Highlights from 25 Years of the Geo PT Programme: What Can be Learnt for the Advancement of Geoanalysis

How to cite:

For guidance on citations see FAQs.

© 2022 The Authors

https://creativecommons.org/licenses/by-nc/4.0/

Version: Version of Record

http://dx.doi.org/doi:10.1111/ggr.12424

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online’s data policy on reuse of materials please consult the policies page.
Highlights from 25 Years of the GeoPT Programme: What Can be Learnt for the Advancement of Geoanalysis

Thomas C. Meisel (1)*, Peter C. Webb (2) and Alessandra Rachetti (1)

(1) General and Analytical Chemistry, Montanuniversität Leoben, Leoben 8700, Austria
(2) Faculty of Science, The Open University, Walton Hall, Milton Keynes MK7 6AA, UK
* Corresponding author. e-mail: thomas.meisel@unileoben.ac.at

Data submitted over the past 25 years to GeoPT, the highly successful proficiency-testing programme for the geochemical analysis of geological materials, organised by the International Association of Geoanalysts, provide a valuable resource that permits detailed investigation of contrasting results associated with different sample preparation and measurement principles. Highlighted issues include the following: recurring problems with the dissolution of the refractory minerals zircon and chromite, which produce a large dispersion in data obtained when acid digestion is involved; issues related to different XRF sample preparation methods, whereby a significant divergence of pressed powder pellet results compared with those from fused glass discs is observed; high relative dispersion of data both at low mass fractions, and those higher than normally found in silicate rocks, due to incomplete method validation, in particular due to an overconfidence in estimating reporting limits and to the employment of limited working ranges. In addition, an example for Sr in an ancient pegmatite with extremely high Rb abundance is presented, where ICP-MS results amount to only a third of the XRF results, severely underestimated due to the disregard of the radiogenic ingrowth from $^{87}$Rb decay. Recommendations are made both for improving data quality and the selection of test materials for future GeoPT rounds.

Keywords: proficiency test, GeoPT, sample preparation, zircon, chromite, $^{87}$Sr/$^{86}$Sr.

Received 27 Sep 21 – Accepted 17 Feb 22

During 25 years of GeoPT, a wealth of data has been accumulated that has led to individual reports for 63 rounds of GeoPT and numerous publications including round-specific and test material-specific research papers (Thompson et al. 1996, 1997, 2015; Kane 2005, Potts et al. 2015; Webb et al. 2019). GeoPT operates according to a detailed protocol (International Association of Geoanalysts 2020), and currently supports the PT requirements of well over 100 geoanalytical laboratories. The results from almost twenty rounds have provided a basis for reference data sheets to be produced for a range of reference materials marketed by IAGeo Ltd. In addition, GeoPT has been a part of the International Association of Geoanalysts’ protocol for characterisation of certified reference materials since 2003 (Kane et al. 2003), and more recently, a strategy has been proposed to permit certification based on GeoPT assigned values (Potts et al. 2019).

A total of 182888 submitted measurement results for eighty-one analytes have been accumulated during sixty-three rounds for which consensus values have been assigned to sixty-eight analytes. With this treasure at hand, ‘data mining’ makes it possible to discuss changes of data quality over time, improvements in instrumentation, calibration issues and recurring issues with sample preparation. The systematic and standardised acquisition of metadata (i.e., details of the contributed measurement procedures) since Round 35 (2014) makes it possible to study and identify the most influential factors contributing to instances of unusually large between-laboratory bias and poor data precision.

In Figure 1, the historic development of submitted measurement results is displayed, which includes the number of assigned and provisional values. The diagram shows the general trend towards an increasing number of measurement results submitted to each round and also a trend towards higher numbers of assigned and provisional values conferred. Obvious exceptions are Rounds 6A limestone CAL-S (Potts et al. 2000), 12 serpentinite GAS (Potts et al. 2003), 14 alkali granite OShBO (Potts et al. 2004), 20...
In the vast majority of cases, submitted values for a measurand approximate to a random sample from a normal distribution symmetrically disposed around the consensus value with data derived from a varied range of measurement principles and sample preparation methods. As a result, only small differences between the potential estimators of the consensus value are observed. A typical example of such a data set is shown in Figure 2. The median, robust mean and the mode are within $\pm 0.25$ RSD % of the assigned value, which is compatible with the characteristics of a near-normal data distribution.

However, there are quite often a number of data sets in each round that do not conform to this ideal, and the following account describes and discusses in detail a number of irregular and anomalous data distributions for some of the GeoPT test materials, in particular for many of the rounds that yielded a smaller than usual number of attributed assigned and provisional values. Our investigations should provide valuable lessons for analysts to avoid presenting data of inferior quality.

Observations on GeoPT data distributions

The following account identifies and illustrates a variety of anomalies in GeoPT data distributions that we investigate using procedural metadata supplied by participants. An appreciation of these metadata (subsequently identified by the shorthand coding keyed below) is essential for the following discussions. They include methods of measurement and preparation of samples as follows:

Measurement principles and sample preparation procedures commonly used for GeoPT

Measurement principles include the following:

- INAA – Neutron activation analysis, with no (NO) sample preparation necessary;

A wide range of spectroscopic methods with abbreviations used in the subsequent text:

- freq_submitted
- 1500
- 2000
- 2500
- 3000
- 3500

Figure 1. Number of assigned and provisional values per round versus GeoPT round over the past 25 years. A slight increase in the number of assigned and provisional values and the number of submitted data measurement results (size of circle) over time is visible. Exceptions are limestones: CAL-S (06A) and ML-2 (30A), ultramafic rocks: serpentinite GAS (12), komatiite OPY-1/OKUM (20) and unserpentinised harzburgite HARZ01 (38A) as well as highly fractionated plutonic rocks: alkali granite (probably a pegmatite) OShBO (14) and Separation Lake pegmatite OU-9 (23).
ICP – inductively coupled plasma-atomic emission spectroscopy (ICP-AES) and inductively coupled plasma-optical emission spectrometry (ICP-OES)
- ICP-MS – inductively coupled plasma-mass spectrometry
- DCP – direct current plasma-atomic emission spectroscopy
- AAS – atomic absorption spectroscopy
- PHOT – photometry
- VOL – volumetric analysis or titrimetry

