PROFICIENCY TESTING: KNOWING HOW FAR YOU CAN TRUST YOUR DATA

Michael Wiedenbeck[1]

Proficiency testing (PT) is a cornerstone of good analytical practice, providing one of the few means of really testing the quality of a lab's analytical output. Participation in a routine PT scheme can form part of lab accreditation, and for many industrial applications (e.g. ore grade determinations for the mining industry) it is often expected that mineral assay laboratories track their PT results. As an example, National Instrument 43-101, which is Canada's "Standards of Disclosure for Mineral Projects", requires the written disclosure of the methods used to verify quantitative data related to new mineral prospects seeking public funding. Hence, proficiency testing has become a *de facto* component of ore grade assessment whenever a mining company applies for listing on the Toronto stock exchange. Clearly, proficiency testing is not only of research importance but also plays a central role in the world's mining industry.

So what exactly is a proficiency test? A laboratory will receive a material for analysis and it will submit its results to the PT organizers, who will, in turn, evaluate the submitted result against some benchmark believed to represent the true concentration of the analyte of interest. The critical point about a proficiency testing scheme is that the participating laboratories are required to perform their analyses using routine procedures – no special sample preparation methods, no special efforts towards improved data quality, no extra steps in the quality assurance process are allowed. Thus, if a laboratory's results closely match the benchmark values then one can have confidence that the analytical methods being used on a routine basis are trustworthy. However, if the report from the PT organizers shows significant discrepancies from the set benchmark then the given laboratory needs to take steps to investigate a so-called "out of control" situation.

The International Association of Geoanalysts (IAG) has operated a whole-rock PT programme since 1996, the first of which involved 49 laboratories reporting concentration results on 51 elements from the Threlkeld microgranite of Cumbria (UK) (Thompson et al. 1996). This round of PT testing involved the dispatch of packets of milled whole-rock powders (Fig. 1) where the participating labs were given three months to report their results. Over the following 20 years, the number of participating labs has now grown to over 100. Likewise, over the years the analytical methods used for such analyses have evolved, such that today's datasets are dominated by solution inductively coupled plasma (ICP) and X-ray fluorescence (XRF) methods – largely gone are the days of neutron activation, wet chemical determinations and atomic absorption analyses. The evaluation of PT data often relies on the experience of the programme organizers, with target benchmarks often based on either the robust mean or the median values of the submitted dataset. So how do we know that these target values are close to the true concentration values of the many various elements? This topic was addressed by Potts et al. (2015) who compared the target values derived from

FIGURE 1   Sample sachets from earlier rounds of the IAG's GeoPT proficiency testing programme. Each unit contains around 50 g of whole-rock powder.

1   Deutsches GeoForschungsZentrum GFZ
   Potsdam, Germany
    michael.wiedenbeck@gfz-potsdam.de
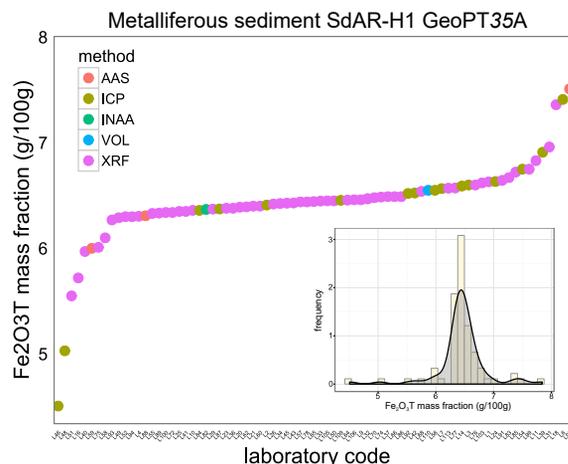
FIGURE 2   Data distribution plot for iron content – as Fe_2O_3 (total) on y-axis – in the metalliferous sediment SdAR-H1, from the July 2014 round of the GeoPT programme. **INSET**: Equivalent histogram plot along with a kernel density curve meant to highlight the Gaussian-like distribution of the central population. Analytical method abbreviations are as follows: AAS (atomic absorption spectroscopy), ICP (inductively coupled plasma), INAA (instrumental neutron activation analysis), VOL (volumetric), XRF (X-ray fluorescence).

the IAG's GeoPT dataset to those derived for the same materials based on painstaking certification projects: few significant differences were found between six such datasets.

Now I would like to present several interesting data sets that reveal the presence of analytical problems which might, were in not for the participation in a proficiency testing programme, have gone unnoticed.

