REVIEW

JAAS
Journal of Analytical Atomic Spectrometry

Sampling as a source of measurement uncertainty: techniques for quantification and comparison with analytical sources

Michael H. Ramsey

Environmental Geochemistry Research Group, T. H. Huxley School of Environment, Earth Science and Engineering, Imperial College, London, UK SW7 2BP

A tutorial review is presented of current methods for the estimation of measurement uncertainty due to primary sampling. Current terminology used in the description of uncertainty and analytical data quality is reviewed and explained. One basic method for the estimation of uncertainty in sampling is described in detail with a worked example of its application to a test dataset. This method employs the taking of a proportion of samples in duplicate, with the further duplication of chemical analysis on these samples. Robust analysis of variance (ANOVA) is applied to estimate the total measurement uncertainty and also to quantify the contributions to that uncertainty which arise from the processes of primary sampling and chemical analysis. The ANOVA program and test data are available electronically to enable application of the methodology. The assumptions and limitations of this basic method are discussed, including its inability to estimate sampling bias. More sophisticated methods are discussed that include the estimation of the contributions to uncertainty from systematic errors in both sampling and analysis. Other approaches to the estimation of uncertainty from sampling, from both sampling theory and geostatistics are compared with these methods. The comparison is made between sampling and chemical analysis as the two sources of uncertainty, relative to each other, and relative to the overall variance of the measurements. Fitness-for-purpose criteria are given and discussed for the ideal maximum and minimum values of the proportion of the measurement variance to the total variance, and the relative contributions of the sampling and analytical variances.

Keywords: *Measurement uncertainty; sampling; analytical data quality; robust analysis of variance; fitness-for-purpose; sampling bias*

Measurement uncertainty has become an important concept in analytical science that unifies many previously disparate strands of information on data quality. Chemometric techniques play a crucial role in estimating values for the overall uncertainty and also in the separation and quantification of the various components of uncertainty. These components include not just those arising from the chemical analysis, but also those arising from the sampling procedure that is used to select the primary sample from the sampling target. Such sampling targets could be for example, a stockpile of process material, a site of contaminated land or a batch of foodstuff. In the case of environmental and geochemical investigations, primary sampling is often the main source of uncertainty and dominates analytical sources such as instrumental determination or the sub-sampling of test materials within the laboratory.

Analytical chemists generally need to recognise sampling as the first step in the measurement process, and to include its contribution in the estimation of uncertainty. This will give more realistic estimates of the measurement uncertainty than consideration of only the lab-based analytical procedures. The inclusion of the sampling step becomes particularly important in deciding acceptable levels of uncertainty arising from chemical analysis. If the uncertainty from sampling is very large for example, then there comes a point where reduction in uncertainty from chemical analysis makes a negligible impact and is therefore not cost-effective. Only when analysts know the contribution of uncertainty from sampling can they decide what is a reasonable contribution from analysis.

This paper explains the terminology relating to both uncertainty and data quality, and tries to clarify their interrelationship and what they both mean for analytical measurements. It also reviews the options for estimating uncertainty that arise from the process of measurement. The established methods for quantifying the analytical component of the uncertainty will be described briefly to act as a comparison with the more recent methods that are being developed for the components that arise from the process of primary sampling. For this purpose sub-sampling of a test material within the laboratory will be considered as part of the analytical component.[1]

The main aim of this paper will be to describe the practical implementation of the chemometric techniques rather than to discuss the details of the field studies,[2–4] or the implications of the findings for either the interpretation of results[5] or for the objectives of geo-analytical science,[6] that have been discussed elsewhere. The assumptions upon which the calculations are made will be discussed as will the limitations that these impose on the validity of the results. Alternative approaches to the estimation of uncertainty and its components will also be discussed.

The chemometric techniques for the comparison of the components of the uncertainty that arise from both the sampling and the chemical analysis are described. These techniques can separate the components, indicate whether the combined uncertainty is excessive, and also indicate which of the components is most dominant and may need improvement. This later objective involves an assessment using 'fitness-for-purpose' criteria.

TERMINOLOGY OF UNCERTAINTY AND DATA QUALITY

Measurement uncertainty has been defined informally[7] as 'the interval around the result of a measurement that contains the true value with high probability'. This is somewhat more

comprehensible than the formal definition of 'uncertainty of measurement' given by the International Organisation for Standardisation (ISO) as: 'A parameter associated with the result of a measurement, that characterises the dispersion of the values that could reasonably be attributed to the measurand.'[8] This definition utilizes the relatively new and often misunderstood term of measurand. This term can be defined formally as the 'particular quantity subject to measurement'.[9] For analytical chemistry, it can be defined informally as the true value of the analyte concentration in a specified segment of the material, and is not a synonym for 'analyte'. The 'true value' is always unknown, but when estimated as the 'accepted reference value' it acts as a reference point for the estimation of uncertainty, error, accuracy and bias. There is some confusion over the use of terms such as 'conventional true value' and 'reference value' in the literature.

The uncertainty of a measurement is quite different from the error. Error is 'the result of a measurement minus the true value of the measurand'[9] and thus contains both a random and a systematic component. Bias is 'the difference between the expectation of the test result and an accepted reference value'.[10] In practice this expectation is estimated as the mean of a large number of measurements and is essentially systematic. Trueness is a lack of bias, and is formally 'the closeness of the agreement between the average value obtained from a large series of test results and an accepted reference value'.[10] Precision is 'the closeness of agreement between independent test results obtained under prescribed conditions',[11] and describes random error. Examples of such conditions are 'repeatability' and 'reproducibility'. The terms 'bias' and 'precision' can be used therefore to describe the quality of methods of measurement, whereas the terms 'uncertainty' and 'error' refer to individual measurements. A clearer understanding of these terms and their relationship with uncertainty may be gained from a graphical representation, either in one or two dimensions (Fig. 1).

