

## Cassiterite fingerprinting by LA-ICP-MS

Cite this: *J. Anal. At. Spectrom.*, 2013, **28**, 1247

Hans-Eike Gäbler,\* Sönke Rehder, Andreas Bahr, Frank Melcher and Simon Goldmann

Cassiterite ( $\text{SnO}_2$ ) has been specified as a 'conflict mineral' by a U.S. Government Act. This requires all companies that use tin in their products to verify that this tin does not originate from conflict regions in the Democratic Republic of the Congo where an ongoing violent conflict is fuelled by those minerals. Possible strategies to meet these requirements are based on written documents which certify the origin of the minerals. Direct chemical analyses of cassiterite, with the aim to confirm the documented sources, support those strategies in the case of doubt of the available documents. The German Federal Institute for Geosciences and Natural Resources (BGR) developed a geochemical fingerprinting method for cassiterite based on LA-ICP-MS. Single grains from cassiterite concentrates are analysed by LA-ICP-MS. Samples in question are checked against a database where data from samples of known origin are stored. The Kolmogorov–Smirnov statistic is used to demonstrate that independent samples taken from the same location show empirical cumulative distribution functions of element concentrations which are more similar to each other compared to those of samples taken from different locations. This gives confidence that cassiterite fingerprinting based on a comparison of a cassiterite sample in question with a database can be successful. The Wilcoxon rank sum test is used to establish a 'ranking of similarity' between a sample in question and samples from a database. This is used to assess whether the declared origin of the sample in question is credible or not.

Received 28th March 2013

Accepted 20th May 2013

DOI: 10.1039/c3ja50106j

[www.rsc.org/jaas](http://www.rsc.org/jaas)

### Introduction

Cassiterite ( $\text{SnO}_2$ ), the most important ore mineral for tin, is one out of four ore mineral commodities which have been specified as 'conflict minerals' by section 1502 of the Dodd-Frank Wall Street Reform and Consumer Protection Act.<sup>1</sup> This Act requires all companies that use 'conflict minerals' in their products to conduct due diligence to verify that these minerals did not originate from conflict regions in the Democratic Republic of the Congo (DRC), where an ongoing violent conflict is fuelled by illegal mining, taxation, and trade of those minerals. If minerals originate from conflict regions the companies have to determine whether or not they benefited armed groups.<sup>2</sup> The companies have to report their efforts to determine the mine or location of origin of the 'conflict minerals' with the greatest possible specificity. A possible strategy to meet Dodd-Frank's reporting requirements is a process based on a 'closed pipe' principle in which the ore mined from a single site is traced along its secure closed supply chain to the end customer's equipment.<sup>3</sup> Alternatively, tagging systems using security seals to trace back the origin of the minerals are applied.<sup>4</sup> Direct analyses of minerals with the aim to identify the regional sources of the materials are

measures to support those strategies in case of doubt/uncertainty of the accompanying documents or to serve as independent proofs of origin.

Cassiterite deposits are formed by magmatic processes related to the intrusion of evolved magmas<sup>5</sup> and hydrothermal processes.<sup>6</sup> Simply spoken, cassiterite either crystallises from melts (pegmatite deposits) or from hot aqueous solutions (vein deposits). Pegmatite deposits form from residual melts and are enriched in incompatible elements (such as tin) during fractionation. In contrast, cassiterite-bearing vein deposits are tabular bodies of variable extent. They have been formed by interaction of the hydrothermal solution and the host-rock.

In general, cassiterite does not appear as a homogeneous 'rock' but as heterogeneously distributed cassiterite grains (<1 mm up to several cm in diameter) within a solid matrix of low commercial value.<sup>7</sup> The task of the miner or a mineral processing plant is to concentrate these valuable cassiterite grains from the less valuable matrix. Those concentrates which still do contain matrix minerals from less than one weight percent up to several tens in weight percent are called cassiterite concentrates and are traded commercially. These concentrates are the samples for which a geochemical fingerprint is needed. Cassiterite grains are not homogeneous in composition. Element concentrations vary within single grains and to a greater extent between grains of the same origin.

As a consequence of this the following points have to be taken into account if a geochemical fingerprint from cassiterite

Bundesanstalt für Geowissenschaften und Rohstoffe, Stilleweg 2, 30629 Hannover, Germany. E-mail: [hans-eike.gaebler@bgr.de](mailto:hans-eike.gaebler@bgr.de); Fax: +49 511 643 3662; Tel: +49 511 643 2586

concentrates shall be deduced. First of all, only cassiterite grains should be analysed, because the ratio between cassiterite grains and barren matrix mineral grains within one concentrate is not only source specific but also the result of the miner's or mineral processing plant's skills to separate cassiterite from barren material. Second, the heterogeneity of the cassiterite grains within one concentrate must be met by analyzing a sufficient number of single cassiterite grains and third, the applied statistics for the comparison of samples should be based on the samples' distribution of element concentrations.

LA-ICP-MS has been applied for discriminating and fingerprinting steel,<sup>8</sup> glass,<sup>9,10</sup> diamond,<sup>11,12</sup> gold,<sup>13</sup> columbite-tantalite (coltan),<sup>14,15</sup> cannabis crops,<sup>16</sup> cotton fibre,<sup>17</sup> polycarbonate headlamp lenses<sup>18</sup> or bone and teeth samples.<sup>19</sup> A similar geochemical fingerprinting method for cassiterite based on LA-ICP-MS data shall be applied if doubt arises on the declared origin of a cassiterite concentrate. To dispel or confirm this doubt, LA-ICP-MS data from a concentrate in question are checked against a database where LA-ICP-MS data from concentrates of known origin are stored.

A prerequisite for the successful application of LA-ICP-MS for cassiterite fingerprinting is that it can be demonstrated that two independently taken cassiterite concentrates (samples) from the same deposit are more similar to each other than two cassiterite concentrates from different locations. This would give confidence that an identification of the origin can be successful, if LA-ICP-MS data of samples in question are compared with a database. From a forensic point of view a comparison problem has to be solved rather than a classification problem.<sup>20</sup>

One aim of this work is to check the above-mentioned prerequisite. The measure of similarity is given by the Kolmogorov–Smirnov statistic. Another aim is to use the analytical results of a cassiterite concentrate in question to verify the origin which is given in the accompanying document. A statistic obtained from the Wilcoxon rank sum test is used to rank cassiterite concentrate samples from a database according to their similarity with a sample in question. This ranking procedure may be used to evaluate the analytical results in terms of a geochemical fingerprint for cassiterite concentrate samples.

