Biotite chemistry as an indicator of hydrothermal deposit types and fluid sources: Insights from big data compilation, multivariate statistical analysis, and machine learning

Abu Saeed Baidya [a,b,*], Giridas Maiti [c], Subhronil Mondal [d], Dewashish Upadhyay [a]

[a] Department of Geology and Geophysics, Indian Institute of Technology, Kharagpur, Kharagpur 721302, India
[b] School of Earth and Environmental Sciences, University of St. Andrews, Queens Terrace, St Andrews KY16 9TS, United Kingdom
[c] Institute of Applied Geosciences, Karlsruhe Institute of Technology, Karlsruhe 76131, Germany
[d] Department of Earth Sciences, Indian Institute of Science Education & Research Kolkata, Mohanpur 741246, Nadia, India

ARTICLE INFO

Keywords:
Biotite
Fluid sources
Deposit types
Discriminator
PLS-DA
Hydrothermal mineralization

ABSTRACT

Identification of a deposit type in an early stage of exploration can save money and time. However, despite several studies, there is a lack of simple deposit classifiers based exclusively on the major and trace element chemistry of minerals, which is relatively easier to measure. Furthermore, the identification of fluid source is crucial for any ore-genetic study, as it exerts a first order control on physicochemical characteristics of fluid, which ultimately control the alteration mineralogy, metal solubility and mobilization, and metal association in a deposit. The stable isotope composition (e.g., $\delta^{34}S$, $\delta^{13}C$, $\delta^{18}O$, $\delta^{11}B$, etc.) of rocks and minerals is conventionally used to constrain the source of ore-forming fluids, however, the potential of mineral chemistry for constraining fluid source is not well-explored. In this study, we compile the major element composition (wt% of oxides) of a large number (n 1932) of hydrothermal biotite from various ore settings and execute multivariate statistical analysis and machine learning-based predictive modeling to explore the potential of using biotite chemistry as an indicator of deposit type and fluid sources. The atoms per formula units (i.e., apfu values), and calculated parameters (e.g., Xphl, chlorine intercept value, $\log(fH_2O/fHCl)^{fluid}$) is used to perform principle component analysis (PCA) and partial least square discriminant analysis (PLS-DA). The results show that among the different ore settings, iron-oxide copper-gold, Archean gold, porphyry, and magmatic-hydrothermal rare earth metals deposits and among the fluid sources, (meta-)evaporitic, magmatic-hydrothermal, and metamorphic fluid sources can be reliably discriminated using biotite chemistry. We also show that skarn, hydrothermal platinum group elements, and carbonatite-related rare earth elements deposits may not be discriminated using biotite chemistry. Similarly, mafic rock-derived, carbonatite-derived, and meteoric fluid sources may not be discriminated by biotite chemistry. We propose multiple binary and ternary classifier diagrams, which can be used as discriminator of deposit types and fluid sources. Furthermore, we have developed four different PLS-DA predictive modeling, two for deposit types and two for fluid sources, which can predict deposit types and fluid sources with good accuracy. We provide a Windows© based program that can be used to discriminate deposit types using our PLS-DA predictive modeling. The binary and ternary classifiers and the Windows© programs are tested with recently published biotite data, and it has been shown that these classifiers and programs can efficiently separate deposit types and fluid sources. We also discuss about the limitations of the classifiers.

1. Introduction

Hydrothermal deposits supply a large proportion of critical, strategic, and economic metals required by our society. The important ones in this category include porphyry, skarn, volcanic-hosted massive sulfide, sedimentary exhalative, orogenic gold, iron-oxide copper-gold (IOCG) and other affiliated deposits, hydrothermal platinum group of elements, and carbonatite-related hydrothermal rare earth element deposits. These deposits are characterized by varied and different alteration types and sequences, metal associations, metal concentration and

* Corresponding author at: School of Earth and Environmental Sciences, University of St. Andrews, Queens Terrace, St Andrews KY16 9TS, United Kingdom.
E-mail address: abusaeed09@gmail.com (A.S. Baidya).

volume, source-transport-sink system, geologic timescales, and tectonic settings (Meinert et al., 2003; Richards, 2003; Lyons et al., 2006; Groves et al., 2010; Leach et al., 2010; Barton, 2014; Cooke et al., 2014; Belford et al., 2015; Xie et al., 2019; Gaboury, 2019; Barrote et al., 2020; Groves et al., 2020). Due to fundamental differences in the nature of these deposits, the localization of ore metals in them also varies significantly, requiring different exploration strategies. Therefore, the correct identification of the deposit type during the early stages of exploration can potentially reduce exploration cost and time. In this regard, a simple discriminator of deposit types based on mineral chemistry could be useful (Huang et al., 2019; Gregory et al., 2019).

Additionally, information on fluid source, which may be unique to a deposit type, is crucial for better understanding hydrothermal deposits. This is based on the reasoning that fluid source exerts a strong control on physicochemical characteristics of fluids. For example, evaporite-derived fluid is more Na- and Cl-rich and CO_2-poor compared to metamorphic fluids expunged from meta-carbonate rocks (Derome et al., 2007; Yardley and Cleverley, 2015). Similarly, magmatic fluids usually have higher temperatures compared to meteoric fluids. The physicochemical conditions (T, fO_2, pH, halogen fugacity, etc.) of the fluid ultimately controls its metal and ligand carrying capacity, the mechanisms of transport and precipitation, alteration types and patterns, mineral association and composition, and metal association/distribution in space. A wide range of fluids including magmatic, metamorphic, (meta-) evaporitic, basinal, meteoric, and biogenic matter-derived having different physicochemical signatures are responsible for alteration and mineralization in hydrothermal settings on our planet. Mineral halogen signature e.g., Cl/B ratio of scapolite and stable isotope composition e.g., B and S isotope composition of tourmaline and pyrite, respectively are noted to be useful proxies in identifying fluid sources (Bernal et al., 2017; del Real et al., 2020; Hammerli et al., 2017, 2014; Kelly et al., 2020; Xavier et al., 2008). As mineral chemistry is strongly dependent on fluid physicochemical conditions, it may be a potential proxy for the fluid sources, however, it has rarely been explored in detail. In this context, biotite which might have a direct connection with multiple fluid sources including meta-evaporitic and magmatic sources (Marshall and Oliver, 2006; Moshefi et al., 2018), could be useful and must be tested.

Several studies noted a direct relation between mineral chemistry and physicochemical conditions of hydrothermal fluids. For example, the Ti concentration of biotite and the Al concentration of chlorite strongly correlates with their crystallization temperatures (Andersson et al., 2018; Bourdelle, 2021; Moshefi et al., 2018; Raza et al., 2021; Rezaei and Zarasvandi, 2020), whereas the halogen composition of apatite and biotite has been shown to be controlled by the halogen composition and halogen acid fugacities (e.g., fHCl) of fluids genetically related to them (Andersson et al., 2018; Bath et al., 2013; Moshefi et al., 2020; Palma et al., 2019). As the physicochemical conditions of fluid varies across deposit type, minerals that precipitate in different deposit types may have a different chemistry. This suggests the potential of using mineral chemistry as a discriminator of deposit type. The earlier studies that have attempted to use mineral chemistry to discriminate different ore deposit types have been reasonably successful. For example, the Co/Ni ratio of pyrite has been used to discriminate hydrothermal and synsedimentary/diagenetic deposits on the basis that Co is more soluble in hydrothermal solutions compared to Ni and high Co content (>1 wt%), which could be an indicator of IOCG type deposits (Reich et al., 2016; Baidya et al., 2021b; and references cited there). Similarly, a plot of V vs. Cr in magnetite has been used to distinguish iron-oxide apatite (IOA) deposits from magmatic Fe–Ti–V, porphyry, and IOCG deposits based on the observation that magnetite related to IOA deposits has lower Cr, but higher V concentrations compared to those from IOCG deposits (Knipping et al., 2015). One of the main reasons for using minerals such as pyrite or magnetite as discriminators of ore or geologic settings is that they are abundant in a wide variety of deposits and incorporate a large number of elements in their crystal lattice. Like magnetite and pyrite, biotite is a common phase in many of the hydrothermal systems mentioned above, however, its usefulness as a discriminator of hydrothermal deposit types has never been tested.

Biotite is a structurally and compositionally complex, ferromagnesian, hydrous phyllosilicate of the mica group with general formula $XY_3Z_4O_{10}(W)_2$, where X site can be occupied by K, Na, Ba, Ca, Cs, Ag, Cu^+, H_3O^+, NH_4^+; Y site by Al, Mg, Fe^{2+}, Fe^{3+}, Ti, Li, Cr, Mn, V, Zn, Co, Cu^{2+}, Ni, Pb^{2+}; Z site by Si, Al, Be; and W site by $(OH)^-$, F^-, Cl^-, O^{2-} (Rieder et al., 1998; Li et al., 2020). A number of earlier studies have used hydrothermal biotite for constraining physicochemical conditions of ore-forming fluids (e.g., temperature, halogen fugacity ratios, and redox conditions), constraining the timing of mineralization (Ar-Ar dating), fingerprinting the evolution of halogen fugacity of fluids, differentiating mineralized and barren zones, and tracking fluid-rock interactions and fluid pathways (Baidya et al., 2021a; Bath et al., 2013; Moshefi et al., 2020; Parry et al., 2001; Rezaei and Zarasvandi, 2020; Selby and Nesbitt, 2000; Shan et al., 2021; van Ryt et al., 2019). Despite the multitude of data being generated on biotite chemistry, its use as an indicator of fluid source has never been tested. Furthermore, although some studies noted the compositional differences in biotite present in the different ore-forming systems (Monteiro et al., 2008; van Ryt et al., 2019; Baidya et al., 2023), the potential usefulness of biotite chemistry as a discriminator of ore-settings is poorly explored.

The advancement of multivariate statistical analysis and machine learning tools in the past decades has made handling of multidimensional, large geochemical data sets comparatively easier. In particular, dimensionality reduction techniques like principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE), prediction tools like random forests, classification methods such as partial least-square discriminant analysis (PLS-DA) are used to identify patterns in large data, to find relations between data classes and variables, and to predict classes (Gregory et al., 2019; Lindsay et al., 2021; Makvandi et al., 2016b; Petrelli and Perugini, 2016). For example, Gregory et al. (2019) used Random Forest predictive tool to identify deposit types based on pyrite trace element chemistry, whereas a combination of PCA and partial least-square discriminant analysis (PLS-DA) has been conducted on major and trace element data of magnetite, tourmaline, and scheelite to explore the role of mineral chemistry as provenance indicators for classifying and differentiating deposit types and for mineral exploration (Huang et al., 2019; Sciuba et al., 2020, 2021). Similar machine learning algorithms/multivariate statistical analysis on hydrothermal biotite chemistry can be attempted because of the availability of a large quantity of data. Among the statistical methods mentioned above, PLS-DA has dominantly been used as a classification tool for geochemical data (Makvandi et al., 2016b; Huang et al., 2019). Moreover, PLS-DA can also be used as a predictive tool (e.g., Lee and Jemain, 2019), however, it has never been tested on a dataset of mineral chemistry.

In this study, we perform multivariate statistical analysis, namely PCA and PLS-DA, on a compiled dataset containing the major and minor element chemistry of biotite from various hydrothermal settings to understand if biotite composition could be used as an indicator and discriminator of hydrothermal fluid sources and deposit types. We propose several binary and ternary classifiers and provide a Windows© based executable program that implements our PLS-DA-based predictive modeling tool for discriminating fluid sources and deposit types. We also identify deposit and fluid source classes that cannot be discriminated using our proposed model.

2. Materials and methods

2.1. Data collection and classification

A total of 1932 major and minor element analyses of biotite, measured on an electron probe micro analyzer (EPMA), is compiled from published works spanning a range of deposit types and fluid source signatures (Table 1, complete data set is available in the Electronic

Supplementary Material 1 or ESM 1; global distribution of the deposits is provided in ESM 3). While compiling the dataset, the following limitations were encountered: 1) data that is not presented in the published paper or provided in the respective supplementary material could not be included; 2) some paper reported only representative biotite data instead of the complete data set, and in such cases, we used the representative data on the assumption that the provided dataset is a close approximation of the complete dataset; 3) in some papers, only averages of multiple analyses is provided instead of analysis of the individual grains and in these cases, we considered the averages as individual points; 4) care was taken to include only those datasets for which the deposit type and source of fluid are definitively known or have been interpreted or proposed; 5) both biotite and phlogopite analyses have been included, although we eventually show that biotite chemistry is more useful for building the discriminators.

Under "deposit type", 7 sub-groups are considered: porphyry- (n = 456), skarn (n = 75), iron-oxide-copper gold (IOCG, n = 419), magmatic-hydrothermal rare earth element-phosphate (MH REE, n = 275), hydrothermal REE related to carbonatites (n = 65), Archean Au (n = 336), hydrothermal platinum group of elements (hydrothermal PGE, n = 163) deposits, and non-mineralized systems (n = 147; Table 1). We want to emphasize that the MH REE deposit category is unique and based on a single deposit in Sweden (Andersson et al., 2018). It a granite-hosted, vein-style, hydrothermal rare earth deposit that was not categorized by the authors. Hence, we have preferred to consider it as a separate category. We have not considered deposits and prospects separately for simplicity. We also have not considered the compositional variation of biotite within a single deposit. A hydrothermal deposit commonly contains biotite of different types belonging to different alteration zones, stages, and types characterized by a significant compositional difference. For example, biotite from different K ± Fe ± Mg alteration stages in an IOCG deposit and from potassic and phyllic alteration zones in porphyry deposits show a significant compositional variation (Baidya et al., 2023; Moshefi et al., 2020). The "no-

Table 1
Data sources and other relevant details.

Deposit location	Deposit type	Fluid Type	Reference
Blackbird, Idaho Cobalt belt, USA	IOCG	(Meta-)evaporitic	Nash and Connor, 1993
Mt. Isa, Australia	IOCG	(Meta-)evaporitic	Marshall and Oliver, 2006
Rakkurijarvi, Sweden	IOCG	(Meta-)evaporitic	Smith et al., 2007
Sossego, Carajas Province, Brazil	IOCG	(Meta-)evaporitic	Monteiro et al., 2008
Sin Quyen, Vietnam	IOCG	Magmatic-hydrothermal	Li and Zhou, 2018
Jatoba, Carajas Province, Brazil	IOCG	(Meta-)evaporitic	Veloso et al., 2020
Khetri Copper Belt, India	IOCG	(Meta-)evaporitic	Baidya et al., 2021a
Olserum-Djupedal, Sweden	MH REE phosphate	Magmatic-hydrothermal	Andersson et al., 2018
Cantung, Canada	Skarn	Magmatic-hydrothermal	Zaw and Clark, 1978; van Middelaar and Keith, 1990
Mt. Lindsay, Australia	Skarn	Magmatic-hydrothermal	Kwak, 1983
Sangan, Iran	Skarn	Magmatic-hydrothermal	Boomeri et al., 1997; Golmohammadi et al., 2015
Somma-Vesuvius Complex, Italy	Skarn	Magmatic-hydrothermal	Gilg et al., 2001
Los Santos, Spain	Skarn	Magmatic-hydrothermal	Timón et al., 2007
Duluth Complex, USA	Skarn	Magmatic-hydrothermal	Benkó et al., 2018
Bingham, USA	Porphyry	Magmatic-hydrothermal	Lanier et al., 1978; Parry et al., 1978; Bowman et al., 1987
Santa Rita, USA	Porphyry	Magmatic-hydrothermal	Jacobs and Parry, 1979
Sungun, Iran	Porphyry	Magmatic-hydrothermal	Hezarkhani and Williams-Jones, 1998
Dalli, Iran	Porphyry	Magmatic-hydrothermal	Ayati et al., 2008
Miduk, Iran	Porphyry	Magmatic-hydrothermal	Boomeri et al., 2009
Sarchesmeh, Iran	Porphyry	Magmatic-hydrothermal	Boomeri et al., 2010
Dulong, Tibet	Porphyry	Magmatic-hydrothermal	Li et al., 2012
Maher-Abad, Iran	Porphyry	Magmatic-hydrothermal	Siahcheshm et al., 2012
Darreh-Zar, Iran	Porphyry	Magmatic-hydrothermal	Parsapoor et al., 2015
Alter, Argentina	Porphyry	Magmatic-hydrothermal	Maydagán et al., 2016
Yuanzhuding, South China	Porphyry	Magmatic-hydrothermal	Mao et al., 2018
Urumieh-Dokhtar, Iran	Porphyry	Magmatic-hydrothermal	Zarasvandi et al., 2018
Bangpu, Tibet	Porphyry	Magmatic-hydrothermal	Tang et al., 2019
Copper Flat hydrothermal system, USA	Porphyry	Magmatic-hydrothermal	Wallace and Maher, 2019
Bingham, USA	Porphyry	Magmatic-hydrothermal+Meteoric	Lanier et al., 1978; Bowman et al., 1987
Sungun, Iran	Porphyry	Magmatic-hydrothermal+Meteoric	Moshefi et al., 2018
Dalli, Iran	Porphyry	Magmatic-hydrothermal+Meteoric	Ayati et al., 2008
Malanjkhand, India	Porphyry	Magmatic-hydrothermal+Meteoric	Panigrahi et al., 2008
Sarchesmeh, Iran	Porphyry	Magmatic-hydrothermal+Meteoric	Boomeri et al., 2010
Maher-Abad, Iran	Porphyry	Magmatic-hydrothermal+Meteoric	Siahcheshm et al., 2012
Kahang, Iran	Porphyry	Magmatic-hydrothermal+Meteoric	Afshooni et al., 2013
Spirit Lake pluton and Margaret, USA	Porphyry	Magmatic-hydrothermal+Meteoric	Iveson et al., 2016
MacLellan, Manitoba, Canada	Archean Au	Metamorphic	Samson et al., 1999
East Repulse, Australia	Archean Au	Metamorphic	Bath et al., 2013
Malartic, Canada	Archean Au	Metamorphic	Gaillard et al., 2018; de Souza et al., 2021
Pampalo, Hatti Schist Belt, Finland	Archean Au	Metamorphic	Kuikka, 2018
Detour Lake deposit, Canada	Archean Au	Metamorphic	Dubosq et al., 2019
Geita, Tanzania	Archean Au	Magmatic-hydrothermal	van Ryt et al., 2019
Stillwater Complex, USA	Hydrothermal PGE	UM-derived fluid	Boudreau et al., 1986
Bushveld Complex, South Africa	Hydrothermal PGE	UM-derived fluid	Boudreau et al., 1986, Willmore et al., 2000
Duluth Complex, USA	Hydrothermal PGE	UM-derived fluid	Saini-Eidukat et al., 1990
East Kimberly, Australia	Hydrothermal PGE	UM-derived fluid	Boudreau and Hoatson, 2004
Skaergard, East Greenland	Hydrothermal PGE	UM-derived fluid	Sonnenthal, 1992
Great Dyke, Zimbabwe	Hydrothermal PGE	UM-derived fluid	Boudreau and Hoatson, 2004
Sudbury, Canada	Hydrothermal PGE	UM-derived + meteoric fluid	Hanley and Mungall, 2003
Atlantis Bank	No mineralization	Evolved ocean water	Nozaka et al., 2019
Fen Complex, Norway	Carbonatite-related REE	Carbonatite-derived fluid	Andersen and Austrheim, 1991
Gardar Province, Greenland	Carbonatite-related REE	Carbonatite-derived fluid	Finch et al., 1995
Bayan Obo, Mongolia	Carbonatite-related REE	Carbonatite-derived fluid	Smith, 2007, Liu et al., 2018

Table 2
Accuracy of different fields in the proposed binary and ternary classifiers.

