Multivariate Analysis of an LA-ICP-MS Trace Element Dataset for Pyrite

**Lyron Winderbaum · Cristiana L. Ciobanu ·
Nigel J. Cook · Matthew Paul · Andrew Metcalfe ·
Sarah Gilbert**

Abstract Application of multivariate statistics to trace element datasets is reviewed using 164 multi-element LA-ICP-MS spot analyses of pyrite from the Moonlight epithermal gold prospect, Queensland, Australia. Multivariate analysis of variance (MANOVA) is used to demonstrate that classification of pyrite on morphological and other non-numeric factors is geochemically valid. Parallel coordinate plots and correlation cluster analysis using Spearman's coefficients are used to discover unexpected elemental relationships without making assumptions a priori. Finally, principal component analysis and factor analysis are used to demonstrate the presence of sub-classes of pyrite. Corroborated with geological data, statistical analysis provides evidence for successive generations of hydrothermal fluids, each introducing specific metals, and for partial or complete replacement of different minerals. The data permit reinterpretation of Moonlight as a telescoped system where epithermal-Au ($\pm$ base metals) is superposed onto early porphyry-Mo mineralization.

L. Winderbaum · A. Metcalfe
School of Mathematical Sciences, University of Adelaide, North Terrace, Adelaide, SA 5005, Australia

C.L. Ciobanu · N.J. Cook (✉) · M. Paul
Centre for Tectonics, Resources and Exploration (TRaX), School of Earth and Environmental Sciences, University of Adelaide, North Terrace, Adelaide, SA 5005, Australia
e-mail: nigel.cook@adelaide.edu.au

S. Gilbert
CODES, University of Tasmania, Private Bag 126, Hobart, Tasmania 7001, Australia

1 Introduction

Statistical analysis has been widely applied to treat large datasets of multivariate data for the past 30 years (Bailey and Krzanowski 2012). There are, however, few studies which apply (a) such techniques to a dataset for a single mineral such as pyrite, and (b) critically evaluate the applicability of several different statistical methods to such a dataset. Interest in multi-trace element datasets for single minerals has expanded since techniques such as Laser-Ablation Inductively Coupled Plasma Mass Spectroscopy (LA-ICP-MS) became available and high-quality standards were developed for reliable quantitative data at low cost (Ciobanu et al. 2009; Cook et al. 2009a, 2011; Danyushevsky et al. 2011; Large et al. 2007; Maslennikov et al. 2009; Sung et al. 2009; Thomas et al. 2011; Ye et al. 2011). In mineral deposit studies, refractory minerals such as pyrite are targeted since unless they have totally recrystallized, a sequence of geological events will be encoded within their zoned, or otherwise heterogeneous, trace element chemical patterns (Cook et al. 2009b; Sung et al. 2009). Mineral trace element data also have applications for exploration vectors that can assist in locating new resources, for genetic interpretation and for traceability of concentrate.

LA-ICP-MS trace element data are commonly presented with relatively low-level statistical treatment: as arithmetic or weighted means, variance data, or as bi- or tri-element correlations (Ye et al. 2011). Application of trace element distributions to derive genetic or exploration information is based on classification of mineral populations defined by grain morphology, habit, mineral association or other non-numerical characteristics (Cook et al. 2009b). The statistical reliability of such an approach is rarely tested. One of the issues that impacts the statistical texture of the dataset is single-grain heterogeneity revealed by trace element mapping (Large et al. 2009). This contribution addresses this gap, presenting statistical analyses in a multi-element LA-ICP-MS dataset for pyrite. We assess the reliability, and geochemical validity, of classifying populations on textural criteria using MANOVA, and evaluate the usefulness of a variety of multivariate statistical techniques; hierarchical agglomerative cluster analysis; parallel coordinate plots; principal component analysis; and factor analysis as exploratory methods for visualizing data, and discovering unexpected relationships.

2 Geological Context, Pyrite Textures and Trace Element Data

The 164 spot analyses used for this study derive from 13 samples taken from three drillholes intersecting Moonlight, a recently discovered epithermal-style Au prospect within the Late Devonian to Early-Mid Carboniferous Pajingo orefield (Bobis et al. 1995; Richards et al. 1998; Baker et al. 2006), located in the Drummond Basin, northwest Queensland, Australia. At Moonlight (Paul 2010), mineralization is located within altered andesites, brecciated epiclastic rocks and crosscutting chalcedony veins. Pyrite is a minor, yet ubiquitous component of the veins and altered rocks. Moonlight is a satellite prospect located 2 km southeast from the Vera
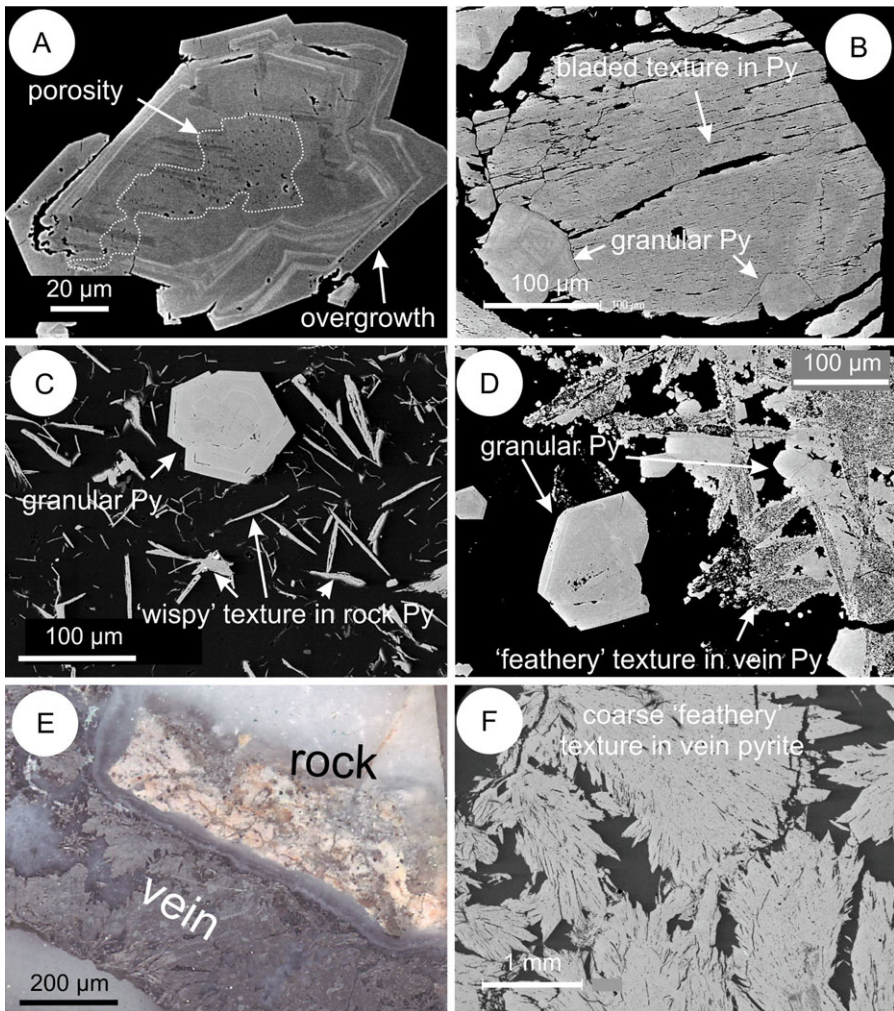
Fig. 1 Texturally distinct pyrite (Py) types (back-scattered-electron images): (**A**) Zoned GR pyrite; (**B**) GR pyrite retaining replacement texture; (**C**) 'wispy' pyrite (RR) coexisting with GR pyrite; (**D**) co-existing GV- and 'bladed' RV-pyrite; (**E**) 'feathery' (RV) pyrite. (**F**) Photomicrograph showing 'feathery' vein pyrite

Nancy deposit where mining is currently ongoing. Gold occurs as fine-grained native gold/electrum as well as invisible gold (Cook and Chryssoulis 1990) within As-bearing pyrite.