for which solid materials can be converted into solutions by:
- AD – acid digestion with mineral acids mainly HCl, HNO₃, HClO₄ and HF in different combinations
- SI – sintering with Na₂O₂ at ~ 480 °C
- FM – fusion with Na₂O₂ at ~ 700 °C or fusion with other reagents such as carbonates, hydroxides or fluorides (NH₄F) and others (FM)
- FM_AD – fusion with LiBO₂ (lithium metaborate) or a combination of LiBO₂/Li₂B₄O₇ at ~ 1000 °C and dissolution of the resulting bead with acids
- AD_FM – a combination of acid digestion with fusion of the residues with typically LiBO₂ (AD_FM)
- FA – fire assay (e.g., NiS and flux) and for:
  - XRF – X-ray fluorescence spectrometry with energy dispersive (ED-XRF) and wavelength dispersive analysis (WD-XRF) and
  - LA-ICP-MS – a laser ablation system coupled to an ICP-MS

sample preparation of solids is by:
- FD – fusion with LiBO₂ (lithium metaborate) or a combination of LiBO₂/Li₂B₄O₇ at ~ 1000 °C
- PP – pressed powder pellets with and without binder (wax, methyl methacrylate resin, cotton linters (cellulose) or polyvinylpyrrolidone-methyl cellulose), or
- PF – powder on BoPET film, for example ‘Mylar’.

Figure 2. A near-normal distribution of the Al₂O₃ mass fraction in SdAR-H1 (GeoPT35A). Symmetrical distributions with an unbiased method mix are the norm amongst submitted GeoPT results. The central green horizontal line represents the assigned value, the parallel ‘tramlines’ the GeoPT fitness-for-purpose criteria (IAG 2020): The solid black lines ‘Quality 1’ (‘pure’ geoscience research laboratories) and the dashed lines the ‘Quality 2’ (‘applied’ geochemical laboratories). Both fitness-for-purpose criteria are derived from the Horwitz function (see IAG 2020 for details). To emphasise the symmetry of the normal distribution, a histogram (yellow) and a kernel density function (grey) are shown in the right diagram. Legend explanations can be found in the text. ‘n.a’ stands for ‘not available’.

© 2022 The Authors. Geostandards and Geoanalytical Research published by John Wiley & Sons Ltd on behalf of International Association of Geoanalysts.
Effects of incomplete digestion

Since its inception in the 1980s, the capabilities of ICP-MS have led to the dominance of this measurement principle for trace element analysis. The very high selectivity and sensitivity of the method allow reproducible measurement results of mass fractions below 1 mg kg\(^{-1}\). These undisputed advancements in measurement technique place greater demands on the procedural blank to account for contamination issues due to the purity and the quantities of reagents used for sample digestion. For this reason, acid digestion (AD) techniques using sub-boiled nitric, hydrochloric as well as hydrofluoric acid are presumably advantageous over fusion techniques. However, such procedures can be prone to the incomplete digestion of refractory mineral phases. The persistent problem of incomplete digestion has been the reason for the development and description of methods using high temperature acid attack, fusion and a combination of both (Hu et al. 2010, Cotta and Enzweiler 2012, Zhang et al. 2012, Okina et al. 2016, Bokhari and Meisel 2017, Magaldi et al. 2019). The effects and consequences of incomplete digestion were recognised in Round 31 of GeoPT involving zircon in a river sediment, a detailed discussion of which is given in Potts et al. (2015). The most common refractory/resistant minerals that are known to be difficult to digest are chromite (chrome spinel FeCr\(_2\)O\(_4\)) and zircon (ZrSiO\(_4\)). In ultramafic rocks, chromite is the main carrier of Cr, whereas zircon is the carrier mineral for Zr and Hf, and in some cases also for the heavy rare earth elements (HREE), in felsic plutonic rocks and derived sediments.

A typical example of a chromite-bearing GeoPT test sample that suffers from sample preparation issues is the unserpentinitised harzburgite, provided as the supplementary GeoPT38A (aka HARZ01) test material in Round 38. Reported chromium mass fractions ranged from 37 to about 4500 mg kg\(^{-1}\), with AD data dominating the range below 2000 mg kg\(^{-1}\) (Figure 3) illustrates this in ordered distribution and ridgeline plots. This clearly indicates that incomplete digestion is a significant factor degrading the Cr data supplied by a number of laboratories. During the last 25 years of GeoPT rounds, only three typical ultramafic rocks were issued, namely serpentinite GAS (GeoPT12), komatite OPY-1 (GeoPT20), which was actually the IAG certified reference material OKUM, and modified harzburgite HARZ01 (GeoPT38A). Although chromite is to be expected in ultramafic rocks, only for GeoPT38A do the Cr data seem to reveal problems with incomplete sample digestion using mineral acids (Figure 4). Notably, the harzburgite HARZ01 is very unusual as it is fresh and almost completely unserpentinitised, with a loss on ignition (LOI) of < 0.5 g/100 g, whereas GAS and OPY-1/OKUM test materials are completely serpentinitised, with LOI values of 13 and 4.7 g/100 g, respectively, which is more typical of most peridotites and komatites. Obviously, there is a connection between serpentinisation and the resistance of chromite to acid decomposition. During the serpentinisation/alteration process, chrome spinel transforms into other spinels with a ferrichromite rind containing large amounts of magnetite (also a spinel Fe\(_2\)O\(_4\) or FeFe\(_2\)O\(_4\)). These replacement minerals are more likely to be easier to digest in hydrochloric acid and do not appear to cause dissolution problems during sample preparation. It is also interesting to note that the Ni mass fraction data submitted to Rounds 12, 20 and 38A have regular distributions that appear to be unaffected by processes associated with any particular form of sample preparation or measurement principle. Hence, chromite cannot be a major host of Ni (Beckett-Brown and McDonald 2018) in HARZ01 (GeoPT38A).

Three clear examples of serious problems associated with the incomplete digestion of zircon are test materials GeoPT34 (Granite GRI-1 which is actually USGS RM G-3), as recognised by Webb et al. (2014), GeoPT35 (Tonalite TLM-1) and GeoPT37A (Blended sediment SdAR-L2) where Zr mass fractions reported range from 1 to 800 mg kg\(^{-1}\) (Figure 5). Figure 5 also demonstrates that although most of the AD results for these test materials appear to be affected by incomplete digestion, some AD results are compatible with the consensus of other methods suggesting that AD methods using more rigorous dissolution techniques, as identified in Potts et al. (2015), remain fully viable.