From a single measurement it is impossible to estimate the bias or precision of a method [Fig. 1(a)]. The difference between the measured value and the 'true' value of the concentration is called the 'error'. From a single measurement however, it is not clear whether this error is caused by a systematic or a random error in the method. The ISO definition of accuracy is 'the closeness of the agreement between a test result and the accepted reference value.'[9] This is the term used to express this error of a single measurement, which cannot necessarily be identified as being of a systematic or a random nature. 'Accuracy' and 'error' have very similar meanings, but 'accuracy' is a lack of 'error' for single measurements. This is equivalent to 'trueness' being a lack of 'bias' for methods, or their results.

When any further measurements are added to the first, it becomes possible to estimate the systematic component (*i.e.*, bias) from the random component of the error in the method (*i.e.*, precision). It is possible that the error in the first measurement is almost entirely random [*e.g.*, Fig. 1(b)] or equally possible that it is almost entirely systematic [*e.g.*, Fig. 1(c)].

In the former case [Fig. 1(b)] all four measurements are made with a method that is not very precise, but the average value of which is close to the 'true' value (*i.e.*, the bias of the method is small). In the latter case [Fig. 1(c)], although the method is quite precise, the average value is a large distance from the 'true value' (*i.e.*, the bias of the method is large). The ideal case is where the bias of the method is small, the precision is good (*i.e.*, a small value) and all measurements therefore have a small error [Fig. 1(d)]. In the case where the precision of the method is poor (*i.e.*, a large value) and the bias is also large, then the individual measurements can have either large or small errors [Fig. 1(e)]. Such differences between the errors of individual measurements made by one method is to be expected for any method whatever the precision and bias.

Estimation of the uncertainty of the measurements in each of these cases (Fig. 1) illustrates the various components of uncertainty and the difference between the terms uncertainty and error. Where the method has a large precision value [Fig. 1(b)] the uncertainty of the measurements is dominated by this component and is also a large value. Where the precision and the bias of the method are both small values [Fig. 1(d)] the uncertainty is also small. Where both the precision and the bias of the method have large values [Fig. 1(e)] the uncertainty has a large value, based on a combination of both components.

A more contentious case is where the precision of the method

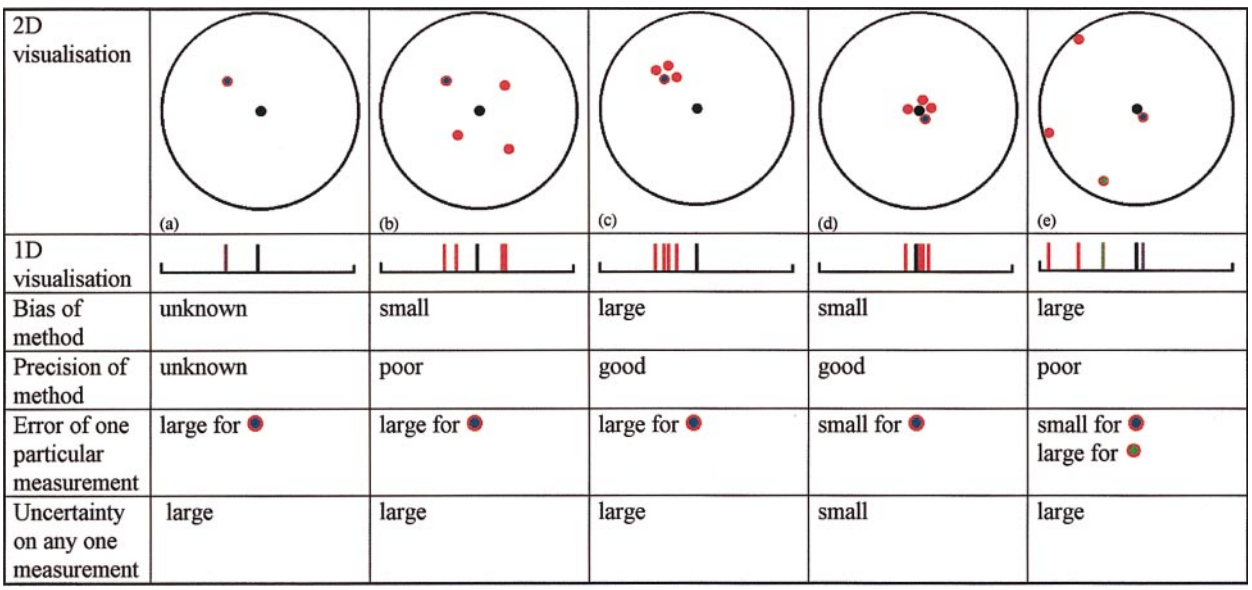

Fig. 1 Diagrammatic explanation of the difference between the terms bias, precision, error and uncertainty. (a) From a single measurement (●) of analyte concentration it is impossible to estimate the bias or precision of a method of measurement, however the error of that single measurement is simply its difference from the true value of the analyte concentration (●). (e) If the method of measurement has a large value of bias and precision, it is still possible to have a single measurement with a small error (●), but it will have a large value of uncertainty. See text for further discussion.

is a small value, but the bias is large [Fig. 1(c)]. In this case the range containing the true value (*i.e.*, the uncertainty) is large. In the ISO definition of uncertainty, it is assumed that all systematic errors have been corrected.[8] If the bias of the method were known with specified uncertainty for the samples analysed, then the measurements could be corrected. In this case the uncertainty would be small, but larger than for the uncorrected measurements by the amount caused by the uncertainty in the correction. The problem with this approach when applied to chemical analysis is that systematic errors in a method are not always known and cannot therefore always be corrected. Sometimes such errors are suspected but their exact size is unknown, perhaps because the reference materials used to estimate them are not well matched to the composition of the samples. In such cases the systematic component of the uncertainty must be incorporated in order to give a realistic estimate of the uncertainty. When inter-organisational trials are used to estimate the uncertainty by a 'top-down' method (see below) the uncorrected analytical bias of each participant automatically becomes incorporated in the estimated uncertainty.