First, I present an example of what might be called a well-behaved dataset for iron in a sediment powder (Fig. 2), which is the norm for such work. The data are plotted in ascending order in terms of the reported Fe_2O_3 (total) concentration. What one sees from this plot is that the data closely approximate a Gaussian distribution (Fig. 2 INSET). However, this dataset also shows a few obvious outliers at both the high and the low ends of the mass fraction spectrum. In this dataset, the robust mean of 6.46 g/100 g and the median value of 6.45 g/100 g are in excellent agreement, indicating that there is no significant skewing. Two things are noteworthy: 1) There were a large number of analytical techniques used to obtain the 79 values contained in this report; 2) There appears to be a slight bias of the ICP results towards higher concentrations compared to the XRF data, albeit with the two lowest reported concentration values being from ICP laboratories. Obviously, those dozen or so laboratories that produced results at the extremes of this data spectrum need to investigate their measurement procedures. Nonetheless, the overall distribution of results for iron in this sediment powder looks satisfactory.
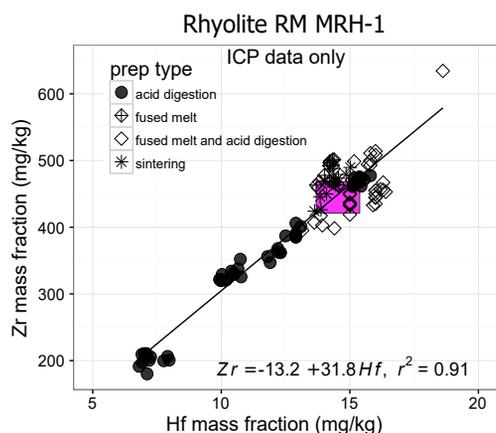
In contrast to the dataset shown in FIGURE 2, I present a dataset for strontium concentrations in the Separation Lake pegmatite (Ontario, Canada) that were reported during the August 2008 round of GeoPT (Fig. 3). Not only do the data show an obvious non-Gaussian data distribution, but there are two distinct plateaus, with the low abundance plateau value of ~8 mg/kg being exclusively populated by laboratories using ICP technology. Furthermore, nearly all members of this group of laboratories used a mass spectrometer as their signal detection system. The second clustering of data around ~35 mg/kg total Sr is dominated by XRF data, but this group also contains a number of analysts reporting results based on ICP technology, the majority of whom recorded their data using an emission spectrograph. The explanation here is easy to understand, in that a common approach for determining total Sr mass
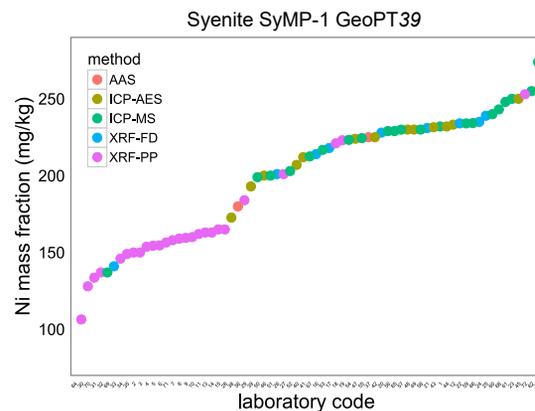
Separation Lake pegmatite OU-9 Geo*PT*23
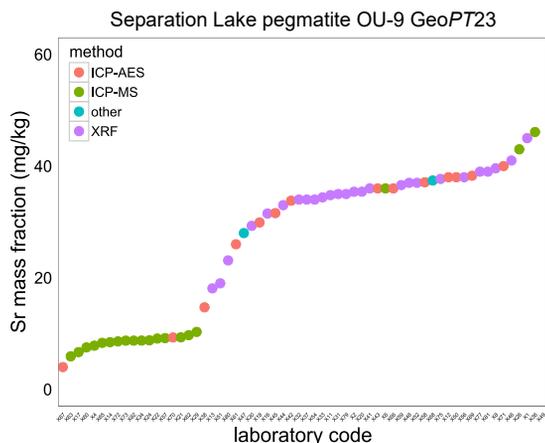
FIGURE 3 Results reported for the strontium mass fraction from the Separation Lake pegmatite (Ontario, Canada). Note the strong bias between the ICP results depending on whether a mass spectrometer (MS) or atomic emission spectrograph (AES) signal detection system was used. Other abbreviations as for FIGURE 2.

Syenite SyMP-1 GeoPT*39*

FIGURE 5 Ni concentration data reported for the SyMP-1 syenite showing a non-Gaussian data distribution. Analytical method abbreviations are as follows: AAS (atomic absorption spectroscopy), ICP–AES (inductively coupled plasma atomic emission spectroscopy), ICP–MS (inductively coupled plasma mass spectrometry), XRF–FD (X-ray fluorescence using fused glass disc sample preparation), XRF–PP (X-ray fluorescence using pressed powder sample preparation).

fraction using mass spectrometry relies on the determination only of the ^{88}Sr content in conjunction with the supposedly known isotopic composition of natural strontium. The Separation Lake pegmatite has an age of 2.64 Ga (Tindle et al. 1998) and a bulk rock rubidium concentration of 2,501 ± 22 mg/kg as defined by the robust mean of the proficiency testing data from 64 laboratories. Considering such a high Rb/Sr ratio, a late Archean age and that the half-life of ^{87}Rb is 48 Ga, one can understand that this material's strontium budget is dominated by radiogenic ^{87}Sr. This enrichment in ^{87}Sr will go undetected if the analytical method assesses only the ^{88}Sr in the sample. This dataset should have been a wake-up call for many ICP–MS analysts who might otherwise have continued to make this oversight were it not for their participation in a proficiency testing programme.