The diverse pyrite textures (Fig. 1) indicate that the epithermal system was multi-phase in terms of brecciation and input of hydrothermal fluids (Paul 2010). Sampling was focused on pyrite from veins and disseminations in lithologies with different alteration in order to constrain (i) Au deportment in pyrite and association with other elements and (ii) the mineralization sequence. The dominant textural pyrite type is granular. This is always characterized by compositional zona-

Table 1 Classification of pyrite grains

Symbol	Morphology	Location	Class	Number
O	Granular	Rock	GR	98
Δ	Granular	Vein	GV	22
O	Replacement	Rock	RR	14
Δ	Replacement	Vein	RV	30
				Total = 164

tion patterns that show not only reabsorption and overgrowth but are also demar-cated or crosscut by pores (Fig. 1A), sometimes associated with trails of inclusions. Whereas zonation is attributable to mineral growth, the latter clearly relate to su-perimposed brecciation. Some zoned grains retain a bladed appearance indicating pseudomorphic replacement of a pre-existing mineral (Fig. 1B). Such a texture is common for pyrite that replaces marcasite and/or pyrrhotite (Ramdohr 1980). This is often accompanied by the presence of relict inclusions suggesting the identity of the precursor minerals; such inclusions, however, are not seen in the present sam-ples.

More typical replacement morphologies observed for Moonlight pyrite are char-acterized by unusual habits, that is, wispy and feathery. These morphologies were interpreted to mimic earlier minerals with lamellar habit and which are also ob-served in the samples: bladed calcite, molybdenite and possibly chlorite (Fig. 1C, D; Paul 2010). Although granular and replacement pyrites are dominant in the rock and veins (Fig. 1E, F), respectively, they nevertheless coexist in the two litholo-gies; in veins, granular pyrite is clearly later. Feathery pyrite always has a dirty appearance outlining the lamellar morphology and this is attributable to retention of dusty inclusions. Titanium-oxide inclusions (rutile and ilmenite) are common in the granular pyrite in the rock, whereas inclusions of base-metal sulfides are present in granular pyrite from the veins. Fine-grained Ag-minerals and gold are present in both.

Multi-element LA-ICP-MS spot analysis and element mapping of pyrite were car-ried out on an Agilent HP 7700 and HP 4500 Quadrupole instruments, respectively (CODES, University of Tasmania). Methodologies followed Danyushevsky et al. (2011) with details given in the online Appendix A. Results of the 164 LA-ICP-MS spot analyses are given in Table 1 of online Appendix B. For the purpose of this paper, we classify pyrite by two criteria: appearance (granular, G, or replacement, R) and location (rock, R, or vein, V), giving four simplified classes: GR; GV; RR; and RV (Table 1). The largest dataset is for GR which is also the earliest generation of pyrite. All the other categories are later than GR, either because of their location (RV and GV in the vein) or as replacement type in the rock (RR). Certain trace elements were commonly measured at concentrations below minimum detection limit (mdl) in part of the population (percentage of spot analyses below the mdl in brackets): Cr (81), Nb (70), Te (63), Re (60), U (52), W (49), Bi (48), Sn (48), Ga (45), In (27), Mn (24), V (20), Cd (15). Values of half the mdl are reported in Table 2 and in Table 1 of online Appendix B. However, for statistical analyses values below the mdl there is no

Table 2 Statistical data for the dataset divided by category

Category	Precious metals and chalcophile elements												Chalcogens			Granitophile elements						Siderophile elements								
	Au	Ag	As	Sb	Cu	Pb	Bi	Tl	Zn	Cd	In	SUM	Se	Te	SUM	Sn	Mo	W	Re	U	SUM	Mn	Ga	Co	Ni	Ti	V	Cr	Nb	SUM
GR (n = 97*)																														
Mean	34	81	12932	224	473	172	3.4	17	30	2.4	0.05	**13969**	73	2.2	**75**	0.22	138	1.5	0.05	0.21	**140**	32	0.26	298	137	249	6.8	1.8	0.39	**725**
S.D.	51	110	8430	288	525	300	9.2	31	125	4.5	0.16		97	7.8		0.29	742	4.4	0.11	0.61		87	0.51	867	403	832	20.1	2.0	1.00	
Median	10	53	11449	102	325	63	0.1	6.2	3.3	0.8	0.02		36	0.2		0.08	0.34	0.04	0.01	0.01		4.33	0.05	67	32	7.0	0.3	1.1	0.01	
GV (n = 22)																														
Mean	198	129	23058	193	723	43	0.09	14	43	44	0.36	**24446**	117	0.53	**117**	0.55	791	0.24	0.05	0.54	**792**	0.47	0.09	8.3	15	0.97	0.4	1.0	0.01	**26**
S.D.	124	177	7330	441	472	52	0.38	32	50	153	0.69		26	0.51		0.70	1627	0.86	0.08	2.5		0.57	0.10	15	28	0.5	0.8	0.2	0.00	
Median	172	61	23435	59	671	22	0.01	0.6	23	1.6	0.09		118	0.23		0.22	26	0.03	0.01	0.005		0.22	0.1	1.3	2.03	1.07	0.05	0.99	0.01	
RR (n = 14)																														
Mean	59	567	7692	755	296	45	0.02	109	12	4.2	0.01	**9540**	108	0.90	**109**	0.31	3157	0.28	0.34	0.07	**3158**	1.2	0.10	0.22	4.7	1.1	1.1	1.5	0.01	**10**
S.D.	47	187	2801	325	70	63	0.01	153	13	5.2	0.01		51	0.78		0.37	1548	0.40	0.38	0.10		1.0	0.14	0.16	12.4	0.5	1.4	0.5	0.00	
Median	40	620	7720	666	317	31	0.02	68	7.4	2.1	0.01		84	0.55		0.20	3105	0.17	0.21	0.03		0.68	0.05	0.20	0.3	1.1	0.73	1.3	0.01	
RV (n = 30)																														
Mean	100	896	17957	1190	396	983	0.01	127	22	125	0.13	**21797**	118	1.4	**120**	1.2	7261	0.31	0.33	0.01	**7263**	2.0	0.12	61	254	1.5	1.7	1.1	0.01	**321**
S.D.	89	521	11430	779	230	941	0.01	114	66	378	0.29		29	1.93		3.1	8171	0.34	0.36	0.02		2.0	0.19	71	234	0.94	2.0	0.35	0.002	
Median	92	818	18847	984	414	1019	0.01	101	5.0	1.5	0.03		116	0.44		0.45	5699	0.27	0.36	0.01		1.1	0.06	65	356	1.4	1.3	1.0	0.01	

*One analysis excluded due to exceptionally high measured Ti

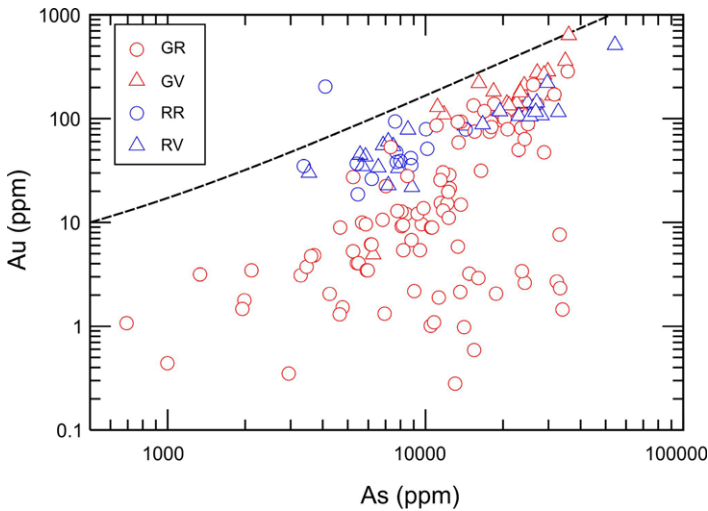Calculated using half mdl values for all measurements < mdl

Fig. 2 Au versus As plot. The solubility limit for Au in pyrite (Reich et al. 2005) is shown as a *dashed-line*. Below this line, Au occurs in solid solution, whereas above, a component of Au is present as inclusions (Au^0)

substantial evidence that the element in question is present and the value is therefore set to 0. This makes negligible difference to analyses.

