amc technical brief

Editor: M. Thompson Analytical Method Committee AMCTB No 25 October 2006

How good were analysts in the good old days before instrumentation?

Gravimetric analysis was in widespread use until about 1960. It could be executed with high precision. Under ideal conditions it is not difficult to obtain a repeatability RSD better than one part in a thousand in the determination of a major constituent, a precision higher than most modern instrumental methods achieve. But the high precision of gravimetry under ideal conditions is one thing, and the uncertainty involved in routine gravimetric analysis is another. Just how good was gravimetric analysis when it involved multiple complex separations and very considerable manipulative skill, as in the classical schemes for the analysis of silicate rocks?

Fortunately we are in a position to answer that question, because we have on record the detailed results of an interlaboratory study conducted by the Department of Geology of the Massachusetts Institute of Technology (Fairbairn, 1951) [1]. The test materials were two rocks and the major elements were determined gravimetrically.

Fairbairn's interlaboratory study

The two rocks, a granite and a diabase, were each crushed to a fine powder and thoroughly mixed. Portions were then distributed to 34 different analysts, all experts in silicate analysis. The main purpose was to establish reference values for the suite of major and minor elements, a certification study. Out of this came the first generally available reference materials, the famous G-1 and W-1 distributed by the US Geological Survey. The materials were intended to be used subsequently as reference materials for the routine analysis of silicate rocks by the less precise "rapid methods" that were coming into fashion, mostly colorimetric and spectrographic. G-1 and W-1 remained in use for a number of years before stocks were exhausted.

The initial results of the study, however, caused a shudder of consternation in the analytical world. The degree of agreement between the expert participants was far worse than suggested by the repeatability precisions that could be obtained within a single laboratory. Of course, this was only the first attempt: better (but still not good) reproducibility precision was obtained in a subsequent study. With the benefit of a lofty hindsight, we can now regard this situation as entirely normal, having seen the evidence of thousands of collaborative trials and proficiency tests. At the time, however, it was an unpalatable shock.

A modern benchmark

So just how good (or bad) were these gravimetric results? We can get a good idea of this by treating the study as if it were a proficiency test. An ideal benchmark is readily available: the current proficiency test Geo*PT*, which is concerned with the analysis of (mostly) silicate rocks. The quantitative standard of performance set by any scheme is the standard deviation for proficiency σ_p (informally known as the "target value"). This σ_p is used in the calculation of a z-score, $z = (x - x_a)/\sigma_p$, from a participant's result x and the assigned value x_a. In Geo*PT*, the value of σ_p regarded as fit for purpose for practitioners of "pure geochemistry" is one half of that given by the Horwitz function, namely $\sigma_p = 0.01 \times c^{0.8495}$, where both σ_p and c are expressed as mass fractions [2]. Participants in Geo*PT* use XRF, ICP-AES or ICP-MS methods for determining major and minor elements, so it is of considerable interest to compare modern instrumental z-scores with those attributed retrospectively to Fairbairn's results. The outcome is shown overleaf.

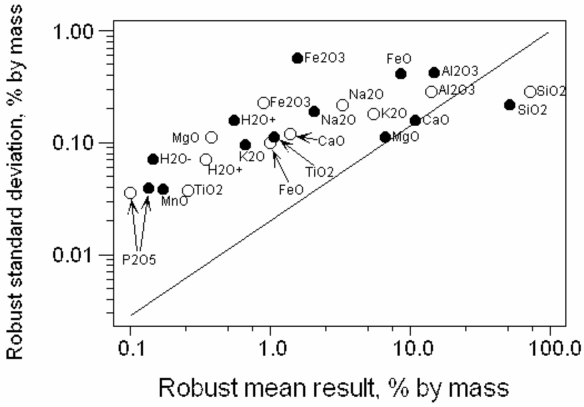

Figure 1. Robust statistics for the Fairbairn data (first attempt). Open circles show results for G-1, solid circles for W-1. The line shows the fitness function $\sigma_p = 0.01 \times c^{0.8495}$.

The classical scheme for silicate rocks (simplified)

Silica was determined first, by precipitation and dehydration with hydrochloric acid after the rock had been decomposed by fusion with sodium carbonate. The residue was ignited and weighed as SiO_2. Iron and aluminium (with some coprecipitants) were precipitated together from the filtrate by ammonia, as a mixture of hydrous Fe_2O_3 and Al_2O_3 and weighed after ignition. Iron in the residue was determined by titrimetry, and aluminium by difference, with small corrections based on the separate colorimetric determination of the coprecipitated minor elements. Calcium was precipitated from the filtrate as oxalate and weighed as CaO after ignition. Magnesium was precipitated from the filtrate as $MgNH_4PO_4.6H_2O$ and weighed as $Mg_2P_2O_7$ after ignition. Analysts used either gravimetry or flame photometry for the determination of sodium and potassium. (Iron(II) was determined on a separate test portion by titrimetry.) Precipitations were often repeated to reduce coprecipitation. These sequential operations took about two weeks for the routine suite of analytes (SiO_2, Al_2O_3, Fe_2O_3, FeO, CaO, MgO, Na_2O, K_2O, TiO_2, MnO, P_2O_5, and H_2O) although several samples could be handled simultaneously.