## Experimental

### Samples

This study comprises a database of 185 cassiterite concentrate samples from 155 locations in 20 countries worldwide with special emphasis on Central Africa (110 concentrate samples). For 43 concentrates from 13 locations the database contains at least one independently sampled concentrate from the same location (hereinafter referred to as 'brother samples'). For Central Africa brother samples from the following locations have been available (number of brother samples is given in parentheses): Rwanda: Nemba (2), Gahapfu (3), Remera (3), Kiyanja (2), Ruhanga (2), Nyagasagara pegmatite deposit (6), Nyagasagara vein deposit (3) and Democratic Republic of the Congo: Manono (3). For South America brother samples from

Bolivia: Japo (2), Caracoles (3), Huanuni (4), Peru: San Rafael (8) and Brazil: Oriente Novo (2) have been available.

### Sample preparation

The samples for cassiterite fingerprinting were taken from mineral concentrates which were obtained from ore-bearing rocks by mineral processing. In this study they originated from artisanal small scale mining where gravity concentration was used for mineral processing as well as from industrial processing plants where magnetic separation was used additionally. From each concentrate a polished section (30 × 30 mm size) was prepared by mounting representative grain aliquots in epoxy resin, grinding, and stepwise diamond polishing (minimum diamond grain size: 0.25 μm) on either lead or cloth laps. The polished sections were used for mineral liberation analysis (MLA)<sup>21</sup> and LA-ICP-MS.

### Grain identification

Each polished section contains about 100 up to several thousand single mineral grains. Depending on the efficiency of the mineral processing step different degrees of purity (varying from 10% to 95% purity) concerning the mineral cassiterite are obtained. LA-ICP-MS analysis is done only on cassiterite grains so it is necessary to identify the mineralogy of the single grains in each polished section prior to LA-ICP-MS analysis. For many polished sections information from mineral liberation analysis (MLA) is available indicating the mineralogy of each single grain together with the grain coordinates within the polished section.<sup>21</sup> The positions of the cassiterite grains within the polished section can be transferred to the laser ablation system. In cases where no MLA analysis is available a quick LA-ICP-MS method (see below) is applied prior to the main LA-ICP-MS analyses to identify cassiterite grains.

### Laser ablation ICP-MS

A sector field ICP-MS instrument (Element XR, ThermoFisher Scientific) coupled to a 193 nm excimer laser ablation system (New Wave UP193-FX) was used. Both instruments were run by instrument specific control software. A trigger signal sent out by the laser control software started ICP-MS data acquisition. Polished sections of mineral concentrates were used as samples. About 40 to 50 mineral grains previously identified as cassiterite were ablated. The laser system was adjusted to ablate a line (size 50 μm × 175 μm) at 10 Hz and energy densities between 4 and 7 J cm<sup>-2</sup>. The line was ablated eight times by moving the 50 μm spot with a velocity of 5 μm s<sup>-1</sup>. Prior to this cleaning of the area to be ablated was carried out by moving a 75 μm spot with a velocity of 15 μm s<sup>-1</sup> along the line (10 Hz, energy density between 4 and 7 J cm<sup>-2</sup>). Helium was used as carrier gas (200 mL min<sup>-1</sup>).

The sector field ICP-MS instrument was used in low ( $M \Delta M^{-1} = 400$ ) and medium ( $M \Delta M^{-1} = 4000$ ) resolution mode. The plasma power was set to 1200 W. Cool, auxiliary and sample gas flows were adjusted to 16, 0.80, and 1.1–1.3 L Ar min<sup>-1</sup>, respectively. In low resolution mode the isotopes <sup>69</sup>Ga, <sup>73</sup>Ge, <sup>74</sup>Ge, <sup>75</sup>As, <sup>88</sup>Sr, <sup>89</sup>Y, <sup>93</sup>Nb, <sup>95</sup>Mo, <sup>137</sup>Ba, <sup>139</sup>La, <sup>140</sup>Ce, <sup>142</sup>Ce, <sup>141</sup>Pr,

$^{146}\text{Nd}$ ,  $^{147}\text{Sm}$ ,  $^{153}\text{Eu}$ ,  $^{155}\text{Gd}$ ,  $^{159}\text{Tb}$ ,  $^{163}\text{Dy}$ ,  $^{165}\text{Ho}$ ,  $^{166}\text{Er}$ ,  $^{169}\text{Tm}$ ,  $^{172}\text{Yb}$ ,  $^{175}\text{Lu}$ ,  $^{178}\text{Hf}$ ,  $^{181}\text{Ta}$ ,  $^{182}\text{W}$ ,  $^{201}\text{Hg}$ ,  $^{204}\text{Pb}$ ,  $^{206}\text{Pb}$ ,  $^{207}\text{Pb}$ ,  $^{208}\text{Pb}$ ,  $^{209}\text{Bi}$ ,  $^{232}\text{Th}$ , and  $^{238}\text{U}$  were analysed, whereas in medium resolution mode  $^{23}\text{Na}$ ,  $^{24}\text{Mg}$ ,  $^{27}\text{Al}$ ,  $^{29}\text{Si}$ ,  $^{44}\text{Ca}$ ,  $^{45}\text{Sc}$ ,  $^{47}\text{Ti}$ ,  $^{51}\text{V}$ ,  $^{52}\text{Cr}$ ,  $^{55}\text{Mn}$ ,  $^{56}\text{Fe}$ ,  $^{59}\text{Co}$ ,  $^{60}\text{Ni}$ ,  $^{62}\text{Ni}$ ,  $^{63}\text{Cu}$ ,  $^{66}\text{Zn}$ ,  $^{68}\text{Zn}$ ,  $^{90}\text{Zr}$ ,  $^{111}\text{Cd}$ ,  $^{113}\text{In}$ ,  $^{118}\text{Sn}$ , and  $^{121}\text{Sb}$  were analysed. The Element XR instrument was equipped with a detector system which combined a secondary electron multiplier (SEM) with a single Faraday collector to give an extended linear dynamic range. Signal intensities from about 5 to  $10^{11}$  counts per second (cps) can be measured with this detector system. The detection system was set up in such a way that for all isotopes both the SEM detector and Faraday detector were used depending on the count rate. Switching between the detector modes was done automatically. For most trace elements in low resolution mode small mass windows of 22% of the nominal mass range of the analysed isotope were used. For elements which are supposed to give intensities large enough to be detected by the Faraday detector mode, broader mass windows of 120% were used (Nb, Hf, Ta, W, Hg, Pb, Bi, Th, and U). A broader mass window was necessary to use the automated detector switching mode for which the detection of a peak shoulder is necessary. For the elements analysed in medium resolution mode a mass window of 100% was used with the exception of Si and Sn where a broader mass window of 150% was used. A single scan over the whole mass range in both resolution modes was done in 12 s. The integration time of the different isotopes of a single scan was set between 0.05 s and 0.2 s. For isotopes which were expected to give lower intensities, longer integration times were selected. 30 scans were run for the analysis of one line. During the first 140 s of the data acquisition the laser beam path was blocked by a shutter to analyse the background signal. The raw data (sequence of calibration materials and samples, intensities of all isotopes of all scans) were exported to a data-handling software tool implemented in JAVA using the libraries JFreeChart,<sup>22</sup> commons Math,<sup>23</sup> and POI.<sup>24</sup> The mean of the signal from scan 2 to 11 was used as the background signal. The median of scans 14 to 28 was used as the signal from the sample from which the background was subtracted to get net intensities.