The dataset shows that pyrite contains Au (mean 69.8 ppm) and up to 50 000 ppm As. Au concentrations vary within each sample and between samples; maximum values exceed 600 ppm. The lack of observed visible gold in the pyrite grains analyzed is consistent with the position of points on the Au versus As plot (Fig. 2) relative to the empirical solubility limit of Au in pyrite (Reich et al. 2005). All except two points plot below the solubility line. Further indirect constraints on the size of Au inclusions come from the Au signal on the time-resolved downhole LA-ICP-MS depth profiles which are typically smooth. Our previous work linking the appearance of the LA-ICP-MS spectra to the observed Au distribution at the micron- to nanoscale (Ciobanu et al. 2011, 2012) are consistent with the above interpretation where Au is present either in the lattice or as nanoparticles.

Basic statistical treatment (Table 2) shows that Ag, Sb and Tl concentrations are markedly higher in the RR and particularly in the RV category; Pb and Cd are anomalous in RV but not in RR. Molybdenum concentrations are also two orders of magnitude higher in the vein categories than the respective rock counterparts. Such high concentrations of these elements are unusual for pyrite. Concentrations of Co and Ni (means 201 and 155 ppm, respectively) vary over four orders of magnitude, spiking to values of over 6000 ppm Co and 2800 ppm Ni. Both elements are generally higher in the innermost zone of GR pyrite. Co and Ni concentrations are significantly lower in RR pyrite; Ni, however, is highest in RV pyrite, as is the Ni/Co ratio. Among the siderophile elements, GR pyrite contains some higher values of Ti and also V, Cr and Nb albeit at much lower concentrations. Element maps of a zoned GR pyrite (Fig. 3) show a core enriched in Co, Ni, Cu, Ag and Sb overgrown by an As- and Au-rich zone. Core-reabsorption and distinct geochemical patterns indicate at least
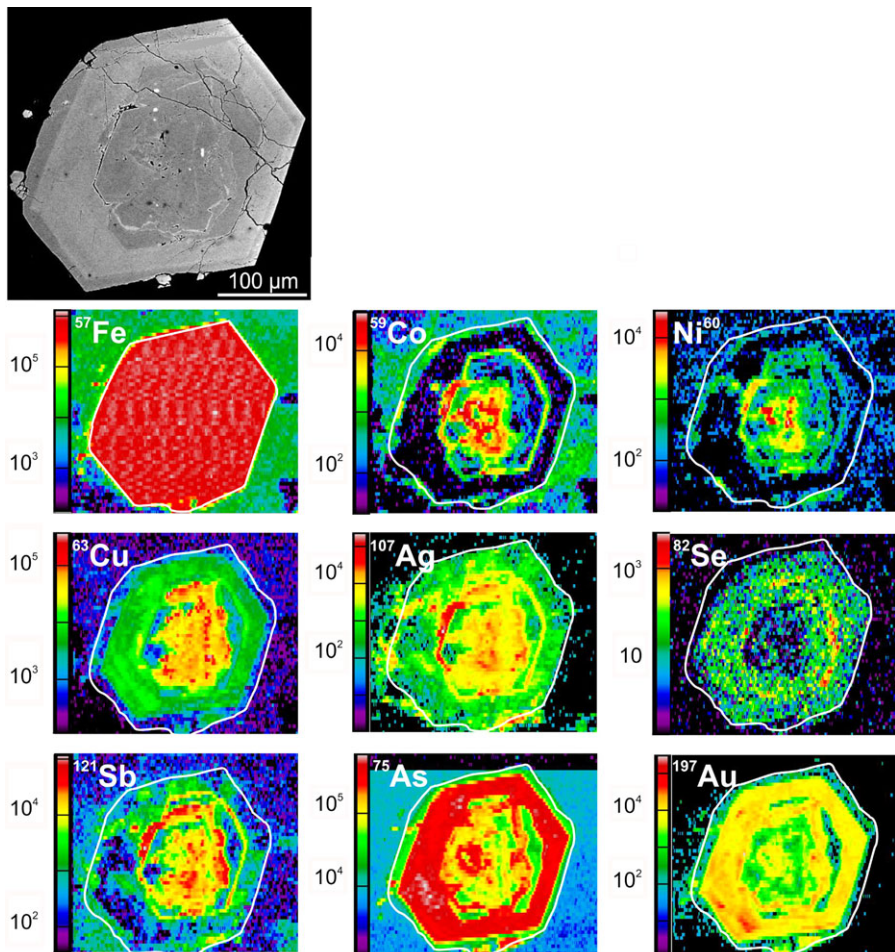
Fig. 3 Back-scatter electron image (*top*) and LA-ICP-MS element maps of zoned GR pyrite. Scale represents counts per second

two growth cycles. Element mapping of feathery RV pyrite (Fig. 4) shows a contrasting pattern in which all elements, except Cu, are homogeneously distributed. Ni is dominant over Co and Hg is also present.

3 Multivariate Statistical Analysis

We emphasize the application of statistical analysis to a geological problem in its ability to provide a succinct description of the data, highlight relationships among variables and potentially provide insight into underlying geological processes (which may not be apparent otherwise). The mathematical formalism is only briefly introduced. Further detail is given by Winderbaum (2011), based on methodology devel-
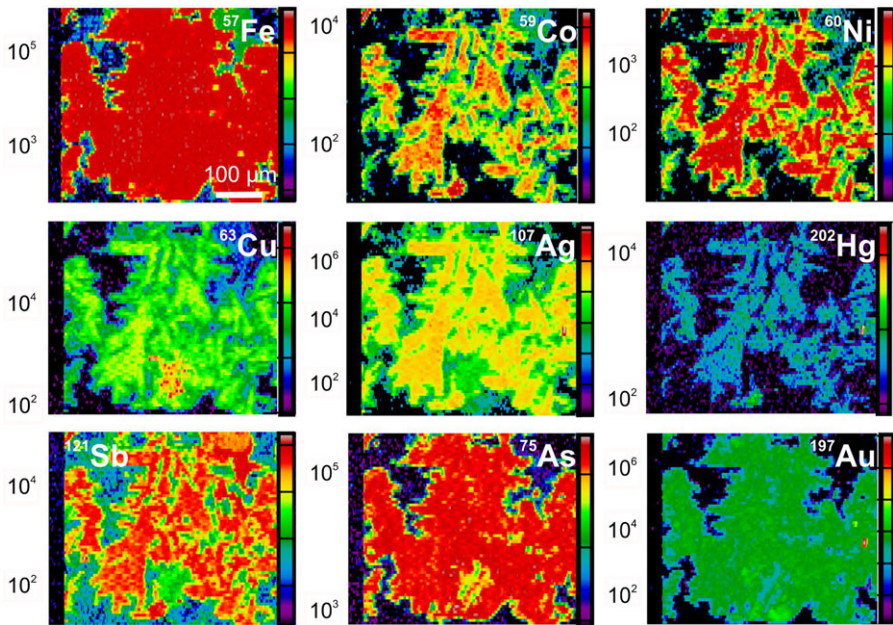
Fig. 4 LA-ICP-MS element maps of a "dirty"-feathery RV pyrite. Scale represents counts per second

oped in Koch (2012) but that is also available in most multivariate statistics textbooks (Manly 1994; Everitt and Dunn 2001).

