Distinguishing Ore Deposit Type and Barren Sedimentary Pyrite Using Laser Ablation-Inductively Coupled Plasma-Mass Spectrometry Trace Element Data and Statistical Analysis of Large Data Sets

Daniel D. Gregory,1,2 Mathew J. Cracknell,3 Ross R. Large,3 Peter McGoldrick,3 Stephen Kuhn,3 Valery V. Maslennikov,4 Michael J. Baker,3 Nathan Fox,3 Ivan Belousov,3 Maria C. Figueroa,3 Jeffrey A. Steadman,3 Adrian J. Fabris,3 and Timothy W. Lyons2

1Department of Earth Sciences, Earth Sciences Centre, 22 Russell Street, Toronto, Ontario M5S 3B1, Canada
2Department of Earth Sciences, University of California, Riverside, California 92521, USA
3Australian Research Council (ARC) Research Hub for Transforming the Mining Value Chain, Centre for Ore Deposit and Earth Sciences (CODES), University of Tasmania, Private Bag 79, Hobart, Tasmania 7001, Australia
4Institute of Mineralogy, Urals Branch, Russian Academy of Sciences, 456301 Miass, Chelyabinsk District, Russia
5Geological Survey of South Australia, Department of the Premier and Cabinet 4/101 Grenfell Street, Adelaide, South Australia 5000, Australia

Abstract

Faced with ongoing depletion of near-surface ore deposits, geologists are increasingly required to explore for deep deposits or those lying beneath surface cover. The result is increased drilling costs and a need to maximize the value of the drill hole samples collected. Laser ablation-inductively coupled plasma-mass spectrometry (LA-ICP-MS) analysis of pyrite is one tool that is showing promise in deep exploration. Since the trace element content of pyrite approximates the composition of the fluid from which it precipitated and the crystallization mechanism, the trace element characteristics can be used to predict the type of deposit with which a pyritic sample is associated. This possibility, however, is complicated by overlapping trace element abundances for many deposit types. The solution lies with simultaneous comparison of multiple trace elements through rigorous statistical analysis. Specifically, we used LA-ICP-MS pyrite trace element data and Random Forests, an ensemble machine learning supervised classifier, to distinguish barren sedimentary pyrite and five ore deposit categories: iron oxide copper-gold (IOCG), orogenic Au, porphyry Cu, sedimentary exhalative (SEDEX), and volcanic-hosted massive sulfide (VHMS) deposits. The preferred classifier utilizes in situ Co, Ni, Cu, Zn, As, Mo, Ag, Sb, Te, Tl, and Pb measurements to train the Random Forests. Testing of the Random Forests classifier using additional data from the same deposits and sedimentary basins (test data set) yielded an overall accuracy of 91.4% (94.9% for IOCG, 78.5% for orogenic Au, 81.1% for porphyry Cu, 93.6% for SEDEX, 97.2% for sedimentary pyrite, 91.8% for VHMS). Similarly, testing of the Random Forests classifier using data from deposits and sedimentary basins that did not have analyses in the training data set yielded an overall accuracy of 88.0% (81.4% for orogenic Au, 95.5% for SEDEX, 90.0% for sedimentary pyrite, 73.9% for VHMS; insufficient data was available to perform a blind test on porphyry Cu and IOCG). The performance of the classifier was further improved by instituting criteria (at least 40% of total votes from the Random Forests needed for a conclusive identification) to remove uncertain or inconclusive classifications, increasing the classifier’s accuracy to 94.5% for the test data (94.6% for IOCG, 85.8% for orogenic Au, 87.8% for porphyry Cu, 95.4% for SEDEX, 98.5% for sedimentary pyrite, 94.6% for VHMS) and 93.9% for the blind test data (85.5% for orogenic Au, 96.9% for SEDEX, 96.7% for sedimentary pyrite, 84.6% for VHMS).

The Random Forests classification models for pyrite trace element data can be used as a predictive modeling tool in greenfield terrains by providing an accurate indication of ore deposit type. This advance will assist mineral explorers by allowing early implementation of predictive ore deposit models when prospecting for ore deposits. Furthermore, the ability of the classifier to accurately identify pyrite of sedimentary origin will allow researchers interested in paleoenvironmental conditions of ancient oceans to effectively screen prospective samples that are affected by a hydrothermal overprint.

Introduction

The correct classification of ore deposits in the early stage of an exploration project can greatly enhance the efficiency of exploration, as it allows for the early application of predictive geologic models. This improvement is especially important when exploring beneath cover due to the increased costs of drilling deep drill holes and when the surface geology or geochemistry fails to reveal details about the deposit at depth. For example, minor disseminated pyrite in a sericite alteration zone intersected in a drill hole under cover could be related to a porphyry Cu outer halo, a volcanic-hosted massive sulfide (VHMS) system footwall alteration zone, a high-sulfidation epithermal Au zone, or barren pyrite unrelated to an ore system. Each of these mineralization types demands a different approach to exploration. Knowing which type is present can save exploration time and money.

© 2019 Gold Open Access: this paper is published under the terms of the CC-BY license.

ISSN 0361-0128; doi:10.5382/econgeo.4654; 16 p.
Digital appendices are available in the online Supplements section.

Submitted: December 20, 2018 / Accepted: March 15, 2019
Laser ablation-inductively coupled plasma-mass spectrometry (LA-ICP-MS) allows for the determination of the trace element content of individual minerals. These data are useful because different ore deposit types have different fluid sources, metal sources, and depositional mechanisms, all of which can significantly affect the trace element content of the minerals that precipitate from them (Gregory et al., 2014; Tardani et al., 2017). Furthermore, these trace elements can be preserved in their mineral hosts during successive hydrothermal and metamorphic events. In this study we focus on pyrite because it is present in many different types of ore deposits, its trace element content can be preserved up to midgreenschist facies (Large et al., 2009), and there are large data sets available that provide (background) trace element contents of pyrite formed in sedimentary environments without hydrothermal inputs (Large et al., 2014, 2015a; Gregory et al., 2015a).

To achieve our objective, LA-ICP-MS analyses of pyrite from a series of different deposit types (iron oxide copper-gold [IOCG], orogenic Au, porphyry Cu, sedimentary exhalative [SEDEX], VHMS deposits, and barren sedimentary pyrite) were used to train a Random Forests classifier to predict deposit type using pyrite LA-ICP-MS analyses. The utility extends to the palaeoceanography community, because the presence or absence of hydrothermal overprints/contributions are often unclear (Gregory et al., 2017), thus eroding confidence in reconstructions of ancient conditions in the oceans.

Random Forests, a supervised classification algorithm, has proven to be an ideal choice for accurately predicting categories from multivariate input features across a wide range of data sets (Fernández-Delgado et al., 2014), but it has only rarely been applied to economic geology problems. While notable exceptions exist, such as identifying zones of hydrothermal alteration and host-rock types (Cracknell et al., 2014) and modeling of mineral prospectivity (e.g., Rodriguez-Galiano et al., 2014; Carranza and Laborte, 2015), many other opportunities remain untested. Additionally, only one previous study (O’Brien et al., 2015) used Random Forests analysis of the trace element contents of individual mineral phases (i.e., gahnite), despite the large amount of multielement geochemistry data generated in recent years by LA-ICP-MS. In this contribution we provide a proof of concept—that is, we show how the Random Forests method can be used to classify ore deposit type both as an exploration tool and as a means of identifying samples most representative of primary marine conditions uncompromised by secondary overprints.

