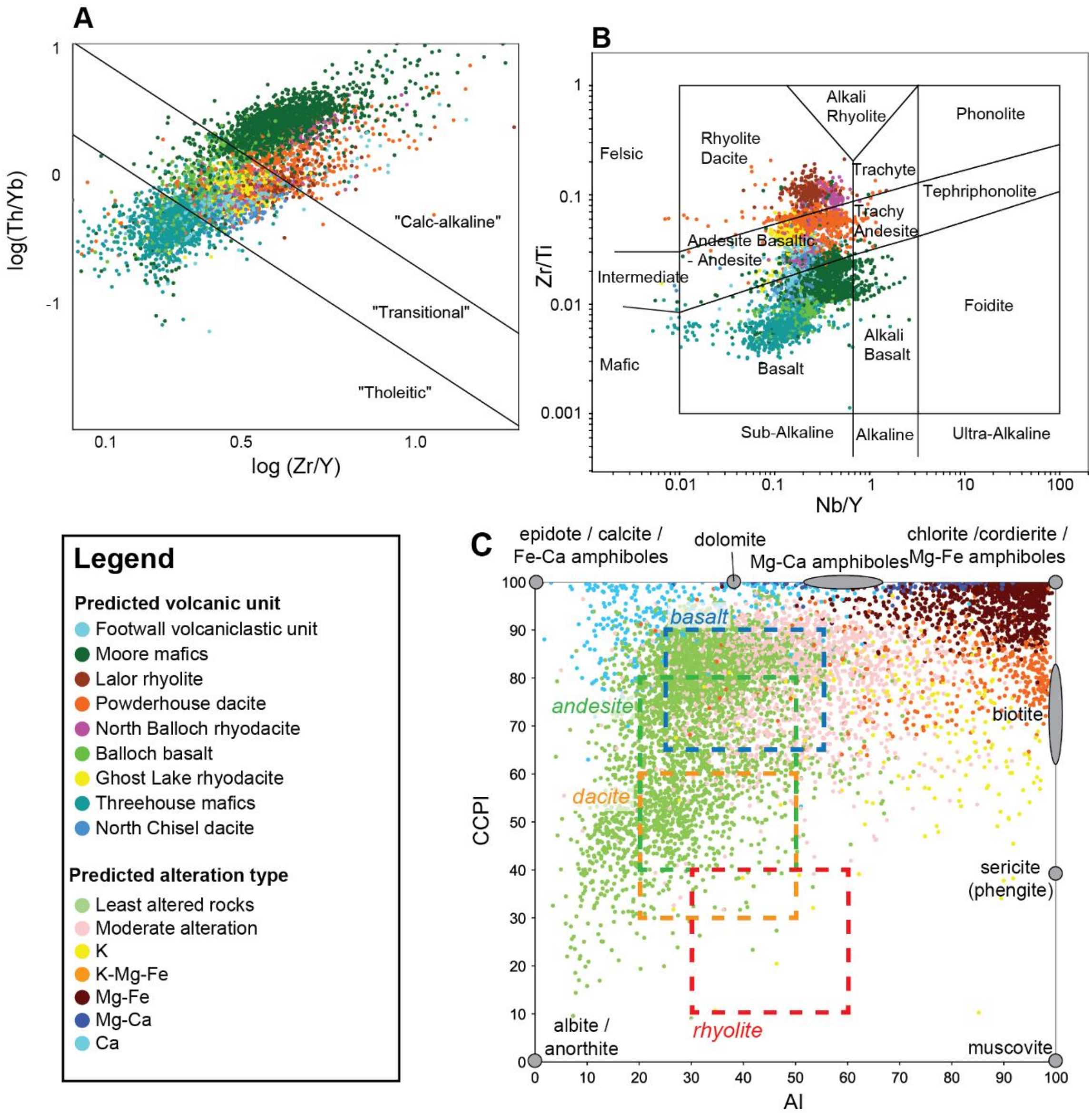

Figure 4: Geochemical diagrams showing the results of the classification of unlabelled samples. A: Zr/Y-Th/Yb diagram from Ross and Bédard (2009) indicating the magmatic affinity of volcanic units; B: Nb/Y-Zr/Ti diagram from Pearce (1996), modified after Winchester and Floyd (1977) showing the differentiation and alkalinity of volcanic units; C: Box plot diagram from Large et al. (2001) showing the geochemical signature of the alteration types. Main minerals associated with alteration assemblages are indicated. Fields of unaltered volcanic rocks are from Gifkins et al. (2005). AI = 100(MgO+K$_2$O)/(MgO+K$_2$O+Na$_2$O+CaO) and CCPI = 100*(Mg+FeO)/(MgO+FeO+Na$_2$O+K$_2$O)*