3.1 Multivariate Analysis of Variance (MANOVA)

We have samples, which will be assumed random, from four natural populations; GR, GV, RR and RV. Let $\mu_{i,j}$ denote the population mean for element i in class j ($i = 1, \ldots, 26$ correspond to Au, ..., Nb (online Appendix B) and $j = 1, 2, 3, 4$ correspond to GR, GV, RR and RV). MANOVA tests the null hypothesis that the mean (concentration of element) vectors are the same across the four populations, specifically

$$\text{Model 1 hypothesis:} \quad \begin{pmatrix} \mu_{1,1} \\ \vdots \\ \mu_{26,1} \end{pmatrix} = \cdots = \begin{pmatrix} \mu_{1,4} \\ \vdots \\ \mu_{26,4} \end{pmatrix}, \tag{1}$$

$$\text{Model 2 hypotheses:} \quad \begin{pmatrix} \mu_{1,1} \\ \vdots \\ \mu_{26,1} \end{pmatrix} = \begin{pmatrix} \mu_{1,4} \\ \vdots \\ \mu_{26,4} \end{pmatrix} \& \begin{pmatrix} \mu_{1,2} \\ \vdots \\ \mu_{26,2} \end{pmatrix} = \begin{pmatrix} \mu_{1,4} \\ \vdots \\ \mu_{26,4} \end{pmatrix}$$

$$\& \begin{pmatrix} \mu_{1,3} \\ \vdots \\ \mu_{26,3} \end{pmatrix} = \begin{pmatrix} \mu_{1,4} \\ \vdots \\ \mu_{26,4} \end{pmatrix}. \tag{2}$$

Table 3 Summary of MANOVA results

	Au	Ag	As	Sb	Cu	Pb	Bi	Tl	Zn	Cd	In	Se	Te	Sn	Mo	W	Re	U	Mn	Ga	Co	Ni	Ti	V	Cr	Nb
Model 1: manova(Y-class)																										
(all)	***	***	***	***	*	***	***	***	***	*	*	***		***	***		***		***	°	***	***	***	°	*	**
Model 2: manova(Y-(GV+RV+RR+GR))																										
GV	***	*		***	***	**	**	*	***	*		***	***			°			***		***	**	***	*	°	°
RV	***	***	**	***		***	***	***		*		***		***	***	***	°		***		*	***		*		**
RR	***	***	°	***		**	**	***				***		***	***	***	°		***		***	***		°	*	

Significance codes [Pr($> F$)]: ***0–0.001, **0.001–0.01, *0.01–0.05, °0.5–0.1, no symbol > 0.1
Summaries produced from summary.aov() in R (www.r-project.org/), as is the notation for the MANOVA models Model 1 and Model 2 above. Pr is for probability and F is standard test statistic, as in Manly (1994) and Everitt and Dunn (2001). Class is a 4-level factor variable, with levels corresponding to GR, GV, RV or RR. Y is the multivariate responses consisting of the log-transformed concentrations of all 26 trace elements. Each of GR, GV, RV and RR are binary indicator vectors for their respective classes

These are preferable to 26 univariate tests of hypotheses of equal means because it allows for multiple comparisons without resorting to the highly conservative Bonferonni inequality. Fitting Model 1 provides evidence that the four vector means differ at the 1 % significance level. More specifically, Model 2, along with the other three pair-wise comparisons (not shown), provide evidence that all four vector means differ from each other at the 1 % significance level. This is taken into consideration in further analyses. Summaries of these results (Table 3) show that only Te, W, U, Ga and V, all elements with very low concentrations, are not significant in differentiating the classes (Model 1). Model 2 hints at more complex patterns, including that Cu and Zn are important in the differentiation between GR and GV but not between GR and the replacement classes.

3.2 Parallel Coordinate Plots

An important step in analysis is to visualize the data. With two-dimensional data (which is where only two trace elements are measured) this is easily done by producing a scatterplot. If the dimension remains small, multiple scatterplots can be produced, one for each element pair. All these scatterplots need to be considered simultaneously in order to obtain a representation of the multivariate nature of the dataset, since the proximity of two datapoints in one scatterplot no longer implies they are similar as d-dimensional geochemical vectors. This quickly becomes impractical for moderate numbers of elements, specifically for a dataset on d trace elements, there are m (Eq. (3)) such pairs

$$m = \binom{d}{2} = \frac{d!}{2!(d-2)!} = \frac{d(d-1)(d-2)!}{2!(d-2)!} = \frac{1}{2}(d^2 - d). \tag{3}$$

In these data $d = 26$, so $m = 325$. One method to overcome this difficulty in visualizing higher dimensional datasets is Parallel Coordinate Plots (Inselberg 1985). In a parallel coordinate plot the value (either concentration or log-transformed concentration) of each element is represented on a vertical line, labeled at the x-axis. Each spot

analysis is then represented by a piece-wise linear line joining d points, each on the vertical line of the corresponding element, at a height corresponding to the value of that spot analysis on that element. This gives an accurate representation of the data in the multi-dimensional (26-dimensional in this case) data-space, providing an initial visualization of the multivariate nature of the data.

We present the raw data in terms of GR and (GV+RR+RV) classes (Fig. 5A, B). It is notable that of the 26 elements, As, Mo and Ti occur at concentrations that range over the largest number of orders of magnitude. In Ti, two anomalous values (both in GR) constitute most of this variance. Arsenic consistently has the highest concentration and sets the scale of the plot, but in contrast high values of Mo are notably absent in the GR class, as observed from Table 2. Also Co, Cu, Zn, Se and Mn are visible on Fig. 5A but not on Fig. 5B; Cd, Tl and Ni are visible in Fig. 5B but not Fig. 5A.

The elements in both figures showing greatest variability are both chalcophile and granitophile in character but a pronounced variability in siderophile elements is exclusive to GR. Such differences are difficult to see because of the large variability in scale within the raw data. Log-transformed data are considered (Eq. (4)), addressing the issues of scale, and influence of outlying values

$$x_{ik}^{\text{(log-transformed)}} = \ln\left(x_{ik}^{\text{(raw)}} + 1\right). \tag{4}$$

$x_{ik}^{\text{(raw)}}$, as in Eq. (4) is the kth observation's concentration of element, i. $x_{ik}^{\text{(log-transformed)}}$ is similarly defined for the log-transformed concentration. The addition of a constant is a standard means of dealing with zero-values attributed to non-negative variables when taking logarithms. The choice of 1 for the constant added before taking the logarithm avoids negative values on the log-scale (which would confuse interpretation) and curtails the influence of zero and low-concentration values.

Parallel coordinate plots of the log-transformed data for GR, GV, RR and RV (Fig. 5C–F) reveal relationships not easily visible in Table 2. The Au–Ag relationship is notable, in that GR, RR and RV all show a consistent pattern of significantly higher Ag than Au, while GV shows the opposite trend. Siderophile concentrations (Ti, V, Co, Mn) in GR are markedly higher than other classes. The Ni–Co pair show a distinctive relationship clearly visible in all the classes except RR, where this relationship is markedly absent. Re is apparent only in granular classes, and not replacement classes.