The concentrations of the elements were calculated based on external calibration against the calibration material NIST SRM 610 with the values of Pearce *et al.*<sup>25</sup> using  $^{118}\text{Sn}$  as the internal standard. The Sn concentration of the samples was calculated from the known elemental stoichiometry ( $\text{SnO}_2$ ) of cassiterite.<sup>26</sup> Detection limits were calculated from the observed variability of the background intensities. Three times the standard deviation of the background intensities was used to define the detection limit. Detection limits for most elements can be taken from the 5th percentile column of the summary statistic in Table 1.

The quick LA-ICP-MS method for grain mineralogy identification consisted of a continuous monitoring of the isotopes  $^{44}\text{Ca}$ ,  $^{55}\text{Mn}$ ,  $^{56}\text{Fe}$ ,  $^{93}\text{Nb}$ ,  $^{118}\text{Sn}$ ,  $^{181}\text{Ta}$  and  $^{182}\text{W}$  in medium resolution mode of the ICP-MS while the LA system ablated about 60 to 100 single grains sequentially. Prior to the application of the quick method 60 to 100 grains were selected visually according to optical mineral properties (such as reflectance and colour) using the microscope of the laser ablation system and stored as

**Table 1** Summary statistics of element concentrations from 8014 cassiterite grains (185 cassiterite concentrates from 138 locations from 20 countries worldwide)

Element	5th percentile [mg kg <sup>-1</sup> ]	50th percentile [mg kg <sup>-1</sup> ]	95th percentile [mg kg <sup>-1</sup> ]
Al	7	108	1540
As	<0.2	0.3	4.4
Ba	<0.1	<0.1	1.4
Bi	<0.01	0.01	0.37
Ca	<7	<7	35
Cd	<0.15	<0.15	0.20
Ce	<0.02	<0.02	0.72
Cr	<0.7	0.7	7.0
Cu	<0.03	<0.03	0.82
Dy	<0.05	<0.05	0.21
Er	<0.01	0.01	0.19
Eu	<0.01	<0.01	0.04
Fe	121	1605	9368
Ga	0.13	2.62	34.09
Gd	<0.13	<0.13	0.21
Ge	<0.20	0.20	1.99
Hf	<0.1	39.6	439.2
Ho	<0.01	<0.01	0.05
In	<0.2	0.2	44.8
La	<0.01	0.01	0.42
Lu	<0.01	0.01	0.17
Mg	<0.1	<0.1	23.5
Mn	<0.4	12.0	1005.5
Mo	<0.01	<0.01	0.39
Na	<13	<13	<13
Nb	<1	990	19992
Nd	<0.02	<0.02	0.39
Pb	0.02	0.41	7.21
Pr	<0.01	0.01	0.09
Sb	0.2	0.4	113.0
Sc	<0.04	4.65	136.14
Sm	<0.03	<0.01	0.13
Sr	<0.02	0.02	1.20
Ta	<1	880	43532
Tb	<0.06	0.09	0.24
Th	0.001	0.007	0.722
Ti	15	805	5532
Tm	<0.01	<0.01	0.05
U	0.11	2.21	23.21
V	<0.09	2.59	95.35
W	0.8	26.5	3165.7
Y	<0.01	0.03	1.09
Yb	<0.01	0.01	0.66
Zn	<0.6	0.9	46.3
Zr	0.2	392.5	1965.1

a sequence in the laser ablation software. During the quick method each grain was ablated for 2 s by a laser spot (50  $\mu\text{m}$ ) with the settings given above followed by a washout period of 10 s with no ablation. While the laser system ablated grain after grain separated by the 10 s washout periods the ICP-MS system continuously monitored the intensities of the above given isotopes. The mineralogy of each grain was indicated by large intensities of the isotopes of the mineral's main components. Only grains which gave large  $^{118}\text{Sn}$  intensities were regarded as cassiterite grains and used for LA-ICP-MS analyses for major and trace elements.

## Statistics

For each sample multi element analyses of about 40 to 50 grains were available. Hence, if 45 elements are analysed, each sample consists of 1800 ( $45 \times 40$ ) up to 2250 ( $45 \times 50$ ) data. Element-specific empirical distribution functions were used to describe the concentrations of a given element within a sample. The Kolmogorov–Smirnov statistic (K–S  $D$ , maximum distance between two empirical cumulative distribution functions) was used to compare two samples for a given element.