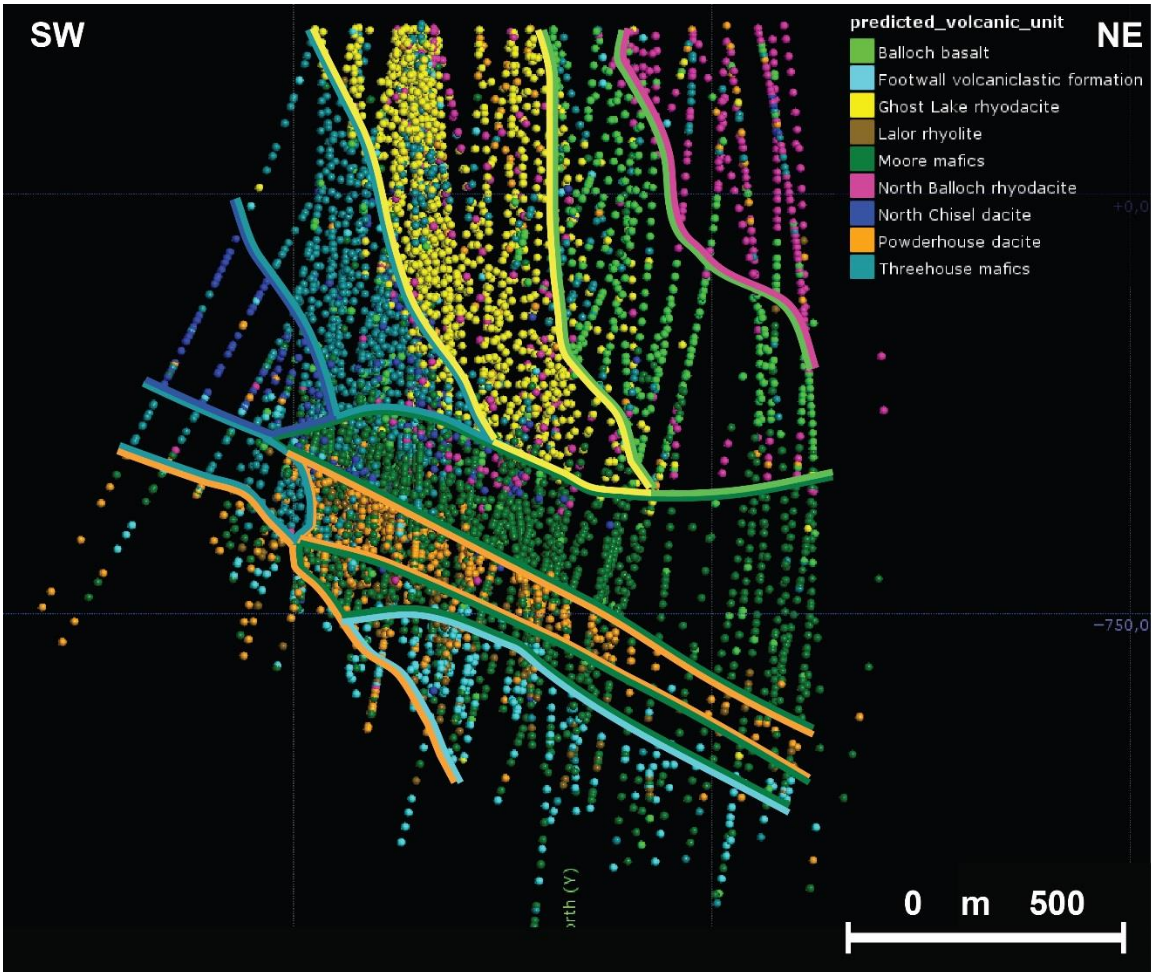

Figure 5: View on the spatial distribution of samples coloured by predicted volcanic unit generated with the Leapfrog Geo software. Approximate location of lithological contacts are presented as coloured lines. A total of 234 drillholes and the 7335 samples are plotted on this approximatively 1.5 km-thick section.

The distribution of the predicted alteration types in a box-plot diagram (Figure 4C) illustrates the very distinct geochemical signature of alteration types with minor to moderate overlap. Predicted least-altered samples are mainly distributed within or close to the fields of least-altered rocks. Predicted moderately-altered samples have a distribution spanning from the least-altered samples to the intensely-altered samples with high CCPI and AI values, illustrating the transition between weak and intense alteration. Predicted Mg-Ca altered samples have high CCPI (>95) and high to moderate AI (>50) values illustrating the presence of chlorite, carbonates and Ca amphiboles. Predicted Mg-

Fe, K-Mg-Fe and K-altered samples have mostly high AI values (>80) with variable CCPI values reflecting the different mineralogical assemblages (Table 4). Mg-Fe altered samples are enriched in chlorite, cordierite and Mg-Fe amphiboles, K-Mg-Fe altered samples are enriched in biotite, and K-altered samples have significant concentrations of muscovite. Samples predicted as Ca-altered have high CCPI values (>80) with moderate to low AI values (<60). Ca-altered samples have a distribution distinct to that of samples affected by other alteration types.

Most samples located in the hanging wall and to the SW of the deposit are predicted as least-altered, with a minority of moderately-altered samples (Figure 6). Intensely-altered samples (Mg-Ca, Mg-Fe, K-Mg-Fe and K alteration types) are located at depth, and to the NE, which corresponds to the location of the ore lenses and their footwall. Moderately-altered samples form a diffuse halo around intense alteration zones. K alteration is more present at the top of the alteration zone, with Mg-Ca alteration located beneath it, and K-Mg-Fe alteration forming the transition toward Mg-Fe alteration zone located to the NE. This geometry corresponds to that described in Caté et al. (2015). Predicted Ca-altered samples are located at the southwestern contact between altered and least-altered zones.