TECHNIQUES FOR ESTIMATING MEASUREMENT UNCERTAINTY

For chemical analysis, two main strategies have been proposed for the estimation of measurement uncertainty. They both need to be evaluated as options for the application to primary sampling. In the 'bottom up' approach the random error from each individual component of a method is quantified separately as a standard deviation (s). The overall uncertainty is then estimated by summing the individual errors by their variances (s^2).[8,12] The alternative 'top down' approach uses inter-laboratory trials (such as proficiency tests or collaborative trials) to estimate the total uncertainty of a measurement. In this method, many selected laboratories ($n>8$) analyse the same sample, by the same analytical method.[1] The scatter of the measurements reported by all of the laboratories is then used to derive an overall estimate of uncertainty. The 'bottom up' approach has the limitation that it requires that all of the sources of uncertainty need to be identified. It is relatively easy to consider the obvious sources of error which are explicit parts of the method (*e.g.*, weighing, volumetric additions). However, the most important source of uncertainty may not be explicit in the method (*e.g.*, laboratory temperature) and it is therefore easily overlooked especially by inexperienced practitioners. Furthermore, it can be a long and expensive procedure to quantify all the component errors if the method is to be applied rigorously. Once the main component of the uncertainty has been identified however, then the variance of this component often dominates to such an extent that future monitoring can concentrate on this one source of variance.

The benefits of the 'top down' approach can be appreciated from the discrepancies that are often evident between laboratories in inter-organisational trials. These differences are often larger than can be accounted for by the individual estimates of uncertainty within each lab. This is because the 'bottom up' approach used by individual labs tends to give an underestimate of the uncertainty. The limitation of the 'top down' approach is that it depends on the selection of laboratories that contribute. If the laboratories all use a similar source of calibration, they may all be equally biased and therefore give an underestimate of the uncertainty. Alternatively, one laboratory may have gross errors, atypical of the application of the method as a whole, and these will cause an overestimate of the uncertainty.

Estimating uncertainty in sampling

For the analysis of many primary materials, the main problem with estimating measurement uncertainty by the methods described above is that they ignore the uncertainty arising from primary sampling. The term 'primary sampling' is used to differentiate it from the sub-sampling of test materials that happens within the laboratory, which can easily be estimated as part of analytical precision.[1] It is often quoted that an analysis can never be of better quality than the sample upon which it is made. What has been lacking however is the means of estimating the size of the uncertainty which is introduced by sampling. Methods devised originally for estimating analytical uncertainty have been adapted however, to the estimation of sampling uncertainty.[5]

A more holistic approach is to consider primary sampling and chemical analysis as just two parts of the same 'measurement' process, and to quantify their combined contribution to the uncertainty. The overall measurement uncertainty can therefore be considered to have contributions from four components. These are the random and systematic errors arising from the procedures of both primary sampling and chemical analysis. In terms of the quality of the methods employed, these four components can be quantified as sampling precision, analytical precision, sampling bias and analytical bias.

Taking the 'bottom up' approach to estimating the total measurement uncertainty we can review the methods available for the estimation of these four components. Analytical precision can be measured by the use of analytical duplicates[13] or in combination with sampling precision using a balanced design of sampling and analytical duplicates.[14,15] A balanced design is achieved when there are equal numbers of analytical replicates on each of the sampling replicates, as is the case in the test data discussed below. Analytical bias is usually estimated by the analysis of certified reference materials.[16] New methods have been proposed for the estimation of sampling bias. For a single sampler it is possible to measure the bias between the measurements by different protocols.[2] Bias against an 'accepted reference value' as specified by ISO, could be achieved in theory by the use of a reference sampling target. This is the sampling equivalent of a reference material for the estimation of analytical bias.[17] Such a target has not yet been created, but in principle it would be given a certified value and uncertainty of some parameter (such as the mean concentration of an element) by a certification trial, analogous to the procedure currently used to certify a reference material.

Taking the 'top down' approach it is possible to use measurements from inter-organisational sampling trials, such as sampling proficiency tests[4] and collaborative trials,[3] to estimate uncertainty[5] which can incorporate the uncorrected bias in the measurements from any of the participants.

Four methods for the estimation of measurement uncertainty from primary sampling have recently been proposed and tested.[5] They correspond to the four possible combinations of numbers of samplers (*i.e.*, people taking samples) and numbers of sampling protocols employed, which are:- Method 1, for single sampler/single protocol; Method 2, for single sampler/multiple protocols; Method 3, for multiple sampler/single protocol; and Method 4, for multiple sampler/multiple protocols.

Only the first of these techniques will be discussed in detail here with an explanation of how to implement the method and how the chemometric technique of analysis of variance (ANOVA) is used to estimate the uncertainty and its components. Application of the other three techniques is discussed elsewhere.[5]

Example of estimation of uncertainty due to sampling and analysis

The most straightforward method for the estimation of uncertainty due to sampling is applicable to the case of a single

sampler applying a single sampling protocol. It gives an estimate of the random components of the procedures (*i.e.*, sampling and analytical precision) but requires separate estimates of the systematic component (sampling and analytical bias).