Element-specific  $p$ -values calculated from the Wilcoxon rank sum test were used as a statistic to set up a ‘ranking of similarity’ for a sample in question. For this ranking the sample in question was compared to each sample from the database by applying element-specific Wilcoxon rank sum tests. The lowest rank was assigned to the ‘most similar’ sample from the database while the highest rank was assigned to the ‘least similar’ one. This ranking was done to account for the similarity of two compared samples but also to account for the rarity of the observed distributions of the element concentrations within the population of cassiterite concentrates.

Linear discriminant analysis (LDA) was used to distinguish ‘brother comparisons’ (comparisons of two samples from the same location) from ‘non-brother comparisons’ (comparisons of two samples from different locations). The data vectors in this analysis consisted of the element-wise K–S  $D$  values from pairs of samples and not of any sample characteristics. Two groups had to be discriminated: the group of all brother comparisons and the group of all non-brother comparisons. In discriminant analysis the multivariate distribution for each class is estimated. If prior probabilities for class membership and the costs for misclassification are assumed equal, an unknown sample is assigned to the class with the highest value of the density function. In order to estimate the distributions (multivariate normal) in linear discriminant analysis, the means were calculated for each class separately; the covariance matrix, however, was calculated from the combined dataset.

In order to visualize multivariate data principal components analysis (PCA) was applied. The input data were zero centred and scaled to have unit variance.

All calculations were done using R.<sup>27</sup>

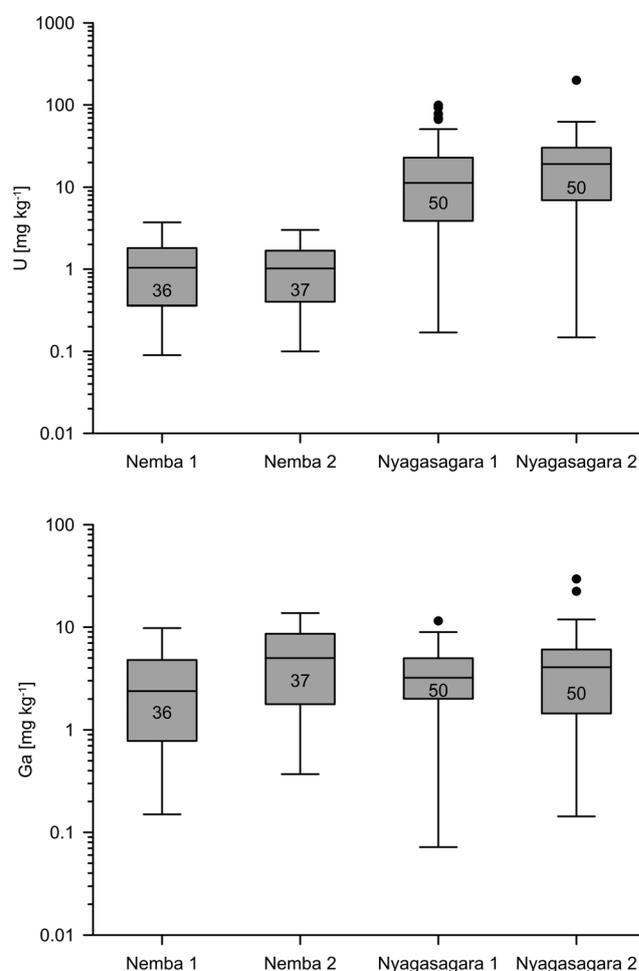
## Results and discussion

### LA-ICP-MS data

The applied non-matrix matched calibration strategy (glass calibration material NIST SRM 610; internal standard Sn) implicates that the resulting element concentrations may deviate systematically from the accurate values of the analysed cassiterite grains.<sup>28,29</sup> Non-matrix matched calibration with NIST SRM 610 of Fe-based samples by a laser system comparable to the one used in this study (193 nm ArF laser) was reported by Možná *et al.*<sup>29</sup> This approach can be regarded as an example for a very extreme case of matrix differences between calibration material and sample and resulted in bias from reference content in the range of 13–62% for different elements. Matrix-matched calibration materials for cassiterite are not

available so far. Natural cassiterite crystals are usually very heterogeneous in composition and do not have the necessary elevated concentrations for all elements to make those crystals suitable for usage as calibration material. However the application of LA-ICP-MS data for cassiterite fingerprinting does not necessarily require very accurate data as long as the applied analytical technique (LA-ICP-MS, 193 nm ArF laser) is not substituted by an alternative one. If different analytical techniques will be used in the future to provide data for cassiterite fingerprinting it is essential to calibrate all methods based on reliable calibration materials.

The 185 cassiterite concentrate samples of this study sum up to 8014 analysed individual cassiterite grains (45 elements each) with 360630 single element analyses. Summary statistics are given in Table 1. Within this database 43 cassiterite concentrate samples exist for which brother samples are available (1971 cassiterite grains, 88695 single element concentrations). As an example for the nature of the data, Box-Whisker plots for two elements (U, Ga) of four samples from two locations in Rwanda (Nemba and Nyagasagara pegmatite deposits) are given in Fig. 1. The typical variability of element concentrations of



**Fig. 1** Box-Whisker plots of U and Ga concentrations in four cassiterite concentrates from the pegmatite deposits Nemba and Nyagasagara. Denotations ‘1’ and ‘2’ indicate independent samples from the given location. The numbers within the boxes indicate the number of cassiterite grains analysed.

different grains from the same sample is in the range of one to two orders of magnitude. Brother samples give similar distributions in their element concentrations. For some elements these distributions differ between different locations (U in the given example) while for other elements they do not (Ga in this example).

Substitution of trace elements into the cassiterite structure depends on the ionic radius and the valence; simple (*e.g.*,  $\text{Ti}^{4+}$ ,  $\text{Zr}^{4+}$ ,  $\text{Hf}^{4+}$  for  $\text{Sn}^{4+}$ ) and complex substitutions (*e.g.*, coupled substitution of  $\text{Nb}^{5+}$  and  $\text{Sc}^{3+}$  for  $2\text{Sn}^{4+}$ ) are possible. Trace element concentrations in cassiterite depend on different factors, namely the availability of trace elements in the melt, the element-specific distribution coefficients (cassiterite – melt) and for redox-sensitive elements the redox conditions of the melt. The similar distribution of Ga in both the Nemba and Nyagasagara deposits may be related to similar source rocks of the parental melts, namely Kibaran-age metasedimentary rocks. The different U concentrations in cassiterite from Nyagasagara and Nemba may be caused by different oxidation states of the related melts.