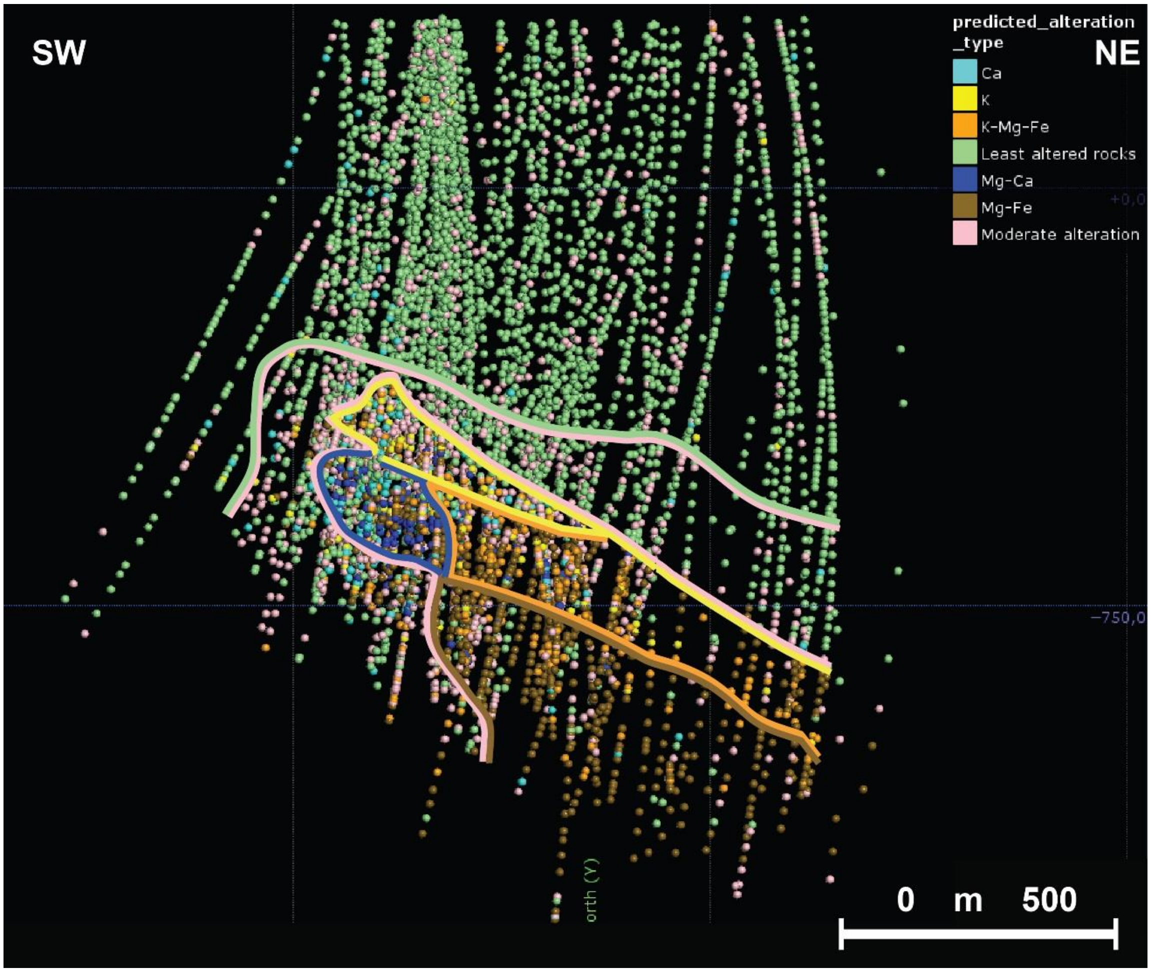

Figure 6: View on the spatial distribution of samples coloured by predicted alteration type generated on the Leapfrog Geo software. Approximate location of alteration zones are presented as coloured lines. A total of 234 drillholes and the 7335 samples are plotted on this approximatively 1.5 km-thick section.

5. Discussion

5.1. Classification results for each label

The geochemical dataset was classified by two thematically-distinct training sets, one for volcanic units (the "protolith") and one for the alteration assemblages. Classes defined for the training set of the volcanic units are based on geochemical signature, preserved volcanic textures and spatial distribution. Classes defined for the training set of the alteration units are based on visual differentiation of distinct mineralogical assemblages.

The initial discrimination of volcanic units is partially based on the geochemical signature, and the classification pattern is well retrieved with a F1 score close to 0.9. This score is likely to be higher to what would have been obtained from a diagram(s)-based classification such as those presented in Figure 2. The use of a large spectrum of elements with a significant classification power instead of a restricted set of the best elements or ratios slightly increases the classification success. Contrary to a machine learning-based multivariate classification, the use of such a large number of elements would not be practical in “manual” classification, especially on a large number of samples, such as the 7335 samples from this study. The confusion matrix for the classification of volcanic units (Table 8) and the related 3D view (Figure 5) demonstrate that accurate classification of spatially-coherent volcanic units was obtained, and that results are consistent with previously published geological cross-sections.