This method is based on the taking of duplicate samples for some proportion of the sample increments (*e.g.*, 10%). Duplicate chemical analyses are then made on both of these sample duplicates, in a balanced design (Fig. 2). This experimental design was originally suggested for the estimation of sampling and analytical precision by Miesch[18] and then again by Garrett.[14] The most important aspect is the selection of how the duplicate samples are taken. They are not taken at exactly the same place, but separated by a distance that reflects the separation that might have occurred by a totally independent interpretation of the sampling protocol. In the investigation of contaminated land, for example, application of a rapid survey technique may place a duplicate sample two metres away from an original sample.[2] For sampling targets with rapid temporal variability (*e.g.*, river water) the sampling duplicates will need to also be separated in the time of their collection. The period of separation should reflect the uncertainty in the time of collection that could arise when independent samplers use the same specified sampling protocol.

The three components of the variability can be separated using classical analysis of variance (ANOVA) which is available in most statistical software packages. Classical ANOVA is based on three assumptions (discussed below) and is strongly affected by a few outlying values. These problems can be largely overcome by the use of robust analysis of variance,[15,19] but this is not usually available within statistical packages.

Robust ANOVA can be implemented using a specifically written computer program called ROBCOOP4.EXE, that has been adapted from a published program.[20] This program has been validated on simulated test data of known nominal variance and proportion of added outliers.[15] This program and the test data used below (MUTEST.DAT) are available from the JAAS web site *via*:- http://www.rsc.org/jaas.

In order to use this program, the four measurements made on the two duplicate analyses (A1 and A2) of the duplicate samples (S1 and S2) must be arranged in the format shown for the test data in Table 1. The headings such as 'S1A1' (*i.e.*, sample 1, analysis 1) and 'S2A1' (*i.e.*, sample 2, analysis 1) are included for information only and should not appear in the data file for computation. The measured concentration values can be entered into the columns of a spreadsheet program and can then be exported and stored as an ASCII or TEXT data file, an example of which can be seen in the file MUTEST.DAT.

A good general principle when applying any chemometric technique is that the analyst should have an intuitive appreciation of whether the result of a computation is reasonable. It is advisable therefore to visually inspect the raw measurements to predict a reasonable outcome for the ANOVA analysis. Each line in Table 1 shows the four measurements of Pb (in $\mu g\ g^{-1}$) in soil at one of the 18 locations sampled for this purpose. The analytical duplicate measurements (*e.g.*, S1A1, S1A2) generally show good agreement with differences of around 10% relative to the mean value. The agreement between the duplicate samples at the same location (*e.g.*, average S1 against average S2 for one site) varies by up to a factor of about $\times 2$ (± 50 to 100%). This variation, whilst substantial, is much less than the variation between the locations, which varies by a factor of approximately $\times 10$.

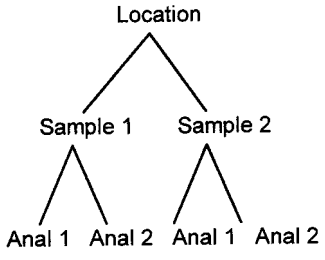

Fig. 2 Balanced experimental design for the estimation of random components of measurement uncertainty by Method 1.

Table 1 Format of the analytical measurements of test dataset MUTEST.DAT for estimation of measurement uncertainty using analysis of variance (ANOVA)*

S1A1†	S1A2†	S2A1†	S2A2†
155.6	156.0	173.2	170.0
201.2	184.8	163.2	161.2
139.2	142.0	229.2	222.0
389.0	408.0	754.0	812.0
187.0	191.0	124.0	121.0
314.4	316.8	231.2	226.4
462.0	528.0	848.4	859.2
286.8	288.0	492.4	490.4
176.4	184.8	232.0	238.0
157.2	151.6	121.2	120.4
264.0	277.6	266.0	264.8
447.2	450.0	374.0	371.2
372.8	373.6	392.4	379.6
113.6	116.4	85.2	83.2
586.4	583.2	528.4	418.8
209.6	178.8	137.2	145.2
271.2	256.0	284.4	299.6
113.2	114.8	127.2	120.4

* Four measurements of Pb (in $\mu g\ g^{-1}$) in soil at 18 sites, with duplicate chemical analysis (of A1, A2) of two field samples (S1, S2), corresponding to the experimental design in Fig. 2. † Headings (*e.g.* S1A1, S2A1) are for information only and should not appear in data file for computation. Visual inspection of the variation shows that the analytical precision is good and although there are differences of a factor of 2 between the field duplicates, these are much smaller than the variability between sites, which ranges over a factor of ten.

The program ROBCOOP4.EXE is a compiled FORTRAN program which can be run in MS-DOS, without a compiler, by following the input requirements shown below (in bold). This input assumes that the program is in current drive and directory, and that the data is, for example on drive 'a:'.

ROBCOOP4
Enter file name of raw data **a**:**MUTEST.DAT**
#elements **1**
#sites, #replicate samples per site, #replicate analyses per sample **18,2,2**
---- element 1

The structure of the test data file (18,2,2) corresponds to that shown in Table 1. The results of the ANOVA calculations are written in a file RESULTS.LIS that is created by the program ROBCOOP4.EXE. This file can be viewed using a word-processing program. It is overwritten every time the ANOVA program is run so it should be renamed and saved after each run if required.