Potential homogeneity issues

The high sensitivity of modern ICP-MS instruments has enabled the reduction of commonly used test portion sizes from 0.1–0.3 g to a range from 50 to 100 mg. Because mineral grains containing much higher mass fractions of trace elements than the matrix as a whole are very commonly present as accessory minerals in geological samples and are liable to exhibit the so-called ‘nugget effect’ (see Meisel et al. (2001) for an explanation), the influence of sampling heterogeneity on the uncertainty budget of the measurement result for such small test portions cannot be ignored (Meisel et al. 2001).

However, it is not straightforward to estimate the mass of a test portion that would produce a relative standard deviation for sampling of ± 1% (Ingamells and Switzer 1973), in other words, the test portion size for which the influence of sampling would be negligible (which is equivalent to the sampling constant). The influence of all other sample preparation steps...
other than sampling itself needs to be very small or at least constant. This situation can be achieved only within one laboratory using a measurement procedure that assures complete digestion of a range of sample masses. Such an attempt can be seen in Bokhari and Meisel (2017). GeoPT data provide only a limited opportunity to study heterogeneity issues as only one measurement result is submitted per laboratory and because sample digestion issues are more likely to be severe for geological sample powders potentially exhibiting the nugget effect.

For the Modified River Sediment SdAR-1 (GeoPT31), Potts et al. (2015) previously reported that the range of discrepancies seen among the AD results (with much lower values for the analytes Y, Yb and Zr) far exceed those of the fusion results using similar test portion sizes of material and therefore cannot be attributed to any form of heterogeneity. The same shape of data distribution for SdAR-1 (GeoPT31) is also observed for the material from GeoPT34 (granite GRI-1/USGS G-3) and GeoPT35A (Metalliferous River Sediment, SdAR-H1) in Figure 6. All of these GeoPT test materials are obviously zircon bearing and the data reveal serious issues with AD procedures employed in many laboratories. In all three test materials, the hypothesis that the low values are caused by the nugget effect as a single influence factor of the observed bias can be rejected as zirconium mass fractions based on AD remain persistently low with increasing test portion size. The effect of heterogeneity associated with the nugget effect would result in a trumpet-shaped distribution of mass fraction data versus mass of test portion.

Figure 3. Ordered data distribution and ‘ridgeline’ plots of chromium mass fractions in HARZ01 (GeoPT38A) demonstrating measurement bias in nine results obtained from solution preparation by AD. It is also noteworthy that results for a significant number of samples prepared as PP fall below the consensus provided by other methods. The shapes of the ridgeline plots are representations of data density functions.
extending above and below the central value and tapering from a dispersed distribution towards convergence at higher sampling masses.

**Method validation issues**

GeoPT test materials with unusual high mass fractions for some selected analytes: Calibration range problems:

Most terrestrial rocks have a limited mass fraction range for Nb and Ta and a more or less constant (Nb to Ta) mass fraction ratio (Green 1995). This statement is supported by the range of GeoPT assigned values with 91% of Nb values below 64 mg kg\(^{-1}\), 95% of Ta values less than 7 mg kg\(^{-1}\) and a slope of the fit to the Nb/Ta consensus value distribution of 17.2.

As calibration standards for XRF are mainly selected from a range of well-known and well-characterised reference materials, and if those used routinely for calibration have a similar and limited range of Nb and Ta values, the valid calibration range is inherently biased towards these relatively low mass fractions. As a consequence, GeoPT rounds involving highly fractionated rocks with Nb/Ta ratios outside the typical range of 10 to 20 and with unusually high mass fractions, will most likely lie outside the typical calibration range established with relatively unfractonated rock reference materials as measurement standards.

Figure 7 shows the extraordinary large mass fraction dispersion of highly fractionated GeoPT test materials (GeoPT03, 14, 23, 30) and in particular those with Nb > 100 mg kg\(^{-1}\) and Ta > 20 mg kg\(^{-1}\). Note how most of the Nb and Ta values cluster around the slope of 17.2, while in the highly fractionated rock types of GeoPT Rounds 03, 14, 23 and 30 with unusually high mass fractions of Nb and Ta, the scatter of the data implies that there is a large variation for many measurement results which is considered here to be mostly due to exceeding the range of many calibrations. Similarly, high U and Th (Figure 8) mass fraction values in highly fractionated rocks (syenites GeoPT39 and 39A and phosphate rock GeoPT46A) that exceed the normal ranges of \(\text{U} < 20 \text{ mg kg}^{-1}\) and \(\text{Th} < 50 \text{ mg kg}^{-1}\) also exhibit large scatter and non-normal distributions expressed in the ordered distribution GeoPT data plots (Figure 9).

It would be expected that at higher mass fractions, better counting statistics and better signal to noise ratios would result in a decrease of the scatter of interlaboratory data, expressed as relative precision with increasing mass fraction, whereas absolute precision would increase proportionally. The empirical Horwitz function (see Horwitz 1982) takes account of this fact and acts as a reference for the fitness-for-purpose criteria for the GeoPT programme. However, in some of the examples presented the data dispersions observed for Nb-Ta, Th-U and also Ce-Pb at high mass
fractions (Figure 10) tend to act counter to this trend towards reduced relative precision. For mass fractions of Nb, Ta, Th, U, Ce and Pb that are much higher than normally found in geological samples, the unexpectedly poor relative precision of the submitted data is explained here by inadequate control on calibration range and as a consequence the working range (for definition see Magnusson and Örnemark 2014).