The Powderhouse dacite, Lalor rhyolite and Moore mafics are sometimes misclassified or inverted in the confusion matrix. These three units have a distinct signature in geochemical diagrams (Figure 2), which should lead to only very few misclassifications. However, these units are hosting or are located immediately below massive sulphide ore lenses. They are thus affected by the most intense hydrothermal alteration. Such alteration produces important mass changes and potential modifications to the relative concentrations of “immobile” elements, leading to misclassifications. The introduction of a significant number of altered samples in the training set could help the model better predict volcanic units in altered lithologies. The use of variables unaffected by relative mass changes due to alteration (e.g., Pearce element ratios, Stanley and Madeisky, 1994; or other immobile element ratios Barrett and MacLean, 1994) can also limit the influence of alteration on the classification. However, these misclassifications represent a very low percentage of the total of samples from these units, and do not significantly affect the overall classification scores. A simple spatial analysis can help quickly identify such miss-classified samples.

In Figure 5, a minority of samples are classified as part of the North Chisel rhyodacite or the Powderhouse dacite in the volume dominantly occupied by samples from the Ghost Lake rhyodacite. They can reasonably be considered as misclassified samples due to their location. The addition of location information as a predictor variable would potentially increase classification success rates in relatively simple geologic environments, but it could bias classification results and prevent previously unrecognized occurrences of volcanic units in more complex geologic environments.

The initial discrimination of alteration types is based on visual estimation of the mineralogy. The mineralogical composition of rocks is directly related to their geochemical composition (e.g., Verma et al., 2003; Piché and Jébrak, 2004), which suggests a multivariate classification model based on lithogeochemistry should perform well on mineralogy-derived alteration types. Classification success scores close to 0.75 validate this hypothesis, but these scores are significantly lower than that obtained for the classification of volcanic rocks. Misclassification occurs between compositionally adjacent classes, especially between least-altered and moderately-altered rocks. This can result from errors in the labelling of training data, related to the fact that mineral concentrations in rock samples are mostly estimated visually from macroscopic observations. Also, the geochemical composition of both least-altered and moderately-altered rocks is strongly dependant on the composition of the volcanic protolith. Both alteration types are heterogeneous and have gradational transitions, which leads to important overlaps of the geochemical compositions of both classes (e.g., Figure 3 and Figure 4C). Finally, the heterogeneous nature of the alteration, even locally, might induce further variability in the geochemical composition of samples of each class, even though samples were carefully chosen to be representative.

5.2. Choice of the algorithm

Overall, the SVM algorithm is the best performer for the classification of rock types from geochemical data, closely followed by ensemble methods (RF and GTB). The relative difference in success rate between algorithms changes from the classification of volcanic units to that of alteration types, which suggests that the best-performing algorithm might change for other classification exercises. The relative performance of algorithms might change with larger training datasets.

5.3. Choice of variables

Element ratios and alteration indices are used to facilitate the interpretation of geochemical data using diagrams. This transformation is necessary for “manual” classification as the human brain cannot process simultaneously more than two to three variables (with each variable representing one element or a combination of elements). However, by combining different elements and reducing the number of variables, the classification power of the data decreases. It is illustrated by the better performance of untransformed elements compared to element ratios and alteration indexes used in diagrams. As a general rule, the inclusion of more elements tends to increase the classification power of predictive models. Thus, the use of multivariate classification is likely to outperform diagram-based classification given a large enough training dataset. On the other hand, as shown by the similar success rates of predictive models using the restricted and extended variable sets, most of the classification power of chemical elements is concentrated within a restricted set of elements. The addition of more elements to the predictive variables does not significantly increase the classification success rate. Using previous work on geochemical classification of rock units or alteration styles (e.g., Irvine and Baragar, 1971; Pearce and Norry, 1979; Barrett and MacLean, 1994; Verma and Agrawal, 2011), the best discriminating elements can be included in the set of predictive variables depending on the classification exercise. Further variable selection can be performed by calculating the contribution of each variable in predictive models (e.g., feature importance in RF models).

Opening the compositional geochemical data using a CLR-transformation does not show a significant difference in classification success rates. Thus, untransformed elements seem the best suited for classification, as further interpretation of the results is more intuitive.

For the classification of volcanic units, the relative concentration of least mobile elements is still affected by alteration, even though it is less significant than for mobile elements (e.g., Barrett and MacLean, 1994). This could have an effect on the classification success rates for the most intensely-altered rocks (e.g., Moore mafics and Powderhouse dacite at Lalor). Dividing all elements by an immobile element (e.g., TiO_2 or Zr) would provide variables completely independent of the effect of alteration (Barrett and MacLean, 1994), and increase classification success rates in the most altered rocks.

Alteration is based on enrichment/depletion of elements in rocks. Using Pearce element ratios instead of raw elements would provide variables more sensitive to relative concentration changes between elements resulting from the alteration. This could increase the classification power of predictive models.

5.4. Success rate

The f1 score for the classification of volcanic units is close to 0.9 (Table 7), and both precision and recall scores are above 0.7 for all volcanic units (Table 8). These scores can be considered as high enough for relying on the predictive model of the lithology for 3D geological modelling. The low misclassification rate is unlikely to have a significant effect on further use of the classification results for 2D or 3D modelling (see Figure 5).

The f1 score for the classification of alteration types is close to 0.75. This indicates scores high enough for classification results to be reliable for geological modelling, but around 25% of the samples are likely to be misclassified. Thus, care should be taken in the interpretation of the results, and during 2D or 3D modelling of the alteration zones. Because of the misclassification of adjacent alteration types, and the progressive nature of hydrothermal alteration, boundaries between alteration zones should be seen as “progressive” or “soft” boundaries compare to the “sharp” or “discrete” boundaries between volcanic units.