### Supervised classification

The concept of supervised classification can be thought of as linking input features to target classes via a discrimination function \( y = f(x) \). Input features \( x \) are represented as \( m \) vectors of the form \( [x_1, \ldots, x_m] \), and \( y \) is a finite set of \( c \) class labels \( [y_1, \ldots, y_c] \). Given \( N \) instances of \( x \) and \( y \), supervised classification attempts to train a classification model \( f^* \) based on a limited number of training samples (Gahegan, 2000; Hastie et al., 2009; Kovacevic et al., 2009).

In general, there are three stages to supervised classification: (1) data preprocessing, (2) classifier training, and (3) prediction evaluation (Cracknell and Reading, 2014). Data preprocessing involves compiling, correcting, and transforming inputs to a representative set of features containing information relevant to the classification problem (Guyon, 2008; Hastie et al., 2009). Classifier training usually requires the adjustment and selection of one or more parameters, specific to a given supervised classifier, that optimize performance on a given set of input features and target classes (Guyon, 2009). The selection of relevant features necessarily reduces the dimensionality of the input data, thus speeding up processing time while also facilitating interpretations of the relationships between categories and features (Cracknell et al., 2014). Prediction evaluation is vital for assessing the validity of classification outcomes and is typically carried out using a test data set not previously seen by the classifier. An assessment of test data and blind test classifications through a confusion matrix and standard classification metrics—such as overall accuracy, recall, and precision—provides an unbiased indication of the performance of trained classifiers (Congalton and Green, 1998).

### Random Forests

Random Forests (Breiman, 2001) is an ensemble supervised classifier that generates predictions based on a majority vote cast by multiple randomized decision trees, known as a forest. Randomness is introduced by randomly subsampling a number of input features to split at each node of a decision tree and by bagging (bootstrap aggregation). Bagging (Breiman, 1996) generates training data for a single decision tree by sampling, with replacement, a number of samples equal to the number of instances in the training data. The Gini index is used by Random Forests to determine a best split threshold at each node of a decision tree. The Gini index is defined as

\[
Gini(t) = \sum_{c=1}^{C} g_c (1 - g_c),
\]

where \( g_c \) is the probability or the relative frequency of class \( c \) at node \( j \) and is given by

\[
g_c = \frac{n_c}{n},
\]

where \( n_c \) is the number of samples belonging to the class \( c \), and \( n \) is the total number of samples within a particular node. For each candidate split, the threshold that defines maximum reduction in class heterogeneity of the resulting child nodes is selected (Breiman, 1984; Waske et al., 2009).

In addition to a label indicating a predicted class for a given sample, Random Forests produces class membership probabilities. These occur in the form of a vector \( p \) comprising probabilities for individual predictions representing the proportion of decision trees that predict candidate classes.

### Data and Methods

Data preprocessing was primarily executed in standard spreadsheet software (Microsoft Excel), with Random Forests classifier training and prediction evaluation conducted in the open source data mining software platform Orange version 3.18 (Demšar et al., 2013).

**LA-ICP-MS data sources and preprocessing**

This project arose from two major programs of pyrite analysis funded by the Geological Survey of Western Australia (Belousov et al., 2016) and the Geological Survey of South
Australia (D. Gregory, unpub. report, 2015), where pyrite from a large number of ore deposits in both states was analyzed. Additional data from various ore deposits have been analyzed subsequently, leading to the current database of 3,579 pyrite analyses (Figs. 1, 2). LA-ICP-MS data has been provided from a number of different sources, including published peer reviewed manuscripts (Maslennikov et al., 2009, 2017; Large et al., 2014, 2015b; Revan et al., 2014; Gregory et al., 2015a, b, 2016, 2017; Gadd et al., 2016), project reports (G. Davidson, unpub. report, 2005; D. Gregory, unpub. report, 2015), Ph.D. theses (Maier, 2011), and new, previously unreported data from the Chalkidiki porphyry Cu district, Greece, and the Lady Loretta SEDEX deposit, Australia.

All pyrite analyses except those taken from Gadd et al. (2016) were conducted at the LA-ICP-MS facility located at the University of Tasmania, Australia; however, spot size and the number of standards varied. Detailed analytical procedures are available in the references in Table 1. All samples (except for the Gadd et al., 2016, data, which lacked Te and Au) were analyzed for Co, Ni, Cu, Zn, As, Mo, Ag, Sb, Te, Au, Tl, and Pb, and these are the elements emphasized here. When analyses were below detection limits, either half the detection limit was used or the value was inserted from the referred literature source. Because Gadd et al. (2016) did not report Te or Au, we used average values for these elements from the Lady Loretta SEDEX deposit. These data were assumed to be reasonable estimates, as these elements are commonly below detection in SEDEX deposits. Analyses were conducted on 2.5-cm-diameter polished laser mounts. Beam size varied from 10 to 100 µm, depending on the size of pyrite analyzed and the goals of the relevant study. For each analysis, background was measured for 30 s prior to a 40- to 60-s laser ablation period. The analyses were conducting in a pure He atmosphere, and Ar was added to the gas stream prior to injection into the ICP-MS to improve aerosol transport. No correction was applied for doubly charged species, because these species were kept at low levels (below 0.2%). Standards were analyzed at the start and end of each sample change and approximately every 25 analyses in between. The standard STDGL2b2 (Danyushevsky et al., 2011) was used to analyze the elements of interest (except those taken from Gadd et al., 2016).

The locations, pertinent references, and number of analyses used for Random Forests training, testing, and blind testing are given in Table 1. To limit the influence of trace elements from microinclusions of other minerals that might be included during the ablation of pyrite, the data was screened to ensure that no analyses had higher than 1% Zn, 2% As, 1% Cu, 1% Ni, and 2% Co. Also, for analyses on which matrix corrections were preformed, samples with higher than 20% matrix were removed. This combination of newly acquired and compiled data yielded a total of 3,579 analyses from 70 different deposits and sedimentary units. Of these, 2,898 analyses from 43 individual deposits/sedimentary formations were used to train and initially test the Random Forests classifier to identify five distinct ore deposit types: IOCG, orogenic Au, porphyry Cu, SEDEX, and VHMS. In addition to these mineral deposit types, barren sedimentary pyrite was included as a class in

Fig. 1. Sample location map for the entire data set.
the training data in an attempt to avoid misclassification of nonmineralized pyrite as from an ore deposit. The remaining 681 analyses from 27 different deposits/sedimentary formations were used as blind tests of the trained classifier. These data are referred to as blind because analyses from these deposits/sedimentary formations were not present in the training or test data sets.

Data distributions

The geometric mean, multiplicative standard deviation, median, and median absolute deviation (MAD) values of element concentrations for the different ore deposit types from the training and total data sets are provided in Tables 2 and 3. The geometric mean and the median are both presented, because they provide robust summaries of the data, depending on their distributions. Where data are log-normally distributed, the geometric mean and multiplicative standard deviation provide a more useful summary. However, when data are not log-normally distributed, the median and MAD are more appropriate (Reimann and Filzmoser, 2000).

