PhotonAssay™ – techniques and workflows to establish fitness for purpose

R Sterk[1], J Nessler[2], M van de Ven[3], M Knorsch[4] and D Woestenburg[5]

1. Principal Consultant, RSC, Dunedin 9600, New Zealand. Email: r.sterk@rscmme.com
2. Consulting Exploration Geologist, RSC, EMEA. Email: j.nessler@rscmme.com
3. Senior Resource Consultant, RSC, Dunedin 9600, New Zealand. Email: m.vandeven@rscmme.com
4. Consulting Geoscientist, RSC, Perth WA 6005. Email: m.knorsch@rscmme.com
5. Consulting Statistician, The Hague, Netherlands. Email: dionwoestenburg@hotmail.com

ABSTRACT

Analytical techniques based on gamma activation analysis (GAA) are rapidly gaining popularity in the mining industry for analysis of Au, Cu, and Ag. Since being introduced commercially by Chrysos Corporation in 2018 under the brand name PhotonAssay™ (CPA), 130 companies have reported Au exploration results to major stock markets using data collected with this technique, and 12 companies have used such results to support a publicly reported mineral resource estimate (MRE).

While it has been established that, under best-practice operating conditions, results from CPA are accurate and precise, test results may not prove accuracy for every style of mineralisation and any individual CPA instrument. Therefore, many operators carry out their own test work, comparing CPA assays with results obtained by conventional techniques (eg fire assay). The authors' experience with comparative studies around the world is that they are often not designed for rigorous scientific testing, and hence do not meet the objective of drawing robust conclusions on the quality of CPA data. Additionally, the issue of materiality, key to every mining and exploration project, is rarely discussed, and studies do not define a data quality objective (ie when is something fit for purpose).

In this paper, we add detail to previously published recommended testing programmes, setting out the sample selection process, and detail all critical steps during sample preparation, including the collection of true duplicates for precision benchmarking.

Based on several real-world examples, we then provide a step-by-step process to assess the validity of the testing data, accuracy and precision, and whether any observed differences are both statistically significant and material to the project.

INTRODUCTION

Gamma activation analysis (GAA) is a non-destructive method that uses X-rays for the rapid analysis of Au, Ag, and Cu in geological and metallurgical process samples (Tickner *et al,* 2017; Tremblay *et al,* 2019). Since its commercial introduction by Chrysos Corporation (Chrysos) in 2018, GAA has emerged as a fast and cost-effective alternative to conventional analytical techniques, such as lead fire assay (FA) or aqua regia digest followed by atomic absorption spectrometry (AAS) or inductively coupled plasma mass spectrometry (ICP-MS). Gamma activation analysis is now widely known under its Chrysos brand-name: PhotonAssay™.

Frequently cited benefits of Chrysos PhotonAssay™ (CPA) include: the ability to analyse larger samples (400–600 g) of coarse crushed (2–3 mm) material; a highly automated process increases sample throughput (~70 samples per hour, per unit) that minimises sample-handling mistakes; the lack of ecotoxic chemicals such as lead and cyanide; and an ~80 per cent lower carbon and energy footprint compared to a standard 30–50 g FA (Chrysos, 2022, 2025). CPA has wide applicability, spanning the entire mining value chain from exploration, resource development, grade control, and processing plant control (Tremblay *et al,* 2019; Dominy *et al,* 2024).

As with any emergent technology, there is a requirement to prove that data produced are of comparable or superior quality to that produced by conventional approaches. There are several published case studies that focus on the comparison of results from CPA and FA in different settings (Tickner, Lannan and Preston, 2021; Dominy *et al,* 2024; Hitchman *et al,* 2024). Comparative studies can provide justification for a switch to CPA and may be undertaken by companies independently or in collaboration with commercial laboratories or Chrysos. However, the authors have found that company-led comparative studies commonly lack a sound statistical foundation. Most studies do not address the important relationship between the size of the study and the confidence required for decision-making. As a result, many such studies do not succeed in reaching an objective conclusion; instead, subjective statements are made that do not aid decision-making.

Junior explorers and major mining companies are adopting CPA, and market announcements on major exchanges are starting to rely on results obtained through CPA. The authors used data from opaxe (www.opaxe.com) to analyse 130 stock market releases that announced results from CPA, including results from 12 companies that used CPA Au data to support a publicly reported MRE up until 2026. The review indicates that most companies relied on subjective conclusions to support the switch to CPA from FA.

This paper adds to the growing literature on best-practice for vetting CPA by providing a standard-practice, step-by-step guide to reliably test the performance of CPA compared to other analytical techniques for Au, and we present a simple flow chart to statistically test and determine whether CPA is fit for purpose.

CPA can be used to analyse a range of elements; however, it is mostly used for Au analysis. Therefore, this paper centres on the comparison between Au analyses obtained by CPA and FA. Our comparative workflow and statistical methods described are applicable to other elements such as Ag and Cu. The authors note that Chrysos is not the only provider of GAA analysis for geological samples (eg Baltic Scientific Instruments). However, the authors have only reviewed Chrysos' facilities, and therefore this paper is focused on PhotonAssay™. The principles underpinning the proposed test programme and described in this paper are transferrable.

HOW CPA WORKS

Gamma Activation Analysis relies on high-energy (>6 MeV) X-rays to excite transitions in nuclides of target elements, and detect the characteristic gamma-rays that are emitted from the nuclides when they decay from their excited state (Tickner *et al,* 2017). The gamma-ray count is proportional to the concentration of the element in the sample, and can be used inversely to determine the element concentration. Matrix or particle-size effects are largely insignificant, as the highly penetrative X-rays allow for the measurement of the entire sample volume contained within a 320 mL plastic jar, and the gamma-ray signal is emitted by the excited Au atoms even when Au is contained in solid particles.

Calibration steps are required to ensure that the CPA instrument delivers consistent results. System-inherent variations (eg changing activation energy or detector efficiency) are corrected by measurement of a reference disc with bromine-salt, placed underneath each jar. Bromine activates in a very similar way to Au, but produces a distinct lower-energy signal that can be used to correct for signal variance.

While the gamma-ray signal is directly proportional to Au concentration, the normalisation slope, referred to by Chrysos as the k-value, is determined by analysis of so-called *k-cal standards*. These standards are custom-made Au-bearing glass frits, with a superior chemical and physical stability to pulverised rock certified reference materials (CRM), and are analysed on a daily basis. Additionally, a monthly check of an instrument's performance is carried out by analysing a suite of high-quality CPA-specific CRM (*super k-cal*). After every major change of components, maintenance and/or calibration events, a change test suite is analysed for confirmation of the required operating performance that includes blanks, glass frit jars with various fill levels, and CRM of a wide range of grades, masses, densities, and matrices.

Elevated concentrations of U, Th, and Ba interfere with Au analysis, as they increase the background of the gamma-ray signal. For example, above combined U+Th concentrations of 15 ppm, or Ba concentrations of 3500 ppm, the lower detection limit of the CPA increases from 0.01 ppm to 0.03 ppm, and measurement precision decreases from 11.5 per cent to 18 per cent at 0.1 ppm but there is minimal impact on precision for Au concentrations above 0.3 ppm (Chrysos, 2023).

DATA QUALITY PRINCIPLES

Quality and materiality

A logical outcome of comparative CPA-FA studies is an objective statement on the *quality* of the CPA data, with quality conventionally expressed in terms of the accuracy and precision of the data. However, the scope of such studies is rarely specific about the data quality objective (DQO). In other words, it is not clear *how* accurate and precise the data need to be to support the transition to CPA. For instance, Hitchman *et al* (2024) in their study on the feasibility of transferring to CPA at the Fosterville deposit, Australia, state that 'success would be achieved if photon assaying [….] is analytically accurate and precise' (p 341), without stating what would trigger such a statement to be validated; ie when is something accurate enough and precise enough? In other words, when is something good *enough*?

Defining a DQO is critical in any assessment of data quality (Abzalov, 2008; Pitard, 2013; Dominy *et al,* 2024) and part of formulating a DQO is translating the quality requirements into a testable hypothesis. In most studies the authors reviewed, an explicitly stated DQO is absent, but it is implied that the CPA data need to be *at least as good as the quality of FA data*. However, such a statement also comes with caveats. There is an intuitive margin of error where the results would still satisfy the DQO. For example, if the precision of FA duplicate pairs is 12.1 per cent (as defined by the root-mean-square coefficient of variance (CV) approach (Stanley and Lawie, 2007; Abzalov, 2008)), then if the precision of CPA duplicate pairs is 12.5 per cent (ie CPA has slightly worse precision), this would not necessarily lead to a rejection of the hypothesis that the CPA data are at least as good as the FA data; it is within an intuitive (subjective) margin of error.

Well-known statistical tools can provide guidance; Student t-tests and equivalent tests remain the gold standard in many disciplines and industries to indisputably state differences between populations using well-known conventions. Key CPA papers such as Tickner, Lannan and Preston (2021), in their first comparison of results from standard FA versus the results of CPA, use such tests to conclude that CPA is accurate. Arguably, if no statistically significant difference is found, one can state that one method is as accurate as the other. However, a small, yet statistically significant difference between the data for two different techniques does not always mean that the data are not suited to the purpose or objective; such an outcome may occur, especially with a large data set. This concept is not just true in the mining industry, but also in other industries, and the concept of statistical significance versus practical significance is a well-studied subject area (Gelman and Stern, 2006; Peeters, 2016; Miller, 2023).