For both classification exercises, the significant standard deviations of the f1 scores obtained by cross-validation (Table 6 and Table 7) indicate that the small size of the training set introduces a significant bias in the classification models. These relatively high standard deviations are likely to decrease with an increasing training dataset size. Thus, a larger geochemical dataset would produce more stable prediction models and might increase success scores.

6. Conclusions

A series of supervised predictive models have been tested on rocks of the Lalor deposit by varying the target variable (i.e., volcanic units and alteration types), the predictive variables (Table 5) and the machine learning algorithms. The results have a series of implications for the use of multivariate supervised classification methods on lithogeochemical datasets.

1. Using controlled training sets, classification models of lithologies and hydrothermal alteration using lithogeochemical data can be obtained with machine learning. High success rates can be attained, and the performances are probably higher than those achieved by manual classification based solely on lithogeochemistry.
2. The classification success is strongly dependant on training data quality and quantity. The training data must be representative of the local geology and include enough occurrences of each class (i.e., volcanic units and alteration types).
3. Several machine learning algorithms are suitable for supervised multivariate lithogeochemical classification. The best performing algorithm changes from a case to another, and a careful selection based on success scores should be completed.
4. No complex feature engineering (transformation of the data) is necessary to obtain high predictive power from chemical element concentrations. A selection of elements adapted to the labels and based on knowledge of geochemical processes can be done to reduce the number of variables.

Acknowledgements

This study is part of the Geological Survey of Canada Targeted Geoscience Initiative and its Lalor research activity, conducted in collaboration with the Institut national de la recherche scientifique, centre Eau Terre et Environnement (INRS-ETE). The Lalor research was also conducted in very close collaboration with Hudbay who granted access to the sites and to data, and provided financial support. The Manitoba Geological Survey provided in kind support and expertise about the district. The work presented in this contribution results was made possible by the Mountjoy Exchange Award of the Geological Association of Canada to Antoine Caté. We are particularly indebted to Hudbay geologists, to Dr. Alan Bailes and to the Laurentian University Snow Lake research group for the insightful discussions on the geology of the Snow Lake area and Lalor deposit. We thank both anonymous reviewers for their helpful comments.

References

References:

Abbaszadeh, M., Hezarkhani, A., and Soltani-Mohammadi, S., 2015, Classification of alteration zones based on whole-rock geochemical data using support vector machine: Journal Geological Society of India, v. 85, p. 500-508.

Aitchison, J., 1982, The statistical analysis of compositional data: Journal of the Royal Statistical Society. Series B (Methodological), v. 44, p. 139-177.

Androutsopoulos, I., Koutsias, J., Chandrinos, K. V., Paliouras, G., and Spyropoulos, C. D., 2000, An evaluation of naive bayesian anti-spam filtering: 11th European Conference on Machine Learning, Barcelona, Spain, 2000, p. 9-17.

Bailes, A. H., Rubingh, K., Gagné, S., Taylor, C., Galley, A., Bernauer, S., and Simms, D., 2013, Volcanological and structural setting of Paleoproterozoic VMS and gold deposits at Snow Lake, Manitoba, Geological Association of Canada–Mineralogical Association of Canada, Joint Annual Meeting, Winnipeg, Manitoba, May 22–24, 2013, Field Trip Guidebook, Manitoba Innovation, Energy and Mines, Manitoba Geological Survey, 63 p.

Barrett, T. J., and MacLean, W. H., 1994, Chemostratigraphy and hydrothermal alteration in exploration for VHMS deposits in greenstones and younger volcanic rocks, *in* Lentz, D. R., ed., Alteration and alteration processes associated with ore-forming systems. Short Course Notes, Geological Association of Canada, p. 433-466.

Bellefleur, G., Schetselaar, E., White, D., Miah, K., and Dueck, P., 2015, 3D seismic imaging of the Lalor volcanogenic massive sulphide deposit, Manitoba, Canada: Geophysical Prospecting, v. 63, p. 813-832.

Bennett, K. P., and Campbell, C., 2000, Support vector machines: hype or hallelujah?: ACM SIGKDD Explorations Newsletter, v. 2, p. 1-13.

Breiman, L., 2001, Random forests: Machine learning, v. 45, p. 5-32.

Carranza, E. J. M., and Laborte, A. G., 2015, Random forest predictive modeling of mineral prospectivity with small number of prospects and data with missing values in Abra (Philippines): Computers & Geosciences, v. 74, p. 60-70.

Caté, A., 2016, Geology of the Paleoproterozoic Zn-Cu-Au Lalor volcanogenic massive sulphide deposit and its gold-rich lenses, Snow Lake, Manitoba, Institut national de la recherche scientifique, 430 p.

Caté, A., Mercier-Langevin, P., Ross, P. S., Bécu, V., Lauzière, K., Hannington, M., and Dubé, B., 2017, Whole-rock lithogeochemistry of the Lalor auriferous VMS deposit, Manitoba, Canada, Open File 8107, Geological Survey of Canada.