Deposit classifiers	Fig. reference	Field under consideration	Outlier factor	Misclassification factors	Field accuracy		
Ternary Cl-F-Ti	Fig. 8C	IOCG	(26/419)*100 = 6 %	(2/36)*100 = 0.6 % (Archean Au)	(3/275)*100 = 1.1 % (MH REE)	(12/456)*100 = 2.6 % (Porphyry)	89.5 %
Ternary Cl-F-Ti	Fig. 8C	MH REE	(29/275)*100 = 10.5 %	(0/336)*100 = 0 % (Archean Au)	(16/419)*100 = 3.8 % (IOCG)	(10/456)*100 = 2.2 % (Porphyry)	83.5 %
Ternary Cl-F-Ti	Fig. 8C	Archean Au + Porphyry	(25/(456 + 336))*100 = 3.2 %	(10/275)*100 = 3.6 % (MH REE)	(5/419)*100 = 1.2 % (IOCG)	–	92 %
IV(F/Cl) -log $(fHF/fHCl)^{fluid}$	Fig. 8B	IOCG	(27/419)*100 = 6.2 %	(6/336)*100 = 1.8 % (Archean Au)	(20/275)*100 = 7.3 % (MH REE)	(3/456)*100 = 0.7 % (Porphyry)	84 %
IV(F/Cl) -log $(fHF/fHCl)^{fluid}$	Fig. 8B	Archean Au + Porphyry + MH REE	26/(336 + 275 + 456)*100 = 2.4 %	(18/419)*100 = 4.3 % (IOCG)			93.3 %
IV(Cl) -log$(fH_2O/fHCl)^{fluid}$	Fig. 8F	Archean Au	(6/334)*100 = 1.8 %	(42/456)*100 = 9.2 % (Porphyry)			89 %
IV(Cl) -log$(fH_2O/fHCl)^{fluid}$	Fig. 8F	Porphyry	(44/456)*100 = 9.6	(6/334)*100 = 1.8 % (Archean Au)			88.6 %
Fluid source classifiers							
Ternary Cl-F-Ti	Fig. 9D	(Meta-)evaporitic	(19/305)*100 = 6.2 %	(11/741)*100 = 1.5 % (Magmatic)	(2/259)*100 = 0.8 % (Metamorphic)		91.5 %
Ternary Cl-F-Ti	Fig. 9D	Magmatic	(122/741)*100 = 16.5 %	(16/305)*100 = 5.2 % ((meta-)evaporitic)	(14/259)*100 = 5.4 % (Metamorphic)		72.3 %
Ternary Cl-F-Ti	Fig. 9D	Metamorphic	(15/(259)*100 = 5.8 %	(0/305)*100 = 0 % ((meta-)evaporitic)	(11/741)*100 = 15 % (Magmatic)		79.2 %
IV(F/Cl) -log $(fHF/fHCl)^{fluid}$	Fig. 9C	(Meta-)evaporitic	18/(305)*100 = 5.9 %	(4/741)*100 = 0.5 % (Magmatic)	(2/259)*100 = 0.8 % (Metamorphic)		92.7 %
IV(F/Cl) -log $(fHF/fHCl)^{fluid}$	Fig. 9C	Magmatic + Metamorphic	(6/(741 + 259))*100 = 0.6 %	(14/305)*100 = 4.6 % (((meta-)evaporitic))			94.8 %
X_{Mg} -log$(fHF/fHCl)^{fluid}$	Fig. 9B	(Meta-)evaporitic	16/(305)*100 = 5.2 %	(4/741)*100 = 0.5 % (Magmatic)	(0/259)*100 = 0 % (Metamorphic)		94.3 %
X_{Mg} -log$(fHF/fHCl)^{fluid}$	Fig. 9B	Magmatic + Metamorphic	(9/(741 + 259))*100 = 0.9 %	(16/305)*100 = 5.2 % ((meta-)evaporitic)			93.9 %
IV(Cl) -log$(fH_2O/fHCl)^{fluid}$	Fig. 9F	Metamorphic	(2/(259)*100 = 0.8 %	(90/741)*100 = 12.1 % (Magmatic)			87.1 %
IV(Cl) -log$(fH_2O/fHCl)^{fluid}$	Fig. 9F	Porphyry	(94/741)*100 = 12.7 %	(2/(259)*100 = 0.8 % (Metamorphic)			86.5 %

mineralization" category is genetically related to the altered ocean floor system and does not include hydrothermal biotite from the non-mineralized assemblages of aforementioned deposit types.

Under the "fluid type" category, the following 8 subgroups have been considered: magmatic (n = 741), carbonatite-derived (n = 65), (meta-) evaporate-derived (n = 305), mafic/ultramafic rock-derived (n = 147), metamorphic (n = 259), modified oceanic (n = 147), magmatic-hydrothermal + meteoric (n = 255), and mafic/ultramafic-derived + meteoric fluid (n = 16) sources (Table 1). In all the deposits considered, biotite has either precipitated from or equilibrated with fluids derived from these sources. Magmatic-hydrothermal or magmatic fluids are related to granite/granitoid/alkaline magmas and have been reported from porphyry, skarn, and IOCG deposits (Chen and Zhou, 2012; Moshefi et al., 2018). Carbonatite-derived, (meta-)evaporate-derived, mafic/ultramafic rock-derived, and metamorphic fluids are related to hydrothermal REE, IOCG, hydrothermal PGE, and Archean Au deposits, respectively (Fig. 1; Willmore et al., 2000; Gilg et al., 2001; Marshall and Oliver, 2006; Smith et al., 2007; Kuikka, 2018). Modified oceanic fluids are related to biotite formed during ocean floor alteration (Fig. 1; Nozaka et al., 2019). The term "modified" is used to emphasize the elevated temperatures (and pressures) of these fluids compared to normal ocean water. Meteoric fluids are not known to precipitate biotite, however, there are some deposit types where late-stage influx of meteoric fluids has been reported and biotite is thought to have precipitated during the mixing of these fluids with magmatic fluids (Panigrahi et al., 2008; Moshefi et al., 2018). This is the reason why magmatic-hydrothermal + meteoric and mafic/ultramafic-derived + meteoric fluids have been considered here and they are reported from porphyry and hydrothermal PGE deposits, respectively (Panigrahi et al., 2008; Hanley et al., 2011). The relationship between the different deposit types and fluid sources is illustrated in Fig. 1.

2.2. Data pre-processing

The biotite major/minor element compositions in the raw dataset are reported as weight percent oxides of Si, Al, Ti, Cr, Fe, Mg, Mn, Zn, K, Ba, Ca, and Na, and elemental concentration of F and Cl. However, in many analyses, the concentrations of some of these oxides/elements are consistently below the detection limit of EPMA, while in others some values are missing. Therefore, subsequent statistical and machine-learning analyses were done only after excluding such data. The selection of the elements was based on the following two criteria: 1) the element or its oxide should have measured values for >60 % of the total number of data points. This is because up to 40 % of missing data can in principle be imputed without compromising the structural integrity of a dataset (Makvandi et al., 2016a). Our compiled dataset usually has values (either values or notes on the values such as below detection limit) and > 90 % data for elements except for Cr, Zn. Ca, and Ba. 2) the measured oxide/element contents should be above the detection limit of EPMA for >60 % of the total measured observations. In the compiled data, concentrations of Cr_2O_3, ZnO, CaO, and BaO are below the detection limit or are not measured for >90 % of the total number of compiled analyses, and therefore not considered. For other elements, 80–90 % data are above detection limit of EPMA for most of the elements.

While performing multivariate statistical analysis on multidimensional data, an incomplete dataset with censored or missing values is not desirable since the heterogeneity cannot be accommodated by the commonly used machine learning programs. Moreover, missing values in a dataset raise significant problems in descriptive statistics and plotting and estimation of parameters. Hence, it needs to be dealt with effectively to retain maximum data. One of the most effective ways to deal with the missing values is imputation, so that the overall

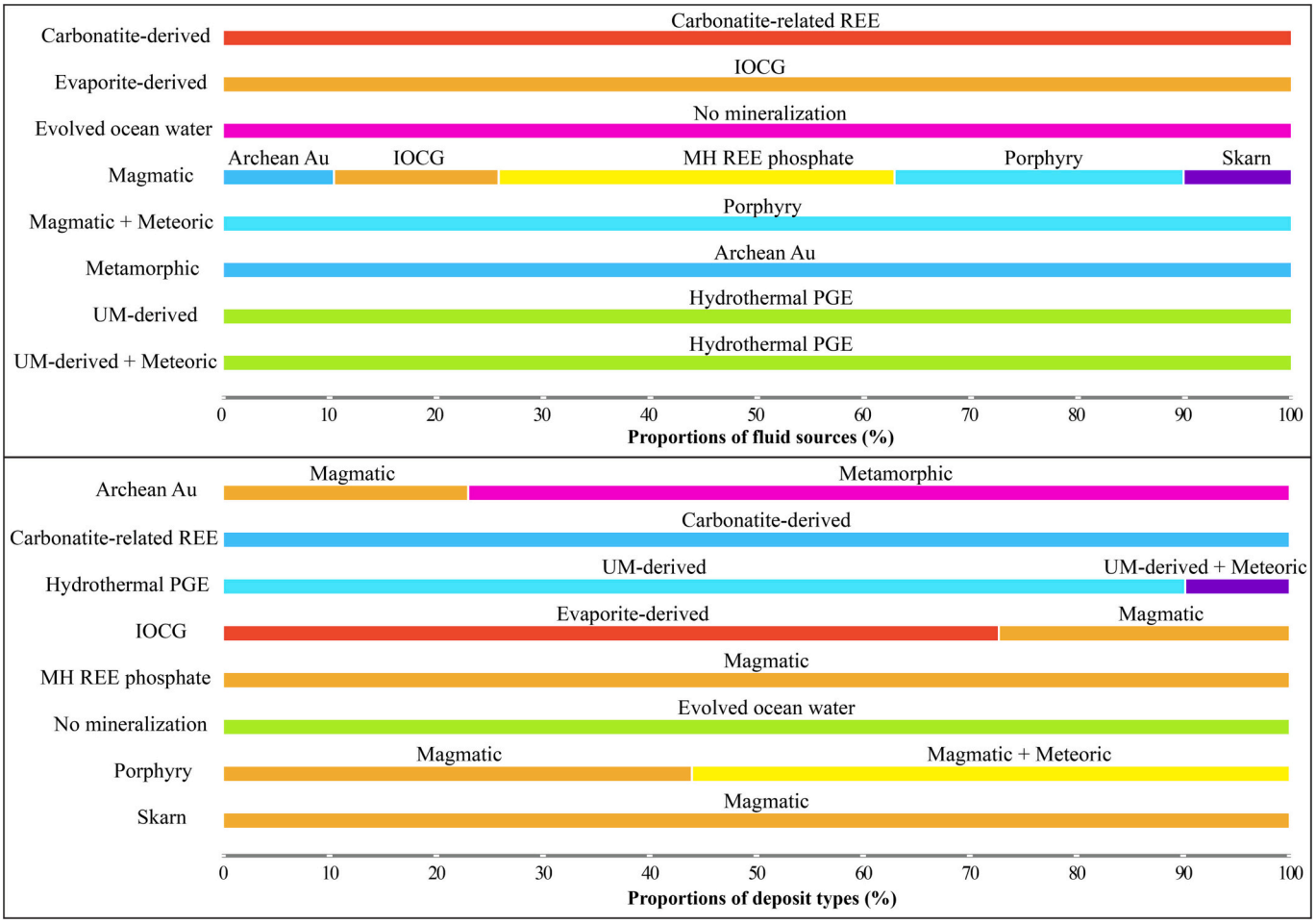

Fig. 1. Bar diagram shows the relationship between fluid sources and deposit types. It is to be noted that one deposit type may contain multiple fluid sources and vice versa. As an example, magmatic fluids are found in IOCG, Archean Au, Porphyry, MH REE, and skarn deposits. On the other hand, IOCG deposits are characterized by (meta-)evaporitic and magmatic fluids whereas porphyry deposits are characterized by magmatic and magmatic + meteoric fluids.

distribution of the feature is not affected by the newly imputed values significantly (Makvandi et al., 2016a and references therein). Imputation of the missing values can be done by detection limit (DL), DL/2, DL/$\sqrt{2}$, average values, or zero (Croghan and Egeghy, 2003; Verbovšek, 2011), but this may introduce biases in the dataset (Dupuis and Beaudoin, 2011; Makvandi et al., 2016a). Alternatively, the imputation can be performed by statistical methods, e.g., by k-nearest neighbor (kNN) method, which has been suggested to be an optimal technique for compositional datasets (Grunsky et al., 2013; Hron et al., 2010; Makvandi et al., 2016a). However, the kNN approach often imputes values above the DL, and it has been suggested that if the imputed values are below the 3 times of DL, then it could be used for statistical analysis (Makvandi et al., 2016a; Sciuba et al., 2020). Recently, Codeço et al. (2021) pointed out that the kNN imputation method has the drawback that it neglects the constraint that the imputed values should be below the DL. Codeço et al. (2021) executed the imputation by two methods namely, DL/$\sqrt{2}$ and kNN, and noted that the latter method returned values >3 times the DL. Consequently, they preferred the imputation of BDL values by DL/$\sqrt{2}$. As in earlier studies, there are two types of missing values in the dataset considered in this study: below detection limit (BDL) and not available/reported (NA) values. For example, in the Khetri area, F and Mn contents in biotite were measured in EPMA but they were not detected i.e., concentrations were BDL of EPMA (400 and 800 ppm, respectively; Baidya et al., 2021a). These values are reported as BDL or 0 in previous studies. On the other hand, NA values are those where EPMA analysis was not done and no data is available, e.g., F was not measured in biotite from the IOCG deposits in the Mount Isa area (Marshall and Oliver, 2006). Initially, the imputation of all BDL values was done by the *impKNNa* function of the *robCompositions* package in R software (Makvandi et al., 2016a). The imputed values often were >3 times the DL, therefore, we preferred imputation by DL/$\sqrt{2}$. For the NA values, imputation is done by the k-nearest neighbor method using the *impKNNa* function. 'k' in the kNN method is a parameter that refers to the number of nearest neighbors to consider for a decision for the group in which a particular data point may belong. There is no structured method to find the best value for "k", and the imputation is usually tested with various values by trial and error assuming that training data is unknown. In general, choosing smaller values for k can be noisy and will have a higher influence on the result while larger values of k will have smoother decision boundaries meaning lower variance but increased biasness. For datasets of current proportions (n = 1932), optimal k values range from 2 to 10 (Makvandi et al., 2016a; Hron et al., 2010). Our data were subjected to imputations using this range of k values but no major differences in the imputed values were noticed. Furthermore, only 75 data among 1932 were imputed by this method, hence we do not think that choice of k values would impart significant changes in the dataset. We have imputed the NA values considering k = 5 to generate the complete dataset.

2.3. *Post-imputation treatment of data*

After the imputation, biotite major oxide data were used for

calculation of structural formula. Different methods are available for calculating the structural formula of biotite, e.g., some studies prefer to calculate the formula based on 22 anions (or 11 anions), some have used 24 anions while a recent study has proposed a machine-learning-based method for the formula recalculation (Dymek, 1983; Henry and Daigle, 2018; Li et al., 2020; Yavuz, 2003; Yavuz and Öztaş, 1997). The published papers from which the present data were collected used different methods for formula calculation making it difficult to compare the computed formulae. To maintain uniformity, the formula for all the analyses should be calculated by the same method. The BIOTERM and MICA+ software, which calculate the biotite formula based on 11 or 22 anions using Dymek (1983)'s method for Fe^{3+} calculation, have been used in some studies. However, we often could not reproduce the Fe^{3+} values obtained by Dymek (1983)'s method in these two softwares. On the other hand, the machine-learning method of Li et al. (2020) may provide some unrealistic values of Fe^{3+} and other tetrahedral, octahedral, and W-site cations/anions for halogen-rich biotite (Baidya and Das, 2022). Hence, we preferred to calculate the structural formulae based on 24 anions (Henry and Daigle, 2018). Oxyanions (O^2) were calculated using the formula O^2 = 2*Ti as recommended by recent studies (Henry et al., 2005; Henry and Daigle, 2018), while OH has been calculated considering full W-site occupancy.

The obtained atoms per formula unit (apfu) values were then used to calculate 12 parameters including mole proportions of phlogopite (Xphl), siderophyllite (Xsid) and annite (Xan); halogen proportion ratios ($\log(X_{Cl}/X_{OH})$, $\log(X_F/X_{OH})$, and $\log(X_F/X_{Cl})$); halogen intercept values (IV(Cl), IV(F), and IV(F/Cl)); and halogen fugacity ratios of fluid ($\log(fH_2O/fHF)^{fluid}$, $\log(fH_2O/fHCl)^{fluid}$, and $\log(fHF/fHCl))^{fluid}$). The Xphl, Xsid, and Xan were calculated using the following formulas:

$$\text{Xphl} = \text{Mg}/(\text{sum of octahedral cations}) \quad (1)$$

$$\text{Xsid} = (3 \quad (\text{Si}/\text{Al}))/1.75^{*}(1 \quad \text{XMg}) \quad (2)$$

$$\text{Xan} = 1 \quad \text{Xphl} \quad \text{Xsid} \quad (3)$$

For the halogen proportion ratios, X_{Cl}, X_F, and X_{OH} are mole proportions of Cl, F, and OH in biotite. The concept of halogen intercept values was developed by Munoz (1984) to account for the effect of Fe and Mg on the incorporation of F and Cl in biotite structure. It is well-established that Cl and Fe on one hand and Mg and F on the other show good positive correlations (the trends are opposite for Mg-Cl and Fe-F pairs) in natural biotite suggestive of strong structural control on halogen incorporation (Munoz, 1984). The intercept values diminish this structural control and express the relative enrichment of F and Cl in biotite. These values are calculated by the following equations:

$$\text{IV(Cl)} = \quad 5.01 \quad 1.93\text{Xphl} \quad \log(X_{Cl}/X_{OH}) \quad (4)$$

$$\text{IV(F)} = 1.52\text{Xphl} + 0.42\text{Xan} + 0.20\text{Xsid} \quad \log(X_F/X_{OH}) \quad (5)$$

$$\text{IV(F/Cl)} = \text{IV(F)} \quad \text{IV(Cl)} \quad (6)$$

Halogen fugacities of fluid are important parameters as they can affect metal-halogen complexation, which in turn, may control precipitation or solution of ore minerals (Munoz and Swenson, 1981). Moreover, large differences in the relative activities of the halogen acids may reflect fundamental differences in the ultimate origin of the ore-forming fluids. These parameter have been noted to be different in different hydrothermal deposits including IOCG, porphyry and Archean Au deposits (Monteiro et al., 2008; van Ryt et al., 2019; Baidya et al., 2023). The halogen fugacity ratios of fluids which equilibrated with biotite or precipitated biotite can be calculated from the halogen concentrations of biotite (Munoz, 1992). The fluid halogen fugacity ratios were calculated using the equations of Munoz (1992) that are based on the revised coefficients for halogen-hydroxyl exchange between biotite and hydrothermal fluid (Zhu and Sverjensky, 1991, 1992). The equations are as follows:

$$\log(fH_2O/fHF)^{fluid} = 1000\Big/(T(2.37 + 1.1\ (\text{Xphl}))) + 0.43 - \log(X_F/X_{OH}) \quad (7)$$

$$\log(fH_2O/fHCl)^{fluid} = 1000\Big/(T(1.15 + 0.55(\text{Xphl}))) + 0.68 - \log(X_{Cl}/X_{OH}) \quad (8)$$

$$\log(fHF)\Big/(fHCl)^{fluid} = 1000\Big/(T(1.22 + 1.65(\text{Xphl}))) + 0.25 - \log(X_F/X_{Cl}) \quad (9)$$

"T" in these equations is the temperature in Kelvin. We have used temperature data available in the studies from which biotite data were collected and/or in other related studies that estimated temperature by independent methods such as fluid inclusion-thermometry or mineral-pair stable isotope data (Marshall and Oliver, 2006). In cases where temperature information was not available, we have calculated temperatures for biotite grains using the empirical method of Wu and Chen (2015), which has been successfully used on hydrothermal biotite by many studies (Rezaei and Zarasvandi, 2020).

For statistical analysis of compositional data, the "effect of closure" that arises from the use of variables with a constant sum must be eliminated (Grunsky, 2010; Whitten, 1995). As the compositional data carry relative concentration, the increase in the concentration of some components results in a decrease in concentration for others. There are different ways to tackle this issue such as using the ratio of features rather than concentration or log-ratio transformation (Aitchison, 1982; Pearce, 1968). Among these closure-removal techniques, the centered log-ratio transformation appears to be the most efficient one for compositional data (for details readers are referred to Makvandi et al. (2016a). For the present study, centered log-ratio transformation was imposed on the calculated apfu values. As the secondary calculated parameters do not have such a closure problem and many of them have negative values (e.g., IV(Cl)), such transformation was not imposed on them. The calculated parameters and the log-ratio transformed apfu values were treated separately for the subsequent multivariate statistical analysis.