In the mining industry, there is arguably a disconnect between statistical test results and the practical ramifications of these results from a materiality perspective. An example of this is the outcome of a classic 'laboratory umpire' study, where a company submits a few hundred samples to another laboratory each quarter to compare the assay results. In the authors' experience many of these studies indeed find a small, yet statistically significant difference using conventional statistical testing methods. However, rarely has a 'small' difference between two laboratories stopped a mine in its tracks or led to re-assaying of all samples. This leads to a dead end and is arguably one of the reasons why many practitioners skip any statistical testing and go straight to subjective 'expert interpretation'.

The issue of materiality is key to every mining and exploration project, in any commodity, and at any part of the grade curve; yet, linking quantitative accuracy and precision thresholds to associated project risk, or even worse, project failure, is difficult. Full model-to-mill reconciliation is the only quantitative comparison available to provide a feedback loop to any decision made in the complex process of data collection, estimation, mining, and processing. It is clear that a small change to a part of a sampling or analytical process will have an almost undetectable impact on the monthly mine reconciliation data. This is not helped by the poor state of most mine reconciliation systems, which are often based on Excel sheets, and which, apart from noise in analytical data, also embed noise from the modelling, mining, handling, and processing steps, and do not have any defined performance thresholds themselves. The consequence is that strictly following the scientific method, ie quantitative thresholding of accuracy and precision using common statistical tests, may not lead to practically relevant conclusions about data quality. Qualitative comments and interpretations that focus on materiality may need be added to the conclusions.

Accuracy

Testing the DQO that the CPA data quality is *at least as good as the quality of FA data* would be possible if the true grade of the material being assayed was known as a basis of comparison. For example, if the true grade of a sample is 5.00 g/t Au, and if the FA result is 4.69 g/t Au, and the CPA result is 4.81 g/t Au, then, objectively, the CPA result is at least as good (ie as accurate) as the FA result. In this example, if the FA result has been set as an acceptable benchmark to allow resource classification to meet production risk objectives, then by inference, the CPA results should be acceptable also. Of course, the true grade of a geological material is never known, and hence making any claim on the accuracy of the CPA data requires asking: Compared to *what*?

Since the true grade of a geological sample is unknown, it is impossible to determine an exact bias of the CPA method, and CPA performance must be compared to the best-known alternative (ie FA to extinction). However, the destructive nature of FA methods complicates this, as the sample can only be analysed once, and comparing CPA results against a benchmark FA analysis from a single laboratory would, in the best scenario, only provide a relative bias, with no guarantee that the benchmark is not biased itself.

In all comparative studies reviewed by the authors, samples analysed by CPA were sent to only one laboratory for routine FA analysis, and these studies can therefore only assess the relative bias (ie the bias between two individual laboratories). This is not necessarily a fatal flaw, as the routine FA laboratories typically have a track record of supporting robust resource estimates, and results from that laboratory have been accepted across a significant period of time as providing fit-for-purpose results. However, this experimental design does put constraints on the conclusions that can be drawn from such a testing programme. As discussed in Tickner, Lannan and Preston (2021), 'if […] the systematic or random errors [of the benchmark FA method] are not negligible compared to the errors of the new assay technique being evaluated [CPA], then any comparison of grade results will include the errors of both methods' (p 2). To illustrate this, if there was a consistent and statistically significant bias introduced by scooping the 50 g pulp from the 300 g lot for FA, this would impact the comparison between FA and CPA. This problem could be addressed by using FA to extinction in the comparison study. However, if the bias was consistent, and these FA values are from the mine's routine laboratory, then this bias would also be present in the data on which the MRE is based, and which were accepted as fit for purpose. Therefore, this bias would have no bearing on the evaluation of quality of the CPA data.

For most comparative studies, there are several additional sources of information available that provide important context to appraise materiality in the comparison of the accuracy of CPA against FA.

- First, if there are frequent umpire tests conducted on the routine FA laboratory itself, then these data are important to set the framework of reference for accuracy. If the tenor of bias observed between the routine FA laboratory and the umpire FA laboratory is similar to the bias observed between the routine FA laboratory and the CPA laboratory, then this provides an argument for accepting the CPA results. In other words, if the routine FA umpire results demonstrate a small and statistically significant bias, then a bias of a similar nature in the CPA versus FA comparison would validate the CPA results by proxy.
- Second, FA laboratories routinely test a large number of FA-certified CRMs. The bias of these CRMs sets a semi-quantitative framework of reference for acceptance of any assay bias, in which the results of a comparative study should be viewed. For instance, if 12 different CRMs routinely demonstrate biases between -3 per cent and +3 per cent, and the assays were accepted in the database, and were used to inform and classify a high-confidence MRE, then, arguably, a bias within that range is also acceptable for the CPA data.

Precision

Precision is a concept, rather than a well-defined quantity (Hyslop and White, 2009). There is no universally accepted metric, and this makes the communication of acceptance levels for precision equally challenging. The reader is referred to the work of Stanley and Lawie (2007), who demonstrate that there are five different variations of precision definitions that are used in geological applications to represent precision: Coefficient of Variation (CV), Relative Precision (RP), Relative Variance (RV), Absolute Relative Difference (ARD), and Half Absolute Relative Difference (HARD).

Ultimately, which one is 'correct' comes down to convention, or one's interpretation of which is the correct correction factor to use. The authors favour expression of the precision of paired grade data by the CV, obtained using the 'root-mean-squared' formula for duplicate data, as presented by Stanley and Lawie (2007), and shown below in Equation 1.

$$RMS\ CV\ (\%) = 100 \times \sqrt{\frac{2}{N}\sum_{i=1}^{N}\left(\frac{(a_i-b_i)^2}{(a_i+b_i)^2}\right)} \qquad (1)$$

This aligns with reviews of precision also presented in Abzalov (2008) and Smee *et al* (2024), and should, in the authors' opinion, be the one and only universally applied measure of average precision of paired grade data.

The authors note that, until 2025, Chrysos itself preferred to present precision values calculated across different parts of the grade range. This is because the data are inherently heteroskedastic; precision decreases abruptly with increasing grade close to the detection limits (Smee *et al,* 2024), which makes the calculation of a single average CV value challenging. In the authors' experience, clipping data below ~3× the level of quantification (LOQ) generates a data set for which the CV is a robust and repeatable precision metric that can be confidently used in comparative studies. The LOQ is the lowest analyte concentration that can be quantitatively detected with a stated accuracy and precision; LOQ = limit of detection (LOD) + 10 x σ_{blank}.

Like the framework of accuracy introduced before, it is again important to establish what *exactly* is compared when precision is calculated in this study. For standard FA, precision is calculated from paired analysis on pulverised samples. These are often referred to as 'pulp duplicates'. For CPA, the final aliquot usually contains material at a larger grain size and of a larger sample mass compared to FA. It is acceptable to compare duplicates created at this final CPA aliquot stage with those created at the final aliquot stage of FA, as both represent the outcome of 'routine' processes.

A DQO for the precision of the CPA data needs to be specified by the geologists managing the technical aspects of the programme. What is the maximum tolerance value of precision, as measured by the CV of duplicate pairs, beyond which the data are not considered fit for purpose? As with accuracy, it can be safely stated that the precision of the CPA data must be at least as good as that of the FA data. Additional information may be contributed by looking at historical duplicate data, particularly when these data can be linked to resource reconciliation.

Another powerful precision DQO can be set using the variogram, which can be generated without the need for extensive duplication of both FA and CPA samples, as long as a sufficient amount of samples have been analysed both by FA and CPA (ie several thousand). This method was applied by Hitchman *et al* (2024) in their review of the performance of FA versus CPA at the very nuggety Fosterville deposit; they demonstrated a lower nugget on normal-score variograms for CPA (~0.4) compared to FA (~0.5), which is objectively better and confirmed CPA's superior performance, in terms of precision, to FA.

Last, precision is dependent on the natural inherent variability of the sample and sampling/analytical errors. Specific to CPA, elevated concentrations of U, Th, and Ba, as well as the 'positional heterogeneity' of Au particles in jars lead to a decrease in precision in CPA analysis (Dominy *et al,* 2024). Therefore, it is important to benchmark the precision of CPA analysis in the full context of all these variables.

COMPARATIVE CPA-FA TEST PROGRAMME

Sample selection

The success of a comparative study between CPA and FA is contingent on the correctness of the sample selection process. For instance, in many of the comparison studies reviewed by the authors, minimum cut-off grades were applied to samples drawn from the database in order to select samples ('original') for testing, which introduces an artificial bias close to the selected cut-off grade. This happens, because the 'original' sample is *always* above that cut-off grade (as that is what it was selected on), but the 'duplicate' *may* return a grade that is lower (Long, 2015). On the contrary, randomised sample selections commonly lead to a very large amount of unmineralised samples, due to the heavily skewed grade distribution in almost all Au data sets, often leading to extreme waste of funds and time working on samples that are not mineralised. The correct approach is to randomly select samples from intervals that were logged as mineralised (Long, 2015) or are contained in modelled ore zones.

Any statistical analysis should only be undertaken on data that are from the same population, as the analysis of precision or accuracy will be flawed if distributions demonstrate any multi-modality. In this context, separate sample populations or domains could be represented by samples from different mineralisation stages, or different parts of a deposit. Even though CPA is relatively agnostic with respect to host rock type, FA is not, and any differences need to therefore be calibrated to relevant domains. For this reason, domains of a project/mine site need to be considered during sample selection, and results must be reviewed separately for each domain (see also approach, comments, and context in Dominy *et al* (2024)). An equal number of samples should be selected from each domain. The authors recommend 200 samples per domain as a reasonable starting point for a robust statistical orientation study.