With the exception of the VHMS and IOCG deposits, the training data set used equal numbers of analyses from each deposit. Therefore, the training data set is less biased by the number of analyses performed on the different deposits (i.e., the classifier will skew toward picking the deposit that has more data points in the training set). VHMS and IOCG deposits did not have sufficient analyses from a variety of deposits to have equal numbers of analyses from each deposit in the training data set. Additionally, of the reported statistics, we assert that the medians of trace element content for the different ore deposit types from the training data set should be used rather than total data set statistics for comparisons in future studies. This is because the training set geometric mean and median attempt to represent equal contributions from the different deposits instead of being overly representative of one deposit from which we have more data.

Random Forests training and evaluation

To train and test the Random Forests classifier, we used a total of 3,579 analyses of pyrite that passed the screening process: 159 IOCG, 416 orogenic Au, 416 porphyry Cu, 863 SEDEX, 1,223 sedimentary pyrite, and 482 VHMS. The pyrite trace element data were then split into three groups for classifier training, testing, and blind testing. The 681 analyses used for the blind test were removed (Table 1): orogenic Au (118 from four deposits), SEDEX (66 from three deposits), sedimentary pyrite (451 from 17 formations/basins), and VHMS (46 from three deposits). From the remaining data, a total of 120 analyses from each ore deposit type were used to train Random Forests. To avoid bias toward classes with more analyses, an equal number of analyses from each deposit were randomly selected, except for VHMS and IOCG deposits, because some
Table 1. Sample Location and Number of Samples Used for Random Forests Test, Training, and Blind Test Data Sets

| Location                        | Deposit type                  | Number of training analyses | Number of test analyses | Number of blind test analyses | Reference                              |
|---------------------------------|-------------------------------|-----------------------------|-------------------------|-------------------------------|----------------------------------------|
| Manxuan, Australia              | IOCG                          | 95                          | 31                      |                               | D. Gregory, unpub. report, 2015         |
| Punth Hill, Australia           | IOCG                          | 25                          | 8                       |                               | D. Gregory, unpub. report, 2015         |
| Darlot, Australia               | Orogenic gold                 | 8                           | 5                       | Belousov et al., 2016         |
| East Repulse, Australia         | Orogenic gold                 | 8                           | 37                      | Gregory et al., 2016          |
| Fortman, Australia              | Orogenic gold                 | 8                           | 22                      | Belousov et al., 2016         |
| Golden Mile, Australia          | Orogenic gold                 | 8                           | 35                      | Belousov et al., 2016         |
| Granny Smith, Australia         | Orogenic gold                 | 8                           | 4                       | Belousov et al., 2016         |
| Lancefield, Australia           | Orogenic gold                 | 8                           | 2                       | Belousov et al., 2016         |
| Mars, Australia                 | Orogenic gold                 | 8                           | 9                       | Belousov et al., 2016         |
| Meekathara, Micky Doolan, Australia | Orogenic gold          | 8                           | 2                       | Belousov et al., 2016         |
| Minjar, Australia               | Orogenic gold                 | 8                           | 2                       | Belousov et al., 2016         |
| Nathans Labourchre, Australia   | Orogenic gold                 | 8                           | 3                       | Belousov et al., 2016         |
| Paddington, Western Australia   | Orogenic gold                 | 8                           | 2                       | Belousov et al., 2016         |
| Songvang, Australia             | Orogenic gold                 | 8                           | 5                       | Belousov et al., 2016         |
| Sunrise Dam, Australia          | Orogenic gold                 | 8                           | 4                       | Belousov et al., 2016         |
| Victory, Australia              | Orogenic gold                 | 8                           | 64                      | Gregory et al., 2016          |
| Cadi, Australia                 | Porphyry                      | 60                          | 40                      |                               |
| Challidiki, Greece              | Porphyry                      | 60                          | 256                     |                               |
| Don, Canada                     | SEDEX                         | 24                          | 97                      | Gadd et al., 2016             |
| HYC, Australia                  | SEDEX                         | 24                          | 316                     | Maier, 2011                  |
| Lady Loretta, Australia         | SEDEX                         | 24                          | 25                      |                               |
| Pelly North, Canada             | SEDEX                         | 24                          | 76                      | Gadd et al., 2016             |
| XY deposit, Canada              | SEDEX                         | 24                          | 163                     | Gadd et al., 2016             |
| Aralka Armadeus basin, Australia | Sedimentary pyrite           | 10                          | 12                      | Large et al., 2014; Gregory et al., 2015a |
| Barney Creek Formation, Australia| Sedimentary pyrite           | 10                          | 5                       | Large et al., 2014; Gregory et al., 2015a |
| Canning basin, Australia        | Sedimentary pyrite           | 10                          | 197                     | Large et al., 2014; Gregory et al., 2015a |
| Carbondale, USA                 | Sedimentary pyrite           | 10                          | 19                      | Large et al., 2014; Gregory et al., 2015a |
| NW Shelf, Late Jurassic, Australia| Sedimentary pyrite         | 10                          | 5                       | Large et al., 2014; Gregory et al., 2015a |
| Hamersley basin, Australia      | Sedimentary pyrite           | 10                          | 123                     | Large et al., 2014; Gregory et al., 2015a, b |
| Perth basin, Australia          | Sedimentary pyrite           | 10                          | 23                      | Large et al., 2014; Gregory et al., 2015a |
| Woody Island siltstone, Australia| Sedimentary pyrite           | 10                          | 12                      | Large et al., 2014; Gregory et al., 2015a |
| Canarvon basin, Australia       | Sedimentary pyrite           | 10                          | 8                       | Large et al., 2014; Gregory et al., 2015a |
| Salmon River siltstone, Australia| Sedimentary pyrite           | 10                          | 6                       | Large et al., 2014; Gregory et al., 2015a |
| Satkinskaya Suite, Russia       | Sedimentary pyrite           | 10                          | 19                      | Large et al., 2014; Gregory et al., 2015a |
| Selwyn basin, Canada            | Sedimentary pyrite           | 10                          | 221                     | Large et al., 2014; Gregory et al., 2015a |
| Kutulaa, Turkey                 | VHMS                          | 15                          | 6                       | Revan et al., 2014            |
| Kyzylkaya, Turkey               | VHMS                          | 15                          |                          | Revan et al., 2014            |
| Lahanos, Turkey                 | VHMS                          | 15                          |                          | Revan et al., 2014            |
| Papua, Australia                | VHMS                          | 10                          |                          | Belousov et al., 2016         |
| Golden Grove, Australia         | VHMS                          | 12                          |                          | Belousov et al., 2016         |
| Scudder, Australia              | VHMS                          | 9                           |                          | Belousov et al., 2016         |
| Yoman-Kay deposit, Russia       | VHMS                          | 46                          | 310                     | Maslennikov et al., 2009, 2017|
| Hill 50, Australia              | Orogenic gold                 | 22                          |                          | Belousov et al., 2016         |
| Wallaby, Australia              | Orogenic gold                 | 23                          |                          | Belousov et al., 2016         |
| Wiluna, Australia               | Orogenic gold                 | 62                          |                          | Belousov et al., 2016         |
| Younami, Australia              | Orogenic gold                 | 11                          |                          | Belousov et al., 2016         |
| Anniversary deposit central, Canada | SEDEX                      | 44                          |                          | Gadd et al., 2016             |
| Anniversary deposit east, Canada | SEDEX                      | 15                          |                          | Gadd et al., 2016             |
| OP, Canada                      | SEDEX                         | 7                           |                          | Gadd et al., 2016             |
| Alun shale, Sweden              | Sedimentary pyrite           | 28                          |                          | Large et al., 2014; Gregory et al., 2015a |
| Cowrie siltstone, Australia     | Sedimentary pyrite           | 28                          |                          | Large et al., 2014; Gregory et al., 2015a |
| Curnanosa, Australia            | Sedimentary pyrite           | 28                          |                          | Large et al., 2014; Gregory et al., 2015a |
| Dead Bullock Formation, Australia| Sedimentary pyrite           | 37                          |                          | Large et al., 2014; Gregory et al., 2015a |
| Doushantuo Formation, China     | Sedimentary pyrite           | 25                          |                          | Large et al., 2014; Gregory et al., 2015a |
| Gordon Group, Australia         | Sedimentary pyrite           | 106                         |                          | Gregory et al., 2017          |
| Jet Rock Formation, UK          | Sedimentary pyrite           | 13                          |                          | Large et al., 2014; Gregory et al., 2015a |
| Johnson Cairns Formation, Australia| Sedimentary pyrite         | 17                          |                          | Large et al., 2014; Gregory et al., 2015a |
| Liucaho Formation, China        | Sedimentary pyrite           | 10                          |                          | Gregory et al., 2017          |
| Lower Keppel Creek Formation, Australia | Sedimentary pyrite       | 48                          |                          | Large et al., 2014; Gregory et al., 2015a |
| Oxford J3, Russia               | Sedimentary pyrite           | 29                          |                          | Large et al., 2014; Gregory et al., 2015a |
| Armadeus basin, Australia       | Sedimentary pyrite           | 9                           |                          | Large et al., 2014; Gregory et al., 2015a |
| Posidonia shale, Germany        | Sedimentary pyrite           | 20                          |                          | Large et al., 2014; Gregory et al., 2015a |
| Que River shale, Australia      | Sedimentary pyrite           | 14                          |                          | Large et al., 2014; Gregory et al., 2015a |
| Railway shale, Australia        | Sedimentary pyrite           | 15                          |                          | Large et al., 2014; Gregory et al., 2015a |
| Vallkyrie Formation, Australia  | Sedimentary pyrite           | 9                           |                          | Large et al., 2014; Gregory et al., 2015a |
| Yeneena basin, Australia        | Sedimentary pyrite           | 15                          |                          | Large et al., 2014; Gregory et al., 2015a |
| Chaely deposit, Turkey          | VHMS                          | 4                           |                          | Revan et al., 2014            |
| DeGrussa, Australia             | VHMS                          | 32                          |                          | Belousov et al., 2016         |
| Kilk, Ural, Turkey              | VHMS                          | 12                          | 10                      | Revan et al., 2014            |
deposits lacked a sufficient number of analyses to have equal numbers of analyses (Table 1). The remaining data (2,178 analyses) were used as the initial test of the classifier. A total of 500 trees were used, and splitting was halted if there were five or fewer instances in the resulting child node.