**Figure 5.** Data distribution plots of zirconium mass fractions in GRI-1/G-3 (GeoPT34), TLM-1 (GeoPT35) and SdAR-L2 (GeoPT37A) according to method of sample preparation demonstrating bias in most of the results obtained from solutions prepared by AD.
GeoPT test materials with exceptionally low mass fractions: problems with data reporting limits: Establishing lower limits of reporting based on appropriate criteria are an integral part of method validation. Both limits of detection (LOD) and quantification (LOQ) are in use but they have different and independent definitions and concepts: they are not interchangeable and should not be confused. The concept of the LOD (a mass fraction where the probability of a false-positive detection is small (e.g., 5%) and a clear distinction of an analyte signal from the background noise is possible (Currie 1968) is often discussed but rarely properly defined (see https://jcgm.bipm.org/vim/en/4.18.html for the correct definition). The LOQ represents a different concept as it is only above the LOQ that the mass fractions can be quantified with acceptable precision. The consequence is that below the LOQ, very high measurement uncertainties (e.g., >30% relative) make a meaningful interpretation of a measurement result more difficult. Unfortunately, the LOQ and its implications seem not to be fully understood and rarely applied across much of the geoanalytical community. Many laboratory systems or analysts censor results beneath their

Figure 6. Plots showing zirconium mass fractions according to methods of sample preparation derived from different masses of test portions alongside data distribution plots of zirconium mass fractions in GRI-1/G-3 (GeoPT34) and SdAR-H1 (GeoPT35A) according to methods of measurement.
recognised limit of detection, but in practice that can allow the submission of extremely variable data at mass fractions well above that limit, because the working range has not been properly established. For a discussion about the definition and use of LOQ refer to Meisel (2012).

Ultramafic rocks are often highly depleted in lithophile elements (e.g., rare earth elements) compared with silica-rich (felsic) crustal rocks and sediments. Because of the low mass fractions, measurements are very challenging and, in most cases, the rapid quantification of all REE was only made possible through the introduction of ICP-MS replacing neutron activation, which was not capable of measuring all of the REE simultaneously, as a routine measurement principle. Closer inspection of the influence of sample preparation on the measurement results of ultramafic GeoPT test materials is possible only for Round 38A (HARZ01), where metadata were supplied in a standardised manner. The clear influence of the sample digestion procedure used on the variation of the results was described in the special report on Round 38A (Webb et al. 2016a) and is demonstrated here in Figure 11 where dominantly ICP-MS results for Yb acquired using AD are relatively coherent, whereas those involving fusion or sintering are drastically dispersed. In circumstances where such low mass fractions are present, it was suggested (Webb et al. 2016a) that fluxes used for fusion and sintering techniques may introduce a relatively high level of contamination. As a consequence, highly fractionated GeoPT test materials.

Figure 7. Tantalum and Nb values reported for forty-nine rounds of GeoPT with sets of data highlighted for the highly fractionated test materials of GeoPT Rounds 03 (granite), 14 (alkali granite), 23 (pegmatite) and 30 (syenite). The insert to the right represents the tight distribution of the assigned and provisional Nb values of all GeoPT samples. The grey line represents the regression line (with slope 17.2) for the assigned and provisional Nb versus Ta values excluding the highly fractionated GeoPT test materials.

Figure 8. Thorium and U values reported for forty-nine rounds of GeoPT with sets of data highlighted for the highly fractionated test materials of GeoPT Rounds 39 (syenite), 39A (nepheline syenite) and 46A (phosphate rock). The insert to the right represents the tight distribution of the assigned and provisional U values of all GeoPT samples. The grey line represents the regression line for the assigned and consensus U versus Th values excluding the highly fractionated GeoPT test materials.
fluctuating mass fraction values arise when subtracting of a large blank signal, which then lowers the signal to noise ratio. Higher LOD and LOQ for samples prepared by fusion or sintering, as well as issues related to incorrect or incomplete blank corrections, would be the immediate consequence, a fact that was clearly ignored for much of the data submitted for GeoPT38A.

Figure 10 provides a further example of failing to establish a validated working range for another measurement principle, namely XRF. The two red data points at the high mass fraction end in the left plot represent cases where the capability to quantify Yb mass fractions by XRF was clearly overestimated. Such observations are commonly encountered for a number of analytes quantified by XRF.

Figure 9. Data distribution plot of thorium mass fractions in Syenite SyMP-1 (GeoPT39) displaying the non-normal distribution of the Th data but also the dominance of the PP data in the lower mass fraction range and AD data in the upper mass fraction range.

Figure 10. Lead and cerium values reported for forty-nine rounds of GeoPT with sets of data highlighted for the highly fractionated test materials of GeoPT Rounds 31 (modified river sediment), 35A (metalliferous sediment), 36A (metal-rich sediment), 39 (syenite), 41A (mineralised stream sediment) and 46A (phosphate rock). The insert to the right represents the tight distribution of the assigned and consensus Ce values of all GeoPT samples.
but not exclusively so, as there are also examples involving ICP-MS and ICP-AES/OES (see also Webb and Potts, this volume). Depending on the sensitivity of the element X-ray signal, which is related to the atomic number, the LOQ may lie in the mass fraction range not far below 10–20 mg kg\(^{-1}\) for many trace elements even with long counting times and not far below 100–200 mg kg\(^{-1}\) for major and minor element oxides, for example TiO\(_2\) where counting times are usually shorter, because the element is expected to present at higher mass fractions in nature. In Figure 12 left plot, the consequence for the data set of an XRF calibration developed for higher mass fraction ranges, without taking an appropriate LOQ into account, is exemplified – the ICP-MS data form a coherent consensus, whereas the XRF data are highly dispersed. Probably, most of XRF TiO\(_2\) data were submitted without taking the LOQs into account. Also, notice that Figure 12 provides a good example of the effect of (unnecessary) rounding of significant figures on the shape of the data distribution, turning it more into a step function.

**Stepwise effects: underestimation of the LOQ:** Stepwise functions in the plotted results have been observed in all GeoPT rounds since the initial one 25 years ago and were discussed in Thompson et al. (2015). Steps in the data distribution plots occur when several of the results for a measurand are identical although independent (i.e., from different laboratories). Such steps (‘plateaus’) in the data distributions are observed frequently for major element oxides at mass fraction ranges below 0.5 g/100 g especially for TiO\(_2\), MnO and P\(_2\)O\(_5\) but may apply to almost any major element oxide measurand including MgO, CaO, Na\(_2\)O and K\(_2\)O, and not only for data provided by XRF. Steps in the data distribution plots are also observed for trace element data below 25 mg kg\(^{-1}\) for XRF measurement results, for example Ga in GeoPT35, and when ICP-MS values are reported with two decimal places only below 0.5 mg kg\(^{-1}\).