Sample preparation and analysis workflow

To correctly review any differences between FA and CPA results, sample preparation and analytical processes should mirror those of standard and routine processes. This includes the equipment used, comminution sizes, sample weights, and protocols for splitting, digestion, and analysis of samples. The authors have reviewed CPA-FA studies where the laboratory equipment was optimised to achieve the 'best possible results'. Though tempting, it is important to note that the comparison is not that of best-behaviour CPA against best-behaviour FA. The point is to establish that routine CPA quality matches that of routine FA.

The authors propose a sample preparation workflow that provides the data and checks and balances needed to confidently compare results from FA and CPA (Figure 1). The following considerations require particular attention.

1. Crush, pulverise, and split samples as per usual sample preparation practices for FA samples, with the aim of comparing any accuracy or precision with routine FA processes (steps 2, 3, 7, 8, 13, and 15 in Figure 1).
2. Carry out two routine CPA analyses per crushed sample (step 5 in Figure 1) and three routine CPA analyses per pulverised sample (step 10 in Figure 1). The duplicate pairs will provide information on the routine CPA precision by comparing the two (note that the calculation of variance needs to be corrected for the use of three pulp pairs versus two crushed pairs). Prevent overfilling the CPA jars or scraping the excess material (coned) off the top (step 4 in Figure 1).
3. The recombined CPA pulp sample (step 11 in Figure 1) was previously split using an LSD splitter from the bowl, which does not represent FA routine, therefore the pulp must be milled again for ~5–10 seconds to 'get back on the flow chart' of the routine original FA process (step 12 in Figure 1). Otherwise, the calculated variance of the FA pulps will not be comparable with the routine process. This is important as the study aim is benchmarking and acceptance-testing based on routine accuracy and precision, hence, the routine process must be honoured.

4. If samples travel between CPA and FA laboratories and potentially segregate, insert another step of 5–10 seconds milling to re-homogenise the sample (between steps 14 and 15 in Figure 1).
5. Carry out two routine FA analyses. The duplicate pair will provide information on the routine FA precision by comparing the two.
6. At the umpire laboratory, ensure that samples are first re-homogenised and subsequently split as per routine umpire laboratory processes. The FA charge weight should be the same as for the primary FA (steps 19–21 in Figure 1).

Quality control

Quality control (QC) protocols are essential for determining that the sample preparation and analytical processes deliver consistent results that are fit for the purpose of informing the comparison between the CPA and FA. Therefore, QC steps should include screen size tests, analysis of certified reference materials (CRMs), and analysis of coarse blank samples, for both CPA and FA. If the geologists managing the technical aspects of the testing programme insist that only client-inserted CRMs can be trusted (instead of relying on laboratory-inserted CRMs, which are arguably fit for purpose; this is a discussion that goes beyond the scope of this paper), then typical client CRM insertion rates of 5 per cent are insufficient, and should be increased to one in five for a comparative study, to generate enough control points for robust statistical analysis (5 per cent of 200 study samples would yield only 10 CRM assays). For smaller sample sets with <400 samples, CRM insertion should be adjusted to obtain at least 25 CRM results per individual CRM for CPA and FA. CRMs should be the same for both CPA and FA. Such insertion rates require a considerable number of CRM jars to not halt the continuous operation of the CPA instrument. The authors recommend inserting three unique medium- to high-grade CRMs. Several CRM manufacturers offer CRMs certified for Au by both CPA and FA, which should be preferred over CRMs certified for just FA.

Heterogeneity – an added bonus

In addition to conclusions on accuracy and precision, the testing programme in Figure 1 allows important information on the heterogeneity and Au deportment to be collected. Having precision averages for two groups of samples of similar size, but of different grain size, allows determination of the sampling constants K and alpha, which in turn can then be used to determine the optimum size of the split and the grain size, while remaining below a defined target threshold of precision. For a worked-out example of how to calculate these factors and use them to make decisions, see Minnitt (2016) and Dominy and Minnitt (2012).

EVALUATION OF RESULTS

This section outlines the series of recommended steps for the review and analysis of paired CPA-FA results, and other routine QC data, to confidently determine the performance of CPA compared to FA (schematised in Figure 2). The authors used assay data from several real-world examples to illustrate different evaluation steps. It should be noted that none of the real-world data sets used here to present workflows are indicative of expected accuracy or precision outcomes; they are used to demonstrate workflows only, and they are not ‘case studies’.

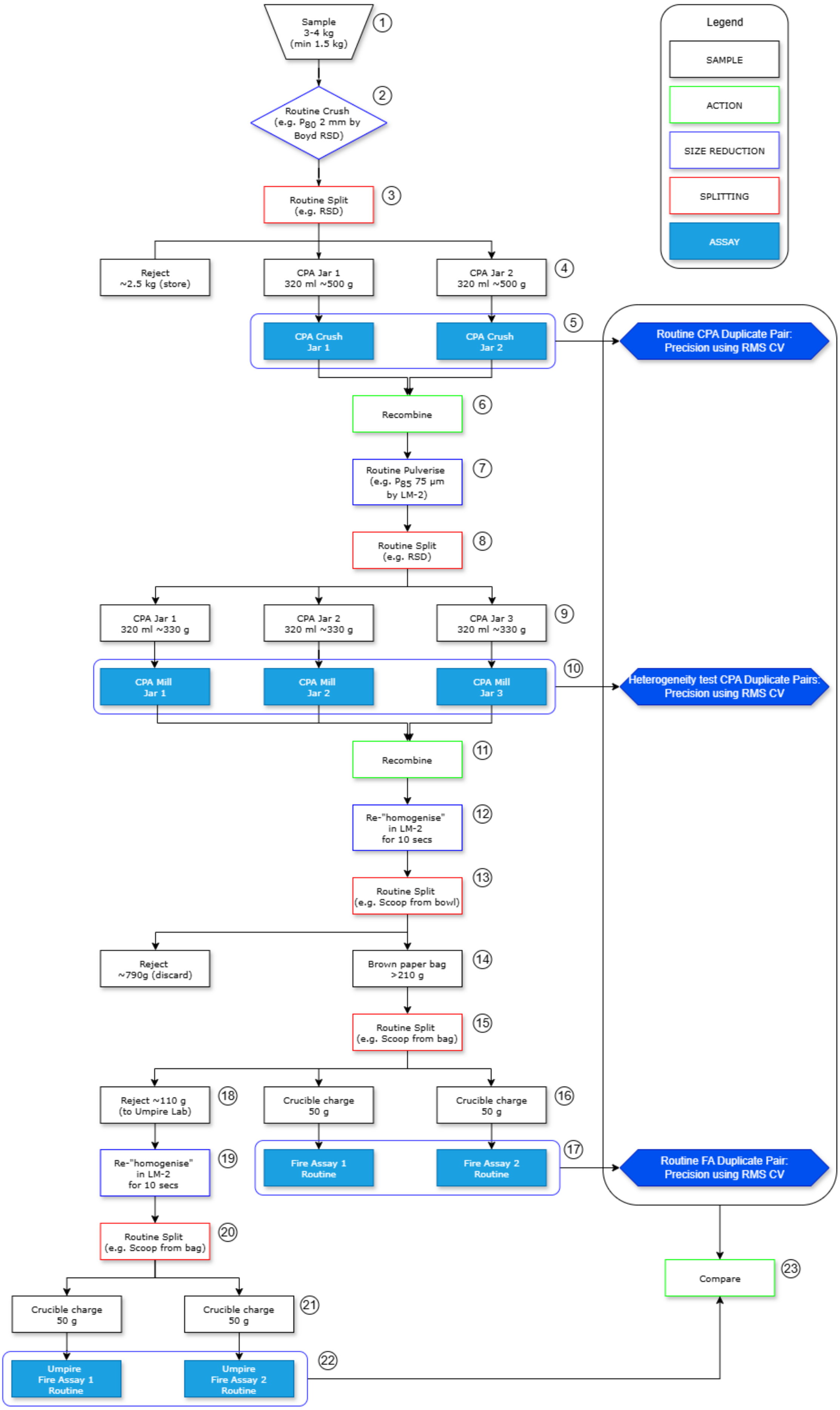

FIG 1 – Sample preparation and analysis flow sheet for comparative CPA versus FA analysis.