2.4. Statistical Methods: PCA and PLS-DA

PCA is a classical, multivariate, non-parametric, unsupervised, dimensionality reduction method that interprets the variability in a high-dimensional interrelated dataset. It reduces the high-dimensional interrelated data to low-dimension by linearly transforming the old variables into a set of new uncorrelated variables called principal component (PC) while retaining the maximum possible variance. We implemented PCA using Scikit-learn Python library (Pedregosa et al., 2011). Before PCA, we performed standardization of the initial datasets to remove biases in the initial variables. For example, if there are large differences in the ranges of initial variables, the variables with larger ranges will dominate over those with small ranges. This may lead to biased results. Mathematically, the standardization is done for each value of each variable by subtracting the mean and dividing it by standard deviation (i.e., *(value - mean) / standard deviation*). A scree test was conducted to determine the number of PCs to retain in our dataset. The scree plot (ESM 2 Fig. S) shows that the first four PCs accommodate >90 % variance in our dataset. The PCA was performed separately on apfu values and calculated parameters (Figs. 2–4 and ESM 2 Figs. S2-S4).

While PCA is a unsupervised method, PLS-DA is a supervised classification method that uses labeled data for predictive and descriptive modeling as well as for discriminative variable selection (Lee et al., 2018; Makvandi et al., 2016b). PLS-DA differs from PCA such that it preserves the maximum co-variance between observations and class labels in its first component which makes it an effective classification method. PCA, on the other hand, preserves the highest variance of the dataset in its first component, and therefore performs well in feature

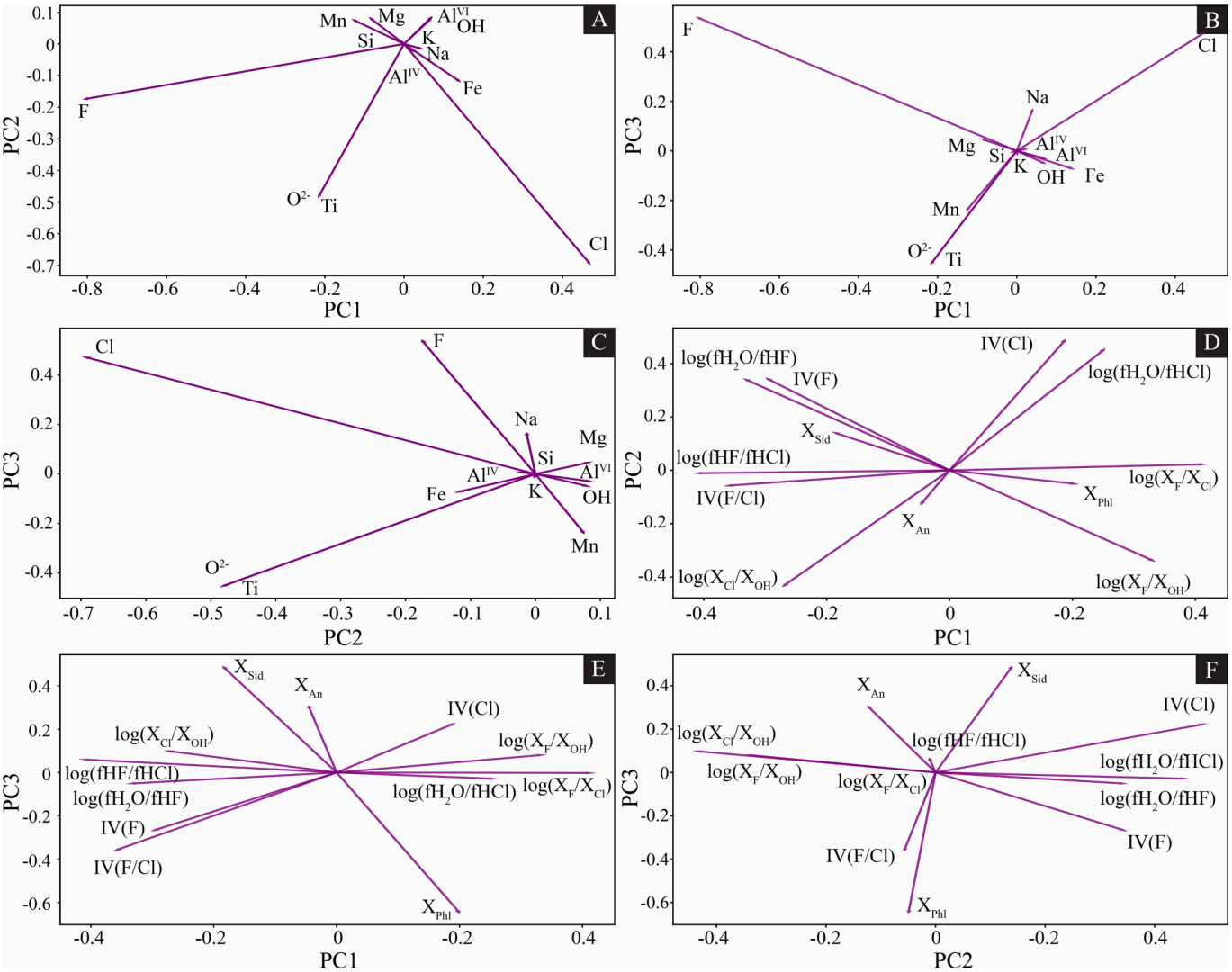

Fig. 2. PCA loading plots show the relationship between principal components and variables. A-C show relationships between apfu values and the first three principal components PC1, PC2, and PC3. D-F show relationships between calculated parameters and PC1, PC2, and PC3. Note that Cl, F, and Ti apfus have the largest arm in PC plots suggesting that they contribute the most for the variation in the dataset (A-C). Similarly, Fe, Mg, and Mn have a moderate while Si and K have very small contribution to the data set variation. D-F shows that halogen fugacity ratios of fluid ($\log(fH_2O/fHCl)^{fluid}$, $\log(fH_2O/fHF)^{fluid}$, and $\log(fHF/fHCl)^{fluid}$) and halogen intercept values (IV(Cl), IV(F), and IV(F/Cl)) have the largest contribution to the dataset variation. Arms of $\log(X_F/X_{Cl})$, $\log(fH_2O/fHCl)^{fluid}$, and IV(F/Cl) are the longest and parallel to PC1 suggesting that parameters related to F/Cl ratio of biotite define the maximum variation in the dataset. The structural parameters such as Xphl, Xsid, and Xan have comparatively lower contributions. Readers are referred to ESM 2 Fig. S1 that summarizes the loading of the apfu values and the calculated parameters on the first four principal components.

selection (Ruiz-Perez et al., 2020). PLS-DA sharpens the separation between groups of observations by rotating PCA components, which results in the maximum separation among classes, and the identification of the variables responsible for the separation of different classes (De Iorio et al., 2008). If there is a dataset of l samples, the data matrix X (input data; N x J matrix) that represents a set of j analytical measurements, and a vector c (output data; N x l) of length l that represents a numerical label for each sample according to its group membership. In such a case, the fundamental PLS-DA equations are:

$$X = TP + E$$

and

$$c = TQ + f$$

where T is called the score matrix, and E and f are called residuals. P and Q are the loading matrices of X and c. To calculate the PLS components, first, the weight vector (w) is estimated by maximizing the covariance between X and c from the following equation: $w = X' \text{ x } c$. Following that, X-scores (t), X-loading (p), and Y-loading (q) are determined sequentially by the following equations:

$$x = \frac{Xw}{\sqrt{w^2}}$$

$$p = \frac{t'X}{\sqrt{t^2}}$$

$$q = \frac{c't}{\sqrt{t^2}}$$

From these values, the first set of PLS components and loading is established. Then the residuals X (X^{res}) and c (c^{res}) of the first PLS component are calculated from the following equations:

$$X^{res} = X - Tp$$

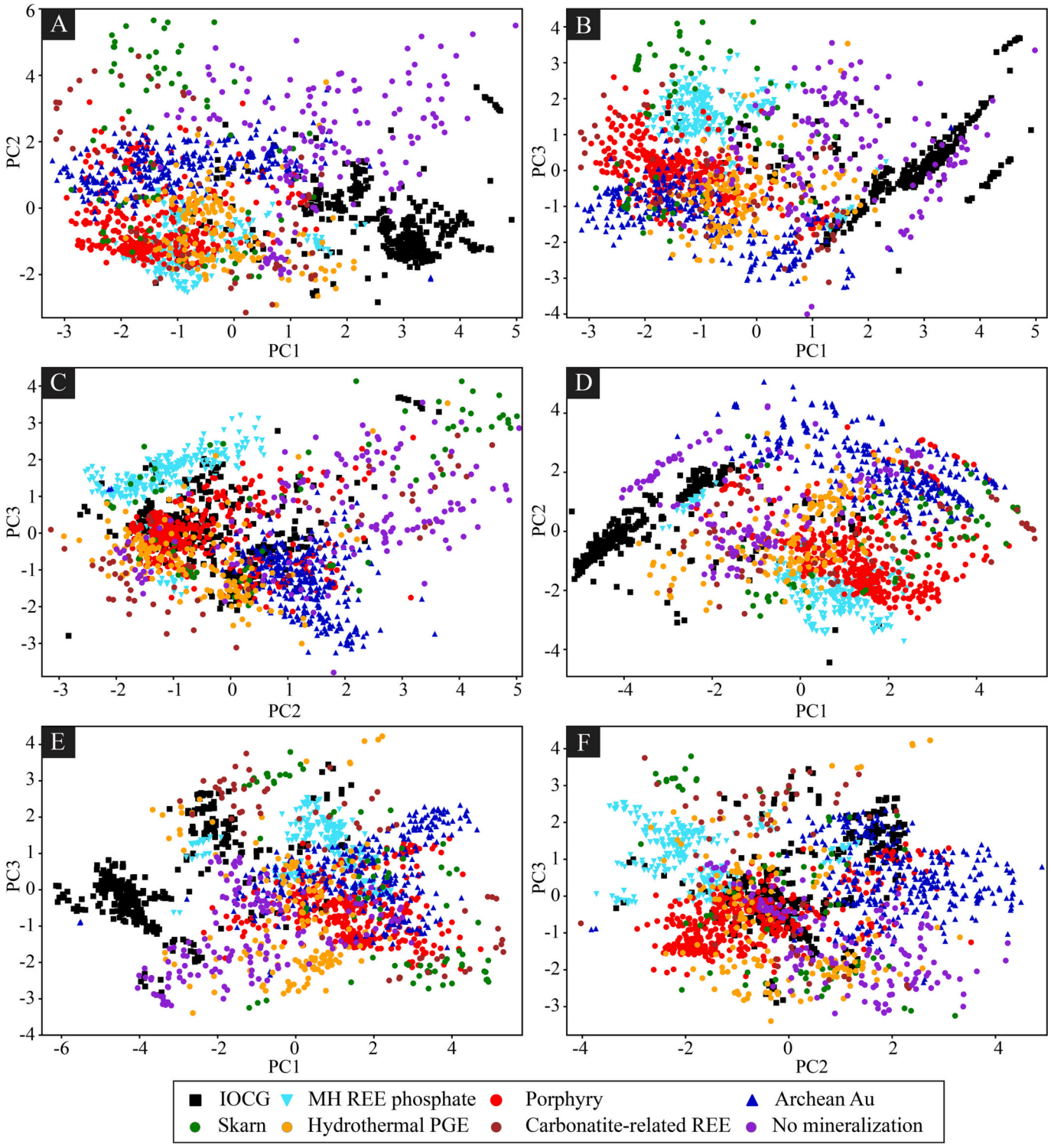

Fig. 3. PCA score plots of hydrothermal biotite from different deposit types. The result of PCA analysis performed with apfu values and calculated parameters are shown in A-C and D-F, respectively. Here, the scores of the first three principal components, namely, PC1, PC2, and PC3 are provided. IOCG group forms two subgroups where large and smaller subgroups are related to (meta-)evaporitic and magmatic-hydrothermal fluid sources, respectively. A shows the separation of IOCG, and B and C show the separation of Archean Au, porphyry, and MH REE deposits. In A, no-mineralization group can be separated to a certain degree from these four deposit types with some overlapping. In A-C, carbonatite-related REE, skarn, and hydrothermal PGE deposits cannot be distinguished from each other and the abovementioned deposit types due to a large variation in their PC scores. ESM 2 Fig. S2 these three deposit types are blurred for visualizing the grouping of other deposit types. D-F show similar pattern as in A-C.

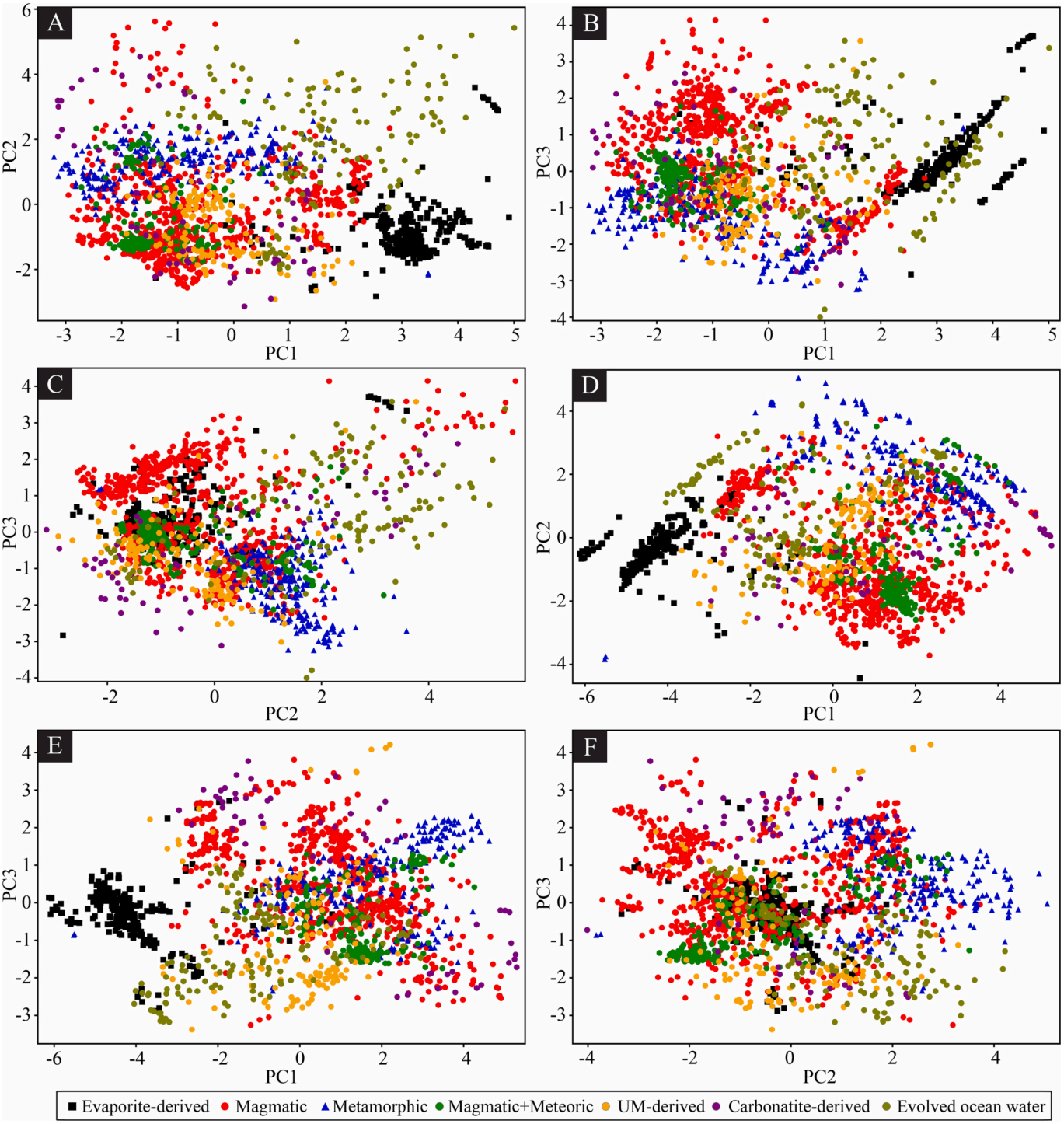

Fig. 4. PCA score plots of hydrothermal biotite genetically related to different fluid sources. The result of PCA analysis performed with apfu values and calculated parameters are shown in A-C and D-F, respectively. Scores of the first three principal components are provided. A-B show the separation of (meta-)evaporitic fluid group from magmatic and metamorphic fluid sources. B-C show the separation of metamorphic and magmatic fluid sources with some overlapping. A-C further shows magmatic+meteoric, carbonatite-derived, UM-derived, and evolved oceanic fluids cannot be distinguished from each other and the abovementioned three fluid sources. In ESM Fig. S3, these four fluid sources are blurred for better visualizing the grouping of the other three fluid sources. D-F show similar pattern as in A-C.

and

$$c^{res} = c - Tq$$

These residuals then become the input data (X) and output data (c), respectively, for constructing the second PLS component. For A numbers of PLS components, these calculation procedures are repeated A times, and those components are then used to construct the desired prediction model. For the prediction modeling, regression coefficient (b) is estimated using the values of w, p, and q by the following equation: $b = w(pw)^{-1}q$. Coefficient matrix B (J x A) is formed with these coefficient values (b). The test set (X^{test}) is reduced into the new dimensions (i.e. A PLS components) via B to produce the predicted values ($\hat{c}$), i.e. c^{pred}. If the training data belonged to T classes, the PLS-DA model would have produced T predicted values ($\hat{c} = \{y^1, y^2, y^T\}$) for each test sample

(X^{test}). For further details of the PLS-DA modeling, readers are referred to Brereton and Lloyd (2014) and Lee et al. (2018).