3.3 Correlation Cluster Analysis (CCA)

We are particularly interested in finding relationships between elements, particularly unexpected relationships. An intuitive place to start in looking for such relationships is correlation coefficients. It is possible (although tedious) to consider all pair-wise correlation coefficients individually (Tables 2, 3, 4 in the online Appendix C), but this approach immediately runs into the same problem previously mentioned in that the number of such comparisons quickly becomes unwieldy (Eq. (3)). Hierarchical agglomerative cluster analysis (Everitt and Dunn 2001) can provide a visualization of correlations in a dendrogram. This is done by replacing the usual Euclidean distance with the correlation based pseudometric $1 - |\rho|$ (where ρ is the coefficient of
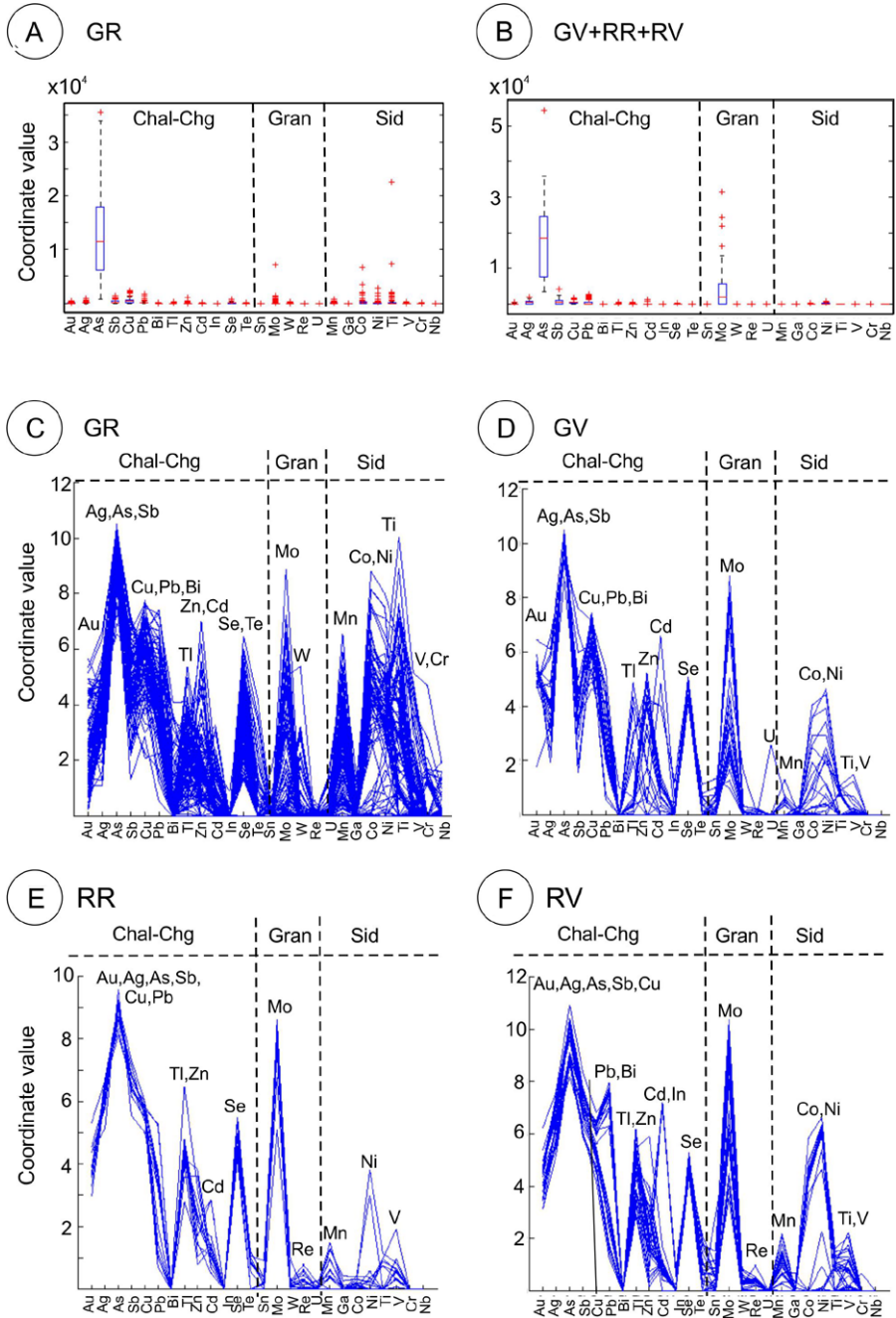
Fig. 5 Parallel coordinate plots. (**A**) and (**B**) show boxplots for the raw data for classes GR and (GV+RR+RV), respectively; (**C**)–(**F**) show the log-transformed data for the four classes GR, RV, RR, RV, respectively. Chal-chalcophile elements; Chg-chalcogens; Gran-granitophile elements; Sid-siderophile
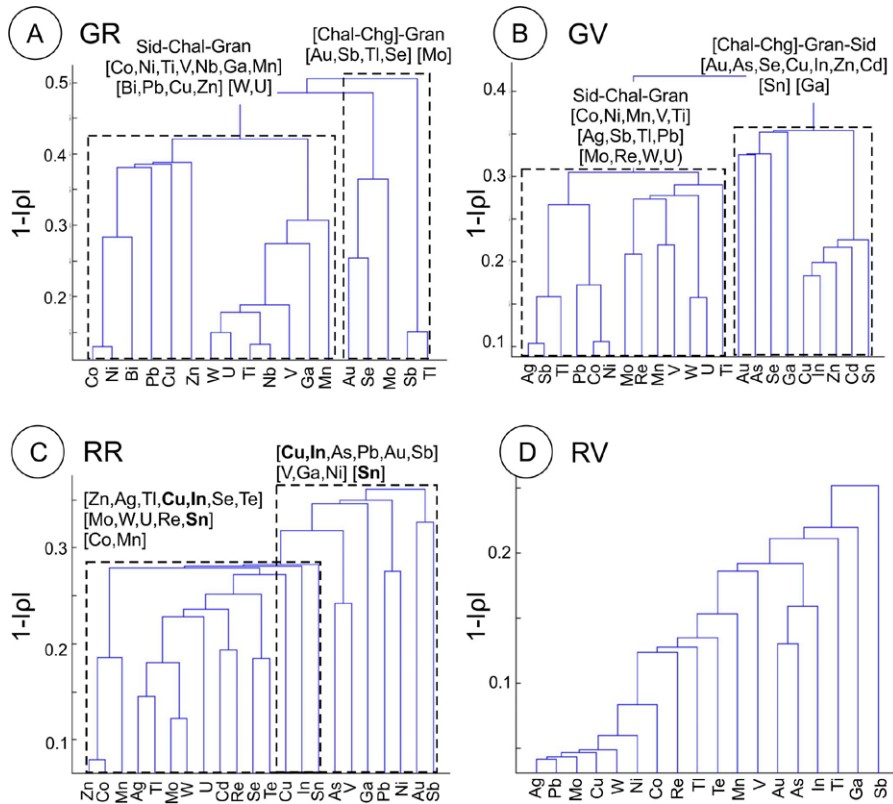
Fig. 6 Dendrograms describing a hierarchical agglomerative cluster analysis scheme by a distance based on Spearman's coefficient (ρ). (**A**)–(**D**) show this analysis for each class (GR, RV, RR and RV, respectively). Abbreviations as for Fig. 5

interest). We thus display all the relationships between all the correlation coefficients in a single, interpretable, diagram.