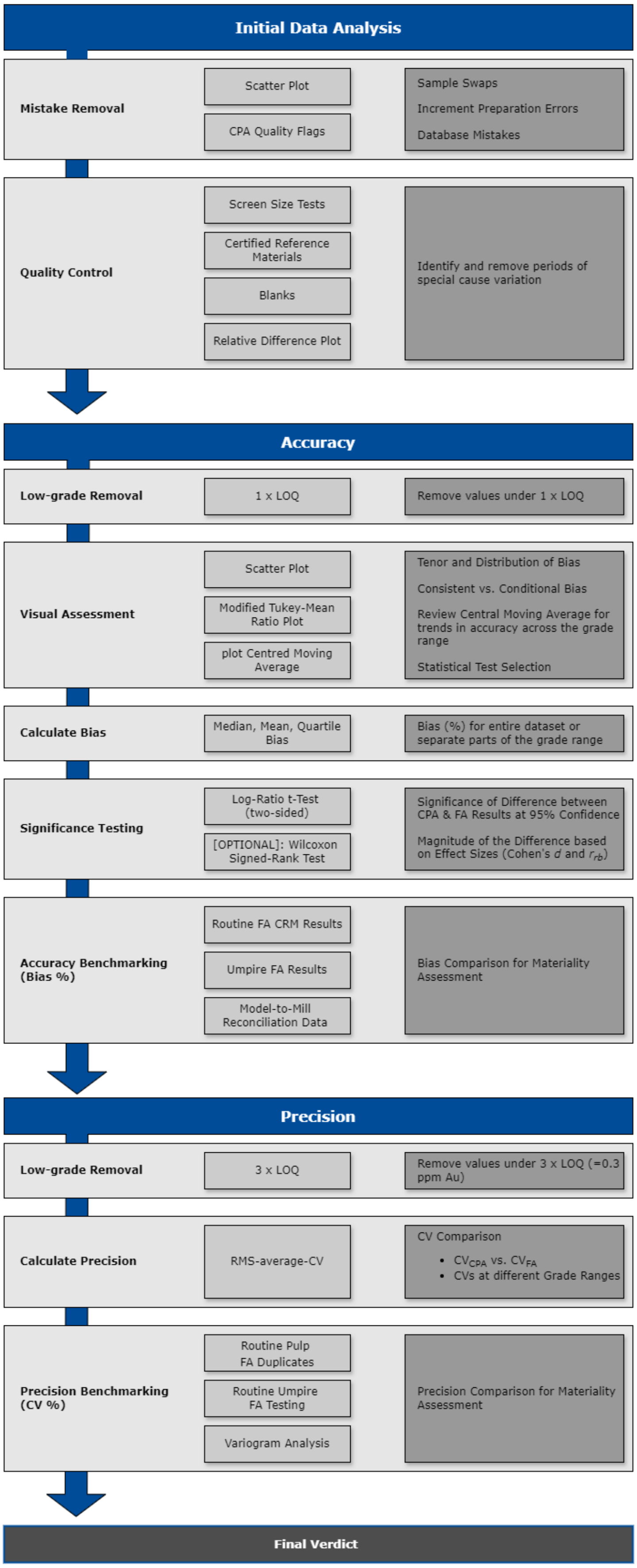

FIG 2 – Schematic data review and analytical process for comparative CPA-FA studies. Each light grey box represents a process step (left) that contains items to use, plot, calculate, or analyse (centre) for the corresponding data assessment(s) (right).

Initial data analysis

Basic validation

As with any statistical data analysis, any true mistakes need to be removed from the data first (Figure 3). Sample swaps, increment preparation errors, and results at the upper detection limit must be identified and eliminated from the data, as these would invalidate the analysis.

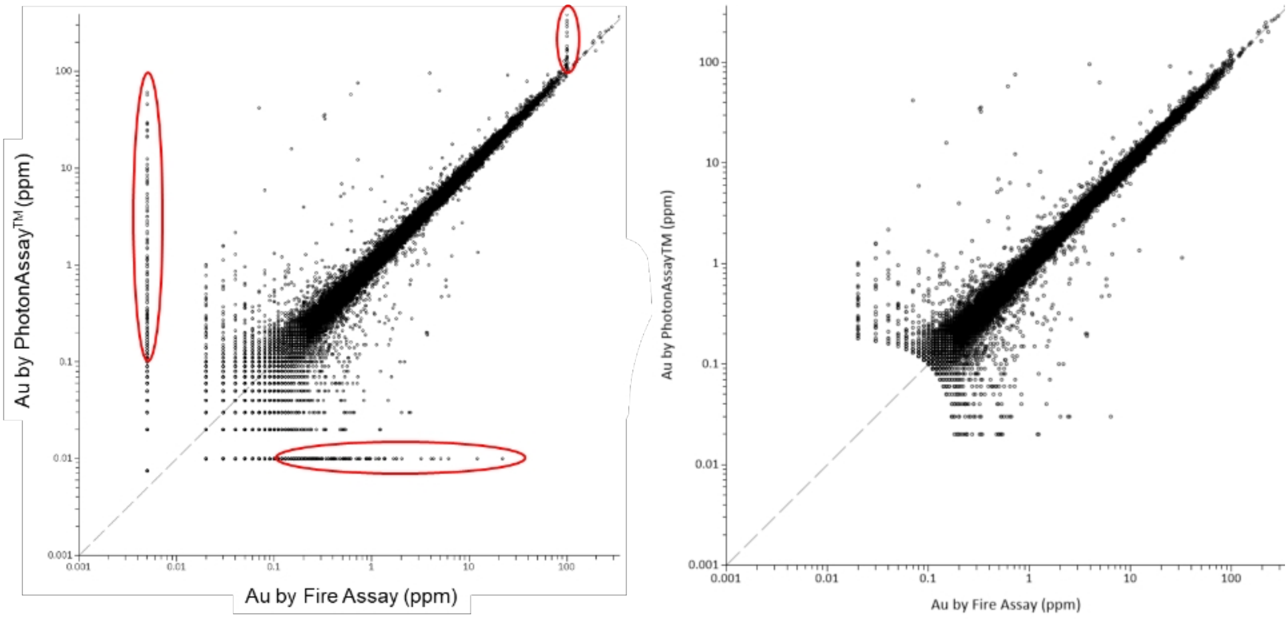

FIG 3 – Scatterplot of raw, unfiltered data, illustrating several data errors (left); and data after cleaning and clipping at geometric pair mean of 0.1 (right).

Following error removal, data dominated by analytical error or limited by low resolution should be excluded from further analyses (Tickner, Lannan and Preston, 2021; Dominy *et al,* 2024). For accuracy assessments, the removal threshold is set at the limit of quantification (LOQ), approximated here as three times the limit of detection (LOD) (3 × LOD = ~0.1 ppm Au). For precision assessments, a higher threshold of three times the LOQ (3 × LOQ = ~0.3 ppm Au) is applied because the analytical system reports values with a resolution of only two significant digits, artificially increasing variance near the LOQ. It may be tempting to simplify the lower-threshold approach and set both at the same level of 3 × LOQ (~0.3 ppm Au); however, it is important to get an indication of bias at low-grades, since resource estimation domaining is often performed at this level, and plant tailing samples may be analysed using CPA. Lowering the accuracy threshold to 3 × LOD (~0.1 ppm Au) allows evaluation of bias at low-grades but comes with the risk of Type-II errors ('false negatives'), given the high artificial variance due to the lack of resolution at these levels.

Quality control

Following data cleaning, it must be demonstrated that the analytical processes (both CPA and FA) were in control and delivered consistent data. Trends or other special-cause variation (as evidenced by results from CRMs, blanks, or screen size tests) would compromise, or even invalidate, the comparison. For CRMs, control plots (Shewhart charts) are used to illustrate the results per CRM, relative to their certified mean and three standard deviation brackets (Figure 4). Additional insights can be gained from applying Westgard rules (Westgard *et al,* 1981; Sterk, 2015) to identify periods of special-cause variation (see 14a, 15u1s, etc, in Figure 4). Any periods of special-cause variation must be identified, and samples from these periods must be removed from the data set, especially if special-cause variation is evident across different CRMs.

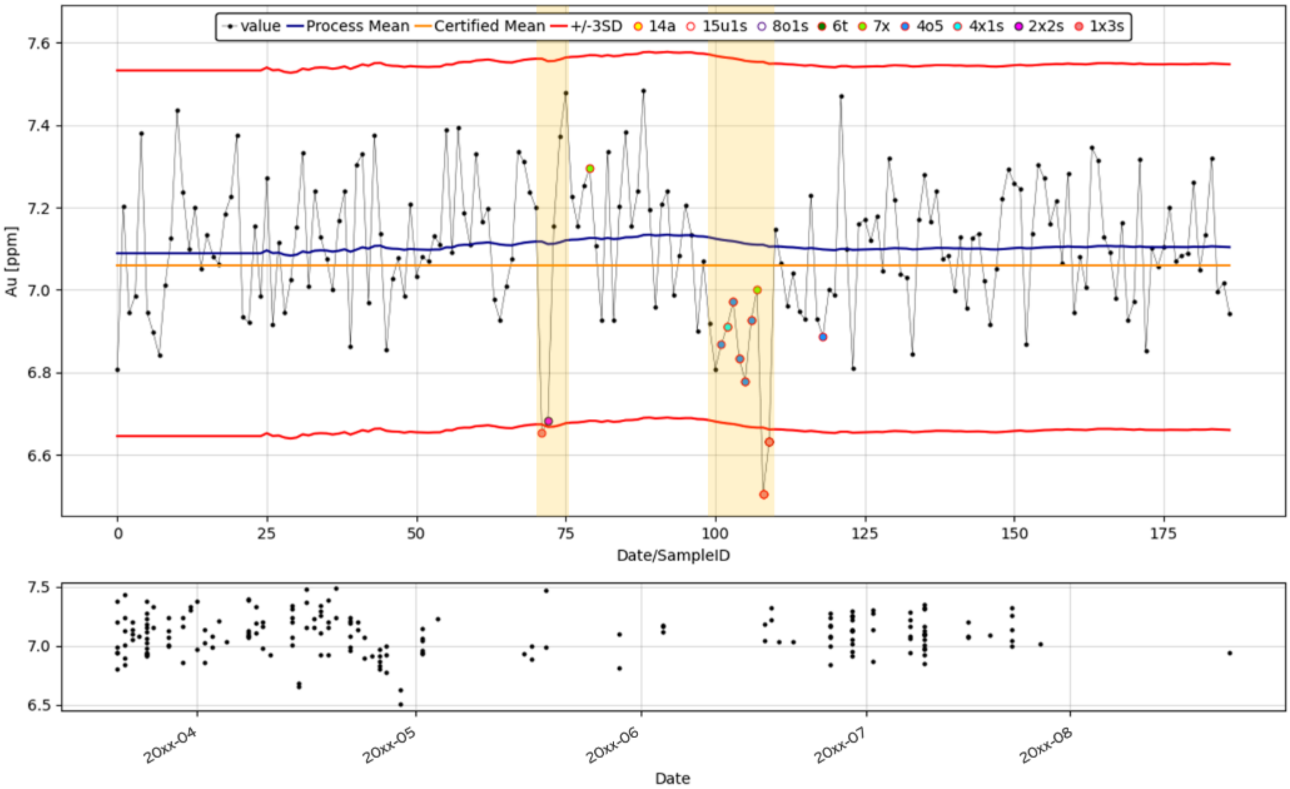

FIG 4 – Example of Shewhart control plot for one Au CRM, with periods of special-cause variation highlighted in yellow.