The output of ROBCOOP4.EXE when applied to the test data in Table 1 (*i.e.*, MUTEST.DAT) is shown in Fig. 3. The last five lines of the output file show the results of the robust ANOVA. The 'mean' value is the robust mean, in the units of concentration used in the input data (in this case $\mu g\ g^{-1}$). The 'Sigma values' are the estimates of standard deviation for the three different sources of variance, again in the same units of concentration. The last line 'sigma (total)' is the total of all these three sources, summed by their variance using eqn. (1), given below. The 'Percent variance' line expresses each of the three variances as a percentage of the total variance. The upper part of the output gives the same items of information but

```
--- element        1
Classical results: Mean =   289.1001
Sums of Squares are -  1989270.4   365656.5  11331.401
Sigma values(geochem, sampling, analysis)    -  155.484   99.999   17.742
Percent variance(geochem, sampling, analysis) -   70.09    28.99    0.91
sigma (total) -  185.715
Robust results:
mean =   273.9382
Sigma values (geochem, sampling,analysis)    -  154.789   56.967   6.662
Percent variances (geochem, sampling,analysis) -   87.93    11.91    0.16
sigma (total) -  165.073
```

Fig. 3 Output of the program ROBCOOP4.EXE when applied to the test data in Table 1 (*i.e.* MUTEST.DAT). The upper five lines give the results of the classical ANOVA and the lower five lines the results of the robust ANOVA. For discussion see text.

calculated using classical ANOVA, for the purpose of comparison. The classical variance estimates are generally higher than the robust estimates. This is because of a small number of outlying measurements, for example the duplicate analyses (528.4 and 418.8 μg g^{-1}) of the second sample taken at the 15th location in Table 1.

Two of the component variances can be classed as measurement uncertainty, and these are the sampling and the analytical variance (s^2_{samp} and s^2_{anal}). The third component is the between-location variance due to real variation of the analyte across the target. This is called the geochemical variance ($s^2_{geochem}$) in this particular case of a geochemical investigation. In this example the sampling uncertainty or within-location variance will be partially due to small scale geochemical variation within the location, but represents the uncertainty in all of the samples that could be taken from that 'location' as specified (in this case within a two metre radius).

All three variances can be summed to give the total variance of the survey. This is the figure that would be calculated when calculating the standard deviation of all the measurements, and can be expressed by:-

$$s^2_{total} = s^2_{geochem} + s^2_{samp} + s^2_{anal} \quad (1)$$

The measurement uncertainty (u) can be estimated using this 'bottom up' approach, from the combination of the sampling and analytical variance, giving the measurement variance as:-

$$u = s_{meas} = \sqrt{(s^2_{samp} + s^2_{anal})} \quad (2)$$

It is usual to increase the confidence interval of the uncertainty by multiplying by a coverage factor (k) to give the expanded uncertainty (U).[8] The is analogous to the quotation of precision using two standard deviations (*i.e.*, $k=2$) for 95% confidence (95.44% to be exact). This gives:-

$$U = ku = 2s_{meas}$$

As relative uncertainty (*i.e.*, expanded uncertainty expressed relative to the concentration) this becomes:-

$$U\% = 200 s_{meas}/\bar{x} \quad (3)$$

For the results on the test data, the uncertainty estimate is 57.36 μg g^{-1}, using eqn. (2). The expanded uncertainty with a coverage factor of 2 is twice this value, and the relative uncertainty is 41.87%, using eqn. (3). This equation assumes that the uncertainty (and hence s_{anal}) is a fixed proportion of the concentration. This is a reasonable assumption when the analyte concentration is well above the detection limit of the analytical method. The standard deviation of the analysis ($s_{anal,c}$) at some concentration (c) has been shown to vary as a function of concentration,[13] by the equation:-

$$s_{anal,c} = s_{anal,0} + \theta c$$

where $s_{anal,0}$ is the standard deviation of the analysis at zero concentration. The constant θ is related to the relative precision that is approached at high concentration (*e.g.*, 0.05 for 5% precision, expressed for one standard deviation). If the analyte concentration is close to the detection limit of the analytical method, it should be possible to include this function for the standard deviation from analysis as $s_{anal,c}$ in eqn. (2) for the estimation of uncertainty.

The calculated value of the uncertainty applies to measurements made on single samples taken in the survey. If n multiple samples are taken at any individual location within the site, the uncertainty on the average for that location is the value given by eqn. (3) divided by $\sqrt{n}$. This is equal to the standard error on the mean value ($s/\sqrt{n}$). For example the estimated relative uncertainty at a location where four measurements have been made would be half ($1/\sqrt{4}$) of the value given by eqn. (3) (20.93% for the test data).

Relative contributions of sampling and analysis to uncertainty

It is clear from eqn. (2) that random errors from both the processes of primary sampling (s^2_{samp}) and chemical analysis (s^2_{anal}) contribute to the overall measurement uncertainty. This basic method should give a realistic estimate of the uncertainty, but the use of ANOVA also allows the individual contributions to the uncertainty to be separated and quantified. Although the absolute size of these components can be useful, it is their size relative to each other and relative to the total variance, that can be most useful in deciding the suitability of the methods of sampling and analysis employed. This can best be appreciated when the component variances are expressed as percentages of the total variance. For the results of the robust ANOVA on the test data (Fig. 3) the 'percent variance' figures show that the analytical variance contributes a very small fraction of the total variance (0.16%). This is well below the suggested maximum percentage for the ideal method in which the analytical variance should not exceed 4% of the total variance.[15] The proportions and the relationship to the ideal values can be communicated more clearly by displaying them as a pie chart (Fig. 4).

Alternatively the analytical variance should ideally contribute less than 20% to the measurement variance, if the measurement uncertainty is not to be limited by the analytical component. For the test data this percentage is 1.3% [0.16/(11.91 + 0.16)], which is well within this ideal value.