Caté, A., Mercier-Langevin, P., Ross, P. S., Duff, S., Hannington, M., Dubé, B., and Gagné, S., 2015, Geology and Au enrichment processes at the Paleoproterozoic Lalor auriferous volcanogenic massive sulphide deposit, Snow Lake, Manitoba, *in* Peter, J., and Mercier-Langevin, P., eds., Targeted Geoscience Initiative 4: Contributions to the understanding of volcanogenic massive sulphide deposit genesis and exploration methods development. Open File, Geological Survey of Canada, p. 131-146.

Caté, A., Mercier-Langevin, P., Ross, P. S., Duff, S., Hannington, M., Gagné, S., and Dubé, B., 2014a, Insight on the chemostratigraphy of the volcanic and intrusive rocks of the Lalor auriferous VMS deposit host succession, Snow Lake, Manitoba, Current Research 2014-6, Geological Survey of Canada, 19 p.

Caté, A., Mercier-Langevin, P., Ross, P. S., and Simms, D., 2014b, Structural controls on geometry and ore distribution in the Lalor volcanogenic massive-sulphide deposit, Snow Lake, west-central Manitoba (part of NTS 63K16): preliminary results from underground mapping, Report of Activities 2014, Manitoba Mineral Resources, Manitoba Geological Survey, p. 104-115.

Cracknell, M. J., Reading, A. M., and McNeill, A. W., 2014, Mapping geology and volcanic-hosted massive sulfide alteration in the Hellyer–Mt Charter region, Tasmania, using Random Forests™ and Self-Organising Maps: Australian Journal of Earth Sciences, v. 61, p. 287-304.

De La Roche, H., Leterrier, J., Grandclaude, P., and Marchal, M., 1980, A classification of volcanic and plutonic rocks using R 1 R 2-diagram and major-element analyses—its relationships with current nomenclature: Chemical geology, v. 29, p. 183-210.

Domingos, P., 2012, A few useful things to know about machine learning: Communications of the ACM, v. 55, p. 78-87.

Duff, S., 2016, Ore types of the auriferous Lalor VMS Deposit, Snow Lake, Manitoba: implications for genesis and post-depositional processes, University of Ottawa, 133 p.

Duff, S., Hannington, M., Caté, A., Mercier-Langevin, P., and Kjarsgaard, I. M., 2015, Major ore types of the Paleoproterozoic Lalor auriferous volcanogenic massive sulphide deposit, Snow Lake, Manitoba, *in* Peter, J., and Mercier-Langevin, P., eds., Targeted Geoscience Initiative 4: Contributions to the understanding of volcanogenic massive sulphide deposit genesis and exploration methods development. Open File, Geological Survey of Canada, p. 147-170.

Flach, P. A., and Lachiche, N., 2004, Naive Bayesian classification of structured data: Machine Learning, v. 57, p. 233-269.

Franklin, J. M., Gibson, H. L., Jonasson, I. R., and Galley, A. G., 2005, Volcanogenic Massive Sulphide Deposits, *in* Hedenquist, J. W., Thompson, J. F. H., Goldfarb, R. J., and Richards, J. P., eds., Economic Geology 100th Anniversary Volume, The Economic Geology Publishing Company, p. 523-560.

Fresia, B., Ross, P.-S., Gloaguen, E., and Bourke, A., 2017, Lithological discrimination based on statistical analysis of multi-sensor drill core logging data in the Matagami VMS district, Quebec, Canada: Ore Geology Reviews, v. 80, p. 552-563.

Friedman, J., Hastie, T., and Tibshirani, R., 2001, The elements of statistical learning: Berlin, Springer, 533 p.

Friedman, J. H., 2001, Greedy function approximation: a gradient boosting machine: Annals of statistics, p. 1189-1232.

Froese, E., and Gasparrini, E., 1975, Metamorphic zones in the Snow Lake area, Manitoba: Canadian Mineralogist, v. 13, p. 162-167.

Galley, A. G., 1993, Characteristics of semi-conformable alteration zones associated with volcanogenic massive sulphide districts: Journal of Geochemical Exploration, v. 48, p. 175-200.

Galley, A. G., Bailes, A. H., and Kitzler, G., 1993, Geological setting and hydrothermal evolution of the Chisel Lake and North Chisel Zn-Pb-Cu-Ag-Au massive sulfide deposits, Snow Lake, Manitoba: Exploration and Mining Geology, v. 2, p. 271-295.

Galley, A. G., Hannington, M., and Jonasson, I. R., 2007a, Volcanogenic Massive Sulphide Deposits, *in* Goodfellow, W. D., ed., Mineral Deposits of Canada: A Synthesis of Major Deposit-Types, Disctrict Metallogeny, the Evolution of Geological Provinces, and Exploration Methods. Special Publication, Geological Association of Canada, Mineral Deposit Division, p. 141-161.

Galley, A. G., Syme, R., and Bailes, A. H., 2007b, Metallogeny of the Paleoproterozoic Flin Flon belt, Manitoba and Saskatchewan, *in* Goodfellow, W. D., ed., Mineral deposits of Canada: a synthesis of major deposit types, district metallogeny, the evolution of geological provinces, and exploration methods. Special Publication, Geological Association of Canada, Mineral Deposits Division, p. 509-531.

Gemmell, J. B., 2006, Geology, geochemistry and exploration implications of hydrothermal alteration associated with epithermal Au-Ag deposits, AMIRA P588 Research Team.