Accuracy

Basic visual assessment

Before accuracy can be calculated and assessed for significance, it is important to assess whether any bias depends (ie is conditional) on the mean concentration. If conditional, an average bias result may not hold across the full range of paired means.

With the cleaned data from periods where processes were in control, a good place to start this assessment is a visual evaluation of basic scatter of CPA and FA measurements (Figure 5, left). It is important to use the log scale for both measurements, allowing relative differences to be interpreted consistently from low to high grades. The authors have found that QQ plots are widely used to describe the accuracy of paired data across the grade range by comparing the quantile distribution of the two measurements. As simple and practical as this approach is, it treats the two measurements as independent, which obscures important information about the distribution of each pair, and therefore its utility is limited.

Due to scale dependence, the absolute numerical difference between CPA and FA is larger when the assay values are higher, and the ratio of the results is a more consistent way to quantify any bias (it is scale-invariant). However, analysing ratios is problematic because ratios are intrinsically asymmetric; all negative differences between original and duplicate are expressed as ratios between zero and one, while positive differences between original and duplicate are expressed as ratios higher than 1.0. A logarithmic transformation solves this issue (eg the logarithm of 2.0 is 0.301; the logarithm of 0.5 is -0.301). A Tukey mean-ratio plot (Figure 5, right) as used in Tickner, Lannan and Preston (2021) plots the geometric mean grade of the pair versus the ratio between CPA and FA (log-transformed to correct for asymmetry), and helps identify systematic deviations. A simple centred moving average function can be plotted to highlight any trends.

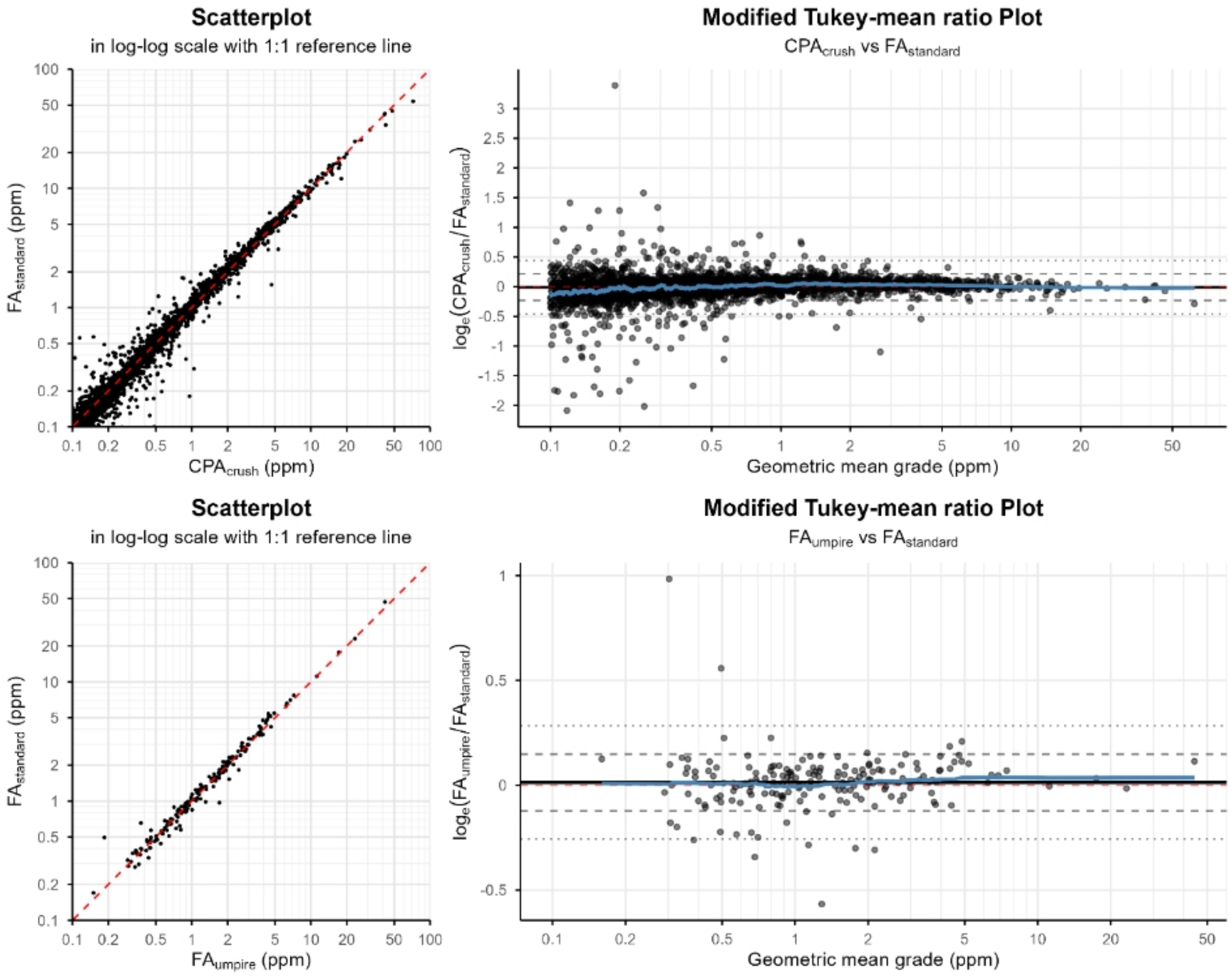

FIG 5 – Scatterplot (left) and modified Tukey plot (right) for inspecting differences between FA and CPA_{crush} versus $FA_{standard}$ (top) and FA_{umpire} versus $FA_{standard}$ (bottom). The black solid horizontal line represents the mean value of the ratio (ie the average bias across the entire grade range, see Bias Calculation section); the horizontal black dashed and dotted lines represent one and two standard deviations from the mean respectively. The blue line is the centred moving average of log ratios by grade level with a window of 101 pairs (edge-truncated). The plots present results from one study and are not indicative of typical CPA outcomes.

When the log-ratio is not conditional to the grade level, the moving average trendline follows a straight horizontal line. If the moving average deviates from a straight horizontal line in the Tukey plot, the bias is conditional and any bias calculation or significance testing should be evaluated separately for different grade ranges. In this case, the calculation of a *single* summary statistic of the average bias across the entire grade range could be misleading, especially if the bias is negative at one end and positive at the other end (in which case the *average* bias could be close to zero, suggesting a high degree of accuracy that is not representative of the entire data set).

Bias calculation

The mean or median bias can be calculated, either across the whole grade range (if the bias is not conditional), or for each part of the grade range where bias is consistent. The bias can be expressed as a percentage using the transformation:

Equation 2 turns the average of the log-ratios back into 'normal space', and by subtracting one leaves the residual, which is the bias, expressed as a percentage.

$$Bias\ (\%) = (e^{\,log(ratio)} - 1)\times 100 \qquad (2)$$

Significance testing

Statistical tests should then be used to determine whether any calculated biases between CPA and FA are statistically significant (Table 1). The log-ratio t-test (Tickner, Lannan and Preston*,* 2021) and Wilcoxon Signed-Rank test (Wilcoxon, 1945; Napier-Munn, 2014) can both be used for the accuracy-significance testing of CPA and FA data, and are both simple to implement using Excel. The log-ratio t-test can be used if the bias is multiplicative in nature and provides more interpretable effect sizes (see below), whereas the Wilcoxon Signed-Rank test is useful when the data have extreme outliers and where the bias is additive in nature, but ignores the magnitude of the differences. The authors recommend using the log-ratio t-test in the first instance and use the Wilcoxon Signed-Rank test to double check results.

TABLE 1

Overview of statistical tests for accuracy suitable for comparison of CPA versus FA measurements.

Test name	Null hypothesis	Assumed data distribution	Unit effect type	Test formula	Effect size
log-ratio t-test (two-sided)	Mean of logarithm of CPA/FA ratios is zero	ratio: lognormal (asymmetrical) log(ratio): normal (symmetrical)	(Log-) ratio*multiplicative*	$Y_{LR} = \log_{\mathrm{e}}\left(\frac{Y_{CPA,i}}{Y_{FA,i}}\right)$ $t = \frac{\bar{Y}_{\mathbf{LR}}}{\hat{\sigma}_{Y_{LR}}\ /\sqrt{n}}$	$Cohen's\ d = \frac{\overline{Y}_{LR}}{\hat{\sigma}_{Y_{LR}}}$
Wilcoxon Signed-Rank test	Sum of ranks of positive differences equals sum of ranks of negative differences	Non-Normal (asymmetrical)	Rank *additive*	$R_i = rank(\lvert D_i \rvert),$ $W^+ = \sum_{i:D_i>0} R_i,$ $W^- = \sum_{i:D_i<0} R_i$ $W = \sum_{i=1}^{n} sign(D_i) \times R_i$	$r_{rb} = \frac{W^+ - W^-}{W^+ + W^-}$

Notes. D = difference; $\hat{\sigma}$ = sample standard deviation (as estimation of the population standard deviation σ); R_i = rank score of the absolute paired difference; W^+ = sum of the positive ranks; W^- = sum of the negative ranks; $sign(D_i)$ = indicator if the difference is positive (= 1), zero (= 0) or negative (= -1); r_{rb} = rank-biserial correlation (effect size of Wilcoxon); log$_e$ = natural logarithm; Y_{LR} = pair log-ratio; n = number of paired observations. Overline above value ($\bar{Y}_{LR}$) indicates the sample mean.