The measurement variance should ideally contribute less than 20% to the total variance,[15] if the measurements are to give a clear representation of the true variation of the analyte across the sampling target (in this case the true geochemical variation). For the test data the measurement variance contributes 12.07% (*i.e.*, 11.91 + 0.16) to the total variance, which is well within this ideal value. If the 20% figure were to be exceeded it does not mean that the measurements are un-usable, but rather that particular emphasis must be placed on consideration of the measurement uncertainty in the

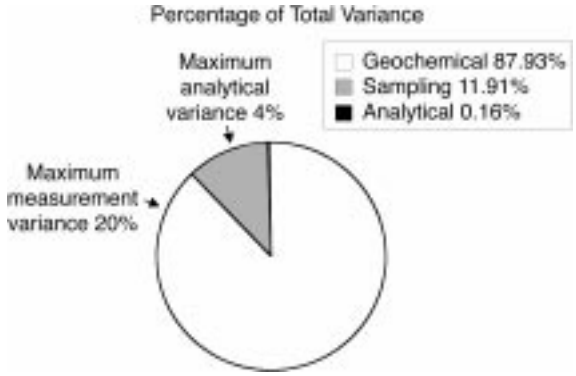

Fig. 4 Diagrammatic representation of the relative contributions of the variance introduced by the primary sampling and chemical analysis to the total variance of the test data set, given numerically in Fig. 3. The measurement variance (sampling + analytical) at 12.07% is well below the suggested ideal maximum of 20% of the total variance,[15] and can therefore be considered fit-for-purpose. The measurement variance and hence uncertainty can be seen to be dominated by the contribution from the sampling (11.91%) rather than from the chemical analysis (0.16%).

interpretation of apparent differences between concentrations at different sampling locations.

The conclusions drawn from the ANOVA are broadly similar to those that can be made from the visual inspection of the raw test data, discussed above, before the application of the computer program.

Where the reduction of the measurement uncertainty is required for future measurements, the information from the ANOVA can be used to identify where the improvements can be made most effectively. If the measurement uncertainty is dominated by the sampling variance, then this is clearly where reductions in variance will produce the greatest improvements overall, because of the summation of squared terms in eqn. (2). This reduction might, for example, be achieved by increasing the mass of the sample by a factor that can be calculated by an equation discussed below. Conversely the dominating measurement variance may come from the chemical analysis, in which case this can usefully be reduced by the selection of a more precise analytical method.

It is possible to modify the experimental design to gain more detailed information on the sources of the uncertainty. This could be implemented by including duplication at another level of the procedure, such as for the instrumental measurements.

There is also a minimum value for the analytical component of 1% of the measurement variance. Below this figure further reductions in analytical uncertainty would not produce any appreciable reductions in the overall measurement uncertainty. Similarly if the total measurement uncertainty accounts for less than 1% of the total variance, then further reductions are unlikely to substantially change the interpretation of the differences between the concentrations at different sampling locations.

These ideal maximum values for the variance proportions are what have become known as 'fitness-for-purpose' criteria. In this case, if the true variation in the analyte is large then the measurement uncertainty required to describe that variation is relatively large. Alternatively, if the true variation in the analyte concentration is small then a much smaller measurement uncertainty will be required. It is also possible to introduce financial considerations in the consideration of fitness-for-purpose. This has been approached using a loss function to apply cost-benefit analysis to the selection of the most appropriate level of measurement uncertainty.[21] In this case the relative expenditure on the sampling and chemical analysis was balanced so that it reflected their respective contribution to the uncertainty. The combined expenditure of the sampling and analysis were also balanced against the financial losses that could arise from excessive uncertainty in the measurement.

LIMITATIONS OF THIS METHOD

The assumptions of classical ANOVA have been summarised[15] as:- 1 That the variances should be independent. An example of this assumption being invalid would be if the presence of a particular mineral in one soil type caused both high sampling variance due to heterogeneity and also high analytical variance due to instrumental effects. 2 Each level of variance should be homogeneous, that is it should not vary systematically within one level. For example, the analytical variance should not vary between different sample types or as a function of concentration. 3 The distribution of errors within each level of variance should be approximately Gaussian.

The first two assumptions may well be invalid to some extent in some instances and should be monitored. The variance of the analyte concentration is assumed not to be a function of concentration. This is known not to be true over large ranges of concentration, but is assumed to be so for the relatively narrow range of values left after the accommodation of high outlying values by the robust ANOVA. The third assumption is very rarely valid in the sampling of natural materials, which often give a few high outlying values. These may be due, for example, to rare particles with high concentrations of the analyte. Log transformation has been suggested to overcome this problem,[14] but the more recent approach applied here has been to use robust statistics. However, the application of robust ANOVA introduces further assumptions. The program is adjusted for a specified maximum incidence of outlying values, in this case 10% of the total population.[15] If there is a higher proportion of outlying values, then this would be expected to lead to somewhat erroneous estimates of the component variances.

Another limitation on the technique is imposed by the number of measurements used for the estimation of the uncertainty. In the example test data, 18 sets of measurements were used, this leads to quite reliable estimates of the component variances. If however only a few locations had duplicate samples then there would be large uncertainties on the estimates of the variance. Whether one variance estimate is significantly greater than another can be evaluated using the *F*-test, with reference to the number of measurements upon which it is based.[14] A practical minimum number of locations at which to take duplicate samples is eight, in order to give a reasonably reliable estimate of the component variances. An 'unbalanced' experimental design has been suggested to make the application of ANOVA more cost-effective.[22]

As was discussed previously, the method described above only estimates the random component of the uncertainty, and does not estimate any systematic errors in the sampling or analysis (*i.e.*, sampling or analytical bias). This limitation can be overcome by the following approaches.