The mean decrease in Gini index, a measure of the contribution of a given variable to correctly classify training data, was used to determine the relevance of different elements during Random Forests classifier training (Fig. 3). This measure of variable importance compares the average total decrease in node impurity (based on the Gini index) when splitting on a given variable, weighted by the proportion of samples in that node. Nickel, As, and Co generated the lowest mean decrease in Gini index values (0.069, 0.069, and 0.062, respectively). To assess if any or all of these elements could be excluded from classifier training, different combinations of these elements were removed from the training data. Random Forests classifiers were also trained with different combinations of these elements removed (Co, Ni, As, Mo, Ag, Sb, Te, Tl, and Pb were chosen as the preferred input variables).

**Results**

**Ore deposit type classification**

Mineralization type classification outcomes for the test data are summarized in Table 4. The Random Forests classifier was run 10 times with different random selections of training data to assess the effectiveness of the classifier with different
Table 3. Summary of Statistics for Entire Data Set

| Deposit | Statistic | Co (ppm) | Ni (ppm) | Cu (ppm) | Zn (ppm) | As (ppm) | Mo (ppm) | Ag (ppm) | Sb (ppm) | Te (ppm) | An (ppm) | Tl (ppm) | Pb (ppm) |
|---------|-----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| IOCG n  | Median    | 159      | 159      | 159      | 159      | 159      | 159      | 159      | 159      | 159      | 159      | 159      | 159      | 159      |
|         | MAD       | 1,694.4  | 61.48    | 2.83     | 0.24     | 2.47     | 0.03     | 0.03     | 0.03     | 0.03     | 0.03     | 0.03     | 0.03     |
|         | GM        | 786.16   | 49.89    | 2.50     | 0.69     | 4.39     | 0.05     | 0.05     | 0.07     | 2.24     | 0.01     | 0.03     | 0.75     |
|         | MSD       | 12.03    | 14.67    | 19.07    | 3.39     | 28.07    | 14.77    | 10.39    | 16.03    | 3.94     | 5.29     | 24.15    | 49.42    |
| Orogenic Au n | Median    | 436      | 436      | 436      | 436      | 436      | 436      | 436      | 436      | 436      | 436      | 436      | 436      |
|         | MAD       | 100.50   | 96.99    | 5.38     | 0.89     | 105.85   | 0.02     | 0.22     | 1.59     | 1.10     | 0.22     | 1.95     | 3.80     |
|         | GM        | 92.36    | 108.05   | 5.25     | 1.46     | 106.62   | 0.03     | 0.19     | 0.95     | 1.06     | 0.23     | 0.02     |
|         | MSD       | 9.22     | 6.24     | 24.73    | 15.64    | 38.35    | 13.32    | 15.33    | 38.62    | 11.36    | 29.98    | 11.91    |
| Porphyry n | Median    | 416      | 416      | 416      | 416      | 416      | 416      | 416      | 416      | 416      | 416      | 416      |
|         | MAD       | 452.81   | 256.89   | 3.37     | 1.65     | 64.62    | 0.31     | 0.17     | 0.16     | 1.57     | 0.05     | 0.01     |
|         | GM        | 264.53   | 176.06   | 5.77     | 2.11     | 83.17    | 0.24     | 0.18     | 0.24     | 1.57     | 0.05     | 0.02     |
|         | MSD       | 9.93     | 8.74     | 15.23    | 5.73     | 8.32     | 7.42     | 7.54     | 9.13     | 6.82     | 6.41     | 5.37     | 14.36    |
| SEDEX n | Median    | 863      | 863      | 863      | 863      | 863      | 863      | 863      | 863      | 863      | 863      | 863      |
|         | MAD       | 77.91    | 409.00   | 394.48   | 82.21    | 898.62   | 22.57    | 11.03    | 57.63    | 0.27     | 0.02     | 20.24    |
|         | GM        | 65.44    | 371.44   | 308.78   | 61.03    | 644.74   | 18.35    | 8.38     | 45.11    | 0.00     | 0.00     | 19.02    |
| Sedimentary n | Median   | 72.82    | 347.80   | 429.97   | 95.10    | 765.34   | 20.57    | 11.47    | 52.44    | 0.31     | 0.02     | 20.86    |
|         | MAD       | 5.24     | 5.88     | 4.07     | 4.75     | 3.35     | 5.35     | 3.92     | 3.95     | 2.13     | 1.60     | 9.90     | 3.56     |
| VHMS n  | Median    | 482      | 482      | 482      | 482      | 482      | 482      | 482      | 482      | 482      | 482      | 482      |
|         | MAD       | 5.33     | 2.94     | 841.51   | 314.29   | 734.17   | 1.79     | 23.37    | 33.98    | 23.94    | 1.25     |
|         | GM        | 4.35     | 1.69     | 605.88   | 190.58   | 771.80   | 1.05     | 15.03    | 30.08    | 18.33    | 0.94     | 1.21     |
|         | MSD       | 31.01    | 12.60    | 7.93     | 11.93    | 7.06     | 15.33    | 12.46    | 12.56    | 22.03    | 10.11    |