The log-ratio t-test evaluates whether the CPA/FA ratio is equivalent to 1, since $\log_e(1) = 0$. The Wilcoxon Signed-Rank test evaluates whether the distribution of the difference between paired CPA/FA data is symmetric around a central value, and tests whether this centre value differs significantly from zero. The results of the selected statistical tests are assessed by evaluating the resulting *p*-value against the selected threshold ($\alpha = 0.05$).

In addition to determining whether a bias is statistically significant, the effect size can be used as a quantitative reflection of the magnitude of a phenomenon (Kelley and Preacher, 2012). There are many different effect sizes, the most well-known one being the correlation coefficient (Pearson's r), and the coefficient of determination (R^2). For the log-ratio t-test, the effect size is commonly reported as Cohen's *d* (Cohen, 1988) and represents a measure of the standardised difference between two means, calculated as the logarithm of the mean ratio (here between CPA and FA measurements) divided by the standard deviation of those differences (eg Lakens, 2013; Hassan *et al,* 2020). Cohen's *d* measures how many standard deviations the mean of the log ratios deviates from a hypothesised difference. Under the null hypothesis, the mean of the log ratios is assumed to be zero, ie no difference between CPA and FA ($\mu_D = 0$).

The effect size metric is important as it can further help with the discussion around the materiality of any differences observed. It can be appraised by comparison against common qualifications used in the literature (Table 2); however, these conventions should be considered as general guidance and do not dictate precise or required interpretations (Thompson, 2007; Lakens, 2013).

TABLE 2

Thresholds for interpreting (absolute) Cohen's *d* effect size (Cohen, 1992).

Effect size	| Cohen's *d* |
small	>0.20
medium	>0.50
large	>0.80

Beyond an average bias and effect size calculation, a graphical examination of effect sizes across grade ranges provides valuable insight into the relevance of the conditional nature of the differences between analytical methods. The authors recommend applying one-sample t-tests on the centred moving averages of the log-ratios using a moving window of 101 paired observations. For each window, Cohen's d can be calculated based on the log-ratio test. The number 101 is a reasonable choice, because for 101 observations per window (100 degrees of freedom), a two-sided t-test becomes significant at α = 0.05 when effect size is small (Cohen's d > 0.20 or < -0.20, see Equation 3):

$$\left|t_{d=.20;n=101}\right| = |d| \times \sqrt{n} = .20 \times \sqrt{101} = 2.01 > t_{crit,.975,df=100} = 1.985 \quad (3)$$

The question then is: how large does the total data set of duplicates (N) above 3 × LOQ need to be to get a meaningful window length of n pairs to result in a statistically reliable assessment of bias across the total grade range? To determine this, simulations were performed following the approach of Tickner, Lannan and Preston (2021). Paired assay data ($N_1 = N_2 = 50$) were generated for a log-normal grade distribution with σ = 2, yielding 14.1 per cent of geometric pair means exceeding 10 units after filtering cases under 0.01 ppm. The authors recommend that more than half the window width (n = 101) lies above this threshold, meaning at least 51 observations. With higher grades being less frequent, this requires a minimum data set of 362 duplicate pairs (≥LOQ) to adequately evaluate grade-dependent bias. For initial scoping studies where lower confidence is accepted, we recommend N ≥ 121 and window length n = 33.

The combination of the centred moving average trend line (Figure 6, left), the plotting of the quartiles of the bias calculations (Figure 6, centre) and Cohen's d across the grade range (Figure 6, right) allows a more-informed interpretation of the bias across the grade range. It allows users to quickly determine in which range the bias is statistically significant and relevant, as per their effect sizes (ie small, medium, large).

It should be noted that due to the symmetrical nature of the centred moving average, information is lost at the minimum and maximum grades. Therefore, edge truncation is applied, and the test is performed on the available observations within each window (keeping at least half of the window's observations).

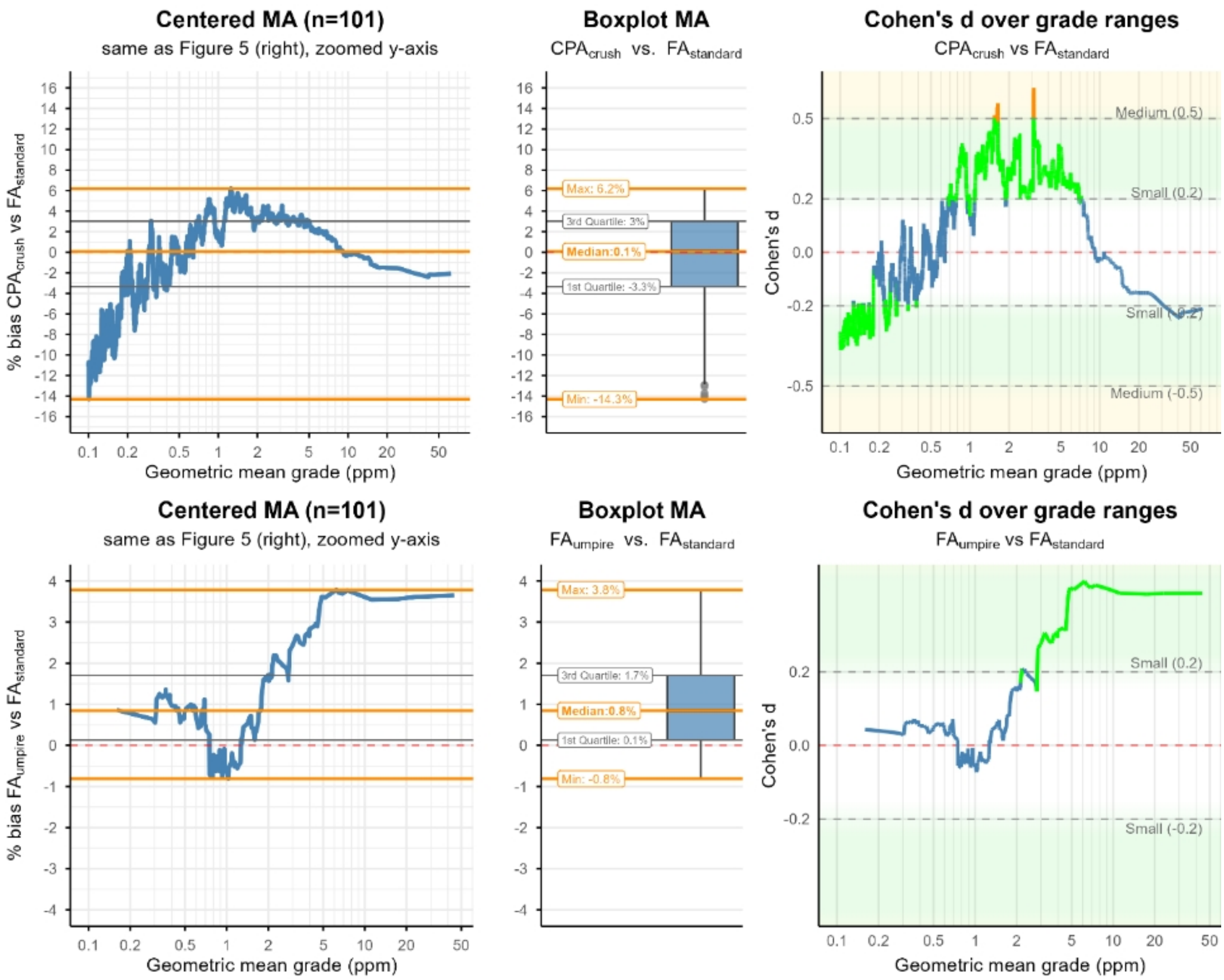

FIG 6 – Visual inspection of mean bias and effect size (Cohen's d) across grade levels for CPA$_{crush}$ versus FA$_{standard}$ (top) and FA$_{umpire}$ versus FA$_{standard}$ (bottom). Left: centred moving averages (MA) with a moving window of 101 observations (edge-truncated). Centre: range and quartile distribution of values by mean grade. Right: effect size and significance across mean grade levels (green (small effect) and orange (medium effect) segments of the line indicate $p < 0.05$, two-sided log-ratio test).

Precision

The precision, expressed as the CV, is calculated from the results of pairwise analysis of CPA duplicate pairs and FA duplicate pairs, using the RMS-CV approach (equation 1), and for pairs with an average grade above 3 × LOQ (Table 3). Precision should be calculated separately for each geological domain. A decision on whether the CPA precision is fit for purpose can then be made based on a simple table-based comparison of precision between CPA crushed, pulverised and FA analyses (Table 3). If the CV for the CPA is better (lower) than that of FA, then it can be stated that the precision of CPA is fit for purpose.

TABLE 3

Precision for three duplicate data groups. In this case, the CPA precision is fit for purpose. NB: These results are from one study and not indicative of broader findings.

Duplicate group	Precision
$\text{CPA}_{\text{coarse}}$	12.1%
CPA_{fine}	8.7%
FA	12.5%

If the CV for CPA is similar but slightly higher than that of FA (Table 4), then a Feltz-Miller test (Feltz and Miller, 1996) can be used to test whether that difference is statistically significant at $\alpha = 0.05$.

TABLE 4

Precision for three duplicate data groups. In this case, whether the CPA precision is fit for purpose depends on the outcome of statistical testing. NB: These results are from one study and not indicative of broader findings.

Duplicate group	Precision
$\text{CPA}_{\text{coarse}}$	11.8%
CPA_{fine}	9.7%
FA	11.2%

MATERIALITY

If biases and precisions calculated between CPA and FA are not statistically significant using the tests and methods described above, then it can be concluded that the method is fit for purpose. However, if statistically significant biases *are* found (which happens often in large data sets and is the rule rather than the exception), or when CPA precision values are statistically significantly higher than FA, this does not necessarily mean that CPA is not fit for purpose. This presents a practical problem that requires further assessment (see earlier comments and literature references in the section on 'Quality and Materiality').