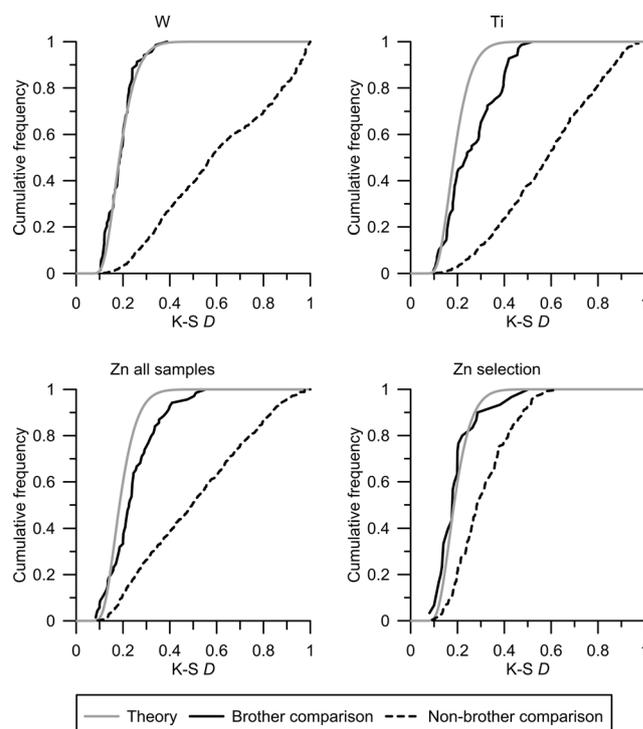
Distributions of element concentrations in cassiterite grains differ significantly from those of man-made materials like glass, steel or polycarbonate headlamp lens materials for which fingerprinting applications based on LA-ICP-MS data are given in the literature.<sup>8,10,18</sup> Elements are significantly more heterogeneously distributed in cassiterite than in those materials. For instance the relative standard deviations analysed in float glass samples range between 1.4% (Fe) and 7.3% (Hf)<sup>10</sup> and in exterior surfaces of polycarbonate headlamp lamps between 12% and 15%.<sup>18</sup> Replicate analyses of the same steel using LA-ICP-MS also give excellent reproducibility.<sup>8</sup> However, grains from the same cassiterite concentrate give relative standard deviations that are much higher (mean RSDs in the range of 100–200%) and the concentrations do not follow a normal distribution. This is the reason why for fingerprinting of cassiterite concentrate samples the distributions of element concentrations and non-parametric statistics are used.

### Selection of elements suitable for cassiterite fingerprinting

Inspection of the LA-ICP-MS data showed that not all analysed elements are suitable to contribute to a cassiterite fingerprint. Elements like Ca, Na, Mo, Y, Th, Ge, Sb, Ba or most of the rare earth elements (REEs) often give concentrations below or close to the detection limit of LA-ICP-MS which excludes them from being used for cassiterite fingerprinting purpose.

For successful application of element concentrations in cassiterite fingerprinting it is essential that two samples from the same deposit can be regarded as two random samples from the same population. To study this, all possible comparisons between brother samples have been evaluated together. From the database, for each element 69 pair-wise comparisons of brother samples were possible. The K–S *D* is considered as a measure for the similarity of samples. So for each element 69 K–S *D* values for brother sample comparisons are available. Element-wise the empirical distribution functions of the 69 K–S *D* values are compared to the theoretical distribution function

of a K–S statistic under the hypothesis that both samples were randomly collected from the same population. They are also compared to the distribution functions which could be calculated from the K–S *D* values of the possible 834 non-brother comparisons. The theoretical distribution function was deduced from an approximation<sup>30</sup> applied to the two-sample case with  $n_1 = n_2 = 40$ . Examples for the obtained distribution functions of the K–S *D* values are given in Fig. 2. For tungsten (W) the empirical distribution function of the brother comparisons is very close to the theoretical curve, while the distribution function of the non-brother comparisons differs significantly from the other two curves. This indicates that for this element brother samples can be regarded as samples from the same population and being representative for that population. Other elements like Ti or Zn do not show this behaviour when all possible brother comparisons from the available data are taken into account. A resurvey of the analytical data reveals that for some elements like Zn, Nb, V, Nb, and Ta some of the brother sample locations do not contain concentrations of these elements which are high enough to be significantly above the detection limit of LA-ICP-MS. These locations were excluded and the distribution functions were recalculated. The effect of this exclusion can be seen from Fig. 2 where for Zn the distribution functions for all locations and for locations with well quantifiable Zn concentrations ('Zn selection') are given (five locations from South America have been excluded). The



**Fig. 2** Cumulative distribution functions of the Kolmogorov–Smirnov statistic (K–S *D*) of the elements W, Ti, and Zn for brother and non-brother comparisons for locations where independent samples are available. A theoretical cumulative distribution function for brother comparisons is also given (for details see the text). 'Zn all samples' refers to the consideration of all possible locations; 'Zn selection' refers to the consideration of locations with Zn concentrations that are significantly above the detection limit of LA-ICP-MS.

empirical distribution function of the remaining brother comparisons now plots closer to the theoretical curve and indicates that Zn can also be used as an element for cassiterite fingerprinting.

The difference between two cumulative distribution functions of element-specific  $K-S D$  values (comparisons of brother *vs.* theory, non-brother *vs.* brother, non-brother *vs.* theory) can be described by the maximum distance obtained between those functions (*e.g.* a  $K-S D$  value for the comparison of two cumulative distribution functions of element-specific  $K-S D$  values). Those  $K-S D$  values for different elements are given in Table 2 and used to select suitable elements for cassiterite fingerprinting. A maximum threshold value of 0.25 for the brother *vs.* theory comparison was defined empirically to pick suitable elements. The concentrations of these elements are distributed in the samples in such a way that the brother samples can be regarded as samples from the same population and being representative for that population. The selected elements are U, Sc, W, Fe, Nb, Ga, V, Zn, Mn, and Ta.