The dataset contains numerous extreme values (common in multi-element datasets, Reimann and Filzmoser 1999), and so the usual Pearson's correlation coefficient is not appropriate; a more robust measure of association is needed. We use Spearman's rank-based coefficient of association (Eq. (5))

$$\rho_{ij} = \frac{\sum_{k=1}^{n}(y_{ik} - \bar{y}_i)(y_{jk} - \bar{y}_j)}{\sqrt{\sum_{k=1}^{n}(y_{ik} - \bar{y}_i)^2 \sum_{k=1}^{n}(y_{jk} - \bar{y}_j)^2}}, \tag{5}$$

where i and j represent the trace elements in question, y_{ik} and y_{jk} are the rank of the kth observation on i and j, respectively (in increasing order of concentration, where tied values are given their average rank) and, clearly, $\bar{y}_i = \frac{1}{n}\sum_{k=1}^{n}k = \frac{1}{2}(n+1)$, $\forall i \in \{1, 2, \ldots, 26\}$.

Spearman's coefficient is a direct measure of monotonic association between two variables and, in contrast to Pearson's correlation coefficient, is not unduly influenced by outliers. GR (Fig. 6A) shows grouping of the majority of siderophile, chalcophile and granitophile elements. Importantly, however, elements such as Au, Sb and Tl join

a chalcophile/chalcogen group that also contains Mo and Se. It is also interesting that this grouping does not join with the dominant chalcophile group of base metals until after the siderophile grouping, implying its association to the former is stronger than the latter. GV (Fig. 6B) shows two clear groupings with strong correlations among geochemically similar elements but significant differences from those in GR, particularly for the ore-forming metals for which there seem to be two signatures: one defined by Ag–Sb–Tl–Pb, the other by Au–As–Se–Cu–In–Zn–Cd. This could be expected given the fact that GV is late relative to GR.

Both granular classes show clear groupings relative to the replacement categories (Fig. 5C, D) which both feature cascade structures. Despite this, the geochemically defined groupings can still be recognized in RR, especially the dominant chalcophile and granitophile groupings, albeit with some overlap between the two. Figure 5 emphasizes the very different signatures of each class and also indicates that the replacement categories are more internally heterogeneous than the granular categories. Replacement of different minerals, each with its own inherited geochemistry, and possibly different sets of mobilized elements, have contributed to this heterogeneity. The varying correlation structures also reflect the different element packages in the replacement and granular categories. Particularly interesting is that the elements with which Au correlates vary from class to class, and that the expected Co–Ni pairing is only obvious for GR and GV (and to some extent also for RV) reflecting the difference observed in the parallel coordinate plots. In the replacement categories, the two elements are separated into different groups or not directly correlated, further stressing different inherited signatures.

3.4 Principal Component Analysis (PCA)

Another method to approach the visualization of such multi-dimensional data is to find interesting planes to project the data into. This is equivalent to producing scatterplots, but rather than considering an infeasible (Eq. (3)) number of scatterplots, we plot a smaller number of potentially interesting (linear) combinations of elements against each other. PCA is a method for finding the (orthogonal) directions that produce maximum variance, which we call the principal components (PCs). These directions are linear combinations of elements. PCA is particularly elegant as these can be estimated directly as the eigenvectors of the sample covariance matrix (Everitt and Dunn 2001); $S = (s_{ij})$ where the entries are as in Eq. (6)

$$s_{ij} = \widehat{cov}(X_i, \ X_j) = \frac{1}{n-1} \sum_{k=1}^{n} (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j). \tag{6}$$

We use log-transformed data (as they are less influenced by outlying values, and is invariant to scale and thus more relevant), so the values in Eq. (6) refer to the data as given in Eq. (4). PC1–PC8 are given in Table 5 of online Appendix D. Notably As is only a minor component of PC1–PC8 emphasizing how despite its high concentrations, its relatively small variability makes it comparatively uninteresting in terms of discussing subgroups of the data.

PC1 and PC2 are loaded by elements of both siderophile and hydrothermal (granitophile and chalcophile) character but their structures are different. In PC1, hydrothermal elements (Mo, Au and Ag) have opposite signs to the siderophiles (Ti, Co, Ni),
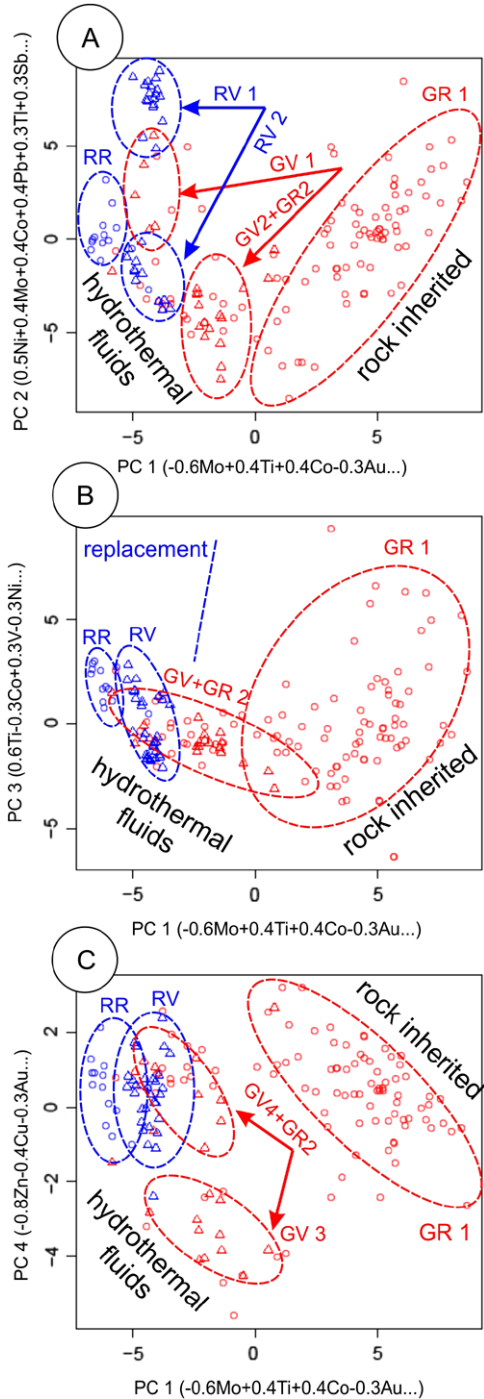
and the combination of low Mo and high Ti, Co separates GR from the replacement classes. In contrast, in PC2 they carry the same signs, and concentrations of Ni, Co, Mo and Pb contribute to differentiate between two sub-classes of RV. PC3 is siderophile-dominant (Ti, V) and does not separate the data into groups. PC4 demonstrates how the cumulative concentration of Zn, Cu, and to a lesser degree Au separate two subgroups of GV. These four PCs (Fig. 7) show interesting trends between the classes which we interpret in terms of inherited signatures and successive hydrothermal overprints. Perpendicular trends of GR and (GV+RR+RV) in Fig. 7A and B show that hydrothermal fluids, carrying Mo and Au, interacted with the signature inherited from the rock (Ti, Ni). PC2 (Fig. 7A) and PC4 (Fig. 7C) show that the hydrothermal fluids evolved during processes of replacement and vein formation (GV and RV split into subgroups in Fig. 7A; GV is split into subgroups in Fig. 7C). Splitting of RV into sub-populations is predictable considering pyrite provenance in this class from replacement of different minerals, giving a group with shared morphology but with subgroups having distinct geochemical signatures (Fig. 7A). In contrast, the homogeneous RR class (mainly wispy pyrite) is consistent with replacement of one mineral only. Splitting of GV into geochemically distinct subgroups (Fig. 7A, C) is attributable to inclusions of sphalerite and chalcopyrite when GV occurs as single grains but not when GV occurs as overgrowths on replaced pyrite (Fig. 1D). On all three diagrams in Fig. 7, a sub-population of GR pyrite is observed to be separated by PC1, in the direction of GV. Splitting of GR is attributable to overprinting by later fluids, introducing elements that coincide with one of the GV sub-populations. All these interpretations are consistent with the textural variation (Fig. 1) and element mapping (Figs. 3 and 4).