A second interesting example involves the MRH-1 rhyolite reference material. The XRF data for zirconium showed a very well-defined mode of ~465 mg/kg, whereas the ICP zirconium results showed a broad spectrum of results with a range of ~200–500 mg/kg. So, why the difference between the XRF and ICP results and the wide range within the ICP results? A clue towards explaining this riddle is that the ICP results are influenced by the sample preparation method, with acid digestion revealing by far the largest scatter in the results. This may come as a

bit of a surprise because one would think that a "rhyolitic glass" is one of the less challenging silicate matrices to convert to a solution. A further clue is the strong correlation found between the reported hafnium and zirconium results (FIG. 4). The slope of the best-fit line of Zr/Hf ≈ 31.8 is within the range that is common for zircon (e.g. Hoskin and Schaltegger 2003). Hence, the MRH-1 data clearly point towards some of the participating labs having unsuspected problems getting complete sample dissolution, and, in particular, being unable to fully digest zircon. Those laboratories that reported low Zr (as well as low Hf) abundances should investigate their sample preparation procedure for this suspected problem.

As a final example, I would like to share data from the most recently completed round of the Geo*PT* programme. The data come from nickel concentrations reported for the SyMP-1 syenite. The results are clearly non-Gaussian in their distribution, with a lower concentration plateau of around 160 mg/kg populated nearly exclusively by XRF data (FIG. 5). Significantly, this group of data consist almost exclusively of laboratories using pressed powder (PP) sample preparation. In contrast, most of the data based on ICP, along with numerous XRF results based on fused disc (FD) sample preparation, cluster around the significantly higher value of ~230 mg/kg. Clearly, there is a method-dependent analytical problem influencing the Ni dataset, and here a bit of knowledge about this syenite's mineralogy is key to understanding this rather perplexing observation. This syenite contains a significant concentration of 0.15 g/100 g of sulfur, much of this is in the form of iron sulfides. It turns out that much of the Ni contained in this material is also housed within these sulfide grains. The XRF measurements need to be corrected for X-ray self-absorption by the sample material, yet the absorption coefficients are routinely based on the bulk composition of the material. In other words, one might suspect that the correction made by many of the XRF laboratories for determining Ni would have been based on a silicate composition, whereas a more appropriate self-absorption correction for Ni measurements on pressed powder pellet samples would use the iron sulfide matrix that is believed to house the nickel (Webb et al. 2016). According to this model, many of the XRF facilities would have under-corrected for self-absorption, resulting in the determination of values for the bulk Ni content of this material that are too low.

Rhyolite RM MRH-1

$$Zr = -13.2 + 31.8\,Hf, \; r^2 = 0.91$$

FIGURE 4 Plot of the Hf vs. Zr results for the MRH-1 rhyolite. These data are only for the labs reporting ICP-based data using either mass spectrometric or optical signal detection. The magenta-coloured field indicates the certified values for this material. Best-fit line and its associated parameters are also shown.

I hope these few case studies help illustrate some intriguing pitfalls in the analysis of bulk rock samples. Certainly, I think these examples put a spotlight on the power of a proficiency testing programme for revealing shortcomings in a laboratory's methods. I would like to end by thanking Peter Webb (Milton Keynes, UK), who has been a long-term member of the GeoPT steering committee, Thomas Meisel (Leoben, Austria), who is President of the International Association of Geoanalysts, and Marcus Burnham (Sudbury, Canada) of the Ontario Geological Survey, all of whom provided invaluable assistance in preparing this toolkit contribution. For those geochemists eager to learn more about assuring the quality of their data, I would recommend you visit the proficiency testing section of the www.geoanalyst.org web page.

Best Regards from Potsdam
Michael Wiedenbeck

REFERENCES

Hoskin PWO, Schaltegger U (2003) The composition of zircon and igneous and metamorphic petrogenesis. Reviews in Mineralogy and Geochemistry 53: 27-62

Potts PJ, Thompson M, Webb PC (2015) The reliability of assigned values from the GeoPT proficiency testing programme from an evaluation of data for six test materials that have been characterised as certified reference materials. Geostandards and Geoanalytical Research 39: 407-417

Tindle AG, Breaks FW, Webb PC (1998) Wodginite-group minerals from the Separation Rapids rare-element Granitic Pegmatite Group, Northwestern Ontario. Canadian Mineralogist 36: 637-658

Thompson M, Potts PJ, Webb PC (1996) GeoPT1. International proficiency test for analytical geochemistry laboratories – report on round 1 (July 1996). Geostandards and Geoanalytical Research 20: 295-325

Webb PC, Thompson M, Potts PJ, Gowing CJB, Wilson SA (2016) GeoPT39 – an International Proficiency Test for Analytical Geochemistry Laboratories – Report on Round 39 (Syenite, SyMP-1) / July 2016. International Association of Geoanalysts, 40 pp

ADVERTISERS IN THIS ISSUE