In this contribution, we used a combination of R and Python programming to generate the score and loading biplots, and the results are shown in Figs. 5–7 and ESM 2 Figs. S5-S6. In addition to these plots, PLS-DA is often used for the predictive-modeling purpose with high accuracy (Nguyen and Rocke, 2002; Tan et al., 2004). We developed our predictive modeling tool based on the PLS-DA using Python. For this, we divided our initial dataset into a training set (80 % data) and a validation set (20 % data), and a Machine Learning algorithm was developed using

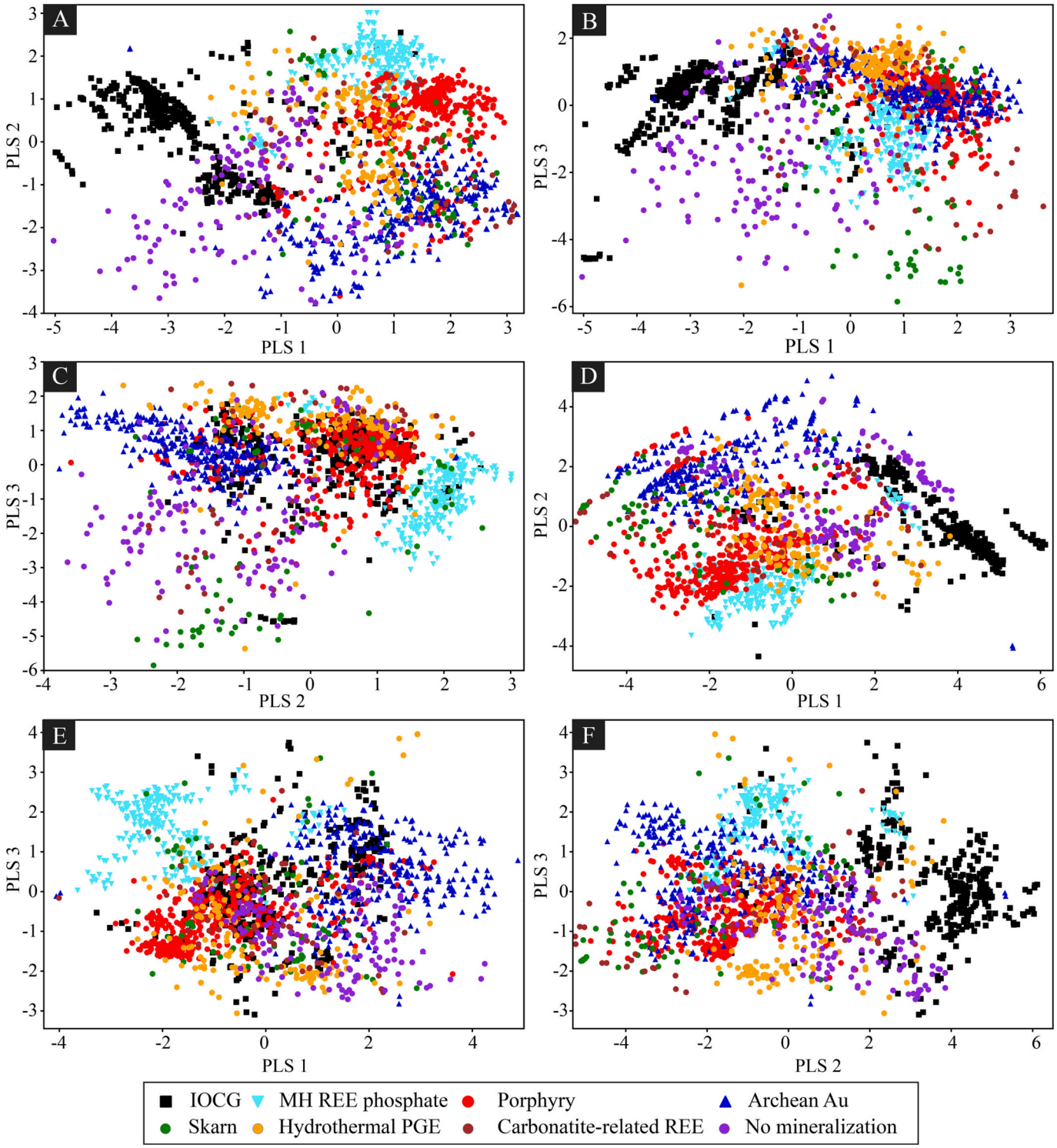

Fig. 5. PLS-DA score plots of hydrothermal biotite from different deposit types. Scores in A-C and D-F are based on apfu values and calculated parameters, respectively. The scores of the first three partial least square regression components are provided. Note that clustering of IOCG, Archean Au, Porphyry and MH REE deposits is better distinguished in PLS score plots based apfu values compared to that based on calculated parameters (A-C and D-F). Also, note that the clustering of these five deposit types is better here compared to PCA score plots in Fig. 2. Carbonatite-related REE, skarn, and hydrothermal PGE deposits cannot be distinguished from each other and the abovementioned deposit types. In ESM Fig. S4, these three deposit types are blurred for better visualizing the grouping of other five deposit types.

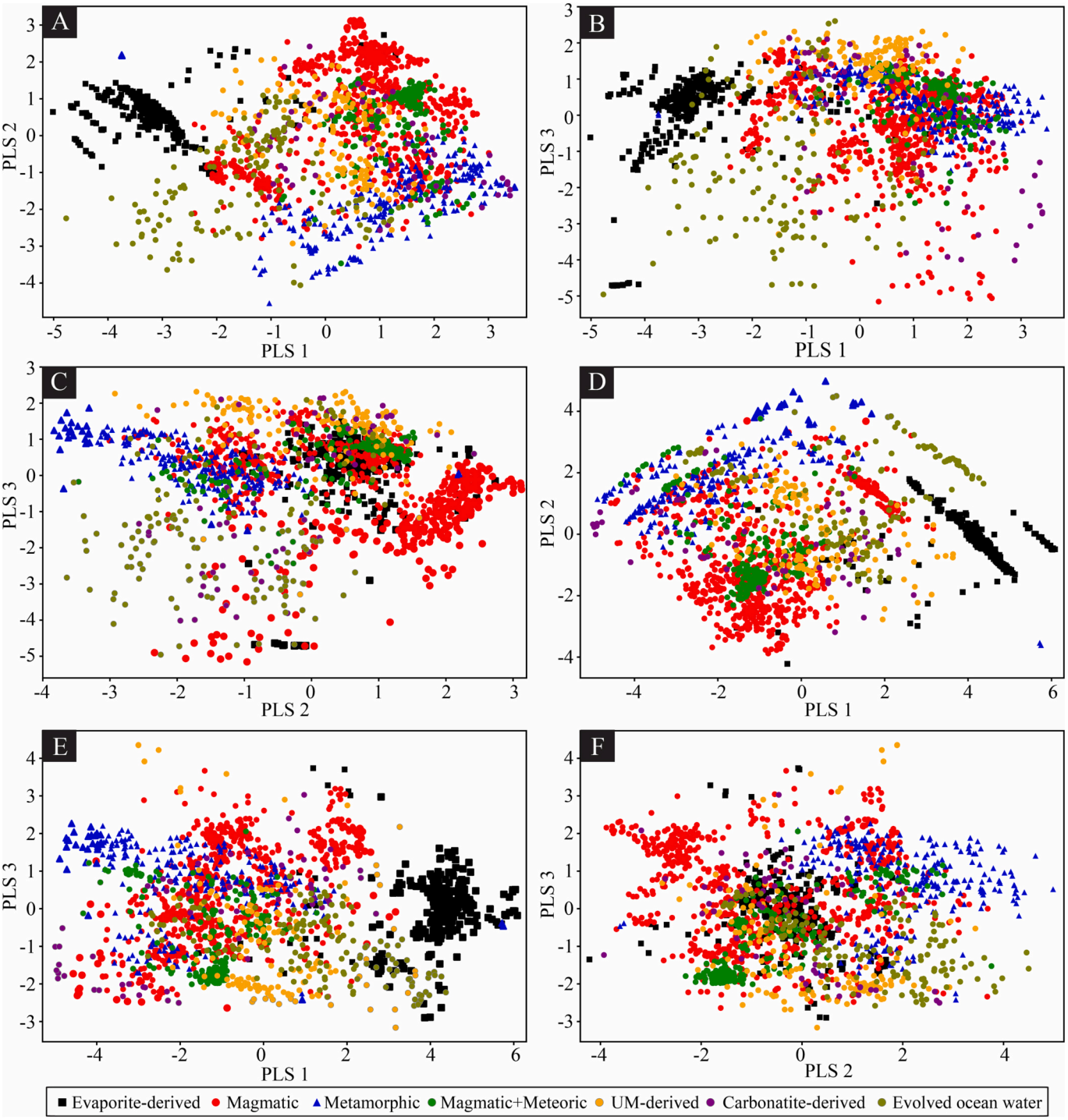

Fig. 6. PLS-DA score plots of biotite related to different fluid sources. A-C and D-F show the result of PLS-DA performed with apfu values and calculated parameters, respectively. Here, the scores of the first three partial least square regression components are provided. Note that (meta-)evaporitic, metamorphic, and magmatic fluid sources show different clusters with minor overlapping and the clustering of the aforementioned fluid sources is better distinguished in PLS score plots based apfu values compared to that based on calculated parameters (compare A-C and D-F). Also, note that the clustering of these fluid sources is better here compared to PCA score plots (compare with Fig. 2). In ESM 2 Fig. S3, four fluid sources are blurred for better visualizing the grouping of (meta-)evaporitic, metamorphic, and magmatic fluid sources.

the training set. The response matrix in the PLS-DA prediction model is a continuous matrix that cannot predict class names as text values. Therefore, to develop our prediction model, we converted the class names from the original data (X) to a binary "dummy" matrix. For our multiple class problems, X is ($n \times p$) the original input data matrix, Y is a factor vector of length n that indicates the class of each sample, and Y^* is the associated binary "dummy" matrix ($n \times K$). Here n is the total number of samples, p is the number of initial variables, and K is the number of classes. For each sample in the dummy matrix (Y^*), we assign value 1 for the column which represents the sample's true class and we assign 0 for the remaining columns. This way the dummy matrix (Y^*) internally records the membership of each observation (Rohart et al., 2017). In the next step, using Y^* as our response matrix we trained the PLS-DA algorithm for our training data set (80 %), and then it is

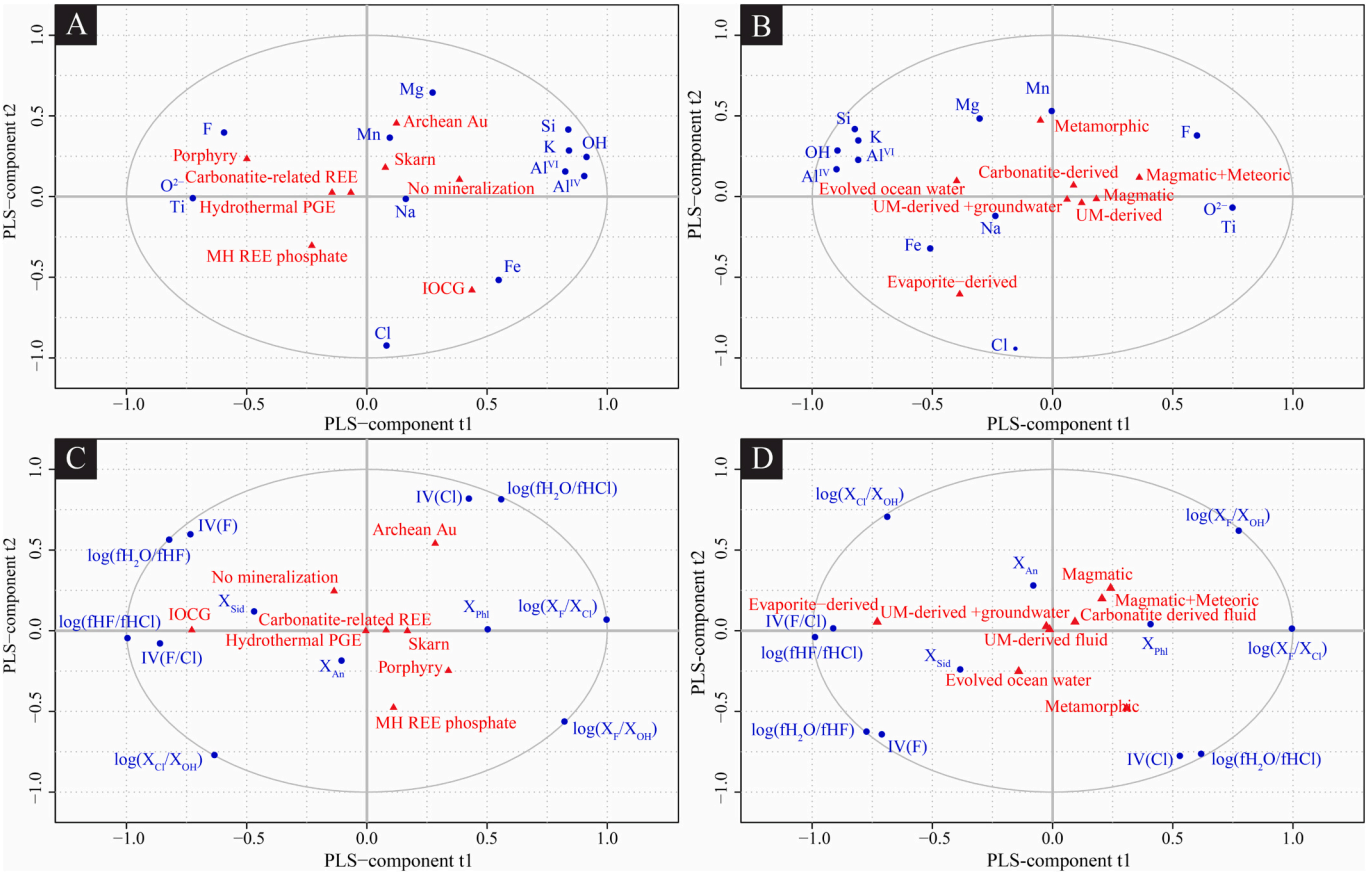

Fig. 7. PLS-DA loading plot showing the relationship between deposit types and fluid sources. Element or parameters which plot at the edge of the ellipse have contributed more in discriminating one class from the others. As an example, in C, it can be seen that halogen parameters (e.g., Cl and F intercept values) have contributed more compare to structural parameters (e.g., Xan) in discriminating deposit types. In A, IOCG, Porphyry, Archean Au, and MH REE deposits plot separately in four different quarters of the graph. A shows the close association of IOCG deposit and Fe and Cl, porphyry deposit and F, and Archean Au deposit and Mn and Mg. B shows the close association of metamorphic fluid and Mn and of (meta-)evaporitic fluid and Fe and Cl. C shows the close association between IOCG deposit and IV(F/Cl) and log(fHF/fHCl)^{fluid}, porphyry deposit and Xmg, and Archean Au deposit and IV(Cl) and log(fH$_2$O/fHCl)^{fluid}. D shows the close association between (meta-)evaporitic fluid and IV(F/Cl) and log(fHF/fHCl)^{fluid} and metamorphic fluid and IV(Cl) and log(fH$_2$O/fHCl)^{fluid}.

validated using the validation data set (20 %). The PLS-DA algorithm for the validation data set produces a response matrix as if it is a continuous one. As we assigned 1 for the true class and 0 for the other classes in defining the binary "dummy" matrix Y^*, in the response matrix of the validation dataset the column with values close to 1 will represent the true class. This PLS-DA classification trick works well in practice, as also demonstrated in many previous studies (Barker and Rayens, 2003; Boulesteix and Strimmer, 2007; Chung and Keles, 2010). Later on, the validation data was used to check how well the training algorithm performed. Furthermore, we collected some recently published data across the deposit types and fluid sources and used them as our test data to cross-validate the performance of our predictive model (Szymańska et al., 2012). Finally, we converted the Python code to a Windows© executable program, which can be used by others for predicting deposit types and fluid types on any new dataset.

3. Results

3.1. Results of PCA and PLS-DA analysis

The results of the PCA and PLS-DA based on the apfu values and calculated parameters are shown in Figs. 2–7 and ESM 2 Figs. S1-S6. Principal components PC1, PC2, and PC3 define 38 %, 25 %, and 17 % variation, respectively, of the dataset containing elemental apfu values accounting for 80 % variance of the dataset (ESM 2 Fig. S1). Among the compositional variables, F, Cl, Ti, and O^2 contribute dominantly to the PC1, PC2, and PC3 whereas Si, Na, K, and Al^{IV} have little influence (Fig. 2A-C). Iron, Mg, Mn, OH , and Al^{VI} have moderate contribution to these PC components where Fe, Mg, and Mn contribute more to the first three PC components compared to the other two (Fig. 2A-C and ESM 2 Fig. S2). On the other hand, when the calculated parameters are used for the PCA, the first three PCs yielded 50 %, 25 %, and 15 % variations, respectively, explaining 90 % variation of the dataset (ESM 2 Fig. S1). Log (X_F/X_{Cl}) and log(X_F/X_{OH}) with loadings >0.3, and log(fHF/fHCl)^{fluid} and IV(F/Cl) with loadings < 0.3 have the highest contribution to PC1 (Fig. 2D, E and ESM 2 Fig. S2). In addition, log(fH$_2$O/fHF)^{fluid} and IV(F) contribute negatively to PC1 to a significant degree (Fig. 2D, E). PC2 is characterized by positive contributions of log(X_{Cl}/X_{OH}) and log(X_F/X_{OH}) (loadings >0.3) while IV(Cl), log(fH$_2$O/fHCl)^{fluid}, IV(F) and log(fH$_2$O/fHF)^{fluid} contribute weakly to PC2 (Fig. 2D, F and ESM 2 Fig. S2). Xphl and Xsid have the highest positive and negative contributions, respectively, to PC3 (Fig. 2E, F).

3.1.1. Clustering of deposit types

In PC and PLS score plots, IOCG, Archean Au, porphyry, and MH REE groups form four different clusters with some overlapping (Figs. 3, 5 and ESM 2 Figs. S3, S5). A more robust distinction of these deposit types is obtained in PLS score plots than in PCA plots (Figs. 3A-C and 5A-C). The distinction of these groups is robust in the PLS score plots irrespective of input data i.e., apfu values and calculated parameters. For example, in PLS score plots based on apfu values, the IOCG group is characterized by negative PLS1 scores whereas the porphyry, Archean Au, and MH REE

groups have dominantly positive PLS1 scores (Fig. 5A, B). Among these three groups, Archean Au type has negative PLS2 scores and the other two types have positive scores with MH REE group having the highest positive PLS2 scores (Fig. 5B, C). Although the no-mineralization group has a wide range of PLS scores, it seems to form a different cluster and show a different trend from other groups (Fig. 5A, C). The separation of this group from others is more prominent when PLS-DA is done with apfu values compared to calculated parameters (Fig. 5A, D). Skarn and carbonatite-related REE groups show a large spread in PC and PLS score plots based on both apfu and calculated parameters and overlap with other groups and hence cannot be discriminated (Figs. 3, 5). Biotite related to PGE deposit have relatively restricted PC and PLS scores but overlap with other types in PC and PLS score plots and cannot be unambiguously classified (Fig. 3A-C).

The PLS-DA loading plots based on apfu values suggest that except for Na, other elements contribute significantly in discriminating the above-mentioned biotite groups (Fig. 7). Si, K, Al^{IV}, Al^{VI}, and OH plot far away from the origin in the loading plots suggesting that they contribute significantly in discriminating different groups (Fig. 7A). However, they plot far away from the biotite groups, hence, their contribution to a particular group cannot be quantified and compared (Fig. 7A). IOCG deposits are characterized by higher concentrations of Fe and Cl and lower concentrations of F (Fig. 7A). Archean Au and porphyry deposits are characterized by high and positive contribution from Mg and Mn, and from F and Ti, respectively (Fig. 7A). MH REE deposit type is characterized by negative contribution of Mg to the PLS2 loading (Fig. 7A). PLS-DA loading plot based on calculated parameters reveals that except for Xan, other parameters contribute significantly in discriminating different groups where halogen related parameters (e.g., $\log(fHF/fHCl)^{fluid}$ and IV(F/Cl)) have the highest contribution (Fig. 7C). Comparatively lesser contribution came from Xphl and Xsid (Fig. 7C). PLS-DA biplot also reveals that the IOCG group is distinguished from the rest by its unique $\log(fHF/fHCl)^{fluid}$ and IV(F/Cl) values, whereas the Archean Au system is different from others for its IV(Cl) and $\log(fH_2O/fHCl)^{fluid}$ values (Fig. 7C). Similarly, biotite from the porphyry systems is characterized by its unique Xphl (Fig. 7C).

3.1.2. Clustering of fluid types

Biotite types genetically related to evaporite-derived, metamorphic, and magmatic fluids, form distinguishable clusters in PC and PLS score plots based on apfu values and calculated parameters (Figs. 4, 6 and ESM 2 Figs. S4, S6). Biotite related to evolved ocean water fluid is also distinguishable to a significant extent from these three major types due to its different trend (Figs. 4A, 6A). The distinction among these fluid sources is better visualized in PLS score plot compared to PC score plots (Figs. 4A-C and 6A-C). Furthermore, the distinction of biotite related to evolved ocean water is better when apfu values are used for the PCA or PLS-DA compared to the calculated parameters. As an example of clustering, we note that in PLS score plot based on apfu values, biotite genetically related to evaporite-derived fluid has negative PLS1 scores while that related to magmatic and metamorphic fluids dominantly has positive PLS1 scores except for the magmatic-hydrothermal ones from IOCG system (Fig. 6A, B and ESM 2 Fig. S6). Magmatic-hydrothermal group dominantly has positive and the metamorphic group mostly negative PLS2 scores with some overlapping (Fig. 6B, C). Biotite crystallized from magmatic + meteoric fluid has similar PC and PLS scores as those that formed from magmatic fluid, therefore cannot be distinguished from each other (Fig. 4A-F, 6A-F). UM- and carbonatite-derived fluids groups are indistinguishable from other deposit types in these score plots (Fig. 4A-F).