Overview of the results

We can get an overview of the outcome by plotting the robust standard deviations of the results against the mean value: if the quality of the analysis conforms to the criterion, the results should be clustered around a plot of $\sigma_p = 0.01 \times c^{0.8495}$. What we find is quite different (Figure 1). Over most of the concentration range (about 0.1-5%) we see the standard deviations averaging about five times the expected value. And these are *robust* statistics—the outliers have already been downweighted! (Above a concentration of 10 % we see results, for example for silica, that are considerably better. However, it is well known that the Horwitz function tends to over-estimate standard deviation in that range.

Z-scores from the Fairbairn data

We can get a more detailed idea of the quality of the data by converting the G-1 and W-1 results into z-scores, according to the GeoPT Protocol, and comparing them with modern z-scores from a typical round of the scheme (from 2003). Some of the results are shown in Figure 2. For silicon we see that the gravimetric analysts could do rather well, with a smaller dispersion and fewer outliers than GeoPT. For aluminium the pattern is reversed, with somewhat greater dispersion for the gravimetric scores and more outliers. For many elements (examples total iron, sodium) we see a much greater dispersion in the gravimetric results. For calcium and magnesium (not shown) there is a dispersion comparable with GeoPT for W-1, but a greater dispersion for G-1, which reflects the lower concentration of both elements in G-1. For titanium (and other minor elements not shown) modern instrumental methods give much better results than the older colorimetric methods.

Conclusions

What can we conclude from this retro-analysis, apart from the obvious facts that (a) the classical analysts did pretty well given the difficult methods at their disposal, and (b) we can usually obtain more accuracy with *far* less effort nowadays? One factor, not immediately apparent, is that there was no attempt to standardise procedures for the Fairbairn study. Every analyst had his or her own little modifications of the basic methods used. Standardised procedures, however, as used in food analysis, are shown by collaborative trials from that period to provide better precision. Another factor was that the accuracy obtainable in silicate analysis was strongly dependent on the concentration of the analyte.

Perhaps the important message for modern analysts is simple and still not fully appreciated in all sectors: repeatability standard deviation seldom provides a realistic approximation to uncertainty. This message has been slow to sink in. Even as recently as ten years ago, many analysts found it difficult to accept that reproducibility standard deviations derived from collaborative trials were usually better estimates of their uncertainty than within-laboratory precisions.

References

1. W H Fairbairn, *A cooperative investigation of precision and accuracy in chemical, spectrochemical and modal analysis of silicate rocks.* Bulletin 980, US Geological Survey, Washington DC, 1951.
2. Geo*PT* Protocol. Can be downloaded from http://geoanalyst.org

This Technical Brief was prepared for the Analytical Methods Committee by the Statistical Subcommittee (Chairman M Thompson).

Other AMC documents and products can be found at
www.rsc.org/amc

CPD Certification I certify that I have studied this document as a contribution to Continuing Professional Development.

Name…………………..

Signature……………………….…….Date………………………….

Name of supervisor…………………………………………………

Signature…………………….……….Date……….........................

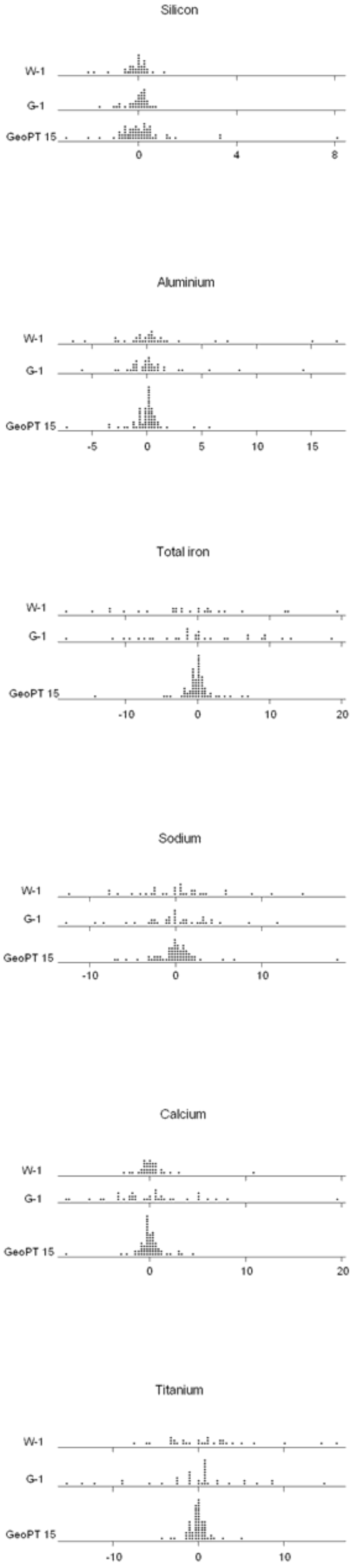

Figure 2. Z-scores of the results of the Fairbairn study compared with those from a recent proficiency test. Scores outside the range ± 20 are not shown.