In Figure 12 right plot, we see artificial steps for MnO in harzburgite at mass fractions of 0.11, 0.12 and 0.13 g/100 g due to excessive rounding to 1–2 significant figures/digits effectively owing to the underestimation of the anticipated quality of the XRF measurement results. Indeed, some XRF software allows for only two figures in the range below 1 g/100 g. These discrete data plateau artefacts are inconsistent with the observation that proficiency-testing data tend to follow a continuous data distribution. For many years, GeoPT participants have been instructed to provide at least one more significant figure than they would routinely submit when reporting, however, little or no improvement in the incidence of such distributions has been observed. GeoPT data submitted with more significant figures are in harmony with the true value being the data distribution, turning it more into a step function. Figure 12 provides a good example of the effect of (unnecessary) rounding of significant figures on the shape of the data distribution, turning it more into a step function.

**Figure 11.** Data distribution plots of ytterbium mass fractions in HARZ01 (GeoPT38A). The left plot shows the measurement principles used over the entire data set while, for better visibility, the right plot shows the distribution of the sample preparation method for the mass fraction range < 0.1 mg kg\(^{-1}\) only with ICP-MS as dominating measurement principle. The right plot demonstrates the disparity between the relatively coherent results supplied by AD data and the variable results by methods involving fusion. The true value is most likely defined by AD data and some fusion values and is about 0.014 mg kg\(^{-1}\).

© 2022 The Authors. Geostandards and Geoanalytical Research published by John Wiley & Sons Ltd on behalf of International Association of Geoanalysts.
with the expected data distribution. This can be seen in the MnO plot of GeoPT38A in the range between 0.12 and 0.13 mg kg\(^{-1}\). However, as mentioned in the previous section, in the case of TiO\(_2\) in test material GeoPT38A, most of the XRF values are much higher and overall more highly dispersed than the ICP-MS values, which provide a plausible consensus. The apparent mode for XRF data due to rounding at 0.01 g/100 g is entirely unjustified and artificial, thus indicating significant problems with both the validation of the working range as well as rounding at low TiO\(_2\) mass fractions. The dispersion of the XRF could be due to the low signal to noise ratio and subsequent issues with background subtraction or insufficient constraints on the calibration at low mass fractions combined with an excessive influence of higher mass fractions on the slope of the calibration function. In the case of GeoPT38A, HARZ01, the best defined mode for TiO\(_2\) is dominated by ICP-MS data and indicates a likely mass fraction of about 0.002 g/100 g, that is 20 mg kg\(^{-1}\) which is truly an unusual example of an extraordinarily highly depleted and analytically challenging rock type.

**XRF-specific issues**

As a measurement principle in geoanalysis, X-ray fluorescence spectrometry (XRF) has been around for more than 60 years. It is fast, cheap and reliable, and works well as long as the sample presented to the instrument is homogeneous, calibrations are fully researched and sufficiently comprehensive, and the assumptions incorporated in the analytical data reduction software reflect the physical realities.

Under the prerequisite that the sample is effectively homogeneous, problems in XRF analysis can arise because of four main reasons: poor sensitivity (in response to prevailing excitation conditions); calibration issues; incorrect and/or inappropriate compensation of characteristic signal interferences and spectral background issues; and incorrect estimations of matrix effects.

These problems described above affect both WD-XRF and ED-XRF systems. Additionally, WD-XRF has to correct appropriately for lines of higher order that stem from Bragg's law multiple order reflections from the diffracting crystal, whereas ED-XRF has to deal with escape and sum peak artefacts associated with X-ray detection characteristics.

Based on these possible pitfalls of XRF measurements an explanation of some data peculiarities in a few GeoPT rounds is attempted. Where available, information about the

![Figure 12. Data distribution plots of TiO\(_2\) and MnO with increasing mass fractions in HARZ01 (GeoPT38A) revealing significant steps in the reported data, mainly XRF data that consists mainly of fused disc (FD) data, due to excessive rounding. The XRF data for TiO\(_2\) are clearly biased compared with the ICP-MS data. For MnO, there is little evidence of bias between the results obtained by different measurement principles. Please note that for TiO\(_2\), values greater than 0.04 g/100 g and for MnO, values below 0.105 g/100 g as well as above 0.145 g/100 g were removed to avoid compressing the mass fraction scale of the diagram.](image-url)
form of sample preparation and the XRF technique used is considered.

Poor sensitivity leading to overconfidence in the detection limit: Not all element mass fractions are adequately determined by XRF. Relatively high LODs and LOQs for low atomic number elements are rooted in reduced sensitivity and low peak to background ratios leading to poor counting statistics. TiO₂ in harzburgite GeoPT38A (Figure 12) is a good example of an element with a mass fraction that is most likely to be outside the reliable working range of the method. Limestone ML-2 (GeoPT30A) is the subject of one of the rounds with fewer than usual results submitted and fewer measurands credited with assigned values. One of the reasons is the low mass fraction of most analytes present other than CaO. The MnO mass fraction of 0.023 g/100 g has a provisional status and is one of the lowest values of the more recent GeoPT test materials. Because of this, the distribution of results for MnO in ML-2 in the data distribution plot (Figure 13) shows the typical steps due to over-rounding, and in addition, dispersion of XRF data is apparent owing to the mass fraction being most likely close to the detection limit range of XRF, that is outside a valid working range. The WD-XRF FD data in particular show a very large dispersion with data tails at the high and low mass fraction ranges, which is to be expected as the measurement uncertainty rises dramatically in the interval between LOD and LOQ.

Background modelling and peak fitting problems: Inappropriate estimation of the shape or measurement of the spectral background, or inaccurate application of overlap corrections between X-ray lines may give rise to incorrect measurement of analyte line intensities. The same outcome can also be caused by missing out corrections required for an element that is present in the measured sample but not in the calibration standard. However, such problems would arise individually and are more likely to result in greater variability of submitted GeoPT data rather than causing systematic bias.