The PLS-DA loading plot suggests that Si, Al^{VI}, Al^{IV}, OH, and K contributed significantly in discriminating different groups while the contribution of Na is small (Fig. 7B). These loading plots indicate that the unique position of the evaporite-derived fluid group in PLS score plot is primarily due to the positive contributions from Cl and Fe and the negative contribution from F (Fig. 7B). Similarly, among the parameters, IV(F/Cl) and $\log(fHF/fHCl)^{fluid}$ values contributed the most to discriminate this group from others. The loading plot further suggests that biotite related to metamorphic fluid is characterized by positive contributions from Mg and Mn while the evolved ocean water group is characterized by negative contribution of Ti (Fig. 7B). Similarly, among the parameters IV(Cl) and $\log(fH_2O/fHCl)^{fluid}$ values contributed the most in discriminating the metamorphic group from others.

4. Discussions

4.1. Choice of input data for multivariate statistical analysis: apfu vs. calculated parameters

In earlier studies, multivariate statistical analysis and machine learning algorithms have mostly been performed on major or trace element compositional data of minerals for classification/discrimination problems (Huang et al., 2019; Saha et al., 2021; Sciuba et al., 2020, 2021). In the present study, apfu values and calculated parameters are used separately for multivariate analyses and discrimination, and we evaluate whether calculated parameters are equally suitable for classification problems and which of the two inputs can classify the fluid sources and deposit types better.

The discrimination of IOCG, porphyry, Archean Au, and MH REE deposits, and no mineralization group and evaporitic, magmatic, metamorphic, evolved oceanic fluid sources in PC and PLS score plots based on apfu values are comparatively better than that based on the calculated parameters. For example, the separation of MH REE and porphyry deposits is better when the input data is elemental apfu values (compare Fig. 5A-C and 5D-F). Furthermore, the no mineralization group shows a separate trend compared to the other four groups in the PLS1 vs. PLS2 score plot when elemental apfu was used, which is not the case when calculated parameters were used (compare Fig. 5A and D). The case is similar for the fluid sources as well (compare Fig. 4A, C and 4D, F).

Intuitively, the better clustering in PCA and PLS-DA with apfu values compared to that with calculated parameters make sense. While calculating the parameters, some of the apfu values were not used, e.g., Na, K, Ti (and therefore, O^2), and Mn, among which, Ti and Mn have contributed significantly to differentiating certain groups. For example, evolved ocean water group is characterized by low Ti and the Archean Au group has strong positive correlation with Mn as evident from PLS loading plots (Fig. 7A, B). Consequently, the apfu values together contain more information than the parameters. On the other hand, for calculating the halogen fugacities of fluids, the temperature has been used, a parameter that has been collected separately from the apfu values for many cases. Therefore, the parameters have some intrinsic properties that are independent of apfu values. Hence, we conclude that rather than using apfu values and parameters separately, a combination of both could be used. Among the considered parameters and apfu values, some are highly correlated which is evident from the PCA and PLS-DA loading plots e.g., Al^{IV} and Si (e.g., Fig. 7A, D). While developing and executing machine learning-based predictive modeling, using only one of the highly correlated variables is recommended. In the present dataset, Al^{IV}, O^2 , Fe, Mg, $\log(X_{Cl}/X_{OH})$, $\log(X_F/X_{OH})$, $\log(X_F/X_{Cl})$, are highly correlated to Si, Ti, Xphl, Xphl, $\log(fH_2O/fHCl)^{fluid}$, $\log(fH_2O/fHF)^{fluid}$, and $\log(fHF/fHCl)^{fluid}$, respectively. Hence, we have discarded the former variables while developing the subsequent PLS-DA predictive modeling. In summary, it is suggested that Si, Al, Al^{VI}, Ti, Mn, Na, K, Cl, F, OH, Xphl, Xsid, Xan, IV(Cl), IV(F), IV(F/Cl), $\log(fH_2O/fHCl)^{fluid}$, $\log(fH_2O/fHF)^{fluid}$, and $\log(fHF/fHCl)^{fluid}$ maybe be used for better results in future statistical/machine learning models with biotite data.

4.2. Biotite chemistry as an indicator of deposit type and fluid sources

In this section, we first propose some binary and ternary diagrams, followed later by PLS-DA predictive models, which can be used to

differentiate deposit types and fluid sources based on biotite composition and then discuss about limitations of our model. It is to be noted that four deposit types (IOCG, porphyry, Archean Au, and MH REE) and three fluid sources (magmatic, evaporitic, and metamorphic) could only be classified based on the proposed binary and ternary classifiers. As the machine learning-based classification models usually provide better results than classical binary and ternary plots, we provide two PLS-DA models to classify deposit types, the no-mineralization group was considered in one model and discarded in the other. Similarly, for discriminating the fluid sources, the evolved ocean water group was considered in one PLS-DA predictive modeling and discarded in the other.

4.2.1. Standard binary and ternary plot-based classifiers

PLS-DA and PCA loading plots suggest impacts of F, Cl, and Ti among the apfu values, and halogen intercept values and fugacity ratios among the calculated parameters are the most significant for the discrimination of fluid sources and deposit types. In addition, these parameters contribute the most to the first three PC scores (Fig. 2). Therefore, these variables are responsible for majority of the variation in the dataset and should constitute the best classifiers. It is pertinent to note that apfu values such as Si, Al^{IV}, Al^{VI}, and OH have notable contribution in discriminating deposit types and fluid sources as evident from their distal position away from the center in the PLS-DA loading plots (Fig. 7A, B). However, the variation in these parameters is small as evident from the PCA loading plots (Fig. 2A-C). Therefore, these apfu values may not be suitable for developing binary and ternary classifiers. In the proposed binary and ternary plots, fields for individual deposit types and fluid sources were drawn by hand based on the distribution of data. In addition, we have provided the accuracy of each field.

It is proposed that the following three classifiers can differentiate four deposit types most efficiently: 1) IV(F/Cl) vs. $\log(fHF/fHCl)^{fluid}$, 2) Ti-Cl-F, and 3) IV(Cl) vs. $\log(fH_2O/fHCl)^{fluid}$ (Fig. 8). We emphasize that one plot may not discriminate all the deposit types, rather more than one of the proposed plots may be required to discriminate them. The first two plots can be used to separate the IOCG deposits from the other three types. The fields of different deposit types are given with respect to the trend for the natural hydrothermal biotite (Fig. 8A, B). As the IOCG deposits are very closely associated with IV(F/Cl) and $\log(fHF/fHCl)^{fluid}$ in PLS-DA loading plot but not with Cl, F, and Ti, it is suggested that the binary classifier (1) can distinguish the IOCG group better than the ternary Ti-Cl-F classifier (Fig. 8B, C). This ternary plot, however, is better capable of separating MH REE deposits from other deposits (Fig. 8C). It is to be noted that these two plots are not suitable for discriminating Archean Au and porphyry deposits from each other, and therefore are not recommended for the same (Fig. 8B, C). As the Archean Au group is located closer to IV(Cl) and $\log(fH_2O/fHCl)^{fluid}$ in the PLS-DA loading plot compared to the porphyry group, these two parameters may discriminate them from each other (Fig. 7C). Hence, the third classifier is chosen to distinguish them. This classifier, however, is not capable of separating IOCG and MH REE deposits from others. It is recommended that the three classifiers should be used in the suggested order while determining the deposit type based on biotite chemistry.

The PLS-DA and PCA score plots are obtained with all apfu values or parameters, which suggests that all parameters have a contribution to the observed clustering. Naturally, when two or three apfus/parameters are chosen as classifiers, the clustering is expected to be weaker with more overlapping between the different deposit/fluid source types. To quantify the degree of overlap and error limit of the proposed classifiers, we introduce a parameter called "Field Accuracy", which is expressed in % and is calculated by the following equation

$$\text{Field Accuracy } (\%) = 100 - \text{Outlier factor } (\%) - \text{Misclassification factor } (\%) \quad (10)$$

Here, the outlier factor (%) of the field/group under consideration is defined by: Outlier factor (%) = (number of data plot outside the field/total number of data in this group) * 100. On the other hand, misclassification factor (%) comes from other groups and is defined by: Misclassification factor (%) = (number of data of another group plotted inside the proposed field/total number of data of that group) * 100. Therefore, the misclassification factor may have more than one value, i.e., a value for each group other than the one whose classification accuracy is being measured. Field accuracy can have the highest value of 100, and in general, the higher the value, the better is the classifier. For example, in the Cl-F-Ti plot, the outlier factor and misclassification factors for the IOCG field are 6 % and 0.6 % (related to Archean Au field), 1.1 % (related to MH REE field) and 2.6 % (related to porphyry field), respectively. The field accuracy for the IOCG group in this plot, therefore, is 89.5 % (i.e., 100–6.0 - 0.6 - 1.1 - 2.6 = 89.5). Similarly, for all the classifiers, outlier factor, misclassification factors, and field accuracy of all the deposit types have been calculated and are listed in Table 1. In general, the accuracy of proposed fields in all three plots is higher than 83.5 %.

Similar to the deposit type, the following four plots are recommended for discriminating fluid sources based on biotite composition: 1) X_{Mg} vs. $\log(fHF/fHCl)^{fluid}$, 2) IV(F/Cl) vs. $\log(fHF/fHCl)^{fluid}$, 3) IV(Cl) vs. $\log(fH_2O/fHCl)^{fluid}$, and 4) Cl-F-Ti (Fig. 9). Here also, the fields are drawn with respect to the trend of natural hydrothermal biotite (Fig. 9A). The first two plots can discriminate (meta-)evaporitic fluids from magmatic-hydrothermal and metamorphic fluids with an accuracy of 94.3 and 92.7 %, respectively (Table 1, Fig. 9B-C). This plot, however, is unable to discriminate metamorphic and magmatic-hydrothermal fluid sources. For that purpose, the classifier (3) is recommended (Fig. 9E, F). The accuracy of metamorphic and magmatic fields in this classifier are 87.1 and 86.5 % respectively (Table 1). The fourth classifier is capable of discriminating these three fluid sources, however, the accuracy of magmatic and metamorphic fields is comparatively lower at 72.3 and 79.2 % respectively (Fig. 9D).

4.2.2. Predicting deposit types and fluid sources using Windows© program

For the PLS-DA predictive modeling, the combination of apfu and calculated parameters discussed in Section 4.1 is used as variables. Our PLS-DA model, developed for predicting five deposit types, furnishes accuracy [= (TP + TN)/(TP + TN + FP + FN)] of 83 % and precision [= TP/(TP + FP)] of 86 %, 82 %, 74 %, 93 %, and 87 % for IOCG, Porphyry, Archean Au, MH REE, and no-mineralization biotite groups, respectively (where TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative) when applied on validation dataset. However, instead of five if we want to predict four deposit types, the accuracy of our validation dataset results increases to 92 % and precision to 96 %, 90 %, 84 %, and 97 % for IOCG, porphyry, Archean Au, and MH REE deposit types, respectively. The lower accuracy and precision in the 5-deposit model compared to the 4-deposit model is due to the consideration of the no-mineralization group, which has some overlap with the other four groups as evident from PLS-DA score plots (Fig. 5, ESM 2 Fig. S5). Similarly, in PLS-DA predictive model, developed for classifying four fluid source types, the validation data gives accuracy of 80 % and precision of 87 %, 82 %, 83 %, and 65 % for evaporitic, evolved-oceanic, magmatic, and metamorphic fluid sources, respectively. However, when we consider three fluid sources, the accuracy increases to 88 % and precision to 95 %, 89 %, and 77 % for evaporitic, magmatic, and metamorphic fluid sources, respectively. It is to be noted that the Windows© program based on these models provide results in the form of a deposit type or fluid source category for each input data.

4.2.3. Limitations of the proposed classification model

There are a few limitations of our model that are related to intrinsic property of the dataset. First, the number of data for the deposit type and fluid source groups is not "perfectly balanced" (Bellinger et al., 2020; Hamid et al., 2022). Deposit types that are separable from each other i.e., IOCG, MH REE, porphyry, and Archean Au deposits, have a data number ratio of 1.53: 1: 1.66: 1.22. Similarly, among the fluid sources,

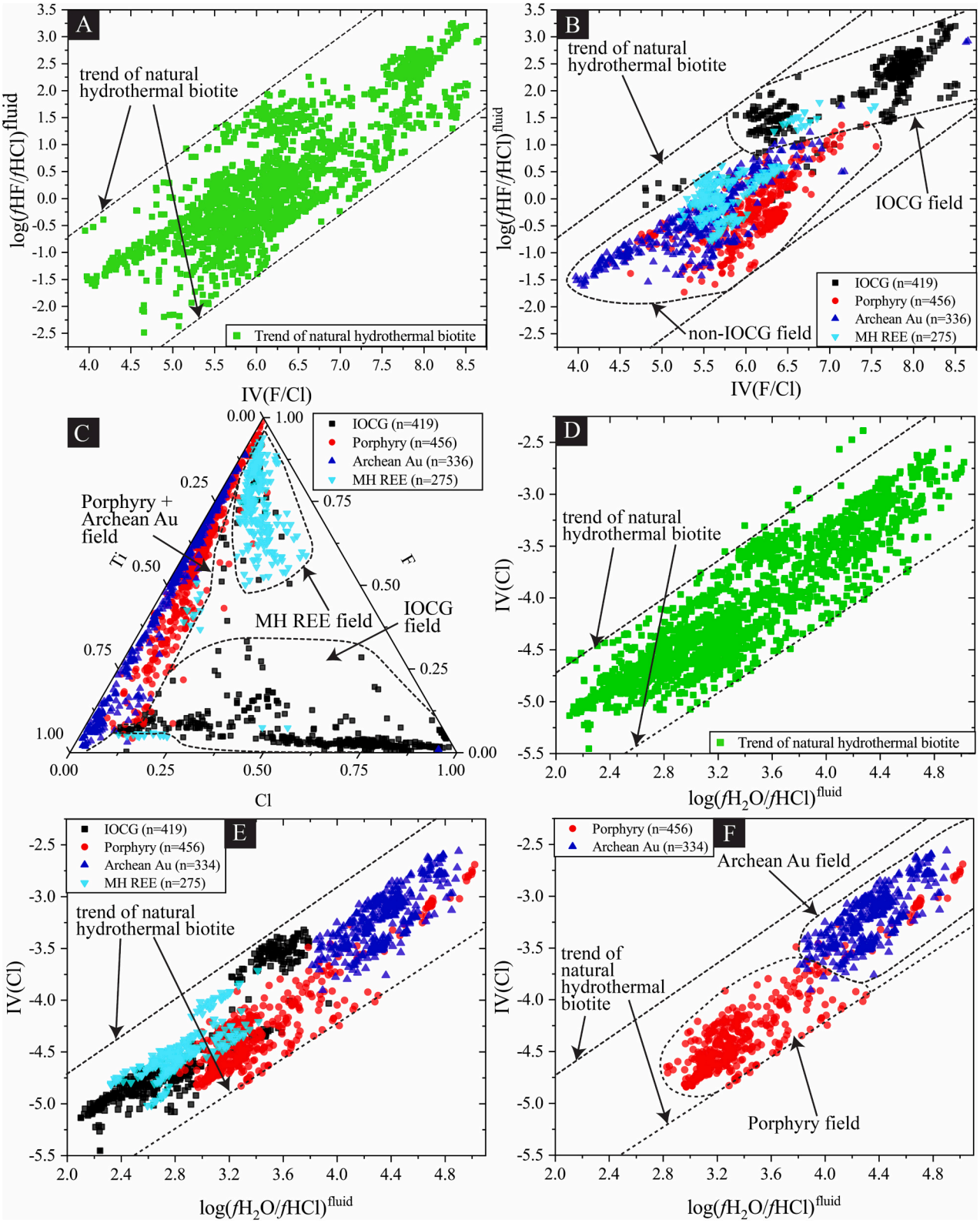

Fig. 8. Binary and ternary classifiers of deposit types. Fields of each deposit type are drawn by hand based on the distribution of the data. In A, the natural trend of hydrothermal biotite is shown in the $\log(fHF/fHCl)^{fluid}$ vs. IV(F/Cl) plot. In B, the $\log(fHF/fHCl)^{fluid}$ vs. IV(F/Cl) plot is capable of separating IOCG deposit from non-IOCG (i.e., porphyry, Archean Au, and MH REE deposits, which cannot be separated from each other in this diagram). In C, IOCG and MH REE deposits are separated from porphyry and Archean Au deposits, where the latter two types cannot be separated from each other. In D, the natural trend of hydrothermal biotite are shown in IV(Cl) vs. $\log(fHF/fHCl)^{fluid}$ plot. E shows that this classifier cannot discriminate IOCG and MH REE deposits. In F, fields of the Archean Au and porphyry deposits are shown.

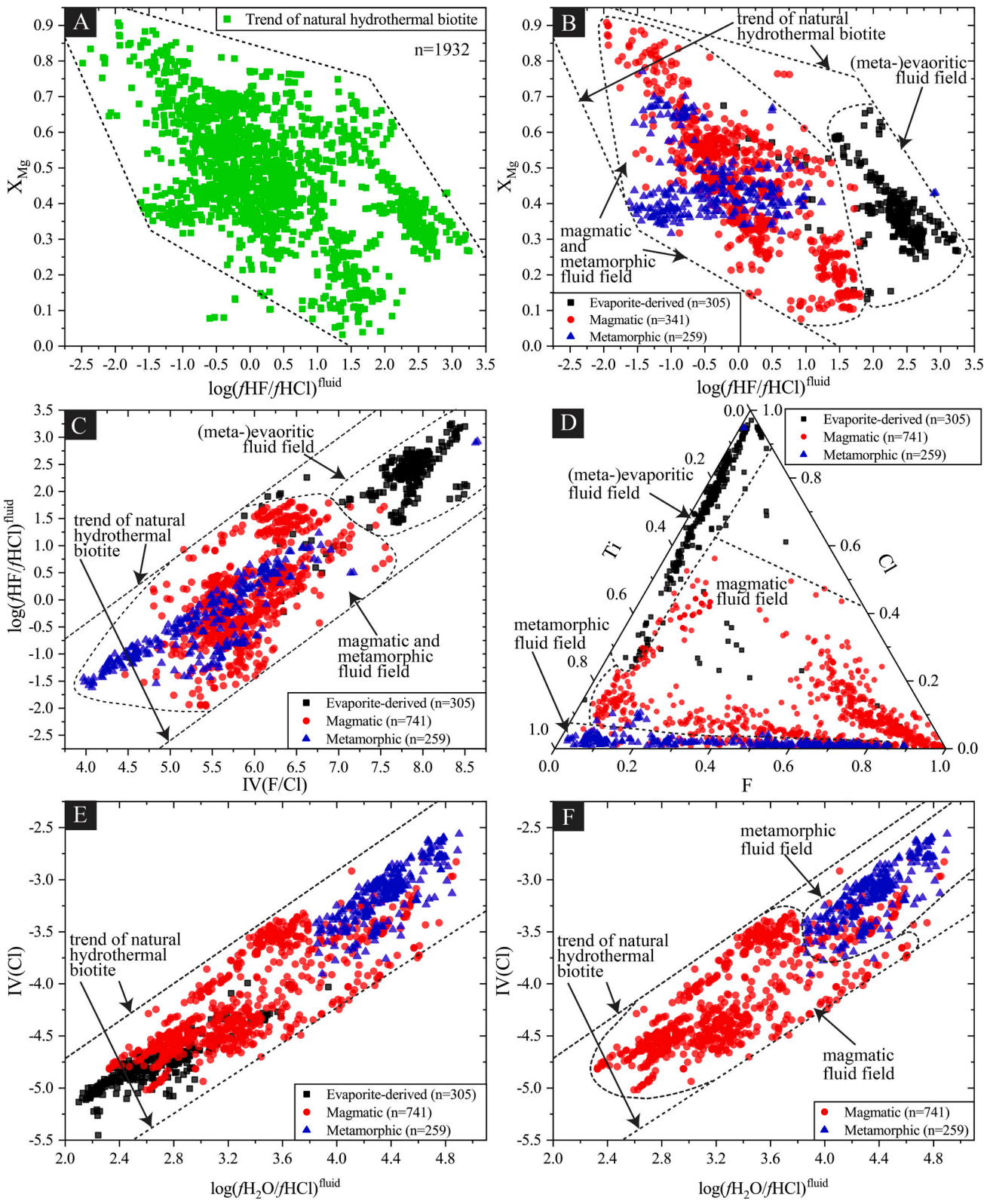

Fig. 9. Binary and ternary classifiers of fluid sources. Fields of each fluid source are drawn by hand based on the distribution of the data. All fiends are given with respect to the natural trend of hydrothermal biotite. Table 2 has information on the accuracy of the proposed fields. In A, the natural trend of hydrothermal biotite is shown in the XMg vs. $\log(fHF/fHCl)^{fluid}$ plot. B-C shows that XMg vs. $\log(fHF/fHCl)^{fluid}$ and $\log(fHF/fHCl)^{fluid}$ vs. IV(F/Cl) plots are capable of separating meta-evaporitic fluids from non-evaporitic fluid sources (i.e., magmatic and metamorphic fluids). In D, magmatic, metamorphic, and meta-evaporitic fluid sources can be separated from each other. In E, the IV(Cl) vs. $\log(fHF/fHCl)^{fluid}$ plot is not capable of separating magmatic and (meta-)evaporitic fluid sources. In F, fields of magmatic and metamorphic fluid sources are shown.

our model could separate magmatic, meta-evaporitic, and metamorphic sources, which have 741, 305, and 259 numbers of data, respectively, with a ratio of 2.86: 1.18: 1. Such ratios are considered "slightly imbalanced" (Bellinger et al., 2020; Hamid et al., 2022). As most of natural datasets are "slightly imbalanced" but suitable for statistical analysis, we think that this issue may not impact our model significantly. Second, in the dataset, data from a locale or a few locales often dominate over other locales under a category. For example, in IOCG category, most of the data belongs to Khetri Cooper Belt and Carajás Province while data belonging to MH REE group is based on one deposit in Sweden. In such cases, the proposed field may be a bit biased. At this moment, there is not a way to resolve this, and it can be resolved when new data will be published making it possible to give equal weight to all the locales. To address these issues, we have left our PLS-DA predictive model "open" to incorporate future data. In our model, it is possible to change the training set and we suggest to update the training set with new data (see below for guidance), which is expected to make the classification more robust.