An alternative approach to check whether two samples from the same deposit can be regarded as two random samples from the same population would be to compare many (>20) cassiterite concentrate samples from the same deposit. Due to the lack of so many samples from the same deposit and the fact that the transferability of the test results from a single deposit to other deposits is difficult, this approach was not adopted. The advantage of the applied procedure is that the nature of different cassiterite deposits and mining techniques are integrated in the evaluation by combining data from different deposits.

**Table 2** Maximum distances of element specific cumulative distribution functions of the Kolmogorov–Smirnov statistic ( $K-S D$ ) for different comparisons given as  $K-S D$  values

Element	Brother comparison <i>vs.</i> theory	Non-brother comparison <i>vs.</i> theory	Brother comparison <i>vs.</i> non-brother comparison
U	0.084	0.755	0.720
Sc	0.092	0.873	0.869
W	0.103	0.809	0.818
Fe	0.116	0.628	0.579
Nb <sup>a</sup>	0.129	0.629	0.579
Ga	0.145	0.781	0.700
V <sup>a</sup>	0.160	0.758	0.692
Zn <sup>a</sup>	0.161	0.490	0.535
Mn <sup>a</sup>	0.212	0.664	0.525
Ta <sup>a</sup>	0.247	0.650	0.654
Al	0.287	0.725	0.584
Zr	0.292	0.830	0.723
Pb	0.294	0.740	0.656
Cr	0.308	0.840	0.692
Sr	0.313	0.742	0.526
Hf	0.315	0.838	0.751
As	0.360	0.783	0.443
Ti	0.365	0.835	0.664
In	0.380	0.843	0.580
Bi	0.407	0.706	0.393
Mg	0.622	0.826	0.350

<sup>a</sup> Locations with concentrations nearby or below the detection limit have been excluded (see text).

### Discrimination of brother and non-brother comparisons using the Kolmogorov–Smirnov statistic

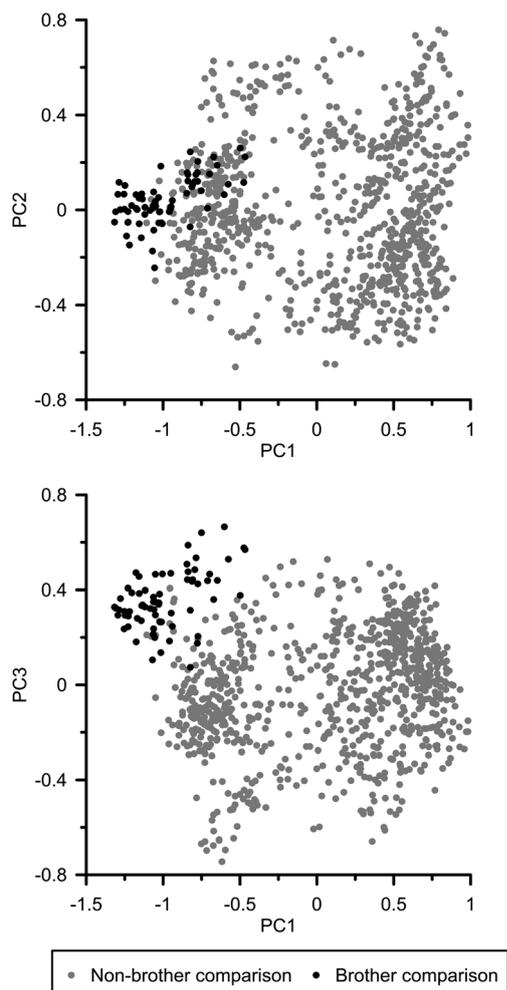
This section investigates whether pair-wise comparisons of samples can be classified in two classes namely, comparisons between brother samples and comparisons between non-brother samples. The Kolmogorov–Smirnov statistic  $K-S D$  is used as a measure for the comparison of two empirical distributions. For each pair-wise comparison of samples one  $K-S D$  value is calculated for each of the 10 elements identified above. The result of one comparison between two samples is a vector consisting of the  $K-S D$  values of the 10 elements. This vector can be used to decide whether the comparison is one between brother samples or between non-brother samples. From this point of view a classification problem has to be solved (does the comparison between two samples belong to the class ‘brother comparisons’ or to the class ‘non-brother comparisons?’). The dataset used consists of vectors of the  $K-S D$  values for the 69 brother and 834 non-brother comparisons; in total 903 ten dimensional vectors with known class information.

A principal component analysis of those vectors is used to visualise the classification problem by plotting the first three main components (Fig. 3). Comparisons of brother samples plot within a different region than comparisons of non-brother samples, with a small region in between where both types of comparisons are plotting.

Linear discriminant analysis (LDA) was used to test the classification ability of the  $K-S D$  values vector concerning the problem whether the comparison belongs to the brother or the non-brother class. The dataset was split into a test set and a training set. The test set consisted of ten comparisons of each class, while the training set consisted of the remaining ones. The selection of the sets was done randomly and repeated 5000 times. As a result misclassifications appeared in 4.3% of the assigned brother comparisons and 7.1% of the assigned non-brother comparisons. This demonstrates that the distribution of element concentrations of cassiterite concentrate samples can be used to verify the declared origin of a sample with an acceptable low degree of uncertainty.

It is noticeable that the misclassification of brother comparisons (4.3%) was lower than that of non-brother comparisons. This can be explained by the fact that LDA is based on the assumption that both classes are normally distributed and have the same covariance matrix. This is not the case for the dataset. The empirical variance of the class ‘brother comparisons’ is smaller than that of the class ‘non-brother comparisons’. This leads to a systematic overestimation of the variability of the brother comparisons class by LDA. So comparisons falling into the ‘border area’ between both classes were rather classified as brother comparisons than as non-brother comparisons, which resulted in lower misclassifications of brother comparisons and higher ones for the non-brother comparisons. From a practical point of view this is a conservative approach because the declared origin of a sample is misleadingly accepted with higher probability than a misleading rejection.