Gibson, H. L., Morton, R. L., and Hudak, G. J., 1999, Submarine volcanic processes, deposits, and environments favorable for the location of volcanic-associated massive sulfide deposits, *in* Barrie, C. T., and Hannington, M. D., eds., Volcanic associated massive sulfide deposits: processes and examples in modern and ancient settings. Reviews in Economic Geology, p. 13-48.

Gifkins, C., Herrmann, W., and Large, R. R., 2005, Altered volcanic rocks: A guide to description and interpretation: Hobart, Tasmania, Australia, Centre for Ore Deposit Research, University of Tasmania, 275 p.

Haaland, D. M., Jones, H. D., and Thomas, E. V., 1997, Multivariate classification of the infrared spectra of cell and tissue samples: Applied spectroscopy, v. 51, p. 340-345.

Hastie, A. R., Kerr, A. C., Pearce, J. A., and Mitchell, S., 2007, Classification of altered volcanic island arc rocks using immobile trace elements: development of the Th–Co discrimination diagram: Journal of Petrology, v. 48, p. 2341-2357.

Hearst, M. A., Dumais, S. T., Osman, E., Platt, J., and Scholkopf, B., 1998, Support vector machines: Intelligent Systems and their Applications, IEEE, v. 13, p. 18-28.

Irvine, T. N., and Baragar, W. R., 1971, A guide to the chemical classification of the common volcanic rocks: Canadian journal of earth sciences, v. 8, p. 523-548.

Ishikawa, Y., Sawaguchi, T., Iwaya, S., and Horiuchi, M., 1976, Delineation of prospecting targets for Kuroko deposits based on modes of volcanism of underlying dacite and alteration haloes: Mining Geology, v. 26, p. 105-117.

Kirkwood, C., Cave, M., Beamish, D., Grebby, S., and Ferreira, A., 2016, A machine learning approach to geochemical mapping: Journal of Geochemical Exploration, v. 167, p. 49-61.

Kishida, A., and Kerrich, R., 1987, Hydrothermal alteration zoning and gold concentration at the Kerr-Addison Archean lode gold deposit, Kirkland Lake, Ontario: Economic Geology, v. 82, p. 649-690.

Kraus, J., and Williams, P. F., 1999, Structural development of the Snow Lake Allochthon and its role in the evolution of the southeastern Trans-Hudson Orogen in Manitoba, central Canada: Canadian Journal of Earth Sciences, v. 36, p. 1881-1899.

Kuno, H., 1968, Differentiation of basalt magmas, *in* Hess, H., and Poldervaart, A., eds., Basalts: New york, Interscience, John Wiley and Sons, p. 623-688.

Lam, J., Tinkham, D. K., and Gibson, H., 2014, Characterization of gold occurrences with respect to metamorphism at the Lalor deposit, Snow Lake, MB: Geological Association of Canada-Mineralogical Association of Canada Annual Meeting, Fredericton, May 20-22, 2014, p. 150-151.

Large, R. R., Gemmell, J. B., Paulick, H., and Huston, D. L., 2001, The Alteration Box Plot: A Simple Approach to Understanding the Relationship between Alteration Mineralogy and Lithogeochemistry Associated with Volcanic-Hosted Massive Sulfide Deposits: Economic Geology, v. 96, p. 957-971.

Le Maître, R. W., 1989, A classification of igneous rocks and glossary of terms. Recommendations of the IUGS Subcommission on the Systematics of Igneous Rocks., Blackwell Publications, 193 p.

Lucas, S. B., Stern, R. A., Syme, E. C., Reilly, B. A., and Thomas, D. J., 1996, Intraoceanic tectonics and the development of continental crust: 1.92–1.84 Ga evolution of the Flin Flon Belt, Canada: Geological Society of America Bulletin, v. 108, p. 602-629.

MacLean, W. H., and Kranidiotis, P., 1987, Immobile elements as monitors of mass transfer in hydrothermal alteration; Phelps Dodge massive sulfide deposit, Matagami, Quebec: Economic Geology, v. 82, p. 951-962.

Menard, T., and Gordon, T. M., 1997, Metamorphic P-T paths from the eastern Flin Flon belt and Kisseynew domain, Snow Lake, Manitoba: Canadian Mineralogist, v. 35, p. 1093-1115.

Mercier-Langevin, P., Caté, A., and Ross, P. S., 2014, Whole-rock oxygen isotope mapping, Lalor auriferous VMS deposit footwall alteration zones, Snow Lake, west-central Manitoba (NTS 63K16), Report of activities 2014, Manitoba Mineral Resources, Manitoba Geological Survey, p. 94-103.

O'Brien, J. J., Spry, P. G., Nettleton, D., Xu, R., and Teale, G. S., 2015, Using Random Forests to distinguish gahnite compositions as an exploration guide to Broken Hill-type Pb–Zn–Ag deposits in the Broken Hill domain, Australia: Journal of Geochemical Exploration, v. 149, p. 74-86.

Pawlowsky-Glahn, V., and Egozcue, J. J., 2006, Compositional data and their analysis: An introduction, Geological Society Special Publication, 264, p. 1-10.

Pearce, J. A., 1996, A user's guide to basalt discrimination diagrams, *in* Wyman, D. A., ed., Trace element geochemistry of volcanic rocks: applications for massive sulphide exploration. Short Course Notes, Geological Association of Canada, p. 79-113.