4.3. Testing of the proposed classifiers and predictive modeling

To test the proposed binary and ternary classifiers and PLS-DA predictive models, some recently-published biotite major oxide data (n = 470) from IOCG, porphyry, Archean Au deposits are collected and discriminated (Chen and Zhou, 2015; Fontaine, 2019; Katz et al., 2020; Khosravi et al., 2021; Liu et al., 2020; de Souza et al., 2021). It is to be noted that these data were not included in the main dataset that was used to build the classifiers or the PLS-DA model. This dataset is named as test dataset hereafter. The fluid sources were interpreted/inferred in most of these studies by independent methods, e.g., in the Lala IOCG deposit, stable oxygen isotope data suggest magmatic-hydrothermal fluid sources related to biotite types (Chen and Zhou, 2012, 2015). It has been noted that biotite in these deposits crystallized either from magmatic or metamorphic fluids. We have not found any additional data from MH REE deposits and (meta-)evaporitic sources, therefore they could not be tested. The imputation of BDL values was done by DL/$\sqrt{2}$ and the biotite formula was calculated based on 24 anions as described in Section 2.3. The temperature data is available in some papers (e.g., Chen and Zhou, 2015) and calculated for other biotite grains using the method of Wu and Chen (2015) assuming the equilibrium between hydrothermal fluid and biotite. Halogen fugacities of fluid were calculated using this temperature. Other parameters were calculated following the method described in Section 2.3.

4.3.1. Binary and ternary classifiers

The results of the testing exercise with proposed binary and ternary classifiers are shown in Fig. 10 and ESM 2 Fig. S7. Among the deposit type classifiers, the proposed Cl-F-Ti ternary plot could not differentiate these three deposit types properly. Most of data from Archean Au and IOCG deposits fall in the respective fields but most of the porphyry data falls in the MH REE field (Fig. 10A, ESM 2 Fig. S7). It is noted that the IOCG test data from the Lala IOCG deposit is characterized by higher concentrations of F relative to other IOCG deposits due to which some of the data points plot outside the proposed IOCG field (Fig. 10A). We observe that biotite related to magmatic-hydrothermal fluids consistently contain F in variable proportions, irrespective of deposit type e.g., biotite from porphyry, skarn, and MH REE deposits have significant to high concentrations of F (Andersson et al., 2018; Golmohammadi et al., 2015; Iveson et al., 2016). The Lala IOCG deposit is considered to have formed from magmatic-hydrothermal fluid without any contribution from evaporitic sources (Chen and Zhou, 2012), which may explain the higher concentration of F in these biotite grains, and therefore the plotting of some points outside the proposed field. The data from the porphyry group dominantly falls in the MH REE field (Fig. 10A). The reasons are yet to be explored. One possible reason could be that the formation of porphyry and MH REE deposits involve similar processes e.g., involvement of magmatic-hydrothermal fluids derived from granitoid magma and similar temperature range. We note that further data from MH REE deposits are required to develop a better classification model to discriminate porphyry and MH REE deposits. In contrast to the Cl-F-Ti classifier, the binary plots are capable of identifying deposit types efficiently (Fig. 10B-C). As expected, IV(F/Cl) vs. $\log(f\mathrm{HF}/f\mathrm{HCl})^{\mathrm{fluid}}$ classifier could separate IOCG deposits from non-IOCG porphyry and Archean Au deposits (Fig. 10B). Subsequently, porphyry and Archean Au deposits could be separated in IV(Cl) vs. $\log(f\mathrm{H_2O}/f\mathrm{HCl})^{\mathrm{fluid}}$ classifier (Fig. 10C). It is noted that there are some outliers in all the cases, however, the bulk of the data plot in the expected fields. In summary, the proposed classifiers, particularly the binary ones are together able to discriminate IOCG, porphyry, and Archean Au deposits efficiently. We think that the ternary classifier is also useful but it may need some modification with new data from MH REE deposits.

As with the deposit types, the fluid sources could also be differentiated using the proposed classifiers. Except for a few outliers, the ternary Cl-F-Ti plot could differentiate metamorphic and magmatic-hydrothermal fluids with high precision and accuracy (Fig. 10D). Both IV(F/Cl) vs. $\log(f\mathrm{HF}/f\mathrm{HCl})^{\mathrm{fluid}}$ and Xphl vs. IV(F/Cl) vs. $\log(f\mathrm{HF}/f\mathrm{HCl})^{\mathrm{fluid}}$ classifiers could separate the metamorphic and magmatic-hydrothermal fluid sources from the (meta-)evaporitic source (Fig. 10E and ESM 2 Fig. S7). Finally, the IV(Cl) vs. $\log(f\mathrm{H_2O}/f\mathrm{HCl})^{\mathrm{fluid}}$ classifier was able to discriminate metamorphic and magmatic-hydrothermal fluid sources (Fig. 10F). To summarize, though there are some outliers, the proposed classifiers may be used to discriminate (meta-)evaporitic, metamorphic, and magmatic-hydrothermal fluid sources.

4.3.2. PLS-DA predictive modeling and Windows© program

We observed that the precision and accuracy of the test data is comparatively lower than the validation data. It is to be noted that we used a "double-check" cross validation following the approach of Szymańska et al. (2012), where model optimization and assessment of model quality are conducted independently. As mentioned before, we initially divided the dataset into training set (80 %) and validation set (20 %). Training set was then used to generate the model and the validation set is used for validation and optimization. The model optimization phase involves selection of the optimal number of principal components (or, more correctly, latent variables in PLSDA). We used the insights gained from PCA analysis in choosing the number of principal components in PLSDA prediction model. To evaluate our model, we use the new test dataset (note that this dataset is not a part of training or validation set) and derive the confusion matrix, which consists of the number of False Positives, False Negatives, True Positives and True Negatives (Westerhuis et al., 2008).

When the PLS-DA predictive model generated integrating four deposit types were implemented on the test data set, a prediction accuracy of 74 % and precision of 68 %, 94 %, and 78 % for IOCG, porphyry, and Archean Au deposit types, respectively are obtained. The lower accuracy and precision of the test dataset may be due to the following: 1) F-enriched nature of IOCG test data from the Lala deposit and 2) similar composition of biotite from porphyry (test data) and MH REE deposits (Fig. 10A; Section 4.3.1). Similarly, applying the PLS-DA predictive model generated for three fluid source types, we find prediction accuracy of 82 % and precision of 93 % for both magmatic and metamorphic fluid sources. Therefore, our model and the Windows© Program that executes it can efficiently discriminate fluid sources.

4.4. Guide to using the proposed classifiers and models

A general guidance for using Windows Application and screen shots of the Application layout are provided in ESM 3. At first, the missing values particularly the "below detection limit" values should be imputed by DL/$\sqrt{2}$. The 'not analyzed' or 'not available' values may be imputed by the statistical method implemented here or by some other methods.

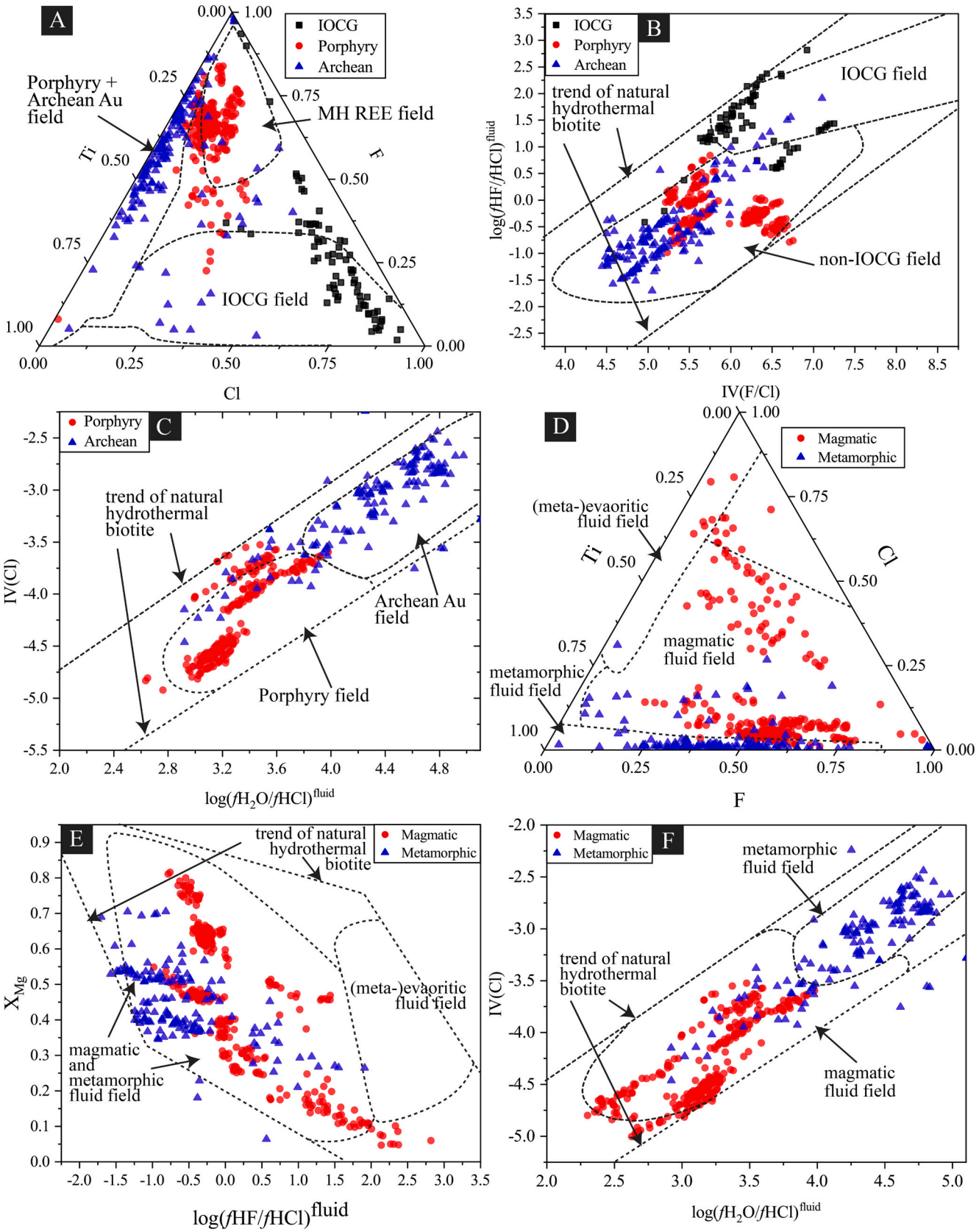

Fig. 10. Testing of the proposed binary and ternary classifiers. In A, new data from the Archean Au and IOCG deposit fall in the right field with a few outliers but data from porphyry deposits fall outside of the expected field. Classifiers in B and C together can discriminate IOCG, Archean Au, and porphyry deposits effectively. In E, biotite data related metamorphic and magmatic fluids fall in the "non-evaporitic" field as expected. The classifiers in D and E are capable of separating magmatic and metamorphic fluid sources efficiently.

Once the complete EPMA dataset is prepared, the formula (i.e., apfu values) should be calculated based on 24 anions. After that, the parameters should be calculated as described in Section 2.3. It is to be noted that information on the crystallization temperature of biotite is required for calculating some of the parameters (e.g., halogen fugacity ratios of fluid). Then the data should be plotted in the proposed binary and ternary classifiers in the recommended sequence to infer the deposit type and fluid sources. Before using the Windows© applications, the 18 variables mentioned at the end of Section 4.1 should be calculated. These variables should be arranged in the same sequence as shown in the *Test_Data.csv* file in ESM 3. Then, the *Test_Data.csv* should be replaced with the new test data for which the user wants to know deposit type and fluid source. It is to be noted that the new dataset should be renamed as "Test_Data.csv". The user can then run the Windows© executable (e.g., *PLSDA_TestData_Prediction_Deposit_Type_4_Groups.exe*) of the predictive model that they want to apply, and it will generate a corresponding .xlsx file as an output data (e.g., *Prediction_using_4_Groups_of_Deposit_Type_Model.xlsx*), which will contain predicted class names against each sample. Interestingly, in this program, users can also update the training data set as well. In other words, with the availability of more hydrothermal biotite data from future publications, the training data could also be modified. To do that, the users should format their data in the order discussed above and replace the present training dataset (e.g., *Deposit_Type_4_Groups.csv*) with the new training data set, keeping the name of the *.csv* file same as in the original folder named *Windows Executables*. It is to be emphasized that if there is a preliminary possibility that an unknown deposit type doesn't belong to the no-mineralization group i.e., not related to ocean floor alteration, it is recommended to use the Windows© application made for 4 deposit types. This would provide better accuracy and precision in the output. Similarly, if the preliminary data suggest the possibility of the absence of evolved ocean water or the presence of magmatic, (meta-)evaporitic, or metamorphic fluids in an unexplored area, the Windows© application with 3 fluid sources is recommended.

5. Conclusions and future scopes

In this work, an attempt is made to evaluate if biotite major element chemistry could be used as a discriminator of deposit types and fluid sources. Multivariate statistical analysis including PCA and PLS-DA conducted with 1932 data suggests that biotite major element chemistry can be used for discriminating IOCG, porphyry, Archean Au, and MH REE deposits and magmatic, (meta-)evaporitic, and metamorphic fluid sources with high precision. The following three classifiers are particularly important to discriminate deposit type: 1) Ti-Cl-F, 2) IV(F/Cl) vs. $\log(fHF/fHCl)^{fluid}$, and 3) IV(Cl) vs. $\log(fH_2O/fHCl)^{fluid}$. The discrimination between MH REE and porphyry deposits remains to be improved. On the other hand, the following four plots are useful in discriminating fluid sources: 1) X_{Mg} vs. $\log(fHF/fHCl)^{fluid}$, 2) IV(F/Cl) vs. $\log(fHF/fHCl)^{fluid}$, 3) IV(Cl) vs. $\log(fH_2O/fHCl)^{fluid}$, and 4) Cl-F-Ti. Upon testing with recently published data, it is observed that these classifiers are efficient. In addition to these standard binary and ternary classifiers, 4 different Windows© programs based on the PLS-DA predictive modeling are provided, which can also be used for predicting fluid sources and deposit types. In two of these programs, biotite chemistry could be used to predict ocean floor alteration (i.e., no-mineralization group) settings and evolved ocean fluids. However, the accuracy and precision of these two models are moderate. It is observed that hydrothermal biotite related to REE (associated with carbonatite), hydrothermal PGE, and skarn deposits have wide compositional range and cannot be discriminated using the present model. In case of fluid sources, biotite major element chemistry may not be useful for discriminating mafic/ultramafic rock-derived, carbonatite-derived, magmatic + meteoric fluids. In future, major and possibly trace element chemistry of other minerals could also be used for this purpose. In this study, PLS-DA-based predictive modeling was used and it is noted that more advanced predictive models such as decision tree, artificial neural network, and support vector machine may be attempted in the future. This study noted major differences in halogen signature across deposit types and fluid sources, however, reasons for such differences are yet to be explored.

CRediT authorship contribution statement

Abu Saeed Baidya: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Validation, Visualization, Writing – original draft, Writing – review & editing. **Giridas Maiti:** Formal analysis, Methodology, Software, Validation, Writing – original draft, Writing – review & editing. **Subhronil Mondal:** Formal analysis, Methodology, Software, Validation, Visualization, Writing – review & editing. **Dewashish Upadhyay:** Project administration, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

We have provided the data in the manuscript.

Acknowledgments

This work is a part of the first author's research as a national post-doctoral fellow (NPDF; Ref. No.: PDF/2020/000938), Science and Engineering Research Board, Govt. of India). The authors acknowledge the help and insight of Eileena Das during data compilation, biotite formula calculation, R-programming development, and manuscript writing. The authors also acknowledge those who shared their data, without which this work would not have been possible. Comments and suggestions from two anonymous reviewers have improved the manuscript significantly. We acknowledge Prof. Stefano Albanese and Dr. Pooria Ebrahimi for editorial handling of the manuscript.

References

Afshooni, S.Z., Mirnejad, H., Esmaeily, D., Haroni, H.A., 2013. Mineral chemistry of hydrothermal biotite from the Kahang porphyry copper deposit (NE Isfahan), Central Province of Iran. Ore Geol. Rev. 54, 214–232. https://doi.org/10.1016/J.OREGEOREV.2013.04.004.

Aitchison, J., 1982. The Statistical Analysis of Compositional Data. J. R. Stat. Soc. Ser. B 44, 139–160. https://doi.org/10.1111/j.2517-6161.1982.tb01195.x.

Andersen, T., Austrheim, H., 1991. Temperature-HF fugacity trends during crystallization of calcite carbonatite magma in the Fen complex, Norway. Mineral. Mag. 55, 81–94. https://doi.org/10.1180/minmag.1991.055.378.07.

Andersson, S.S., Wagner, T., Jonsson, E., Fusswinkel, T., Leijd, M., Berg, J.T., 2018. Origin of the high-temperature Olserum-Djupedal REE-phosphate mineralisation, SE Sweden: a unique contact metamorphic-hydrothermal system. Ore Geol. Rev. 101, 740–764. https://doi.org/10.1016/j.oregeorev.2018.08.018.

Ayati, F., Yavuz, F., Noghreyan, M., Haroni, H.A., Yavuz, R., 2008. Chemical characteristics and composition of hydrothermal biotite from the Dalli porphyry copper prospect, Arak, central province of Iran. Mineral. Petrol. 94, 107–122. https://doi.org/10.1007/s00710-008-0006-5.

Baidya, A.S., Das, E., 2022. "Calculating biotite formula from electron microprobe analysis data using a machine learning method based on principal components regression" by Li et al. (Lithos 356–357 (2020) 105371)-A discussion. Lithos 412–413C.

Baidya, A.S., Pal, D.C., Upadhyay, D., 2021a. Biotite chemistry and mineral association as an indicator of redox conditions in the iron oxide Cu-Au (IOCG) system: Constraints from the Khetri Copper Belt, western India. Ore Geol. Rev. 139, 1–6. https://doi.org/10.1016/j.oregeorev.2021.104544.