The presence of a notable absorption edge from a major component in the sample may distort the spectral background and lead to derivation of higher intensities of analyte lines on the low energy side of the edge. Such problems most likely apply for the Manganese Nodule test material, FeMn-1 of Round 23A, which constitutes a quite unusual matrix for XRF analysis. An exceedingly high content of 45 g/100 g MnO distorts the background in the region where the characteristic K lines of elements with atomic numbers 20–25 and L lines of elements with atomic numbers 50–60 are located. Even elements that usually are determined quite well at moderate mass fractions (e.g., TiO₂ and Ba at 3000 mg kg⁻¹, V at 500 mg kg⁻¹) can be considerably overestimated in these circumstances. It is worthwhile to mention, that WD-XRF measurement results of samples prepared as fusion discs seem to fare better than those prepared as undiluted PP (see examples in Figure 14).

Implications of matrix effects

Matrix matching

Incorrect matrix (enhancement/absorption) correction factors can stem from incorrect assumptions of matrix composition. Standardless (fundamental parameter) quantification software are especially prone to this kind of error, because they rely on the correct estimation of all elements present in the measured sample for their matrix effect calculations, and in geological materials there are often unmeasurable constituents which can only be taken into account using approximations. Such problems will lead to variable XRF measurement results.

For XRF analysis with reference materials as calibrators, providing they are matrix-matched to the measured sample, the intensity and matrix effect issue should be comparable and consequently not have a significant effect on the quality of the measurement results. Background and overlap modelling, even if faulty, as well as enhancement/absorption effects should be similar for both the calibration materials and the samples being measured and the multivariate correction model will be built to take all this into account. Provided the analyte does not exceed the calibration range and there is no significant matrix mismatch or missing elements, it can be expected that derived analyte mass fractions are valid.

Sample preparation issues

It is worth noting that dilution of the sample by fusion with a suitable flux reduces the variation in size of matrix effects and therefore any consequent potential errors. This sample preparation technique ensures the complete dissolution of the analytes and generally guarantees the production of a reliably homogeneous sample. The same does not hold true for samples measured as pressed powder pellets (PP). This form of sample preparation is very commonly applied for trace element measurements, because the sample is easy to prepare, may be hardly diluted at all and LODs and LOQs are usually lower, but PP are prone to inhomogeneity, especially in terms of...
minerological effects as described by Claisse and Samson (1961). When an analyte is mainly present in a dense mineral fraction of the sample, much of its fluorescence radiation is absorbed by the heavier elements within the grain immediately surrounding the emitting atom. The X-ray attenuation correction, however, is calculated based on the average matrix composition, which is considerably less dense than the immediate environment and therefore would cause less attenuation. This results in the underestimation of elements present in those heavy mineral grains, impacting on their measurement results, particularly if that is their dominant form of occurrence.

An example of this mineralogical effect is provided by GeoPT38A harzburgite, where Cr is present in the relatively dense mineral, chromite (Figures 3 and 1S). Most of the XRF powder pellet mass fraction values for Cr are significantly underestimated compared with the likely consensus value based on the mode of the most coherent data derived by other methods including XRF measurement results obtained from fusion discs. Multimodal distributions have also been reported and discussed for Co and Ni in the GeoPT report for Round 39 (Webb et al. 2016b). For both of these measurands, the bimodal distributions are due to results from different XRF sample preparation techniques with powder pellet data yielding lower values (Figure 1S). Localised absorption of analyte X-rays in dense host minerals is likely to be the cause for such results. Mineralogical evidence recently obtained which supports the attenuation hypothesis to explain the data distribution of Zn in GeoPT48 can be found in detailed work by Rushton and Locinska (2022). Through this study, it is demonstrated that zinc in the Monzonite MzBP-1/larvikite sample (GeoPT48) is present either in a fine-grained exsolution phase (gahnite ZnAl2O4) or in sulphides present as secondary minerals in fractures of the host garnet, a spinel that has a high density.

The picture is a little more complex for Nb and Ta in GeoPT23 Separation Lake pegmatite. As demonstrated in Figure 16, consistently lower values for PP compared with fusion discs measured by XRF suggest a mineralogical matrix effect. In this case columbite-tantalite ((Fe, Mn)(Nb, Ta)2O6) or wolgainite ((Mn, Fe)5Sn(Ta, Nb)2O12) are the dense minerals (density between 5.3 and 8.2 g cm−3) that disproportionately attenuate Nb and Ta radiation. Another explanation for the problematic XRF data might be that in WD-XRF Ta α is compromised by a 2nd order Nb Kα interference and Nb Kα also may suffer some (uncorrected) interference from a Y Kβ line.

**ICP-MS specific issues**

Something completely different: The bimodal distribution of Sr mass fraction data for OU-9 the Separation
Lake pegmatite, shown in Figure 17 has two distinct modes, one related to ICP-MS results and one to XRF results. This detailed information on instrumentation and sample preparation method was not available in a readily visualised form at the time of writing the GeoPT23 report; however, the bimodal distribution was obvious to the authors (Webb et al. 2008) who recognised populations of Sr results centred at 9.54 ppm and 35.65 ppm. On preliminary consideration of the methods used to produce these data, the lower value was thought to largely represent ICP-MS data and the higher value to largely represent XRF data. A speculative interpretation for this disparity in the results was given in the report as follows: ‘In this case it is clear that the unusually high Sr results are obtained by XRF on account of a severe Ta interference, therefore the low value consensus is likely to be the more reliable’. Figure 18 clearly disproves this theory where the Sr Kα line in the XRF spectrum of OU-9 is well resolved from Nb and Rb and is interference free. Tantalum X-ray lines are in a different spectral region altogether with no evidence of higher order lines interfering with Sr Kα.

Figure 17 also reveals that the median of the ICP-MS measurement results is much lower than the median of the data produced by other ICP methods (i.e., ICP-OES and ICP-AES). The low mass fractions reported by ICP-MS are obviously related to mass spectrometry as a detection principle and not to sample preparation or calibration issues. The story becomes even more remarkable when taking the Rb/Sr ratio and age of this pegmatite (a highly fractionated rock type) into account. The Separation Lake pegmatite is associated with a pluton dated at 2.64 billion years (Larbi et al. 1999), and has a bulk rock Rb mass fraction reported to be ~2500 mg kg⁻¹ and therefore an extremely high Rb/Sr ratio. ⁸⁷Rb decays to stable ⁸⁷Sr by emission of a negative beta particle (β⁻) and thus the radiogenic ingrowth has an influence on the ⁸⁷Sr isotopic abundance of the test material although the half-life of ⁸⁷Rb is a staggering 48 billion years (4.8 × 10¹⁰ years).