However LDA's distribution assumption is violated. The covariance matrixes of the classes are not equal and the



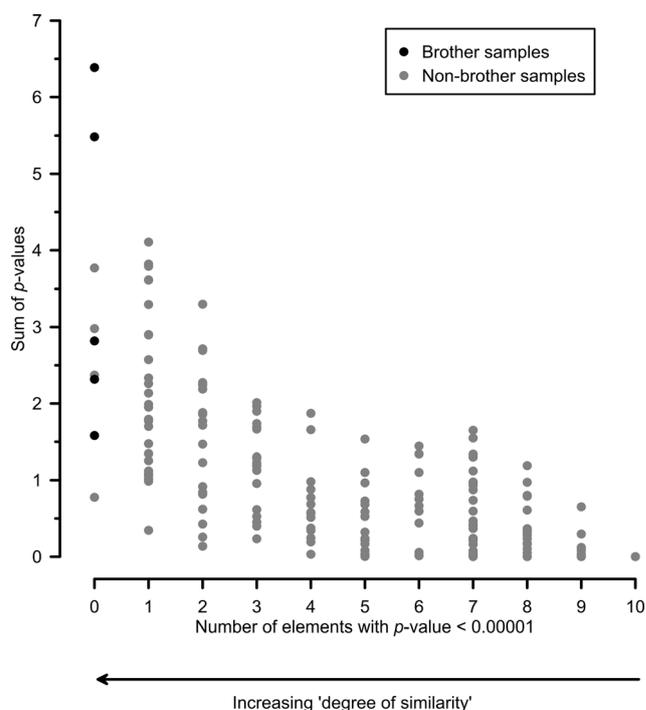
**Fig. 3** Visualisation of the first three principal components (PC1, PC2, PC3) of a principal component analysis of vectors obtained from brother and non-brother comparisons. Each vector results from a pair-wise comparison of two samples. The vectors consist of the Kolmogorov–Smirnov statistics ( $K-S D$ ) of ten selected elements (for details see the text).

distributions show significant deviations from normal distribution (especially for the non-brother comparisons class). Sometimes generalized hyperbolic distributions fit better, but are still not satisfying. In the future, with an extended database of brother samples, the estimation of individual multivariate normal distributions might be an option, but for the moment the estimation of a multivariate normal distribution with 10 dimensions from 69 samples remains questionable.

#### Wilcoxon rank sum test

According to the LDA the distributions of element concentrations within a sample from a cassiterite concentrate seem to be indicative for the origin of this sample. The Wilcoxon rank sum test is a non-parametric statistical test which examines the probability that two samples come from the same population. Using this test a  $p$ -value can be computed which gives the probability that the null hypothesis (“both samples come from the same population”) is misleadingly rejected. Comparing two

samples one  $p$ -value can be computed for each element. If the above identified 10 elements are taken into consideration 10  $p$ -values for every two-sample comparison can be calculated. If a given sample is compared to a number of other samples from a database a ‘ranking of similarity’ between this sample and all the other samples can be established by using these  $p$ -values. First, a threshold value is defined below which a  $p$ -value is considered to be ‘zero’ (e.g., 0.00001; this means that the null hypothesis that both samples come from the same population is misleadingly rejected by a very low probability). Second, for every comparison the number of elements is counted which gave  $p$ -values below the threshold value (‘number of zeros’). A higher degree of similarity is assigned by a lower number of ‘zeros’. If the number of ‘zeros’ is equal for two or more comparisons the sum of the  $p$ -values is used for additional ranking. A higher degree of similarity is indicated by a greater sum of the  $p$ -values. Rank number 1 indicates the sample from the database which is most similar to the test sample and the rank with the highest number indicates the least similar one. Fig. 4 illustrates this for samples from the Nyagasagara pegmatite deposit. One sample from this deposit is tested against the whole database of cassiterite concentrates, which includes five brother samples from this deposit. The results are ranked according to the previously given procedure based on the  $p$ -values. The results for the brother samples are indicated in black while all other samples are indicated in grey. The lowest ranks (highest degree of similarity) are assigned to the comparisons with zero ‘zeros’, the highest ones to comparisons with ten ‘zeros’. All brother samples from the Nyagasagara



**Fig. 4** Visualisation of a ‘ranking of similarity’ based on  $p$ -values calculated from the Wilcoxon rank sum test (for details see the text). One sample of the Nyagasagara pegmatite deposit is tested against the whole cassiterite database. The database contains five independent brother samples from this deposit.

pegmatite deposit appeared at zero 'zeros', thus getting very low ranks indicating that these samples have very similar geochemistry.

If the ranking procedure based on the *p*-values shall be used for cassiterite fingerprinting it is necessary to define threshold values to assess the ranking results or more strictly spoken to define a rank number below which comparisons are accepted as brother comparisons. To do this the 43 cassiterite samples are used for which brother samples are available. Each of those samples is tested against all other samples of the whole database. All samples from the database are ranked according to the above given ranking procedure. A threshold value of 0.00001 is used to consider a *p*-value as 'zero'. In the cases where only two brother samples are available the rank of the brother sample is stored. In cases where more than two brother samples are available all brother samples were deleted from the ranking list except the less similar brother sample (the brother sample with the highest rank number). Now the rank of this brother sample was stored. This was done to simulate the conservative case that only the brother sample with the lowest degree of similarity is available in the database. For all brother sample comparisons the highest rank found by this exercise was rank number 10. All other brother sample comparisons appeared at rank numbers smaller than 10 out of about 180 possible ranks. For a practical application of this procedure it can be deduced that if the rank of a comparison of a sample in question with its 'brother sample' from the database (defined by the declaration of the origin of the sample in question) is higher than 10, the declared origin is questionable and becomes more questionable with higher ranks.

## Conclusion

Cassiterite concentrate analysis by LA-ICP-MS can be used as an independent tool for tracing back the origin of samples, if doubt/uncertainty of information given by document based strategies arises. Samples in question are checked against a database where LA-ICP-MS data from samples of known origin are stored. A first version of a database required in this context is available at BGR and extended continuously. If alternative analytical techniques shall be used in the future to provide data for cassiterite fingerprinting it is essential to calibrate all methods based on reliable calibration materials.

Independent samples taken from the same location show empirical cumulative distribution functions of element concentrations which are more similar than those of samples taken from different locations. This could be demonstrated by samples from 13 different locations where more than one sample has been available. The concentrations of the elements U, Sc, W, Fe, Nb, Ga, V, Zn, Mn, and Ta are used for this.