Baidya, A.S., Sen, A., Pal, D.C., Upadhyay, D., 2021b. Ore-forming processes in the Khetri Copper Belt, western India: constraints from trace element chemistry of pyrite and C-O isotope composition of carbonates. Mineral. Deposita 56, 957–974. https://doi.org/10.1007/s00126-020-01018-z.

Baidya, A.S., Saha, R., Pal, D.C., Upadhyay, D., 2023. Fingerprinting Alteration and Mineralization in the iron Oxide cu-au (IOCG) System Using Biotite Chemistry and Monazite Geochronology: Constraints from the Khetri Copper Belt, Western India. Miner. Depos. https://doi.org/10.1007/s00126-023-01190-y.

Barker, M., Rayens, W., 2003. Partial least squares for discrimination. J. Chemom. 17, 166–173. https://doi.org/10.1002/cem.785.

Barrote, V., Tessalina, S., McNaughton, N., Jourdan, F., Hollis, S.P., Ware, B., Zi, J.-W., 2020. 4D history of the Nimbus VHMS ore deposit in the Yilgarn Craton, Western Australia. Precambrian Res. 337, 1–17. https://doi.org/10.1016/j.precamres.2019.105536.

Barton, M.D., 2014. Iron Oxide(–Cu–Au–REE–P–Ag–U–Co) Systems. In: Turekian, K., Holland, H. (Eds.), Treatise on Geochemistry. Elsevier, Oxford, pp. 515–541. https://doi.org/10.1016/B978-0-08-095975-7.01123-2.

Bath, A.B., Walshe, J.L., Cloutier, J., Verrall, M., Cleverley, J.S., Pownceby, M.I., Macrae, C.M., Wilson, N.C., Tunjic, J., Nortje, G.S., Robinson, P., 2013. Biotite and apatite as tools for tracking pathways of oxidized fluids in the Archean East Repulse Gold Deposit, Australia. Econ. Geol. 108, 667–690.

Belford, S.M., Davidson, G.J., McPhie, J., Large, R.R., 2015. Architecture of the Neoarchaean Jaguar VHMS deposit, Western Australia: Implications for prospectivity and the presence of depositional breaks. Precambrian Res. 260, 136–160. https://doi.org/10.1016/j.precamres.2014.12.019.

Bellinger, C., Sharma, S., Japkowicz, N., Zaïane, O.R., 2020. Framework for extreme imbalance classification: SWIM—sampling with the majority class. Knowl. Inf. Syst. 62, 841–866. https://doi.org/10.1007/s10115-019-01380-z.

Benkó, Z., Mogessie, A., Molnár, F., Hauck, S.A., Severson, M.J., Ettinger, K., 2018. The Influence of Thermal Differences and Variation of cl–F–OH Ratios on cu-Ni-PGE Mineralization in the Contact Aureole of the South Kawishiwi Intrusion. Geosciences, Duluth Complex. https://doi.org/10.3390/geosciences8120474.

Bernal, N.F., Gleeson, S.A., Smith, M.P., Barnes, J.D., Pan, Y., 2017. Evidence of multiple halogen sources in scapolites from iron oxide-copper-gold (IOCG) deposits and regional Na-Cl metasomatic alteration, Norrbotten County. Sweden. Chem. Geol. 451, 90–103. https://doi.org/10.1016/j.chemgeo.2017.01.005.

Boomeri, M., Mizuta, T., Nakashima, K., Ishiyama, D., Ishikawa, Y., 1997. Geochemical characteristics of halogen-bearing hastingsite, scapolite and phlogopite from the Sangan iron skarn deposits, northeastern Iran. J. Mineral. Petrol. Econ. Geol. 92, 481–501. https://doi.org/10.2465/ganko.92.481.

Boomeri, M., Nakashima, K., Lentz, D.R., 2009. The Miduk porphyry Cu deposit, Kerman, Iran: a geochemical analysis of the potassic zone including halogen element systematics related to Cu mineralization processes. J. Geochem. Explor. 103, 17–29. https://doi.org/10.1016/J.GEXPLO.2009.05.003.

Boomeri, M., Nakashima, K., Lentz, D.R., 2010. The Sarcheshmeh porphyry copper deposit, Kerman, Iran: a mineralogical analysis of the igneous rocks and alteration zones including halogen element systematics related to Cu mineralization processes. Ore Geol. Rev. 38, 367–381. https://doi.org/10.1016/J.OREGEOREV.2010.09.001.

Boudreau, A.E., Hoatson, D.M., 2004. Halogen variations in the Paleoproterozoic layered mafic-ultramafic intrusions of East Kimberley, Western Australia: implications for platinum group element mineralization. Econ. Geol. 99, 1015–1026. https://doi.org/10.2113/gsecongeo.99.5.1015.

Boudreau, A.E., Mathez, E.A., Mccallum, I.S., 1986. Halogen Geochemistry of the Stillwater and Bushveld Complexes: evidence for Transport of the Platinum-Group elements by Cl-Rich Fluids. J. Petrol. 27, 967–986. https://doi.org/10.1093/petrology/27.4.967.

Boulesteix, A.-L., Strimmer, K., 2007. Partial least squares: a versatile tool for the analysis of high-dimensional genomic data. Brief. Bioinform. 8, 32–44. https://doi.org/10.1093/bib/bbl016.

Bourdelle, F., 2021. Low-Temperature Chlorite Geothermometry and Related recent Analytical advances: a Review. Minerals. https://doi.org/10.3390/min11020130.

Bowman, J.R., Parry, W.T., Kropp, W.P., Kruer, S.A., 1987. Chemical and isotopic evolution of hydrothermal solutions at Bingham. Utah. Econ. Geol. 82, 395–428.

Brereton, R.G., Lloyd, G.R., 2014. Partial least squares discriminant analysis: taking the magic away. J. Chemom. 28, 213–225. https://doi.org/10.1002/cem.2609.

Chen, W.T., Zhou, M.-F., 2012. Paragenesis, Stable Isotopes, and Molybdenite Re-Os Isotope Age of the Lala Iron-Copper Deposit, Southwest China. Econ. Geol. 107, 459–480. https://doi.org/10.2113/econgeo.107.3.459.

Chen, W.T., Zhou, M.-F., 2015. Mineralogical and geochemical constraints on mobilization and mineralization of rare Earth elements in the Lala Fe-Cu-(Mo, Ree) deposit, SW China. Am. J. Sci. 315, 671–711. https://doi.org/10.2475/07.2015.03.

Chung, D., Keles, S., 2010. Sparse Partial Least Squares Classification for High Dimensional Data. Stat. Appl. Genet. Mol. Biol. 9 https://doi.org/10.2202/1544-6115.1492.

Codeço, M.S., Weis, P., Trumbull, R.B., Van Hinsberg, V., Pinto, F., Lecumberri-Sanchez, P., Schleicher, A.M., 2021. The imprint of hydrothermal fluids on trace-element contents in white mica and tourmaline from the Panasqueira W–Sn–Cu deposit, Portugal. Mineral. Deposita 56, 481–508. https://doi.org/10.1007/s00126-020-00984-8.

Cooke, D.R., Hollings, P., Wilkinson, J.J., Tosdal, R.M., 2014. Geochemistry of Porphyry Deposits. In: Holland, H.D., Turekian, K.K. (Eds.), Treatise on Geochemistry. Elsevier, Oxford, pp. 357–381. https://doi.org/10.1016/B978-0-08-095975-7.01116-5.

Croghan, C., Egeghy, P.P., 2003. Methods of dealing with values below the limit of detection using SAS. South. SAS User Gr. 22, 24.

De Iorio, M., Ebbels, T.M.D., Stephens, D.A., 2008. Statistical Techniques in Metabolic Profiling. In: Balding, D.J., Bishop, M., Cannings, C. (Eds.), Handbook of Statistical Genetics. John Wiley & Sons Ltd., pp. 347–373

Derome, D., Cathelineau, M., Fabre, C., Boiron, M.-C., Banks, D., Lhomme, T., Cuney, M., 2007. Paleo-fluid composition determined from individual fluid inclusions by Raman and LIBS: Application to mid-proterozoic evaporitic Na–Ca brines (Alligator Rivers Uranium Field, northern territories Australia). Chem. Geol. 237, 240–254. https://doi.org/10.1016/j.chemgeo.2006.10.015.

Dubosq, R., Schneider, D.A., Camacho, A., Lawley, C.J.M., 2019. Geochemical and Geochronological Discrimination of Biotite Types at the Detour Lake Gold Deposit, Canada. Minerals 9. https://doi.org/10.3390/min9100596.

Dupuis, C., Beaudoin, G., 2011. Discriminant diagrams for iron oxide trace element fingerprinting of mineral deposit types. Mineral. Deposita 46, 319–335. https://doi.org/10.1007/s00126-011-0334-y.

Dymek, R.F., 1983. Titanium, aluminum and interlayer cation substitutions in biotite from high-grade gneisses, West Greenland. Am. Mineral. 68, 880–899.

Finch, A.A., Parsons, I., Mingard, S.C., 1995. Biotites as indicators of fluorine fugacities in late-stage magmatic fluids: the Gardar Province of South Greenland. J. Petrol. 36, 1701–1728. https://doi.org/10.1093/oxfordjournals.petrology.a037271.

Fontaine, A., 2019. Géologie des minéralisations aurifères de la mine Éléonore, Eeyou Istchee Baie-James, province du Supérieur, Québec. Université du Québec, Canada.

Gaboury, D., 2019. Parameters for the formation of orogenic gold deposits. Appl. Earth Sci. 128, 124–133. https://doi.org/10.1080/25726838.2019.1583310.

Gaillard, N., Williams-Jones, A.E., Clark, J.R., Lypaczewski, P., Salvi, S., Perrouty, S., Piette-Lauzière, N., Guilmette, C., Linnen, R.L., 2018. Mica composition as a vector to gold mineralization: Deciphering hydrothermal and metamorphic effects in the Malartic district, Quebec. Ore Geol. Rev. 95, 789–820. https://doi.org/10.1016/j.oregeorev.2018.02.009.

Gilg, H.A., Lima, A., Somma, R., Belkin, H.E., De Vivo, B., Ayuso, R.A., 2001. Isotope geochemistry and fluid inclusion study of skarns from Vesuvius. Mineral. Petrol. 73, 145–176. https://doi.org/10.1007/s007100170015.

Golmohammadi, A., Karimpour, M.H., Malekzadeh Shafaroudi, A., Mazaheri, S.A., 2015. Alteration-mineralization, and radiometric ages of the source pluton at the Sangan iron skarn deposit, northeastern Iran. Ore Geol. Rev. 65, 545–563. https://doi.org/10.1016/j.oregeorev.2014.07.005.

Gregory, D.D., Cracknell, M.J., Large, R.R., McGoldrick, P., Kuhn, S., Maslennikov, V.V., Baker, M.J., Fox, N., Belousov, I., Figueroa, M.C., Steadman, J.A., Fabris, A.J., Lyons, T.W., 2019. 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. Econ. Geol. 114, 771–786. https://doi.org/10.5382/econgeo.4654.

Groves, D.I., Bierlein, F.P., Meinert, L.D., Hitzman, M.W., 2010. Iron Oxide Copper-Gold (IOCG) deposits through Earth history: Implications for Origin, Lithospheric setting, and Distinction from Other Epigenetic Iron Oxide Deposits. Econ. Geol. 105, 641–654. https://doi.org/10.2113/gsecongeo.105.3.641.

Groves, D.I., Santosh, M., Deng, J., Wang, Q., Yang, L., Zhang, L., 2020. A holistic model for the origin of orogenic gold deposits and its implications for exploration. Mineral. Deposita 55, 275–292. https://doi.org/10.1007/s00126-019-00877-5.

Grunsky, E.C., 2010. The interpretation of geochemical survey data. Geochemistry Explor. Environ. Anal. 10, 27. https://doi.org/10.1144/1467-7873/09-210.

Grunsky, E.C., Drew, L.J., Woodruff, L.G., Friske, P.W.B., Sutphin, D.M., 2013. Statistical variability of the geochemistry and mineralogy of soils in the Maritime Provinces of Canada and part of the Northeast United States. Geochemistry Explor. Environ. Anal. 13, 249–266. https://doi.org/10.1144/geochem2012-138.

Hamid, M.H.A., Yusoff, M., Mohamed, A., 2022. Survey on Highly Imbalanced Multi-class Data. Int. J. Adv. Comput. Sci. Appl. 13 https://doi.org/10.14569/IJACSA.2022.0130627.

Hammerli, J., Spandler, C., Oliver, N.H.S., Rusk, B., 2014. Cl/Br of scapolite as a fluid tracer in the earth's crust: insights into fluid sources in the Mary Kathleen Fold Belt, Mt. Isa Inlier, Australia. J. Metamorph. Geol. 32, 93–112. https://doi.org/10.1111/jmg.12060.

Hammerli, J., Kemp, A.I.S., Barrett, N., Wing, B.A., Roberts, M., Arculus, R.J., Boivin, P., Nude, P.M., Rankenburg, K., 2017. Sulfur isotope signatures in the lower crust: a SIMS study on S-rich scapolite of granulites. Chem. Geol. 454, 54–66. https://doi.org/10.1016/j.chemgeo.2017.02.016.

Hanley, J., Ames, D., Barnes, J., Sharp, Z., Guillong, M., 2011. Interaction of magmatic fluids and silicate melt residues with saline groundwater in the footwall of the Sudbury Igneous complex, Ontario, Canada: New evidence from bulk rock geochemistry, fluid inclusions and stable isotopes. Chem. Geol. 281, 1–25. https://doi.org/10.1016/j.chemgeo.2010.11.009.

Hanley, J.J., Mungall, J.E., 2003. Chlorine enrichment and hydrous alteration of the Sudbury breccia hosting footwall Cu–Ni–PGE mineralization at the Fraser mine, Sudbury, Ontario, Canada. Can. Mineral. 41, 857–881. https://doi.org/10.2113/gscanmin.41.4.857.

Henry, D.J., Daigle, N.M., 2018. Chlorine incorporation into amphibole and biotite in high-grade iron-formations: Interplay between crystallography and metamorphic fluids. Am. Mineral. 103, 55–68. https://doi.org/10.2138/am-2018-6143.

Henry, D.J., Guidotti, C.V., Thomson, J.A., 2005. The Ti-saturation surface for low-to-medium pressure metapelitic biotites: Implications for geothermometry and Ti-substitution mechanisms. Am. Mineral. 90, 316–328.

Hezarkhani, A., Williams-Jones, A.E., 1998. Controls of alteration and mineralization in the Sungun porphyry copper deposit, Iran; evidence from fluid inclusions and stable isotopes. Econ. Geol. 93, 651–670. https://doi.org/10.2113/gsecongeo.93.5.651.

Hron, K., Templ, M., Filzmoser, P., 2010. Imputation of missing values for compositional data using classical and robust methods. Comput. Stat. Data Anal. 54, 3095–3107. https://doi.org/10.1016/j.csda.2009.11.023.

Huang, X.-W., Boutroy, É., Makvandi, S., Beaudoin, G., Corriveau, L., De Toni, A.F., 2019. Trace element composition of iron oxides from IOCG and IOA deposits: relationship to hydrothermal alteration and deposit subtypes. Mineral. Deposita 54, 525–552. https://doi.org/10.1007/s00126-018-0825-1.

Iveson, A.A., Webster, J.D., Rowe, M.C., Neill, O.K., 2016. Magmatic–hydrothermal fluids and volatile metals in the Spirit Lake pluton and Margaret Cu–Mo porphyry system, SW Washington, USA. Contrib. Mineral. Petrol. 171, 1–32. https://doi.org/10.1007/s00410-015-1224-6.

Jacobs, D.C., Parry, W.T., 1979. Geochemistry of biotite in the Santa Rita porphyry copper deposit, New Mexico. Econ. Geol. 74, 860–887.

Katz, L.R., Kontak, D.J., Dubé, B., McNicoll, V., Creaser, R., Petrus, J.A., 2020. An archean porphyry-type gold deposit: the Côté Gold Au(-Cu) deposit, Swayze Greenstone Belt, Superior Province, Ontario, Canada. Econ. Geol. 116, 47–89. https://doi.org/10.5382/econgeo.4785.

Kelly, C.J., Davis, W.J., Potter, E.G., Corriveau, L., 2020. Geochemistry of hydrothermal tourmaline from IOCG occurrences in the Great Bear magmatic zone: Implications for fluid source(s) and fluid composition evolution. Ore Geol. Rev. 118, 103329 https://doi.org/10.1016/j.oregeorev.2020.103329.

Khosravi, M., Christiansen, E.H., Rajabzadeh, M.A., 2021. Chemistry of rock-forming silicate and sulfide minerals in the granitoids and volcanic rocks of the Zefreh porphyry Cu–Mo deposit, Central Iran: Implications for crystallization, alteration, and mineralization potential. Ore Geol. Rev. 134, 104150 https://doi.org/10.1016/j.oregeorev.2021.104150.

Knipping, J.L., Bilenker, L.D., Simon, A.C., Reich, M., Barra, F., Deditius, A.P., Wälle, M., Heinrich, C.A., Holtz, F., Munizaga, R., 2015. Trace elements in magnetite from massive iron oxide-apatite deposits indicate a combined formation by igneous and magmatic-hydrothermal processes. Geochim. Cosmochim. Acta 171, 15–38. https://doi.org/10.1016/j.gca.2015.08.010.

Kuikka, J., 2018. Major and Trace Element Characteristics of Biotite and Chlorite as Proxies for Gold Ore Mineralization. University of Helsinki, Sweden.

Kwak, T.A.P., 1983. The geology and geochemistry of the zoned, Sn-W-F-Be skarns at Mt. Lindsay, Tasmania, Australia. Econ. Geol. 78, 1440–1465. https://doi.org/10.2113/gsecongeo.78.7.1440.

Lanier, G., Raab, W.J., Folsom, R.B., Cone, S., 1978. Alteration of equigranular monzonite, Bingham mining district, Utah. Econ. Geol. 73, 1270–1286.

Leach, D.L., Bradley, D.C., Huston, D., Pisarevsky, S.A., Taylor, R.D., Gardoll, S.J., 2010. Sediment-Hosted Lead-Zinc Deposits in Earth history. Econ. Geol. 105, 593–625. https://doi.org/10.2113/gsecongeo.105.3.593.

Lee, L.C., Jemain, A.A., 2019. Predictive modelling of colossal ATR-FTIR spectral data using PLS-DA: empirical differences between PLS1-DA and PLS2-DA algorithms. Analyst 144, 2670–2678. https://doi.org/10.1039/C8AN02074D.

Lee, L.C., Liong, C.-Y., Jemain, A.A., 2018. Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: a review of contemporary practice strategies and knowledge gaps. Analyst 143, 3526–3539. https://doi.org/10.1039/c8an00599k.

Li, J., Li, G., Qin, K., Xiao, B., Chen, L., Zhao, J., 2012. Mineralogy and Mineral Chemistry of the cretaceous Duolong Gold-Rich Porphyry Copper Deposit in the Bangongco Arc, Northern Tibet. Resour. Geol. 62, 19–41. https://doi.org/10.1111/j.1751-3928.2011.00178.x.

Li, X., Zhang, C., Behrens, H., Holtz, F., 2020. Calculating biotite formula from electron microprobe analysis data using a machine learning method based on principal components regression. Lithos 356–357, 105371. https://doi.org/10.1016/j.lithos.2020.105371.

Li, X.-C., Zhou, M.-F., 2018. The nature and origin of hydrothermal REE mineralization in the Sin Quyen Deposit, northwestern Vietnam. Econ. Geol. 113, 645–673. https://doi.org/10.5382/econgeo.2018.4565.

Lindsay, J.J., Hughes, H.S.R., Yeomans, C.M., Andersen, J.C.Ø., McDonald, I., 2021. A machine learning approach for regional geochemical data: Platinum-group element geochemistry vs geodynamic settings of the North Atlantic Igneous Province. Geosci. Front. 12, 101098 https://doi.org/10.1016/j.gsf.2020.10.005.