An appraisal of materiality of any observed differences can be made by comparison of the results against a specific data quality objective/threshold (DQO) rather than testing for parity using the methods described above.

One of the key challenges with studies of this kind is that often no objective is set before the testing is completed (ie a DQO/tolerance threshold does not exist). Industry benchmarks are equally difficult to find, because reporting codes such as the JORC Code (2012) leave such judgments for the Competent Persons to decide, and papers, such as the one by Abzalov (2008) on benchmarks for precision, are rare. If it is not clear what the thresholds are before the test is started, then it is difficult to expect practical conclusions from any significance test.

However, for both accuracy and precision, there are some options that practitioners can draw from to define a DQO, and the authors recommend the following options:

- Accuracy and precision thresholds/tolerances, as set by the geologists (Qualified/Competent Persons) managing the technical aspects of the programme.
- Industry benchmarks (for instance Table 4 in Abzalov (2011) for precision).
- Inferences of DQOs from other routine FA tests (CRM results and routine umpire tests).

Accuracy

In the absence of a DQO set by the geologist, or DQOs from industry benchmarks for accuracy (the authors are not aware of any), an 'accuracy framework' can be established on the basis of the overall performance of FA CRMs, tested over a long period, as part of the routine FA quality monitoring process. For example, if a bias ranging from -1.0 per cent to +3.0 per cent was determined for eight different CRMs that cover the grade range, and the project or mine has been operating successfully and consistently with these long-accepted analytical biases, this range of bias provides a solid framework of reference for the assessment of the materiality of other analytical biases. In simple words: if one would accept a bias of +3.0 per cent for routine CRMs, it would be reasonable to also accept a bias of 3.0 per cent for CPA versus FA. The performance of CRMs can be included as single points in the Tukey plot (Figure 7, left), or presented by a box and whisker plot, showing the median, quartiles and min-max values of all available CRMs over a long period of time (Figure 7, right).

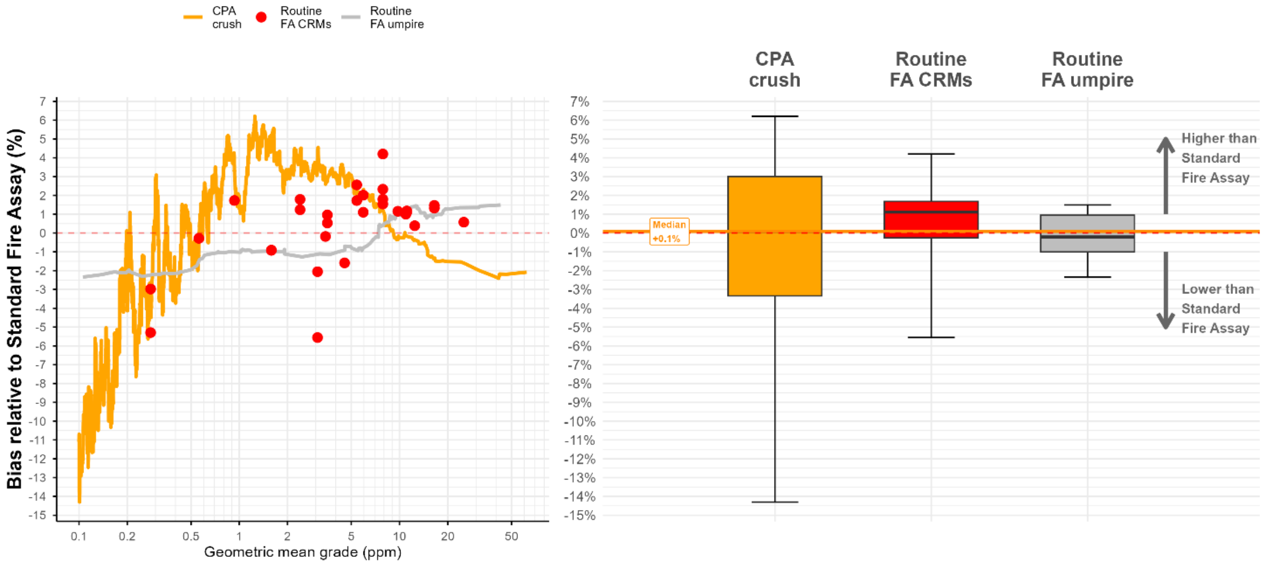

FIG 7 – ‘Accuracy Framework’, showing CPA versus FA, CRMs for standard FA, and FA umpire against standard FA. The boxplots for CPA versus FA and FA_{umpire} versus $FA_{standard}$ represent the distribution of the (geometric) moving average (N = 101, edge-truncated), expressed in percentage difference. Horizontal dashed line is the standard fire assay ‘0 per cent baseline’ that is compared to. Thick orange line (right) is the median for the CPA analysis.

Similarly, the bias determined for the routine FA laboratory process, relative to the overall performance of regular umpire testing, can be referenced in this accuracy framework. If the project or mine has been operating successfully and consistently within, say a -2.5 to +1.0 per cent umpire laboratory bias, then by inference, such a bias becomes acceptable. The testing programme recommended in Figure 1 includes a set of data from a second FA laboratory, in case routine umpire testing is not available for the project. The results of the umpire testing (ie FA_{umpire} versus $FA_{standard}$) can be plotted in a similar fashion as the CPA versus FA and included as a second line in the Tukey plot (Figure 7, left), or by a box and whisker plot, showing the median, quartiles and min-max values (Figure 7, right).

A pragmatic way to review the biases for CPA versus FA, CRMs, and FA_{umpire} versus $FA_{standard}$ together is by plotting them all together as box and whiskers (Figure 7, right). This, of course, obfuscates the trend of the bias across the grade range, but at least allows the practitioner to set reasonable DQOs from this framework to test against.

An example of the accuracy framework is presented in Figure 7 for a fictitious data set. The routine FA performance based on FA CRMs and umpire FA testing demonstrates a median bias of +1.0 per cent and -0.1 per cent, respectively. The range of biases observed by these two references (as defined by the quartiles) is -1.0 per cent to +1.5 per cent, and -1.0 per cent and +1.0 per cent, respectively. The geologist can use these defined metrics to specify sensible DQOs; for instance, in this case it could be set between -1.0 per cent and +1.5 per cent, based on the interquartile ranges for CRMs and routine umpire testing (Figure 7, right).

When DQOs are then set (eg 'the bias must not be outside -3.0 per cent or +3.0 per cent'), based on either industry norms, Competent/Qualified Person requirements, or an evaluation of routine CRM results and umpire test results (this section), two complementary statistical approaches can then be used to make statistically objective claims: (a) one can examine whether the 95 per cent confidence interval of the log ratio (derived from the log t-test) contains the threshold values (eg log(1.03) or log(1/1.03) for a 3.0 per cent difference); or (b) one could use a formal TOST (two one-sided tests) procedure to test for equivalence within a specified range (Napier-Munn, 2014).

Precision

Similarly, other than identifying statistically significant differences between CPA and FA pulp duplicate (precision) performance, further information on the materiality of any difference in the precision of CPA and FA results can be drawn through evaluation against industry benchmarks (Table 3 on page 143 in Abzalov (2013)), or against a DQO set by the geologist (ie 'the precision must be better than 15 per cent'). In other words, the CPA precision may be statistically significantly different than the routine FA pulps, but still good enough as it is well within those thresholds.

To further test the impact and materiality of the CPA precision, experimental variograms based on FA data can be compared against variograms based on CPA data — this is only possible if the test data set is large enough. If the experimental variograms (downhole, major, semi-major, and minor) demonstrate comparable values for each lag value (Figure 8), they would result in near-identical modelled variograms. Therefore, the weights applied during grade interpolation (eg kriging) would result in near-identical estimates. If CPA is used to analyse plant samples, then a similar exercise on a time-based variogram can be performed.

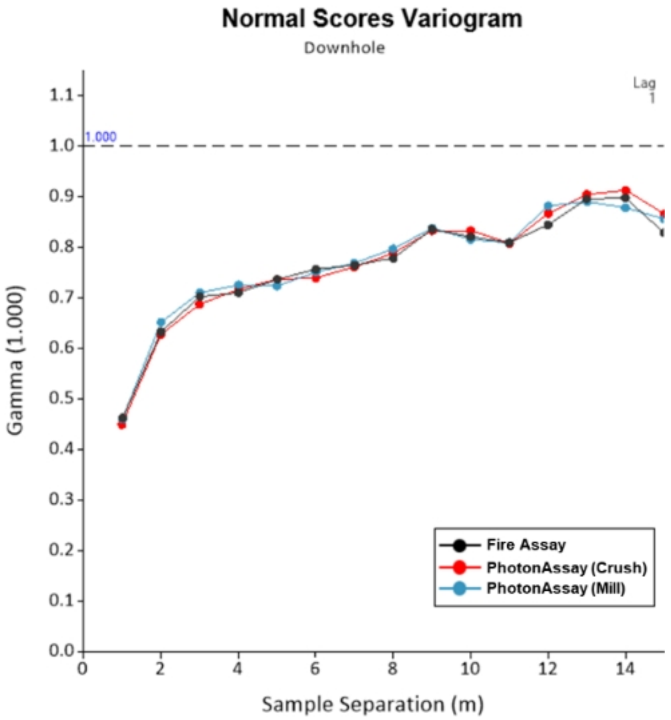

FIG 8 – Example of very similar experimental variograms based on crushed CPA, pulverised CPA, and FA data, demonstrating that results from all three analyses (ie $\text{CPA}_{\text{crush}}$, $\text{CPA}_{\text{pulp (mill)}}$ and FA) would result in near-identical modelled variograms, and thus in near-identical mineral resource estimates (from a *precision* perspective).