Due to the isobaric interference of ⁸⁷Rb⁺ on ⁸⁷Sr⁺ at m/z = 87 and the low isotopic abundance of ⁸⁶Sr⁺, m/z = 88 is the preferred signal for Sr quantification. However, this only works correctly, when the isotopic composition of the calibration solution and that of the measurement test solution are identical or at least very close to each other. But in GeoPT23, the n(⁸⁷Sr)/n(⁸⁶Sr) ratio changed from a typical range of (n(⁸⁷Sr)/n(⁸⁶Sr))₀ = 0.7 to 0.8, to a much higher value of (n(⁸⁷Sr)/n(⁸⁶Sr))₀ = 311 within the 2.64 billion years since the formation of the pegmatite rock. Therefore, the Sr isotopic composition differed greatly in both the calibration and measurement solutions, and this is why the mass fractions measured by ICP-OES/AES and XRF are significantly higher and are thus correct. Both in ICP plasma and by X-ray excitation, all Sr atoms, regardless of their abundance are excited and can be measured simultaneously. Hence, the correct mass fraction of Sr in this GeoPT test material was determined using either XRF or ICP techniques without MS.

Figure 14. Violin plots for Sc and V mass fractions in Manganese Nodule GeoPT23A. XRF analysis data for Sc and V pressed pellets lie generally higher than XRF on fused discs. The distorted background at the root of this problem is clearly more prominent for undiluted PP samples.
Grained Zn-rich minerals (e.g., gahnite ZnAl₂O₄) hosted by magnetite in Monzonite MzBP-1 GeoNi radiation disproportionally. Clear mineralogical evidence to support the hypothesis has been found for fine-thought to be present in iron sulphide grains, which are denser than the average matrix and therefore attenuate the dense mineral grains present in PP, but not in fused discs. Nickel-bearing minerals in SyMP-1 (GeoPT39) PP are thought to be present in iron sulphide grains, which are denser than the average matrix and therefore attenuate the Ni radiation disproportionally. Clear mineralogical evidence to support the hypothesis has been found for fine-grained Zn-rich minerals (e.g., gahnite ZnAl₂O₄) hosted by magnetite in Monzonite MzBP-1 GeoPT48.

To back up this hypothesis, the Rb and the ICP-MS Sr mass fraction data can be used to calculate the amount of radiogenic ingrowth of ⁸⁷Sr through the β⁻ decay of ⁸⁷Rb.

The mass fraction of Sr at the time of the pegmatite formation, i.e., without radiogenic ingrowth, based on the ICP-MS data is w(Sr) = 9.54 mg/kg. The Rb mass fraction is w(Rb) = 2500 mg/kg. The $\frac{ deport{84}Sr}{Sr} nat$ fraction of a natural sample is, $x(84Sr) nat = 0.00056$. The $\frac{ deport{86}Sr}{Sr} nat$ fraction of a natural sample is, $x(86Sr) nat = 0.0986$. The $\frac{ deport{88}Sr}{Sr} nat$ fraction of a natural sample is, $x(88Sr) nat = 0.8258$ and the $\frac{ deport{87}Rb}{Rb} nat$ fraction of a natural sample is, $x(87Rb) = 0.278$. With a molar mass of Rb with $M_{Rb} = 85.47$ g/mol and of natural Sr with $M_{Sr} = 87.62$ g/mol.

### Figure 15
Data distribution plots of Cr, Ni and Zn mass fraction in GeoPT38A, GeoPT39 and GeoPT48. The distributions do not follow the expected normal distribution. Similarly to the distributions revealed in earlier Zr plots (shown in Figure 5), PP data plot largely at the lower ranges of both distributions. One reason for the shape of these distributions are mineralogical effects causing attenuation of the Cr, Ni and Zn fluorescence X-rays emanating from dense mineral grains present in PP, but not in fused discs. Nickel-bearing minerals in SyMP-1 (GeoPT39) PP are thought to be present in iron sulphide grains, which are denser than the average matrix and therefore attenuate the Ni radiation disproportionally. Clear mineralogical evidence to support the hypothesis has been found for fine-grained Zn-rich minerals (e.g., gahnite ZnAl₂O₄) hosted by magnetite in Monzonite MzBP-1 GeoPT48.

### Figure 16
Data distribution plots of Nb and Ta mass fraction in GeoPT23. These distributions do not follow a normal distribution. One reason for this distribution without a point of inflection that would represent a clear consensus are the unusually high mass fractions in this sample which would exceed the upper end of most calibration ranges established in the laboratories. PP data are clearly biased towards low values.
with an assumed initial ratio \( \frac{n(\text{Rb})}{n(\text{Sr})} \) of 0.701 at the time of the pegmatite formation \( t = 2.64 \times 10^9 \) years ago assuming a chondritic development of the source with a decay constant of \( \lambda = 1.47 \times 10^{-11} \text{a}^{-1} \) for \( ^{87}\text{Rb} \) and using the following equations, an estimation of today's \( t = 0 \) \( \frac{n(\text{Rb})}{n(\text{Sr})} \) can be calculated: \( n(\text{Rb}) = n(\text{Sr}) \times \frac{\lambda}{\lambda + \lambda_{\text{nat}}} \) and \( \frac{n(\text{Rb})}{n(\text{Sr})} \) is \( \left( \frac{n(\text{Rb})}{n(\text{Sr})} \right) = \left( \frac{n(\text{Rb})}{n(\text{Sr})} \right)_{t=0} + \left( \frac{n(\text{Rb})}{n(\text{Sr})} \right)_{t=0} \times (e^{\lambda t} - 1) = 30.7 \) this results in today's Sr mass fraction through the radiogenic ingrowth of \( ^{87}\text{Sr} \) of \( w(\text{Sr})_0 = 30 \text{ mg/kg} \).