Liu, S., Fan, H.-R., Yang, K.-F., Hu, F.-F., Rusk, B., Liu, X., Li, X.-C., Yang, Z.-F., Wang, Q.-W., Wang, K.-Y., 2018. Fenitization in the giant Bayan Obo REE-Nb-Fe deposit: Implication for REE mineralization. Ore Geol. Rev. 94, 290–309. https://doi.org/10.1016/j.oregeorev.2018.02.006.

Liu, Y.-P., Zhang, H.-F., Wang, C.-L., 2020. Overprinting by episodic mineralization in the Dongyaozhuang gold deposit, Wutai Mountain, China: Constraints from geology, mineralogy, and fluid inclusions. Geol. J. 55, 5934–5952. https://doi.org/10.1002/gj.3616.

Lyons, T.W., Gellatly, A.M., Mcgoldrick, P.J., Kah, L.C., 2006. Proterozoic sedimentary exhalative (SEDEX) deposits and links to evolving global ocean chemistry. Geol. Soc. Am. Mem. 198, 169–184. https://doi.org/10.1130/2006.1198(10).

Makvandi, S., Ghasemzadeh-Barvarz, M., Beaudoin, G., Grunsky, E.C., Beth McClenaghan, M., Duchesne, C., 2016a. Principal component analysis of magnetite composition from volcanogenic massive sulfide deposits: Case studies from the Izok Lake (Nunavut, Canada) and Halfmile Lake (New Brunswick, Canada) deposits. Ore Geol. Rev. 72, 60–85. https://doi.org/10.1016/j.oregeorev.2015.06.023.

Makvandi, S., Ghasemzadeh-Barvarz, M., Beaudoin, G., Grunsky, E.C., McClenaghan, M. B., Duchesne, C., Boutroy, E., 2016b. Partial least squares-discriminant analysis of trace element compositions of magnetite from various VMS deposit subtypes: Application to mineral exploration. Ore Geol. Rev. 78, 388–408. https://doi.org/10.1016/j.oregeorev.2016.04.014.

Mao, W., Zhong, H., Zhu, W.-G., Lin, X.-G., Zhao, X.-Y., 2018. Magmatic-hydrothermal evolution of the Yuanzhuding porphyry Cu-Mo deposit, South China: Insights from mica and quartz geochemistry. Ore Geol. Rev. 101, 765–784. https://doi.org/10.1016/j.oregeorev.2018.08.016.

Marshall, L.J., Oliver, N.H.S., 2006. Monitoring fluid chemistry in iron oxide-copper-gold-related metasomatic processes, eastern Mt Isa Block, Australia. Geofluids 6, 45–66. https://doi.org/10.1111/j.1468-8123.2006.00127.x.

Maydagán, L., Franchini, M., Impiccini, A., Lentz, D., 2016. Phyllosilicates geochemistry and distribution in the Altar porphyry Cu-(Au) deposit, Andes Cordillera of San Juan, Argentina: applications in exploration, geothermometry, and geometallurgy. J. Geochem. Explor. 167, 83–109. https://doi.org/10.1016/j.gexplo.2016.05.002.

Meinert, L.D., Hedenquist, J.W., Satoh, H., Matsuhisa, Y., 2003. Formation of Anhydrous and Hydrous Skarn in Cu-au Ore Deposits by Magmatic Fluids. Econ. Geol. 98, 147–156. https://doi.org/10.2113/gsecongeo.98.1.147.

van Middelaar, W.T., Keith, J.D., 1990. Mica chemistry as an indicator of oxygen and halogen fugacities in the CanTung and other W-related granitoids in the north American Cordillera. Geol. Soc. Am. Spec. Pap. 205–220.

Monteiro, L.V.S., Xavier, R.P., Hitzman, M.W., Juliani, C., de Souza Filho, C.R., de R Carvalho, E., 2008. Mineral chemistry of ore and hydrothermal alteration at the Sossego iron oxide–copper–gold deposit, Carajás Mineral Province, Brazil. Ore Geol. Rev. 34, 317–336. https://doi.org/10.1016/J.OREGEOREV.2008.01.003.

Moshefi, P., Hosseinzadeh, M.R., Moayyed, M., Lentz, D.R., 2018. Comparative study of mineral chemistry of four biotite types as geochemical indicators of mineralized and barren intrusions in the Sungun Porphyry Cu-Mo deposit, northwestern Iran. Ore Geol. Rev. 97, 1–20. https://doi.org/10.1016/J.OREGEOREV.2018.05.003.

Moshefi, P., Hosseinzadeh, M.R., Moayyed, M., Lentz, D.R., 2020. Distinctive geochemical features of biotite types from the subeconomic Sonajil porphyry-type Cu deposit, northwestern Iran: Implications for analysis of porphyry copper deposit mineralization potential. J. Geochem. Explor. 214, 106543 https://doi.org/10.1016/j.gexplo.2020.106543.

Munoz, J.L., 1984. F-OH and Cl-OH exchange in micas with applications to hydrothermal ore deposits. Rev. Mineral. Geochem. 13, 469–493.

Munoz, J.L., 1992. Calculation of HF and HCl fugacities from biotite compositions : revised equations. Geol. Soc. Am. Abstr. Programs 24, A221.

Munoz, J.L., Swenson, A., 1981. Chloride-hydroxyl exchange in biotite and estimation of relative HCl/HF activities in hydrothermal fluids. Econ. Geol. 76, 2212–2221.

Nash, J.T., Connor, J.J., 1993. Iron and chlorine as guides to stratiform Cu-Co-Au deposits, Idaho Cobalt Belt, USA. Mineral. Deposita 28, 99–106. https://doi.org/10.1007/BF00196334.

Nguyen, D.V., Rocke, D.M., 2002. Tumor classification by partial least squares using microarray gene expression data. Bioinformatics 18, 39–50. https://doi.org/10.1093/bioinformatics/18.1.39.

Nozaka, T., Akitou, T., Abe, N., Tribuzio, R., 2019. Biotite in olivine gabbros from Atlantis Bank: evidence for amphibolite-facies metasomatic alteration of the lower oceanic crust. Lithos 348–349, 105176. https://doi.org/10.1016/j.lithos.2019.105176.

Palma, G., Barra, F., Reich, M., Valencia, V., Simon, A.C., Vervoort, J., Leisen, M., Romero, R., 2019. Halogens, trace element concentrations, and Sr-Nd isotopes in apatite from iron oxide-apatite (IOA) deposits in the Chilean iron belt: evidence for magmatic and hydrothermal stages of mineralization. Geochim. Cosmochim. Acta 246, 515–540. https://doi.org/10.1016/j.gca.2018.12.019.

Panigrahi, M.K., Naik, R.K., Pandit, D., Misra, K.C., 2008. Reconstructing physico-chemical parameters of hydrothermal mineralization of copper at the Malanjkhand deposit, India, from mineral chemistry of biotite, chlorite and epidote. Geochem. J. 42, 443–460. https://doi.org/10.2343/geochemj.42.443.

Parry, W.T., Ballantyne, G.H., Wilson, J.C., 1978. Chemistry of biotite and apatite from a vesicular quartz latite porphyry plug at Bingham, Utah, Utah. Econ. Geol. 73, 1308–1314.

Parry, W.T., Wilson, P.N., Moser, D., Heizler, M.T., 2001. U-Pb Dating of Zircon and 40Ar/39Ar Dating of Biotite at Bingham, Utah. Econ. Geol. 96, 1671–1683. https://doi.org/10.2113/gsecongeo.96.7.1671.

Parsapoor, A., Khalili, M., Tepley, F., Maghami, M., 2015. Mineral chemistry and isotopic composition of magmatic, re-equilibrated and hydrothermal biotites from Darreh-Zar porphyry copper deposit, Kerman (Southeast of Iran). Ore Geol. Rev. 66, 200–218. https://doi.org/10.1016/J.OREGEOREV.2014.10.015.

Pearce, T.H., 1968. A Contribution to the Theory of Variation Diagrams. Contrib. Mineral. Petrol. 19, 142–157.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Bertrand, T., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Petrelli, M., Perugini, D., 2016. Solving petrological problems through machine learning: the study case of tectonic discrimination using geochemical and isotopic data. Contrib. Mineral. Petrol. 171, 81. https://doi.org/10.1007/s00410-016-1292-2.

Raza, M.Q., Absar, N., Pandalai, H.S., Patel, S.C., 2021. Chlorite thermometry and fluid inclusion studies on vein-type Tintini copper deposit, Eastern Dharwar Craton, India: Ore genetic implications. Ore Geol. Rev. 131 https://doi.org/10.1016/j.oregeorev.2021.104058.

del Real, I., Thompson, J.F.H., Simon, A.C., Reich, M., 2020. Geochemical and isotopic signature of pyrite as a proxy for fluid source and evolution in the Candelaria-Punta del Cobre iron oxide copper-gold district, Chile. Econ. Geol. 115, 1493–1518. https://doi.org/10.5382/econgeo.4765.

Reich, M., Simon, A.C., Deditius, A., Barra, F., Chryssoulis, S., Lagas, G., Tardani, D., Knipping, J., Bilenker, L., Sánchez-Alfaro, P., Roberts, M.P., Munizaga, R., 2016. Trace element signature of pyrite from the Los Colorados iron oxide-apatite (IOA) deposit, Chile: a missing link between Andean IOA and iron oxide copper-gold systems? Econ. Geol. 111, 743–761. https://doi.org/10.2113/econgeo.111.3.743.

Rezaei, M., Zarasvandi, A., 2020. Titanium-in-biotite thermometry in porphyry copper systems: challenges to application of the thermometer. Resour. Geol. 70, 157–168. https://doi.org/10.1111/rge.12227.

Richards, J.P., 2003. Tectono-Magmatic Precursors for Porphyry Cu-(Mo-au) Deposit Formation. Econ. Geol. 98, 1515–1533. https://doi.org/10.2113/gsecongeo.98.8.1515.

Rieder, M., Cavazzini, G., D'yakonov, Y.S., Frank-Kamenetskii, V.A., Gottardi, G., Guggenheim, S., Koval, P.V., Müller, G., Neiva, A.M.R., Radoslovich, E.W., Robert, J.

L., Sassi, F.P., Takeda, H., Weiss, Z., Wones, D.R., 1998. Nomenclature of the micas. Clay Clay Miner. 46, 586–595.
Rohart, F., Gautier, B., Singh, A., Lê Cao, K.-A., 2017. mixOmics: an R package for 'omics feature selection and multiple data integration. PLoS Comput. Biol. 13, e1005752.
Ruiz-Perez, D., Guan, H., Madhivanan, P., Mathee, K., Narasimhan, G., 2020. So you think you can PLS-DA? BMC Bioinformatics 21, 2. https://doi.org/10.1186/s12859-019-3310-7.
van Ryt, M.R., Sanislav, I.V., Dirks, P.H.G.M., Huizenga, J.M., Mturi, M.I., Kolling, S.L., 2019. Biotite chemistry and the role of halogens in Archaean greenstone hosted gold deposits: a case study from Geita Gold Mine. Tanzania. Ore Geol. Rev. 111, 102982 https://doi.org/10.1016/j.oregeorev.2019.102982.
Saha, R., Upadhyay, D., Mishra, B., 2021. Discriminating Tectonic setting of Igneous Rocks using Biotite Major Element chemistry a Machine Learning Approach. Geochem. Geophys. Geosyst. 22, e2021GC010053 https://doi.org/10.1029/2021GC010053.
Saini-Eidukat, B., Weiblen, P.W., Bitsianes, G., Glascock, D., 1990. Contrasts between platinum group element contents and biotite compositions of Duluth complex troctolitic and anorthositic series rocks. Mineral. Petrol. 42, 121–140. https://doi.org/10.1007/BF01162687.
Samson, I.M., Blackburn, W.H., Gagnon, J.E., 1999. Paragenesis and composition of amphibole and biotite in the MacLellan gold deposit, Lynn Lake greenstone belt, Manitoba, Canada. Can. Mineral. 37, 1405–1421.
Sciuba, M., Beaudoin, G., Grzela, D., Makvandi, S., 2020. Trace element composition of scheelite in orogenic gold deposits. Mineral. Deposita 55, 1149–1172. https://doi.org/10.1007/s00126-019-00913-4.
Sciuba, M., Beaudoin, G., Makvandi, S., 2021. Chemical composition of tourmaline in orogenic gold deposits. Mineral. Deposita 56, 537–560. https://doi.org/10.1007/s00126-020-00981-x.
Selby, D., Nesbitt, B.E., 2000. Chemical composition of biotite from the Casino porphyry Cu–Au–Mo mineralization, Yukon, Canada: evaluation of magmatic and hydrothermal fluid chemistry. Chem. Geol. 171, 77–93. https://doi.org/10.1016/S0009-2541(00)00248-5.
Shan, P., Cao, M., Jourdan, F., Evans, N.J., Ge, X., Sun, J., Qin, K., Zhao, C., 2021. Biotite 40Ar/39Ar dating and chemical composition inform metallogenesis of Xiaoxi'nancha porphyry Au-Cu deposit, NE China. Ore Geol. Rev. 134, 104140 https://doi.org/10.1016/j.oregeorev.2021.104140.
Siahcheshm, K., Calagari, A.A., Abedini, A., Lentz, D.R., 2012. Halogen signatures of biotites from the Maher-Abad porphyry copper deposit, Iran: characterization of volatiles in syn- to post-magmatic hydrothermal fluids. Int. Geol. Rev. 54, 1353–1368. https://doi.org/10.1080/00206814.2011.639487.
Smith, M., Coppard, J., Herrington, R., Stein, H., 2007. The geology of the Rakkurijärvi Cu-(au) Prospect, Norrbotten: a new Iron Oxide-Copper-Gold Deposit in northern Sweden. Econ. Geol. 102, 393–414. https://doi.org/10.2113/gsecongeo.102.3.393.
Smith, M.P., 2007. Metasomatic silicate chemistry at the Bayan Obo Fe–REE–Nb deposit, Inner Mongolia, China: Contrasting chemistry and evolution of fenitising and mineralising fluids. Lithos 93, 126–148. https://doi.org/10.1016/j.lithos.2006.06.013.
Sonnenthal, E.L., 1992. Geochemistry of dendritic anorthosites and associated pegmatites in the Skaergaard Intrusion, East Greenland: evidence for metasomatism by a chlorine-rich fluid. J. Volcanol. Geotherm. Res. 52, 209–230. https://doi.org/10.1016/0377-0273(92)90141-Y.
de Souza, S.R.C., Botelho, N.F., Jiménez, F.A.C., Dantas, E.L., Reis, M.A., Viana, C.S., 2021. Mineralization and hydrothermal alteration in the Mamão orogenic gold deposit, Andorinhas greenstone belt, Carajás Province, Brazil. J. S. Am. Earth Sci. 112, 103548 https://doi.org/10.1016/j.jsames.2021.103548.
Szymańska, E., Saccenti, E., Smilde, A.K., Westerhuis, J.A., 2012. Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies. Metabolomics 8, 3–16. https://doi.org/10.1007/s11306-011-0330-3.
Tan, Y., Shi, L., Tong, W., Gene Hwang, G.T., Wang, C., 2004. Multi-class tumor classification by discriminant partial least squares using microarray gene expression data and assessment of classification models. Comput. Biol. Chem. 28, 235–243. https://doi.org/10.1016/j.compbiolchem.2004.05.002.
Tang, P., Tang, J., Lin, B., Wang, L., Zheng, W., Leng, Q., Gao, X., Zhang, Z., Tang, X., 2019. Mineral chemistry of magmatic and hydrothermal biotites from the Bangpu porphyry Mo (Cu) deposit. Tibet. Ore Geol. Rev. 115, 103122 https://doi.org/10.1016/j.oregeorev.2019.103122.
Timón, S.M., Moro, M.C., Cembranos, M.L., Fernández, A., Crespo, J.L., 2007. Contact metamorphism in the Los Santos W skarn (NW Spain). Mineral. Petrol. 90, 109–140. https://doi.org/10.1007/s00710-006-0166-0.
Veloso, A.S.R., Monteiro, L.V.S., Juliani, C., 2020. The link between hydrothermal nickel mineralization and an iron oxide-copper–gold (IOCG) system: Constraints based on mineral chemistry in the Jatobá deposit. Carajás Province. Ore Geol. Rev. 121, 103555 https://doi.org/10.1016/j.oregeorev.2020.103555.
Verbovšek, T., 2011. A comparison of parameters below the limit of detection in geochemical analyses by substitution methods. RMZ-Mater. Geoenviron. 58, 393–404.
Wallace, C.J., Maher, K.C., 2019. Phyllic alteration and the implications of fluid composition at the copper flat hydrothermal System, New Mexico, USA. Ore Geol. Rev. 104, 273–293. https://doi.org/10.1016/j.oregeorev.2018.11.009.
Westerhuis, J.A., Hoefsloot, H.C.J., Smit, S., Vis, D.J., Smilde, A.K., van Velzen, E.J.J., van Duijnhoven, J.P.M., van Dorsten, F.A., 2008. Assessment of PLSDA cross validation. Metabolomics 4, 81–89. https://doi.org/10.1007/s11306-007-0099-6.
Whitten, E.H.T., 1995. Open and closed compositional data in petrology. Math. Geol. 27, 789–806. https://doi.org/10.1007/BF02273538.
Willmore, C.C., Boudreau, A.E., Kruger, F.J., 2000. The Halogen Geochemistry of the Bushveld complex, Republic of South Africa: Implications for Chalcophile Element distribution in the lower and critical zones. J. Petrol. 41, 1517–1539. https://doi.org/10.1093/petrology/41.10.1517.
Wu, C.-M., Chen, H.-X., 2015. Revised Ti-in-biotite geothermometer for ilmenite- or rutile-bearing crustal metapelites. Sci. Bull. 60, 116–121. https://doi.org/10.1007/s11434-014-0674-y.
Xavier, R.P., Wiedenbeck, M., Trumbull, R.B., Dreher, A.M., Monteiro, L.V.S., Rhede, D., de Araújo, C.E.G., Torresi, I., 2008. Tourmaline B-isotopes fingerprint marine evaporites as the source of high-salinity ore fluids in iron oxide copper-gold deposits, Carajás Mineral Province (Brazil). Geology 36, 743–746. https://doi.org/10.1130/G24841A.1.
Xie, G., Mao, J., Richards, J.P., Han, Y., Fu, B., 2019. Distal au deposits Associated with Cu-au Skarn Mineralization in the Fengshan Area, Eastern China. Econ. Geol. 114, 127–142. https://doi.org/10.5382/econgeo.2019.4623.
Yardley, B.W.D., Cleverley, J.S., 2015. The role of metamorphic fluids in the formation of ore deposits. Geol. Soc. London Spec. Publ. 393, 117–134. https://doi.org/10.1144/SP393.5.
Yavuz, F., 2003. Evaluating micas in petrologic and metallogenic aspect: I–definitions and structure of the computer program MICA+. Comput. Geosci. 29, 1203–1213. https://doi.org/10.1016/S0098-3004(03)00142-0.
Yavuz, F., Öztaş, T., 1997. BIOTERM—A program for evaluating and plotting microprobe analyses of biotite from barren and mineralized magmatic suites. Comput. Geosci. 23, 897–907. https://doi.org/10.1016/S0098-3004(97)00071-X.
Zarasvandi, A., Rezaei, M., Raith, J.G., Pourkaseb, H., Asadi, S., Saed, M., Lentz, D.R., 2018. Metal endowment reflected in chemical composition of silicates and sulfides of mineralized porphyry copper systems, Urumieh-Dokhtar magmatic arc, Iran. Geochim. Cosmochim. Acta 223, 36–59. https://doi.org/10.1016/j.gca.2017.11.012.
Zaw, U.K., Clark, A.H., 1978. Fluoride-hydroxyl ratios of skarn silicates, Cantung E-Zone scheelite orebody, Tungsten, Northwest Territories. Can. Mineral. 16, 207–221.
Zhu, C., Sverjensky, D.A., 1991. Partitioning of F-Cl-OH between minerals and hydrothermal fluids. Geochim. Cosmochim. Acta 55, 1837–1858. https://doi.org/10.1016/0016-7037(91)90028-4.
Zhu, C., Sverjensky, D.A., 1992. F-Cl-OH partitioning between biotite and apatite. Geochim. Cosmochim. Acta 56, 3435–3467. https://doi.org/10.1016/0016-7037(92)90390-5.