This approach can also be used to test the impact of using crushed versus pulverised samples for CPA (Figure 8). For example, the authors note that even though an RMS CV difference of 12 per cent for FA versus 8.7 per cent for pulverised CPA appears significant, and appears to justify the cost of further pulverising the CPA samples, it is easy to point out that it has a negligible impact on the resource estimate, and therefore likely not worth the additional expense.

FINAL REMARKS

Whether CPA data are fit for purpose for a project should be evaluated in the context of the deposit geology, its objectives, and associated DQOs. The workflow presented in this paper provides practical steps and highlights critical items for consideration when deciding between CPA and conventional methods.

Conclusions on accuracy and precision, and suitability of CPA to replace FA, can only be drawn when clear DQOs are agreed to before the study starts. If third parties are requested to undertake a comparative study, then they can only determine outcomes if they have been provided with such DQOs by the client.

Last, the authors note that fire assaying itself is not a perfect method, and comparing test results from a single laboratory should be treated with caution. Any observed differences need to always be regarded in that light, and DQOs set accordingly. Dominy *et al* (2024), in their conclusions on CPA review, boldly state that 'coarse-gold assaying with FA is flawed', and make reference to work by Royle (1989), Pitard and Lyman (2013), Dominy (2014, 2017), Lyman, Robertson and Day (2016), and Pitard (2017) and Dominy *et al* (2024) to support that statement. A bias at high-grade was also observed by Hitchman *et al* (2024) in their review of the performance of FA versus CPA at the very nuggety Fosterville deposit in Australia, with CPA results being higher than FA at high grades. Screen-fire test work confirmed that this bias was introduced by the sub-sampling of the pulp using routine low charge weights for the fire assay; a well-known issue.

REFERENCES

Abzalov, M, 2008. Quality control of assay data: a review of procedures for measuring and monitoring precision and accuracy, *Exploration and Mining Geology*, 17:131–144.

Abzalov, M, 2011. Sampling Errors and Control of Assay Data Quality in Exploration and Mining Geology, in *Applications and Experiences of Quality Control* (ed: O Ivanov), chapter 31 (IntechOpen: London). https://doi.org/10.5772/14965

Abzalov, M, 2013. Measuring and Modelling of Dry Bulk Rock Density for Mineral Resource Estimation, *Applied Earth Science*, 122(1). https://doi.org/10.1179/1743275813Y.0000000027

Chrysos, 2022. Chrysos PhotonAssay™ Measurement Performance for PAAU02, Performance Note PN-001, Chrysos Corp. internal report, p 2.

Chrysos, 2023. PhotonAssay Measurement Performance for Gold – Impact of, U, Th and Ba, Technical Note TN-106:6, Revision 2, p 6.

Chrysos, 2025. Emissions, Energy and Waste Assessment – PhotonAssay and Fire Assay, GHD internal report, p 51.

Cohen, J, 1988. The effect size, *Statistical Power Analysis for the Behavioral Sciences*, pp 77–83.

Cohen, J, 1992. A Power Primer, *Psychological Bulletin*, 112:155–159.

Dominy, S C and Minnitt, R C A, 2012. Application of Heterogeneity Testing to Coarse Gold Ores, in *Proceedings of the Sampling Conference 2012*, pp 143–160 (The Australasian Institute of Mining and Metallurgy: Melbourne).

Dominy, S C, 2014. Predicting the unpredictable – evaluating high-nugget effect gold deposits, *Mineral Resource and Ore Reserve Estimation Guide to Good Practice*, pp 659–678 (Australasian Institute of Mining and Metallurgy: Melbourne).

Dominy, S C, 2017. Sampling coarse gold-bearing mineralisation – developing effective protocols and a case study from the Ballarat mine, Australia, in *Proceedings of the World Conference on Sampling and Blending*, pp 71–84 (Australasian Institute of Mining and Metallurgy: Melbourne).

Dominy, S C, Graham, J C, Esbensen, K H and Purevgerel, S, 2024. Application of PhotonAssay™ to Coarse-Gold Mineralisation – The Importance of Rig to Assay Optimisation, *Sampling Science and Technology*, January 2024(1):2–30. https://doi.org/10.62178/sst.001.002

Feltz, C J and Miller, G E, 1996. An asymptotic test for the equality of coefficients of variation from k populations, *Statistics in Medicine*, 15:11.

Gelman, A and Stern, H, 2006. The difference between 'significant' and 'not significant' is not itself statistically significant, *The American Statistician*, 60:328–331.

Hassan, A, Huang, D, Mustafa, E K, Mahama, Y, Damos, M A, Jiang, Z and Zhang, L, 2020. Statistical inference and residual analysis for the evaluation of datum transformation models developed on 3D coordinate data, *Journal of Applied Geodesy*, 14:65–75.

Hitchman, S P, Simbolon, S, Symons, D J, Hoare, J P and Carpenter, J B, 2024. Implementation of Chrysos Photon Assay Method at Fosterville Gold Mine, Victoria, in *Proceedings of the International Mining Geology Conference*, pp 341–345 (Australasian Institute of Mining and Metallurgy: Melbourne).

Hyslop, N P and White, W H, 2009. Estimating precision using duplicate measurements, *Journal of the Air and Waste Management Association*, 59:1032–1039.

JORC, 2012. Australasian Code for Reporting of Exploration Results, Mineral Resources and Ore Reserves (The JORC Code) [online]. Available from: <http://www.jorc.org> (The Joint Ore Reserves Committee of The Australasian Institute of Mining and Metallurgy, Australian Institute of Geoscientists and Minerals Council of Australia).

Kelley, K and Preacher, K, 2012. On Effect Size, *Psychological Methods*, 17:137–152. https://doi.org/10.1037/a0028086.

Lakens, D, 2013. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs, *Frontiers in Psychology*, 4:62627.

Long, S, 2015. What reporting Portable XRF or any other Assay Results to 2012 JORC Code Standards should mean, *AIG News*, pp 11–16.

Lyman, G J, Robertson, I G and Day, T, 2016. The simple facts about sampling gold ores, in *Proceedings of Mill Operators Conference 2016*, pp 159–172 (Australasian Institute of Mining and Metallurgy: Melbourne).

Miller, J E, 2023. Beyond Statistical Significance: A Holistic View of What Makes a Research Finding "Important", *Numeracy*, 16:6.

Minnitt, R C A, 2016. Calibrating K and Alpha in Gy's Formula: A New Approach, *Mathematical Geosciences*, 48:211–232.

Napier-Munn, T, 2014. Statistical methods for mineral engineers – How to design experiments and analyse data, Queensland Australia, report, Julius Kruttschnitt Mineral Research Centre.

Peeters, M J, 2016. Practical significance: Moving beyond statistical significance, *Currents in Pharmacy Teaching and Learning*, 8:83–89.

Pitard, F F and Lyman, G J, 2013. Single and multi-stage Poisson processes: a case study for gold exploration, in *Proceedings of the World Conference on Sampling and Blending*, pp 371–385 (Gecamin: Santiago).

Pitard, F F, 2013. Guidelines for acceptable allotted sampling uncertainty, in *Proceedings of the World Conference on Sampling and Blending*, pp 89–98 (Gecamin: Santiago).

Pitard, F F, 2017. From errors to uncertainty – a clarification for proper use by the Theory of Sampling, in *Proceedings of the World Conference on Sampling and Blending 2017*, pp 355–358 (Australasian Institute of Mining and Metallurgy: Melbourne).

Royle, A G, 1989. Splitting gold assay pulps containing coarse gold, *J Leeds Univ Min Assoc*, 89:63–68.

Smee, B, Bloom, L, Arne, D and Heberlein, D, 2024. Practical applications of quality assurance and quality control in mineral exploration, resource estimation and mining programmes: a review of recommended international practices, *Geochemistry: Exploration, Environment, Analysis*, (2024)24(2):geochem2023-046. https://doi.org/10.1144/geochem2023-046

Stanley, C R and Lawie, D, 2007. Average relative error in geochemical determinations: Clarification, calculation and a plea for consistency, *Exploration and Mining Geology*, 16:267–275.

Sterk, R, 2015. Quality control on assays: addressing some issues, *AusIMM New Zealand Branch Annual Conference, Dunedin*.

Thompson, B, 2007. Effect sizes, confidence intervals and confidence intervals for effect sizes, *Psychology in the Schools*, 44:423–432.

Tickner, J, Ganly, B, Lovric, B and O'Dwyer, J, 2017. Improving the sensitivity and accuracy of gamma activation analysis for the rapid determination of gold in mineral ores, *Applied Radiation and Isotopes*, 122:28–36.

Tickner, J, Lannan, O and Preston, R, 2021. Comparing different assay methods for the determination of mineral ore grade, *Journal of Geochemical Exploration*, 226:106780.

Tremblay, C D, Tickner, J, Treasure, D, Oteri, A and Wheeler, G, 2019. PhotonAssay – efficient and bulk gold analysis in the modern world, in *Proceedings of the International Mining Geology Conference 2019*, pp 88–98 (Australasian Institute of Mining and Metallurgy: Melbourne).

Westgard, J O, Barry, P L, Hunt, M R and Groth, T, 1981. A multi-rule Shewhart chart for quality control in clinical chemistry, *Clinical Chemistry*, 27:493–501.

Wilcoxon, F, 1945. Individual comparisons by ranking methods, *Biometrics Bulletin*, 1:80–83.