3.5 Factor Analysis (FA)

Factor analysis is a model-based method (Everitt and Dunn 2001), in which a model as in Eq. (7) is fitted

$$X = Af + \varepsilon. \tag{7}$$

In Eq. (7) X is the $d \times 1$ dimensional data vector, A is the $d \times k$ matrix of factor loadings, f is the $k \times 1$ dimensional vector of factors (also known as latent variables) whose values are not measured directly but which account for much of the measured values, and ε is the $d \times 1$ dimensional vector of errors, representing random variation which makes up the difference between the measurement and the value accounted for by the factors.

The maximum likelihood FA gives an estimate of the 26×2 matrix $A = (a_{ij})$ (the weighting of each underlying factor on each variable—elements, in this case). A is the approximate solution to $V = AA' + D$ (Rubin and Thayer 1982) where V is the variance-covariance matrix of the sample and D is a diagonal matrix representing the variances of the ε. A biplot, as shown in Fig. 8 is a plot of the factor loadings a_{ij} for factor j and element i, with the factors as axes where each element is plotted as the vector (a_{i1}, a_{i2}). Matrix A is not unique, however. The axes can be rotated, in matrix terms by applying the orthogonal transformation U, that is, $(AU)(AU)' = AUU'A' = AIA' = AA'$. This can be useful as factors corresponding to rotated axes are easier to interpret. Computer software displays default axes

Fig. 8 Biplot for maximum likelihood FA on the log-transformed data on class GR. Abbreviations as for Fig. 5

which align best with groups of elements. Models with more than two factors can be fitted but the instructive graphical display is lost and interpretation may be more awkward. The investigator can choose the number of factors to provide a useful physical interpretation.

Meaningful interpretations with basis in geochemical reasoning can be inferred from Fig. 8 (GR only). There are two directions of interest: (i) in the direction of the strong Co–Ni pair (siderophile elements with hydrothermal affinity), coinciding with the direction of the factor labeled component 1; and (ii) perpendicular to direction (i), coinciding with the direction of a group of granitophile elements including Sn, W and V, roughly coinciding with the direction of the factor labeled component 2. There appear to be five broad element groupings in Fig. 8, three agreeing with (i), two in the opposite direction. Each of the three groupings agreeing with (i) have two notable subgroups, one with greater magnitude than the other, in particular base metals (Cu–Pb–Cd–Bi) show affinity to the strong Co–Ni pair, Zn to granitophile elements, and Tl–Sb–Ga to Mn–Ti. This indicates that the subgroups with larger magnitude (Co–Ni, Ti–Mn, and Sn–W–U–V–Nb–Cr) can be largely explained by the factors in this model, while the subgroups with smaller magnitudes (all consisting of chalcophile elements) are also explained by the same factors/processes, but are significantly influenced by other factors not taken into account in this model. The groupings in the opposite direction to (i) (chalcophile elements, chalcogens and granitophile elements with hydrothermal affinity (Mo–Re)) are interesting since one of them (including Mo–Re) seems to have a strong negative component in direction (i) but not in direction (ii), indicating it is affected by the same factor as the Co–Ni pair (direction (i)),

but in an opposite sense, and is not influenced by the factor in direction (ii). The other grouping, (As–Ag) appears to be similarly negatively affected by direction (i), but is positively affected by direction (ii).

Given this distribution, direction (i) is attributable to at least two generations of hydrothermal fluids that have different affinities (hence different signs for Co–Ni and Mo–Re). Base metals (Cu, Pb, Zn, etc.) and Au–Mo define these, emphasizing the opposing trends observed for GR on Fig. 7. Direction (ii) is attributable to pyrite signatures largely inherited from the magmatic rock. Clusters controlled by both directions (As, Ti, Mn, etc.) then reflect modification of signatures during overprinting. This interpretation is based on fluid sources but other factors explaining the distributions, particularly the subgroups of lower magnitude, could include geochemical behavior of each element (Co and Ni for example) and pyrite crystal chemistry which governs element incorporation and release during mineral growth (Ti, Mn, Sb for example).

4 Discussion

In pyrite, elements such as Co, Ni, Au, As, Sb, Se, and probably Mn and Ag, are considered to enter the crystal lattice, whereas others (Nb, Ti, Cr, W, U) are most likely to occur within micro- to nanoscale inclusions of discrete minerals. A degree of correlation between members of each group might be expected but, as borne out by the element maps, grain-scale compositional zoning implies that different parts of the grain are enriched in different elements (for example the Au–As and Ag–Sb distributions in Fig. 3), thus weakening the overall correlation. Many of the strong associations recognized can be directly related to similar geochemical affinity or elements with similar charge and ionic radii (for example Co and Ni, Zn and Cd or Mo and Re). Inclusions of discrete minerals within growing pyrite (for example Ti and Nb in rutile) could also contribute to the high variability observed among the siderophile elements.

The PC plots begin to show the multivariate nature of the spot analyses. They provide not only a representation of the basic statistics (Table 2) but a description of the data that allows for easy interpretation not available in the basic statistics, for example highlighting the Au–Ag relationship as markedly different in GV compared to any of the other groups, and the absence of the Co–Ni interaction in RR pyrite. MANOVA provides formal evidence that the four categories identified a priori (GR, GV, RR, RV) are significantly geochemically distinct, and hints at more complex inter-element relationships (Table 3).

Although the PC plots provide insight into some of the relationships among elements, they are far from comprehensive, and so dendrograms are introduced to systematically consider all pair-wise associations in a single plot, revealing interesting differences in the association structure of the elements in the classes, for example the signatures in granular pyrite, and the absence of such groupings in replacement pyrite, highlighting the heterogeneity in the latter caused by replacement of different minerals.

Multivariate techniques show element associations that assist interpretation of Moonlight mineralization in a context of complex geochemical signatures overlapping one another or discriminating smaller sub-populations. PCA demonstrates the

clear presence of geochemically distinct subgroups within the simple classes of pyrite (Table 1), as expected, thus drawing attention to the heterogeneity in these classes and can even go on to quantify some of the geochemical vectors (PCs) that are capable of discriminating rock-inherited from fluid-introduced signatures. The FA (Fig. 8) visualizes some of these rock-inherited from fluid-introduced signatures in a more interpretable manner, and highlights both the strength of these fluid-based interpretations in explaining elemental composition.

In terms of relevance for a genetic model with application to exploration, there are three groups of elements: base metals (Cu, Pb, Zn, Ag, Sb, etc.); Au–As; and Mo, which are all traceable through the multivariate statistical analysis of pyrite. The signature of core GR pyrite may reflect elements supplied by the fluid as well as those inherited from minerals in the precursor rock, notably Co and Ni. This is reflected in the splitting of the GR class by PCA (Fig. 7) and FA (Fig. 8). Copper and Ag input corresponds to a fluid predating Au and As input (core versus overgrowth patterns in Fig. 3). Geochemical variability in vein pyrite is suggestive of interaction between later fluids and pre-existing rock pyrite, as well as replacement of a wider range of minerals than in the rock. The varied intensity of this interaction generates the heterogeneous GV and RR pyrite. Texture-inferred replacement of molybdenite (among other minerals) by pyrite is strongly supported by multivariate statistics which shows Mo and Tl markedly enriched in one sub-population of RV pyrite. This implies that Moonlight may feature telescoped porphyry-Mo and epithermal-Au ($\pm$ base metal) mineralization, an important finding for this type of mineral system. Geochemical vectors based on the PCAs could be used in development of exploration models for Pajingo and analogous orefields.