The Wilcoxon rank sum test is applied to compare two samples (e.g., the sample in question and a sample from the database with known origin). For each element a *p*-value can be computed which gives the probability that the null hypothesis ("both samples come from the same population") is misleadingly rejected. A degree of similarity between two samples can be deduced by combining the results for all elements. This is

used to assess whether the two samples originate from the same location or not.

## Acknowledgements

The authors thank Peter Rendschmidt for the preparation of the polished sections and Maria Sitnikova for providing numerous MLA maps of cassiterite concentrates. Comments by three anonymous reviewers helped to improve the manuscript.

## Notes and references

- 1 SEC, Dodd-Frank Wall Street Reform and Consumer Protection Act (2010), <http://www.sec.gov/about/laws/wallstreetreform-cpa.pdf>, accessed 27 August 2012.
- 2 Resource Consulting Services, US legislation on conflict minerals, RCS private sector guidance on the Dodd-Frank Act section 1502 (2011), [http://www.resourceglobal.co.uk/documents/RCS\\_DF\\_ACT\\_GUIDANCE\\_APRIL\\_2011\\_lowres.pdf](http://www.resourceglobal.co.uk/documents/RCS_DF_ACT_GUIDANCE_APRIL_2011_lowres.pdf), accessed 27 August 2012.
- 3 Motorola Solutions, Motorola Solutions announces project to source conflict-free tantalum from Democratic Republic of Congo (2011), <http://mediacenter.motorolasolutions.com/Press-Releases/Motorola-Solutions-Announces-Project-to-Source-Conflict-Free-Tantalum-from-Democratic-Republic-of-Congo-36a1.aspx>, accessed 02 October 2012.
- 4 ITRI, ITRI Tin Supply Chain Initiative ITSCi (2009), [http://www.itri.co.uk/SITE/UPLOAD/Document/iTSCi\\_Final\\_Version\\_2\\_English\\_\\_2.10.09.pdf](http://www.itri.co.uk/SITE/UPLOAD/Document/iTSCi_Final_Version_2_English__2.10.09.pdf), accessed 02 October 2012.
- 5 P. Černý, P. L. Blevin, M. Cuney and D. London, in *Economic Geology: 100th Anniversary Volume*, ed. J. W. Hedenquist, J. F. H. Thompson, R. J. Goldfarb and J. P. Richards, Society of Economic Geologists, Littleton, CO, 2005, pp. 337–370.
- 6 C. A. Heinrich, *Econ. Geol.*, 1990, **85**, 457–481.
- 7 W. L. Pohl, *Economic Geology: Principles and Practice*, Wiley-Blackwell, Chichester, 2011.
- 8 R. J. Watling, B. F. Lynch and D. Herring, *J. Anal. At. Spectrom.*, 1997, **12**, 195–203.
- 9 S. J. Bajic, D. B. Aeschliman, N. J. Saetveit, D. P. Baldwin and R. S. Houk, *J. Forensic Sci.*, 2005, **50/5**, 1123–1127.
- 10 P. Weis, M. Dücking, P. Watzke, S. Menges and S. Becker, *J. Anal. At. Spectrom.*, 2011, **26**, 1273–1284.
- 11 C. Dalpé, P. Hudon, D. J. Ballantyne, D. Williams and D. Marcotte, *J. Forensic Sci.*, 2010, **55/6**, 1443–1456.
- 12 L. Coney, A. V. Moila and A. G. Quadling, *S. Afr. J. Geol.*, 2012, **115/1**, 33–46.
- 13 R. J. Watling, H. K. Herbert, D. Delev and I. D. Abell, *Spectrochim. Acta, Part B*, 1994, **49/2**, 205–219.
- 14 H.-E. Gäbler, F. Melcher, T. Graupner, A. Bahr, M. A. Sitnikova, F. Henjes-Kunst, T. Oberthür, H. Brätz and A. Gerdes, *Geostand. Geoanal. Res.*, 2011, **35/4**, 431–448.
- 15 C. Savu-Krohn, G. Rantitsch, P. Auer, F. Melcher and T. Graupner, *Nat. Resour. Res.*, 2011, **20**, 177–191.
- 16 R. J. Watling, *J. Anal. At. Spectrom.*, 1998, **13**, 917–926.
- 17 J. M. Gallo and J. R. Almirall, *Forensic Sci. Int.*, 2009, **190**, 52–57.

- 18 C. D. May and R. J. Watling, *J. Forensic Sci.*, 2011, **56/S1**, S47–S57.
- 19 W. Castro, J. Hoogewerff, C. Latkoczy and J. R. Almirall, *Forensic Sci. Int.*, 2010, **195**, 17–27.
- 20 G. Zadora and D. Ramos, *Chemom. Intell. Lab. Syst.*, 2010, **102**, 63–83.
- 21 Y. Gu, *J. Min. Mat. Char. Eng.*, 2003, **2**, 33–41.
- 22 JFree, JFreeChart, Java chart library (2009), <http://www.jfree.org/jfreechart/>, accessed 02 October 2012.
- 23 Apache Commons, Commons-Math: The Apache Commons Mathematics Library (2009), <http://commons.apache.org/math/>, accessed 02 October 2012.
- 24 Apache POI, Apache POI – the Java API for Microsoft Documents (2009), <http://poi.apache.org/>, accessed 02 October 2012.
- 25 N. J. G. Pearce, J. G. Perkins, J. A. Westgate, M. P. Gorton, S. E. Jackson, C. R. Neal and S. P. Chenery, *Geostand. NewsL.*, 1997, **21**, 115–144.
- 26 H. P. Longerich, S. E. Jackson and D. Günther, *J. Anal. At. Spectrom.*, 1996, **11**, 899–904.
- 27 R Development Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing (2012), <http://www.R-project.org/>, accessed 13 March 2012.
- 28 J. Pisonero, B. Fernández and D. Günther, *J. Anal. At. Spectrom.*, 2009, **24**, 1145–1160.
- 29 V. Možná, J. Pisonero, M. Holá, V. Kanický and D. Günther, *J. Anal. At. Spectrom.*, 2006, **21**, 1194–1201.
- 30 G. Marsaglia, W. W. Tsang and J. Wang, *J. Stat. Software*, 2003, **8/18**, 1–4.