GM = geometric mean, MAD = median absolute deviation, MSD = multiplicative standard deviation

Fig. 3. Gini decrease for elements used in the ore deposit type Random Forests classifier.
random seeds of data (App. 1). Random Forests correctly identified the ore deposit type from pyrite trace element analyses with an overall accuracy of 91.0 ± 0.8%. Recall statistics for individual ore deposit types range from 76.9 ± 5.4 to 95.0 ± 0.8%. IOCG, orogenic Au, and porphyry Cu test data samples were predicted with recalls of 86.9 ± 5.0, 76.9 ± 5.4, and 84.2 ± 2.4%, respectively. SEDEX, sedimentary pyrite, and VHMS test data were predicted with noticeably better recalls of 95.0 ± 0.8, 92.8 ± 2.1, and 94.4 ± 1.8% respectively. These results show that the different random selections for the training data produce similar results. As such the same training data set (the one that produced the results in Table 4) was used for the following experiments.

Table 5 indicates that by removing a small percentage (7.8%) of Random Forests predictions with class membership probabilities of less than 40%, ambiguous classifications can be eliminated. This correction results in an increase in overall accuracy of 3.1% to a total of 94.5%. Similarly, the range of class recalls for individual ore deposit types increased by between –0.3 and 7.0%, with, IOCG, orogenic Au, porphyry Cu, SEDEX, sedimentary pyrite, and VHMS deposits having adjusted individual recalls of 94.6, 85.8, 87.8, 95.4, 98.5, and 94.6%, respectively.

Blind test results indicate the Random Forests classifier generated predictions with an overall accuracy of 88% with class-dependent recalls between 73.9 and 95.5% (Table 6). The orogenic Au, SEDEX, sedimentary pyrite, and VHMS samples were classified with proportions of correct classification of 81.4, 95.5, 90.0, and 73.9%, respectively. More accurate results were again obtained when excluding predictions with maximum class membership probabilities of less than 40% (Table 7). Increases in recall ranged from 1.4 to 10.7% resulting in class recalls for orogenic Au of 85.5%, SEDEX of 96.9%, sedimentary pyrite of 96.7%, and VHMS of 84.6%. Different class membership thresholds were trialed (33, 40, and 50%). The 40% threshold was chosen because it led to an increase in recall rates of 3.1% (importantly, this includes an increase in orogenic Au recall of 7.0% and porphyry Cu recall of 6.7%) while preserving approximately 92.2% of the number of original analyses in the test data.

Table 4. Confusion Matrix for Random Forests Classification of Test Data

| Predicted | IOCG | Orogenic Au | Porphyry | SEDEX | Sedimentary | VHMS | Sum | % correct |
|-----------|------|-------------|----------|-------|-------------|------|------|-----------|
| IOCG      | 37   | 1           | 1        | 1     | 39          |      | 94.9 |
| Orogenic Au | 9    | 156         | 26       | 1     | 6           | 198  | 78.8 |
| Porphyry  | 11   | 30          | 240      | 6     | 9           | 296  | 81.1 |
| SEDEX     | 4    | 7           | 634      | 29    | 3           | 677  | 93.6 |
| Sedimentary | 2    | 6           | 5        | 7     |             | 290  | 316  | 91.5     |
| VHMS      | 2    | 11          | 2        | 4     | 7           | 312  | 2,178 |
| Sum       | 59   | 204         | 282      | 643   | 678         |      |      |          |

Table 5. Confusion Matrix for Random Forests Classification of Test Data when Samples with Less Than 40% of the Votes Are Removed

| Predicted | IOCG | Orogenic Au | Porphyry | SEDEX | Sedimentary | VHMS | Sum | % correct |
|-----------|------|-------------|----------|-------|-------------|------|------|-----------|
| IOCG      | 35   | 1           | 1        | 1     | 37          |      | 94.6 |
| Orogenic Au | 7    | 145         | 16       | 7     | 169         | 118  | 85.8 |
| Porphyry  | 5    | 18          | 216      | 4     | 3           | 246  | 87.8 |
| SEDEX     | 3    | 623         | 25       | 2     | 653         | 95.4 |
| Sedimentary | 1    | 4           | 598      | 4     | 607         | 98.5 |
| VHMS      | 1    | 7           | 1        | 4     | 3           | 279  | 295  | 94.6     |
| Sum       | 48   | 170         | 235      | 631   | 631         | 289  | 2,007|

Table 6. Confusion Matrix for Random Forests Classification of Blind Test Data

| Predicted | IOCG | Orogenic Au | Porphyry | SEDEX | Sedimentary | VHMS | Sum | % correct |
|-----------|------|-------------|----------|-------|-------------|------|------|-----------|
| IOCG      | 3    | 96          | 12       | 7     | 118         |      | NA  |
| Orogenic Au | 7    | 1           | 1        | 63    | 2           | 66   | 95.5 |
| Porphyry  | 14   | 4           | 14       | 9     | 406         | 451  | 90.0 |
| SEDEX     | 4    | 2           | 2        | 34    | 46          |      |      |          |
| Sedimentary | 3    | 27          | 74       | 410   | 45          | 681  |      |          |
| VHMS      | 20   | 105         | 27       | 74    | 410         | 45   |      |          |
increased the recall rates by 1.1%, and the 50% threshold only had an increase in recall rates of 5.2% and required removal of 18.3% of the data. The results of these experiments are in Appendix 2.

A series of Random Forest classifications were rerun with different combinations of Te, Co, As, and Ni removed from the data. This exercise was included because Te has several analyses below detection limits and the data set from Gadd et al. (2016) did not include Te. The value of Co, As, and Ni was tested because these had the lowest mean decreases in the Gini index (Fig. 3). While the removal of one of these elements did not cause large changes in the ability of the Random Forests classifier to predict deposit type in general, it did significantly affect the ability to classify individual deposit types; thus, all these elements were included in the preferred classifier.

It has been proposed that pyrite trace element content is reset when trace elements are forced out of the pyrite lattice at metamorphic grades higher than midgreenschist facies (Large et al., 2009; Thomas et al., 2011). To test this assertion, we put LA-ICP-MS pyrite trace element data (n = 117), and greenstone-related pyrite (py7; n = 20). Of these, sedimentary pyrite and orogenic Au pyrite had 97.5 and 84.9% of the analyses correctly identified. Similarly, these classifications only had 16 and 9% of the analyses removed as inconclusive (received less than 40% of the votes). Py5 had 76% of its analyses removed as inconclusive, and Py3 and Py7 both had only 62.5% of their analyses chosen as the one that had the highest percentage classification. Py4 only had 38% of the analyses removed as inconclusive, and 80% of the analyses were identified as orogenic Au. These results are summarized in Table 9.