5 Conclusions

The methods presented here are valuable tools for investigating large datasets on a mineral with a protracted history of interaction with fluids giving a complex geochemical signature. Analysis of variance (MANOVA) indicates that there is a strong statistical difference in the majority of elements across the four classes. Parallel coordinates provide a first assessment of the scale of variance among a large set of elements whereas CCA can draw attention to element associations that may not be immediately obvious. PCA and FA were able to confirm that observed mineral groupings have distinct geochemical signatures and that some can be further sub-divided.

No single statistical method is, in itself, sufficient to analyze all aspects of the dataset. Dataset size and inherent heterogeneity are both critical factors, especially in cases such as this, where distinctive sub-populations exist. Further studies using multivariate analysis on larger datasets are warranted to demonstrate the robustness of these methods. These conclusions notwithstanding, a sound understanding of inorganic geochemistry, and, in particular, crystal chemistry of the minerals concerned, are essential for correct interpretation of multi-element datasets for minerals.

References

Bailey TC, Krzanowski WJ (2012) An overview of approaches to the analysis and modeling of multivariate geostatistical data. Math Geosci 44:381–393

Baker T, Mustard R, Brown V, Pearson N, Stanley CR, Radford NW, Butler I (2006) Textural and chemical zonation of pyrite at Pajingo: a potential vector to epithermal gold veins. Geochem, Explor Environ Anal 6:283–293

Bobis RE, Jaireth S, Morrison GW (1995) The anatomy of a Carboniferous epithermal ore shoot at Pajingo, Queensland: setting, zoning, alteration, and fluid conditions. Econ Geol 90:1776–1798

Ciobanu CL, Cook NJ, Pring A, Brugger J, Danushevsky L, Shimizu M (2009) 'Invisible gold' in bismuth chalcogenides. Geochim Cosmochim Acta 73:1970–1999

Ciobanu CL, Cook NJ, Utsunomiya S, Pring A, Green L (2011) Focussed ion beam-transmission electron microscopy applications in ore mineralogy: bridging micro- and nanoscale observations. Ore Geol Rev 42:6–32

Ciobanu CL, Cook NJ, Utsunomiya S, Kogawa M, Green L, Gilbert S, Wade B (2012) Gold-telluride nanoparticles revealed in arsenic-free pyrite. Amer Mineral 97: 1515–1518

Cook NJ, Chryssoulis SL (1990) Concentrations of "invisible gold" in the common sulphides. Can Mineral 28:1–16

Cook NJ, Ciobanu CL, Pring A, Skinner W, Danyushevsky L, Shimizu M, Saini-Eidukat B, Melcher F (2009a) Trace and minor elements in sphalerite: a LA-ICP-MS study. Geochim Cosmochim Acta 73:4761–4791

Cook NJ, Ciobanu CL, Mao JW (2009b) Textural control on gold distribution in As-free pyrite from the Dongping, Huangtuliang and Hougou gold deposits, North China Craton (Hebei Province, China). Chem Geol 264:101–121

Cook NJ, Ciobanu CL, Danyushevsky L, Gilbert S (2011) Minor and trace elements in bornite and associated Cu–(Fe)-sulfides: a LA-ICP-MS study. Geochim Cosmochim Acta 75:6473–6496

Danyushevsky L, Robinson P, Gilbert S, Norman M, Large R, McGoldrick P, Shelley M (2011) Routine quantitative multi-element analysis of sulphide minerals by laser ablation ICP-MS: standard development and consideration of matrix effects. Geochem, Explor Environ Anal 11:51–60

Everitt BS, Dunn G (2001) Applied multivariate data analysis, 2nd edn. Arnold, Sevenoaks, 342 pp

Inselberg A (1985) The plane with parallel coordinates. Vis Comput 2(2):69–91

Koch I (2012) Analysis of multivariate and high-dimensional data theory and practice. Cambridge University Press, Cambridge

Large RR, Maslennikov VV, Robert V, Danyushevsky LV, Chang ZS (2007) Multistage sedimentary and metamorphic origin of pyrite and gold in the giant Sukhoi Log deposit, Lena Gold Province, Russia. Econ Geol 102:1233–1267

Large RR, Danyushevsky LD, Hollit C, Maslennikov VV, Meffre S, Gilbert S, Bull S, Scott R, Emsbo P, Thomas H, Singh R, Foster J (2009) Gold and trace element zonation in pyrite using a laser imaging technique; implications for the timing of gold in orogenic and Carlin-style sediment-hosted deposits. Econ Geol 104:635–668

Manly BFJ (1994) Multivariate statistical methods: a primer. Chapman & Hall, London, 215 pp

Maslennikov VV, Maslennikova SP, Large RR, Danyushevskiy LV (2009) Study of trace element zonation in vent chimneys from the Silurian Yaman-Kasy volcanic-hosted massive sulfide deposit (Southern Urals, Russia) using laser ablation-inductively coupled plasma mass spectrometry (LA-ICPMS). Econ Geol 104:1111–1141

Paul M (2010) Geology, geochemistry and mineralogy of epithermal gold ores. Moonlight Prospect, Pajingo, North Queensland. Unpubl Hons thesis, School of Earth and Environm Sci, Univ Adelaide, Australia, 54 pp

Ramdohr P (1980) The ore minerals and their intergrowths. Pergamon, New York, 1207 pp

Reich M, Kesler SE, Utsunomiya S, Palenik CS, Chryssoulis SL, Ewing RC (2005) Solubility of gold in arsenian pyrite. Geochim Cosmochim Acta 69:2781–2796

Reimann C, Filzmoser P (1999) Normal and lognormal data distribution in geochemistry: death of a myth. Consequences for the statistical treatment of geochemical and environmental data. Environ Geol 39:1001–1014

Richards DR, Elliot GJ, Jones BH (1998) Vera North and Nancy gold deposits. Pajingo. Geology of Australia and Papua New Guinea mineral deposits. AusIMM, Melbourne, pp 685–690

Rubin DB, Thayer DT (1982) EM algorithms for ML factor analysis. Psychometrika 47:69–76

Sung YH, Brugger J, Ciobanu CL, Pring A, Skinner W, Nugus M (2009) Invisible gold in arsenian pyrite and arsenopyrite from a multistage Archaean gold deposit: Sunrise Dam, Eastern Goldfields Province, Western Australia. Miner Depos 44:765–791

Thomas HV, Large RR, Bull SW, Maslennikov V, Berry RF, Frazer R, Froud F, Moye R (2011) Pyrite and pyrrhotite textures and composition in sediments, laminated quartz veins, and reefs at Bendigo Gold Mine, Australia: insights for ore genesis. Econ Geol 106:1–31

Winderbaum L (2011) Multivariate analysis of trace elements in pyrite. Unpubl hons thesis, School of Math Sci, Univ Adelaide, Australia. 90 pp

Ye L, Cook NJ, Ciobanu CL, Liu YP, Zhang Q, Liu TG, Gao W, Yang YL, Danyushevskiy L (2011) Trace and minor elements in sphalerite from base metal deposits in South China: a LA-ICPMS study. Ore Geol Rev 39:188–217