Estimation of systematic errors

Analytical bias can be estimated independently by the use of certified reference materials. Ideally this bias should be estimated with several reference materials covering a range of analyte concentrations, and all with a matrix composition perfectly matched with that of all of the samples. Whether the estimated bias value is statistically different from zero can be checked with a *t*-test against the certified concentration value for each reference material.[23] Alternatively a comparison of the bias detected by several reference materials can be made

as a function of concentration.[24] Analytical bias can also be incorporated automatically, if the uncertainty is estimated by the use of a sampling proficiency test, in which all participants analyse their own samples (discussed below).[5]

Sampling bias can also be estimated automatically within a sampling proficiency test. In this case the systematic error of one participant can be regarded as a random error when viewed across all the participants.[5] For use with the more simple method described here, the sampling bias can be estimated separately in principle by two methods. If one sampler applies a specific protocol to a reference sampling target (RST), then the sampling bias can be estimated by comparison with the certified value of the analyte concentration. This method is directly analogous with the use of reference materials for the estimation of analytical bias,[25,17] but has not yet been tested. It depends on establishing a certified value for the candidate RST by a consensus from an inter-organisational sampling trial, which has to incorporate the inevitable heterogeneity of the sampling target.[4]

A second possible method for the estimation of sampling bias is by the use of the sampling equivalent of the analytical 'spike recovery' test. In this case a synthetic RST is prepared by adding a known concentration of the analyte to the sampling target. It is theoretically possible therefore to estimate the bias of any measurement (sampling + analysis) against the known concentration value rather than against a consensus value. Research into this technique is currently being undertaken by the author and co-workers.

Incorporation of systematic errors into uncertainty estimates

The estimates of the analytical or sampling bias can be added into an estimate of uncertainty, but there is no consensus on the best method to be employed for doing so. The suggestion from ISO[8] is that the estimated bias should be used to correct the concentration estimates and that the uncertainty on the estimated bias should be added to the uncertainty as a variance. As was discussed earlier, this is often impractical for many analytical systems where the bias is either unrecognised or effectively unknown at the particular concentration levels found in the samples. Alternatively it has been suggested that the maximum value of bias found should be considered as the range (r) of a rectangular distribution to be added to the uncertainty. The standard uncertainty of a rectangular distribution[12] is given by $r\sqrt{3}$, and this figure can be added to the other estimates of uncertainty by its variance.[26] A second alternative has also been suggested in which the full value of the best estimate of the bias is added to the estimate of the expanded uncertainty when establishing the range of the uncertainty for comparison with threshold values of concentration.[5]

Alternative options for the estimation of uncertainty

Three other experimental methods have been described for the estimation of measurement uncertainty from sampling, as outlined above.[5] They are progressively more complex to implement than the first method described here, but encompass more sources of uncertainty and should therefore give increasingly better estimates of the overall uncertainty of the measurements made. In the inter-organisational sampling trial (Method 3), ANOVA is also used for the estimation of uncertainty but is based on a different experimental design. Details of this application are described elsewhere[3] but one interesting difference is in questioning the benefits of using robust statistics in this case. If one of the nine participants in an inter-organisational trial produces an outlying value of concentration, then this will be down-weighted in the robust estimate of the uncertainty. It may be however that the concentration reported by this one participant is closer to the 'true' value, by perhaps avoidance of a bias that has affected the other participants. In this case a more reliable estimate of the uncertainty would be obtained by using the classical rather than the robust ANOVA.

Theoretically, it is possible to predict the uncertainty of sampling from a mathematical model such as that of Gy.[27] His equation can be simplified as:-

$$s^2 = Cd^3/m \quad (4)$$

Where s is the standard deviation of sampling error, m is the sample mass and d is the size of the largest particles in the sample. The constant C is the product of four factors relating to the particles being sampled which are (i) a liberation factor (to express how much of the analyte will be recovered from any grain of the sample), (ii) a shape factor (*e.g.*, expressing whether grains are spherical or platy), (iii) a particle size distribution factor, and (iv) a composition factor (expressing the analyte concentration in each grain). A readable explanation of sampling theories that includes Gy's, with worked examples is given by Smith and James.[28]

In practice however, such equations are based on very idealised models of the system being sampled. For example, estimates of the four factors within Gy's constant C are difficult to determine with any degree of certainty and may well vary in different parts of the sampling target. This equation has been used to estimate the sample mass required to achieve a specified sampling error and in this role it has been found to often overestimate the mass required,[28] which usefully gives an uncertainty that is better than required. When applied to estimate uncertainty however, it would require a large amount of effort to estimate all the parameters required and would still be unlikely to provide a reliable estimate of the uncertainty. Experimental work would be required to confirm this conjecture.

A very useful role for such equations however, is making recommendations for changing the uncertainty of sampling, once it has been estimated experimentally. For example eqn. (4) indicates that a doubling of sample mass will reduce measurement variance by a factor of two, and therefore uncertainty by a factor of $\sqrt{2}$. In one sense experimental measurements are being used to estimate the variables in the equation which can then be used to predict changes in the uncertainty. Overall therefore, it seems more reliable to measure uncertainty from sampling experimentally than to predict it theoretically.

A very different approach to estimating uncertainty from sampling comes from the study of geostatistics, using a particular technique called kriging.[29] This statistical technique is primarily aimed at estimating analyte concentrations between the sampling locations. It is based on the characterisation of the change in covariance of measurements made at pairs of sampling locations, with increasing distances of separation. The model fitted to this variation is called a variogram. The intercept of the variogram, called the 'nugget', is the variance remaining at zero distance, and will reflect some random components of the measurement uncertainty. The variogram can be used to estimate not only the concentration of intermediate locations within the sampling area, but also a kriging standard error of estimation, which is also an uncertainty. The essential difference between this uncertainty and those discussed above, is that it refers to concentrations between rather than at the sampling locations. If the method described in this paper is used to improve the estimate of uncertainty at the sampling locations (*e.g.*, by including sampling bias), then kriging can be used to improve the estimates of uncertainties between the sampling locations. In this way the two approaches are complementary in their objectives. The construction of a realistic variogram depends on the collection of a large number of samples, spread evenly across the sampling target. This

contrasts with the basic method described here, which is much more adaptable to surveys that are small or uneven in their design.