Table 9. Confusion Matrix for Random Forests Classification of High Metamorphic-Grade Pyrite

| Deposit       | Deposit type      | % correct | Number correct | Number incorrect |
|---------------|-------------------|-----------|----------------|------------------|
| Big Bell      | Orogenic gold     | 80.0      | 8              | 2                |
| Chalice       | Orogenic gold     | 90.0      | 9              | 1                |
| Hunt          | Orogenic gold     | 84.6      | 11             | 2                |
| Junction      | Orogenic gold     | 0.0       | 0              | 0                |
| Kanowna Belle | Orogenic gold     | 94.5      | 11             | 2                |
| Maybell       | Orogenic gold     | 62.5      | 5              | 3                |
| Porphyry      | Orogenic gold     | 55.0      | 11             | 9                |
| Redeemer      | Orogenic gold     | 37.5      | 3              | 5                |
| Total         |                   | 65.2      | 58             | 31               |

Discussion

Conventional X-Y element scatter plots

Conventional element scatter plots of pyrite chemistry have been used with some degree of success to differentiate pyrite from different ore types. However, X-Y scatter plots are less useful when discriminating pyrite from more than two other deposit types. Examples are given in Figure 4 for the pyrite training data set from this study. In general terms, pyrite in the ore zones from medium- to low-temperature hydrothermal deposit types (VHMS and SEDEX) tend to contain higher concentrations of most trace elements compared to pyrite from higher-temperature hydrothermal deposit types (porphyry Cu, IOCG, and orogenic Au). This relationship is illustrated in Figure 4A through C and F (Zn-Cu, Mo-As, Ag-Pb, and Tl-Sb scatter plots). Sedimentary pyrite also contains high concentrations of most trace elements and plots in the same vicinity as data for SEDEX and VHMS deposits. Porphyry Cu, IOCG, and orogenic Au pyrites by comparison generally contain lower levels of Zn, Cu, Mo, Ag, Pb, Tl, and Sb. Commonly, the data for different deposit types exhibit strong overlaps such that it is virtually impossible to distinguish ore type based on simple trace element scatter plots (e.g., Fig. 4D).

By simultaneously using several different elements, Random Forests allows us to go beyond what is possible with...
traditional X-Y plots, but visualization of the distinctions can be challenging. By assessing the overall element concentrations of classified ore deposit types, however, some of the Random Forests decision boundaries can be depicted. For this discussion, we use training data median values, as they are less affected by imbalances in the number of samples from each deposit type compared to complete or test data sets, and they provide a reasonable estimate of the central tendency of populations that are not normally distributed. Copper and Zn can be used to separate SEDEX (medians of 495.49 ppm for Cu and 95.95 ppm for Zn) and VHMS (medians of 1,002.64 ppm for Cu and 150.02 ppm for Zn) deposits from the other deposit types, as they are one to two orders of magnitude more enriched in these elements (Fig. 4). Conversely, distinctly low As values (median 2.15 ppm) can be used to separate IOCG and, to a lesser extent, porphyry Cu mineralization (median 53.36 ppm). Enrichments in molybdenum are known to occur in a number of sedimentary settings, particularly when euxinic conditions are present (Lyons et al., 2003; Tribovillard et al., 2006; Scott et al., 2008; Lyons et al., 2009). Therefore, it follows that high Mo can be used to identify SEDEX and sedimentary pyrite (medians of 23.38 and 28.38 ppm Mo, respectively), both of which formed in marine settings. Similarly, VHMS deposits have low but above detection Mo (median of 0.98 ppm), presumably due to the association of VHMS deposits with seawater and deposition at or near the sea floor. SEDEX (medians of 23.88 ppm for Ag and 963.86 ppm for Pb) and VHMS (medians of 22.00 ppm for Ag and 320.41 ppm for Pb) pyrite is enriched in silver and Pb.

Interestingly, Co and Ni in sulfide minerals, which have long been used to determine pyrite source (Loftus-Hills and Solomon, 1967), were among the lowest ranked elements in terms of mean decrease in Gini index (Fig. 3). Nevertheless, porphyry Cu-related pyrite is enriched in Ni compared to the other deposit types (median of 590.40 ppm), and IOCG is very enriched in Co (median of 1,735.28 ppm). Even Au, which was left out of the favored Random Forests classifier due to concerns about the number of analyses that were below detection limits, is potentially significant for identifying orogenic Au (median 0.16 ppm) and VHMS (median 0.41 ppm) deposit types. However, the strength of the Random Forests method lies with its ability to combine all observations rapidly.

### Ore deposit type predictions

The results of Random Forests predictions for test (91.4% correct predictions) and blind test (88%) data (Tables 4, 6) prove the efficacy of Random Forests analyses of pyrite databases to predict ore deposit type. The classification can be further refined by removing the analyses that did not meet the threshold of obtaining 40% or more of the votes from the Random Forests. This adjustment increased the accuracy of experiments with Au removed to 94.5% with 7.8% of data removed for the test data and to 93.9% with 11.3% of data removed for the blind test data (Tables 5, 7).

The very high proportion of correct predictions (98.5% for test data and 96.7% for blind test data) for sedimentary pyrite is particularly important. Specifically, those data represent the only nonmineralized pyrite samples investigated in this study, suggesting that Random Forests classification is able to accurately discriminate pyrite formed from mineralized systems from that formed at low temperature in the water column and in shallow marine sediments. There is often disagreement in the paleoceanographic community in discussions about whether hydrothermal overprints or ocean conditions are responsible for metal enrichments in the rock record. The Random Forests classifier developed here may facilitate the identification of hydrothermal overprints on sedimentary pyrite in future studies.

As there is a disparate number of analyses from different deposit types, it is possible that the classifier is only working well for the deposits that have larger amounts of data. To test whether this is the case, we checked the individual results of the classifier (with a >40% vote threshold) for each deposit from the test and blind test data set (Tables 10, 11). Of these, all but one of the deposits were conclusively (greater than 65%) correctly identified. The deposit that was inconclusive, the Youanmi orogenic Au deposit, still had 60% of the votes and only had 10 analyses to classify, so it may be that the pyrite trace element content was not accurately represented by the sample. This demonstrates that the Random Forests classifier can identify analyses from the deposits used in developing the classifier.