\( w(\text{Sr})_0 = 9.54 \text{ mg/kg} \ldots \text{mass fraction of Sr based on ICP-MS data i.e. without radiogenic ingrowth} \)

\( w(\text{Rb}) = 2500 \text{ mg/kg} \ldots \text{mass fraction Rb} \)

\( \text{abundance}^{84}\text{Sr}_{\text{nat}} = 0.00056 \)

\( \text{abundance}^{88}\text{Sr}_{\text{nat}} = 0.0986 \)

\( \text{abundance}^{87}\text{Rb}_{\text{nat}} = 0.278 \)

\( M_{\text{Rb}} = 85.47 \text{ g/mol} \ldots \text{molar mass Rb} \)
\[ M_{\text{Srnat}} = 87.62 \text{g/mol} \ldots \text{molar mass Sr} \]

\[ \left( \frac{n(\text{Sr})}{n(\text{Sr})} \right)_i = 0.701 \ldots \text{initial ratio, assuming chondritic development of the source} \]

\[ \frac{n(\text{Rb})}{n(\text{Sr})} = \frac{w(\text{Rb})}{w(\text{Sr})} \cdot \frac{\text{abundance}^{87}\text{Rb}}{\text{abundance}^{86}\text{Sr}} \]

\[ \lambda \cdot 1.47 \times 10^{-11} \text{a}^{-1} \]

\[ t = \text{time since formation in years a which is} \ 2.64 \times 10^9 \]

This results in a mass fraction through the radiogenic ingrowth of \(^{87}\text{Sr}\)

\[ w(Sr_0) = 30 \text{ mg/kg} \]

This approximate calculation supports the hypothesis that the estimated Sr mass fraction \(w(Sr)\) is about 30 mg kg\(^{-1}\), which is close to the median XRF mass fraction, 35.4 mg kg\(^{-1}\). It has to be noted, that the GeoPT23 example is very special as both the \(Rb/Sr\) ratio and the age of the rock are extremely high. Though the variation of \(n(\text{Sr})/n(\text{Sr})\) is limited amongst most rock types, but whenever an unusually high \(w(Rb/Sr)\) ratio is detected in a sample, ICP-MS analysts need to be cautious about the influence of radiogenic ingrowth on the mass fractions of Sr.

**Conclusions**

A review of the data collected over 25 years of GeoPT shows the value of this very successful programme supported throughout by the IAG. Individual participants benefit from the evaluation of their performance through the reports of each round, and the geoanalytical community at large can gain valuable insights into many of the problems associated with the analysis of silicate (and carbonate) rocks. The most obvious problems concern sample preparation. It is the refractory minerals such as zircon and chromite that give rise to persistently lower yields of Zr, Hf and sometimes HREE mass fractions from samples with granitic (also detrital sedimentary) composition. Similarly, lower results are obtained for Cr mass fractions for unserpentinised peridotites when acid attack (as is common in spectrographic methods) is used and the method is not properly validated. In addition, samples containing elements such as Th, U, Nb and Ta at mass fractions exceeding routinely calibrated ranges tend to produce very high dispersion in the data reported. This is not caused by problems with sample preparation, but by disproportionate estimation of mass fractions that are well outside the validated calibration range.

![Figure 18. XRF scan over the region of interest with a FD. No overlap of the Rb and Nb lines with Sr is visible. There are no Ta lines in this wavelength range; hence, Ta, Nb and Rb cannot be responsible for higher Sr values relative to the ICP-MS results.](image-url)
Test materials from the GeoPT programme with very low and very high mass fractions of constituents are not especially well-characterised, yet it is samples of this kind that are most likely to be of considerable and growing interest for method validation and calibration purposes for future applications. Therefore, efforts should be made to provide through GeoPT better tests of performance for materials with more extreme compositions for which reference materials are not generally available from commonly used RM providers, such as the USGS, for example BCR-2 and BHVO-2. Examples of desirable measurands for validation at low mass fractions include TiO₂ at mass fractions, below 0.01 g/100 g, and measurands such as Pb, REE, Zr at high mass fractions, in the range of 1000 mg kg⁻¹ (= 0.1 g/100 g). Test materials must be well characterised and preferably associated with relatively small uncertainties to make them subsequently useful as reference materials or as measurement standards for extending calibration ranges to facilitate the characterisation of peculiar or unusual rock types.

An interesting side aspect of the detailed study of data from GeoPT rounds and the recognition of anomalous distributions of results as displayed in the graduated data distribution chart is the often obvious disparity relating to the applied measurement principles of XRF and ICP-MS. Frequently, the ICP-MS data are more coherent and superior over XRF data, especially in the evaluation of very low mass fractions, but in the case of Sr in Separation Lake pegmatite (GeoPT23), it became clear that most of the ICP-MS data were incorrect and that the ICP-OES/AES and XRF data were essentially correct. In this case, the underlying assumptions for the ICP-MS calibrations were false. ICP-MS calibration relies on the assumption that the isotopic composition of the measurement standards and the analytical sample are identical. However, as demonstrated within this report, Sr also has a radiogenic isotope that is inherently variable. In the case of GeoPT23, the extremely high Rb/Sr ratio and high age of the host rock resulted in a 2.5-fold increase in the initial Sr mass fraction over time from radiogenic ingrowth.

Analytical geochemists can be proud to look back upon 70 years of characterisation of geological reference materials and 25 years of the GeoPT programme. This period has seen incredible improvements in development of instrumentation and analytical applications. However, detailed knowledge of the petrography and mineralogy of rock materials as well as sound training in principles of analytical geochemistry are still essential to generate compositional data that is fit for scientific purposes.

**Acknowledgements**

We want to thank Jenny Cook (IAGeo Limited) for providing a sample GeoPT PT23A. We also want to thank Phil Potts for the thoughtful editorial handling and the anonymous reviewers for their critical and helpful comments that greatly improved the quality of this work.

**Data availability statement**

The data, except the metadata, can be found in the individual GeoPT reports found here http://www.geoanalyst.org/geopt-previous-rounds/.

**References**


references


Rushton J.C. and Locinsk A.M. (2022) Modal and microanalysis of Karara and Meissen Granites, and Laniwikte test samples. BGR.


Reassessment of HF/HNO₃ decomposition capability in the high-pressure digestion of felsic rocks for multi-element determination by ICP-MS. Geostandards and Geoanalytical Research, 36, 271–289.