Pearce, J. A., and Norry, M. J., 1979, Petrogenetic implications of Ti, Zr, Y, and Nb variations in volcanic rocks: Contributions to Mineralogy and Petrology, v. 69, p. 33-47.

Pearce, J. A., W., H. N. B., and Tindle, A. G., 1984, Trace elements discrimination diagrams for the tectonic interpretation of granitic rocks: Journal of Petrology, v. 25, p. 956-983.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, É., 2011, Scikit-learn: Machine Learning in Python: JMLR, v. 12, p. 2825-2830.

Peterson, L. E., 2009, K-nearest neighbor, Scholarpedia, 4. http://scholarpedia.org, p. 1883.

Piché, M., and Jébrak, M., 2004, Normative minerals and alteration indices developed for mineral exploration: Journal of Geochemical Exploration, v. 82, p. 59-77.

Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., and Chica-Rivas, M., 2015, Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines: Ore Geology Reviews, v. 71, p. 804-818.

Ross, P.-S., Bourke, A., and Fresia, B., 2014, Improving lithological discrimination in exploration drill-cores using portable X-ray fluorescence measurements: (1) testing three Olympus Innov-X analysers on unprepared cores: Geochemistry: Exploration, Environment, Analysis, v. 14, p. 171-185.

Ross, P.-S., Bourke, A., Mercier-Langevin, P., Lépine, S., Leclerc, F., and Boulerice, A., 2016, High-Resolution Physical Properties, Geochemistry, and Alteration Mineralogy for the Host Rocks of the Archean Lemoine Auriferous Volcanogenic Massive Sulfide Deposit, Canada: Economic Geology, v. 111, p. 1561-1574.

Ross, P. S., and Bédard, J. H., 2009, Magmatic affinity of modern and ancient subalkaline volcanic rocks determined from trace-element discriminant diagrams: Canadian Journal of Earth Sciences, v. 46, p. 823-839.

Sadeghi, B., and Carranza, E. J. M., 2015, Improving geological logs of drill-cores by correlating with fractal models of drill-hole geochemical data: 17th annual conference of the International Association for Mathematical Geosciences, Freiberg (Germany), 2015, p. 10.

Schetselaar, E., Bellefleur, G., Craven, J., Roots, E., Cheraghi, S., Shamsipour, P., Caté, A., Mercier-Langevin, P., El Goumi, N., Enkin, R., and Salisbury, M., 2017, Geologically driven 3D modelling of physical rock properties in support of interpreting the seismic response of the Lalor volcanogenic massive sulphide deposit, Snow Lake, Manitoba, Canada, *in* Gessner, K., Blenkinsop, T. G., and Sorjonen-Ward, P., eds., Characterization of ore-forming systems from geological, geochemical and geophysical studies. Special Publications: London, Geological Society.

Schetselaar, E., and Shamsipour, P., 2015, Interpretation of borehole gravity data of the Lalor volcanogenic massive sulfide deposit, Snow Lake, Manitoba, Canada: Interpretation, v. 3, p. 145-154.

Schetselaar, E. M., Chung, C.-J. F., and Kim, K. E., 2000, Integration of Landsat TM, gamma-ray, magnetic, and field data to discriminate lithological units in vegetated granite-gneiss terrain: Remote Sensing of Environment, v. 71, p. 89-105.

Stanley, C. R., and Madeisky, H., 1994, Lithogeochemical exploration for hydrothermal ore deposits using Pearce element ratio analysis, *in* Lentz, D., ed., Alteration and alteration processes associated with ore forming systems. Short Course Notes, Geological Association of Canada, p. 193-211.

Tinkham, D. K., 2013, A model for metamorphic devolatilization in the Lalor deposit alteration system, Snow Lake, MB: Geological Association of Canada-Mineralogical Association of Canada Annual Meeting, Winnipeg, May 22-24, 2013, p. 187.

Verma, S. P., and Agrawal, S., 2011, New tectonic discrimination diagrams for basic and ultrabasic volcanic rocks through log-transformed ratios of high field strength elements and implications for petrogenetic processes: Revista Mexicana De Ciencias Geologicas, v. 28, p. 24-44.

Verma, S. P., Torres-Alvarado, I. S., and Velasco-Tapia, F., 2003, A revised CIPW norm: Swiss Bulletin of Mineralogy and Petrology, v. 83, p. 197-216.

Winchester, J. A., and Floyd, P. A., 1977, Geochemical discrimination of different magma series and their differentiation products using immobile elements: Chemical Geology, v. 20, p. 325-343.

Wood, D. A., 1980, The application of a Th Hf Ta diagram to problems of tectonomagmatic classification and to establishing the nature of crustal contamination of basaltic lavas of the British Tertiary Volcanic Province: Earth and planetary science letters, v. 50, p. 11-30.

Zaleski, E., Froese, E., and Gordon, T. M., 1991, Metamorphic petrology of Fe-Zn-Mg-Al alteration at the Linda volcanogenic massive sulfide deposit, Snow Lake, Manitoba: Canadian Mineralogist, v. 29, p. 995-1017.

Zhang, H., 2004, The optimality of naive Bayes: FLAIRS Conference, 2004.