### Effects of metamorphic grade on classifier predictions

To test and assess how high-grade metamorphic overprint will affect the ability of Random Forests to identify ore deposit type, we used analyses from Belousov et al. (2016) that were from upper greenschist or higher-grade metamorphic facies. These data resulted in a total decrease of over 10% effectiveness of the classifier (Table 8) and importantly resulted in 50% of the deposits being inconclusively or misclassified (Table 8) using the initial results or 37.5% after inconclusive analyses (analyses that received less than 40% of
Fig. 4. Scatter plots of trace elements in pyrite used in training data set for the ore deposit type Random Forests classifier: A) Zn versus Cu, B) Mo versus As, C) Ag versus Pb, D) Te versus Au, E) Co versus Ni, and F) Tl versus Sb.
Deposit | Deposit type | % correct | Number correct | Number incorrect |
---|---|---|---|---|
Maxxman | IOCG | 100.0 | 29 | 0 |
Pont Hill | IOCG | 75.0 | 6 | 2 |
Darlot | Orogenic Au | 100.0 | 4 | 0 |
East Repulse | Orogenic Au | 77.4 | 24 | 7 |
Fortnum | Orogenic Au | 75.0 | 15 | 5 |
Golden Mile | Orogenic Au | 82.6 | 19 | 4 |
Granny Smith | Orogenic Au | 75.0 | 3 | 1 |
Lancefield | Orogenic Au | 100.0 | 1 | 0 |
Mars | Orogenic Au | 77.8 | 7 | 2 |
Meekatharra, Prohibition | Orogenic Au | 100.0 | 2 | 0 |
Meekatharra, Micky Doolan | Orogenic Au | 100.0 | 2 | 0 |
Minjar | Orogenic Au | 100.0 | 2 | 0 |
Nathans Labouchere | Orogenic Au | 100.0 | 3 | 0 |
Paddington | Orogenic Au | 100.0 | 2 | 0 |
Songyang | Orogenic Au | 100.0 | 5 | 0 |
Sunrise Dam | Orogenic Au | 100.0 | 2 | 0 |
Victory | Orogenic Au | 91.5 | 54 | 5 |
Chalkidiki | Porphyry | 80.0 | 25 | 7 |
Callia | Porphyry | 89.1 | 183 | 23 |
Don | SEDEX | 100.0 | 93 | 0 |
HYC | SEDEX | 91.9 | 284 | 25 |
Lady Loretta | SEDEX | 100.0 | 25 | 0 |
Pelly North | SEDEX | 100.0 | 72 | 0 |
XY | SEDEX | 96.8 | 149 | 5 |
Armadale basin | Sedimentary pyrite | 100.0 | 12 | 0 |
Barney Creek Formation, McArthur basin | Sedimentary pyrite | 100.0 | 4 | 0 |
Woody Island siltstone | Sedimentary pyrite | 91.7 | 14 | 1 |
Canning basin | Sedimentary pyrite | 98.3 | 174 | 3 |
Carbondale | Sedimentary pyrite | 100.0 | 18 | 0 |
Late Jurassic NW Shelf | Sedimentary pyrite | 100.0 | 5 | 0 |
Perth basin | Sedimentary pyrite | 100.0 | 23 | 0 |
Satinkokaya Suite | Sedimentary pyrite | 100.0 | 19 | 0 |
Hammersley basin | Sedimentary pyrite | 96.5 | 110 | 4 |
Selwyn basin | Sedimentary pyrite | 99.5 | 208 | 1 |
Salmon River siltstone | Sedimentary pyrite | 100.0 | 6 | 0 |
Canarvon basin | Sedimentary pyrite | 100.0 | 8 | 0 |
Kuthular | VHMS | 80.0 | 4 | 1 |
Total | | 94.8 | 1,896 | 111 |

The votes were removed. This suggests that pyrite trace element content can give spurious results in high metamorphic-grade settings. The exact reason for this variation in trace element content is beyond the scope of this study; however, it is interesting to note that the Ni median is higher in the high metamorphic-grade orogenic gold deposits (258 ppm) and lower in the Sb (0.49 ppm; Table 12) more similar to high metamorphic-grade orogenic Au deposits (258 ppm). It is interesting to note that the Ni median is higher in the high metamorphic-grade orogenic Au deposits. This may reflect pyrite dissolution and reprecipitation or recrystallization of the pyrite at high temperatures imparting a chemistry more indicative of magmatic processes.

**Identification of pyrite that has a source not included in the classifier**

One of the limitations of using Random Forests to predict unknowns in a geologic setting is that it will always give an answer that corresponds with the input designations of the training data set. Because there is a wide variety of different deposits and pyrite sources not associated with economic mineral deposits, there is a risk that the classifier will assess everything as coming from a mineralized deposit. To check how a classifier will respond to barren, non-sedimentary pyrite, we used pyrite data from sedimentary pyrite, orogenic gold-related pyrite, and four pyrite generations unrelated to the mineralization from the St. Ives Au district (Gregory et al., 2016). The sedimentary and orogenic Au pyrite was conclusively, correctly identified (note that the orogenic Au pyrite was included in the training data set earlier), while three of the nonmineralized pyrites returned inconclusive results (Table 9). The fourth was incorrectly conclusively identified as orogenic Au. This shows that most barren pyrite can be identified correctly by calculating the proportion of analyses that are inconclusive and by establishing criteria for how many inconclusive identifications are present in a given sample or set of samples. At the same time, it serves as a reminder that this classifier still needs a large number of analyses from many of the deposit types listed, deposit types currently not represented in the classifier, and other types of nonmineralized pyrite before it can be confidently utilized in the mineral exploration industry. Furthermore, it also shows that the classifier has the potential to be used as one of several tools when making decisions regarding priority of drill targets but not as a replacement for traditional tools, such as petrography, when determining the paragenesis of an ore deposit.
Caveats and future work

The pyrite data investigated in this study were obtained from analyses collected over 10 years as part of a number of different projects with contrasting objectives. In addition, the LA-ICP-MS technology has continued to develop over this time, and detection limits for all trace elements vary significantly. This has resulted in a range of detection limits throughout the data, including SEDEX deposits with anomalously high limits for Se, Cd, Au, and Te (Maier, 2011). In the case of the data from Gadd et al. (2016), some of these elements were not analyzed (or reported). Cadmium and Se results were omitted from our training data for this reason but should be included in future analyses, as both these elements accumulate in pyrite and could be useful for discriminating ore deposit type.

Similarly, the optimal Random Forests classifier was refined to not include Au. Tellurium, however, was not omitted from this classifier despite the lack of Te data from SEDEX deposits. The classifier has difficulty identifying orogenic Au mineralization because Te is commonly associated with Au mineralization (Belousov et al., 2016). Because the Random Forests classifier requires all trace elements in the table to contain nonmissing values, the averages from the single SEDEX deposit that had good-quality Te and Au data (Lady Loretta) were used for all the SEDEX analyses. This has probably overestimated the ability of the classifier to identify SEDEX analyses, because the same value for Te was used by all the SEDEX samples. However, because SEDEX pyrite also has distinctly higher Cu, Mo, Sb, Tl, and Pb concentrations compared to most other deposits, it is thought that Te is not particularly important for SEDEX classification. Furthermore, concentrations of Te in SEDEX samples only differ significantly from those in orogenic Au, porphyry Cu, and VHMS samples. To further test this reasoning, the favored classifier test data (omitting Au) was rerun to exclude Te. The results are summarized in Table 13. This experiment showed that, indeed, the SEDEX results were enhanced by substituted Te values; however, the SEDEX analyses without Te were still correctly identified most of the time with a recall of 74.2%.