CONCLUSIONS

A detailed explanation has been given for the application of ANOVA for the estimation of measurement uncertainty. This uncertainty arises both from the chemical analysis but also from the sampling procedures used to select the material from its primary source (*i.e.*, the sampling target). The computer program and a test data set are available electronically. The technique gives an estimate of both the overall measurement uncertainty and also the separate contributions that arise from the processes of sampling and chemical analysis.

A criterion is described that can be applied to deciding whether a particular level of measurement uncertainty is fit-for-purpose. This is based on the percentage of the total variance contributed by the total measurement uncertainty. The ANOVA result makes it possible to compare the contributions to the uncertainty from the sampling and the chemical and to thereby identify where the uncertainty can most effectively be reduced. Further criteria are suggested for the optimum balance between these two sources of uncertainty. Recommendations for the reduction of uncertainty due to sampling are given based on traditional sampling theory.

The ANOVA method is relatively simple to apply, but has limitations especially in requiring further independent estimates of the uncertainty arising from sampling and analytical bias. Methods for this purpose are referred to as well as more sophisticated methods that automatically include contributions of these biases to the uncertainty.

The assumptions, limitations and relative merits of these methods are discussed, together with other potential methods of estimating uncertainty arising from sampling.

The application of these methods should not be limited to sampling of primary materials prior to chemical analysis, but can also be used with *in situ* analytical techniques where sampling and analysis are indivisible. A recent study of *in situ* analysis using portable XRF has illustrated that although measurement uncertainty for certain applications can be high (*e.g.*, 50%) such measurements are still fit for specified purposes.[30] Similar arguments could be made for uncertainties for techniques such as laser ablation based spectroscopy where uncertainties have again been shown to be large for some applications.[31]

The Robust ANOVA program has been adapted from one original written by Prof. B. D. Ripley, Department of Statistics, University of Oxford, UK.

REFERENCES

1 Analytical Methods Committee, *Analyst*, 1995, **120**, 2303.
2 Ramsey, M. H., Argyraki, A., and Thompson, M., *Analyst*, 1995, **120**, 1353.
3 Ramsey, M. H., Argyraki, A., and Thompson, M., *Analyst*, 1995, **120**, 2309.
4 Argyraki, A., Ramsey, M. H., and Thompson, M., *Analyst*, 1995, **120**, 2799.
5 Ramsey, M. H., and Argyraki, A., *Sci. Total Environ.*, 1997, **198**, 243.
6 Ramsey, M. H., *Analyst*, 1997, **122**, 1255.
7 Thompson, M., *Analyst*, 1995, **120**, 117N.
8 *Guide to the expression of uncertainty in measurement*, ISO, Geneva, 1993.
9 *International vocabulary for basic and general terms in metrology*, ISO, Geneva, 2nd edn., 1993.
10 *Statistics, vocabulary and symbols. Part 1. Probability and general statistical terms*, ISO 3534–1 : 1993 (E/F), ISO, Geneva, 1993.
11 *Terms and definitions used in connection with reference materials*, ISO Guide 30: 1992, ISO, Geneva, 1992.
12 *Quantifying uncertainty in analytical measurement*, Eurochem secretariat, Teddington, Middlesex, UK, 1995.
13 Thompson, M., and Howarth, R. J., *Analyst*, 1976, **101**, 690.
14 Garrett, R. G., *Econ. Geol.*, 1969, **64**, 568.
15 Ramsey, M. H., Thompson, M., and Hale, M. J., *Geochem. Explor.*, 1992, **44**, 23.
16 Ramsey, M. H., Thompson, M., and Banerjee, E. K., *Anal. Proc.*, 1987, **24**, 260.
17 Thompson, M., and Ramsey, M. H., *Analyst*, 1995, **120**, 261.
18 Miesch, A. T., *U.S. Geological Survey Prof.*, 1976, Paper 954–A.
19 Ramsey, M. H., *Appl. Geochem.*, 1993, **2**, 149.
20 Analytical Methods Committee, *Analyst*, 1989, **114**, 1693.
21 Thompson, M., and Fearn, T., *Analyst*, 1996, **121**, 275.
22 Garrett, R. G., and Goss, T. I., *Comput. Geosci.*, 1980, **6**, 35.
23 Ramsey, M. H., in *Modern Analytical Geochemistry*, ed. Gill, R., Longman, Harlow, UK, 1997, ch. 2, p. 17.
24 Ramsey, M. H., and Thompson, M., *J. Anal. At. Spectrom.*, 1987, **2**, 497.
25 Ramsey, M. H., *Sampling of environmental materials for trace analysis*. ed. Markert, B., VCH, Weinheim, 1994, p. 106.
26 Kane, J. S., *Analyst*, 1997, **122**, 1283.
27 Gy, P., *Sampling of particulate materials—theory and practice*, Elsevier, Amsterdam, 1979.
28 Smith, R., and James, G. V., *The Sampling of Bulk Materials*, Royal Society of Chemistry, London, 1981.
29 Myers, J. C., *Geostatistical error management: quantifying uncertainty for environmental sampling and mapping*, 1997, Van Nostrand Reinhold, New York.
30 Argyraki, A., Ramsey, M. H., and Potts, P. J., *Analyst*, 1997, **122**, 743.
31 Ramsey, M. H., Coles, B. J., Wilkinson, J. J., and Rankin, A. H., *J. Anal. At. Spectrom.*, 1992, **7**, 587.

Paper 7/06815H
Received September 19, 1997
Accepted December 1, 1997