| Deposit       | Deposit type | % correct | Number correct | Number incorrect |
|---------------|--------------|-----------|----------------|------------------|
| Hill 50       | Orogenic Au  | 66.7      | 12             | 6                |
| Wallaby       | Orogenic Au  | 80.0      | 16             | 4                |
| Wiluna        | Orogenic Au  | 96.8      | 60             | 2                |
| Youanmi       | Orogenic Au  | 60.0      | 6              | 4                |
| Anniversary Central | SEDEX | 95.2      | 40             | 2                |
| Anniversary East | SEDEX  | 100.0     | 15             | 0                |
| SEDEXOP       | SEDEX       | 100.0     | 7              | 0                |
| Curnamonna    | Sedimentary | 94.3      | 33             | 2                |
| Alum shale    | Sedimentary | 100.0     | 22             | 0                |
| Doushantuo Formation | Sedimentary | 93.1      | 67             | 5                |
| Jet Rock      | Sedimentary | 100.0     | 26             | 0                |
| Armadus basin | Sedimentary | 100.0     | 9              | 0                |
| Posidonina    | Sedimentary | 100.0     | 20             | 0                |
| Railway shale | Sedimentary | 92.3      | 12             | 1                |
| Luchapo Formation | Sedimentary | 83.3      | 5              | 1                |
| Gordon Group  | Sedimentary | 100.0     | 13             | 0                |
| Que River shale | Sedimentary | 100.0     | 9              | 0                |
| Oxford J3     | Sedimentary | 100.0     | 29             | 0                |
| Rocky Cape Group Corrie Siltstone | Sedimentary | 100.0     | 26             | 0                |
| Vallyrie Formation, McArthur basin | Sedimentary | 100.0     | 9              | 0                |
| Dead Bullock Formation | Sedimentary | 100.0     | 25             | 0                |
| Togari Group  | Sedimentary | 95.8      | 46             | 2                |
| Yeneena basin | Sedimentary | 100.0     | 8              | 0                |
| Yerrida Group | Sedimentary | 100.0     | 17             | 0                |
| VHMS Chaely   | VHMS        | 75.0      | 3              | 1                |
| VHMS DeGrossa | VHMS        | 80.0      | 20             | 5                |
| VHMS Klik     | VHMS        | 100.0     | 10             | 0                |
| Total         |             | 94.2      | 567            | 35               |

Table 12. Median and MAD for High Metamorphic-Grade Pyrite (from Belousov et al., 2016)

| Deposit                | Statistic | Co (ppm) | Ni (ppm) | Cu (ppm) | Zn (ppm) | As (ppm) | Mo (ppm) |
|------------------------|-----------|----------|----------|----------|----------|----------|----------|
| High metamorphic-grade | Median    | 258.01   | 290.62   | 15.40    | 1.84     | 137.30   | 0.02     |
|                        | MAD       | 217.57   | 173.76   | 15.07    | 1.69     | 136.23   | 0.01     |
| Orogenic gold          | Ag (ppm)  | 1.65     | 0.49     | 6.49     | 0.34     | 0.03     | 12.26    |
|                        | Sh (ppm)  | 1.53     | 0.48     | 6.14     | 0.34     | 0.03     | 9.21     |

MAD = median absolute deviation
The Random Forests classifier developed here, based on the concentrations of Co, Ni, Cu, Zn, As, Mo, Ag, Sb, Te, Tl, and Pb in pyrite, was found to correctly classify both test data and blind test data. These results yielded an overall accuracy for the test and blind test data of 94.5 and 93.9%, respectively, when inconclusive analyses (less than 40% of votes) are not considered. We can conclude that Random Forests classifiers developed from microanalyses of individual minerals are potentially useful for identifying ore deposit type and should be considered a viable geochemical exploration tool, although it should be stressed that this approach should be regarded as a preliminary positive result; before it can be widely applied to mineral exploration additional ore-related and non-ore-related pyrite varieties need to be added to the classifier. Furthermore, we stress that this should be regarded as one of many tools rather than a single stand-alone classification method. Parties who are interested in using the classifier on their own data sets are encouraged to contact the lead author, who can arrange the processing of LA-ICP-MS pyrite data.

By testing how well the classifier can identify ore deposit type on pyrite that has passed through the midgreenschist facies metamorphic window, we have found that at least in some areas the trace element composition of pyrite has been significantly altered such that the classifier can no longer identify the original pyrite type conclusively. This supports the assertion that pyrite chemistry can be altered at these metamorphic grades. These results are also important for fields of geology not interested in ore deposits or exploration for ore deposits. The high degree of effectiveness of the classifier for identifying sedimentary pyrite not associated with hydrothermal fluids has created an additional opportunity for recognizing hydrothermal overprints on sedimentary deposits included in paleoceanographic studies.

Acknowledgments

We would like to acknowledge the Western Australia and South Australia geological surveys for their support of the initial studies that accumulated much of the initial data that this project arose from. We also thank the University of Western Australia Centre for Exploration Targeting (UWA CET) for providing a sample set from Western Australia orogenic gold deposits. Funding for the compilation of additional data and the refining of the classifier was provided by the National Science Foundation Frontiers in Earth System Dynamics (NSF FESD) program and the National Aeronautics and Space Administration (NASA) Astrobiology Institute under cooperative agreement NNA15BB03A issued through the Science Mission Directorate. This study also benefited from data collected as part of the Australian Mineral Industry Research Association (AMIRA) International project P1060, Enhanced Geochemical Targeting in Magmatic-Hydrothermal Systems. The authors gratefully acknowledge Alan Goode and Adele Seymon (AMIRA International) and all the industry sponsors of P1060 for their generous sponsorship of this research. We also thank Artur Deditius and Denis Fougerouse for valuable suggestions on the manuscript.

REFERENCES

Belousov, I., Large, B., Melfire, S., Davyushhevsky, L., Steadman, J., and Beardsmore, T., 2016, Pyrite compositions from VHMS and orogenic Au deposits in the Yilgarn craton, Western Australia: Implications for gold and copper exploration: Ore Geology Reviews, v. 79, p. 474–499.
Scott, C., Lyons, T., Bekker, A., Shen, Y., Poulton, S., Chu, X., and Anbar, A., 2008, Tracing the stepwise oxygenation of the Proterozoic ocean: Nature, v. 452, no. 7186, p. 456–459.
Tardani, D., Reich, M., Dedíttos, A.P., Chrysoulous, S., Sánchez-Alfaro, P., Wrage, J., and Roberts, M.P., 2017, Copper-arsenic decoupling in an active geothermal system: A link between pyrite and fluid composition: Geochimica et Cosmochimica Acta, v. 204, p. 179–204.
Thomas, H.V., Large, R.R., Bull, S.W., Maslennikov, V., Berry, R.F., Fraser, R., Froud, S., and Moye, R., 2011, Pyrite and pyrrhotite textures and composition in sediments, laminated quartz veins, and reefs at Bendigo gold mine, Australia: Insights for ore genesis: Economic Geology, v. 106, no. 1, p. 1–31.
Tribovillard, N., Algeo, T.J., Lyons, T., and Riboulleau, A., 2006, Trace metals as paleoredox and paleoproductivity proxies: An update: Chemical Geology, v. 232, no. 1, p. 12–32.
Waske, B., Benediktsson, J.A., Árnason, K., and Sveinsson, J.R., 2009, Mapping of hyperspectral AVIRIS data using machine-learning algorithms: Canadian Journal of Remote Sensing, v. 35, no. 1, p. 106–116.

Daniel Gregory is an assistant professor in economic geology at the University of Toronto, Canada. He worked as an exploration geologist in the Yukon Territory, Canada, before he moved to Australia to complete his Ph.D. degree in economic geology and geochemistry at the Centre for Ore Deposit and Earth Sciences (CODES), Tasmania. Daniel held postdoc positions at CODES and the National Aeronautics and Space Administration (NASA) Astrobiology Institute at the University of California Riverside (UCR) investigating basin-scale whole-rock geochemistry and mineral chemistry using macro- and nanoanalytical techniques. He focuses on in situ trace element analyses to understand the fluids related to ore deposit formation. Dan is testing machine learning techniques to identify ore deposit style and vector toward economic mineralization.