ISSN 1757- 5958

amc technical briefs

Editor: Michael Thompson Analytical Methods Committee AMCTB 32 July 2008

Optimising your uncertainty—a case study

In analytical measurement, both the sampling procedure and the analytical procedure contribute to the uncertainty of the result. The customer or other end users usually wish to minimise the long term average costs when they make decisions based on the result. How can we ensure that the combined uncertainty obtained is best for the customer's purposes? This Brief describes how it can be done, with the help of an example, the determination of the nitrate content of lettuce

The uncertainty from sampling and analysis

End users of analytical data need to know the composition of a defined portion of material, called the 'target' in sampling terminology. They need this information to support a decision about the target, such as its commercial value or whether it complies with a legal or contractual specification about its composition. The end user needs both the analytical result and an estimate of its uncertainty to support such a decision. A result with a large uncertainty may give rise to different decisions from an identical result with a small uncertainty.

Nearly always we need to take a small sample from the much larger target and conduct the analysis on the sample. It would be helpful if the sample had the same mean composition as the target but, despite our best efforts, it never does. This is because targets are nearly always heterogeneous and sampling methods always imperfect. As an outcome, successive samples from the same target differ in composition, from each other as well as from the target. These differences give rise to uncertainty from sampling u_S.

What the end user needs

After analysis of the sample*, we have a result x with analytical uncertainty u_A. But this result, $x \pm u_A$ applies only to the composition of the sample, not that of the target. It ignores sampling uncertainty u_S. We need to take account of the combined uncertainty $u = \sqrt{u_S^2 + u_A^2}$. The appropriate result for the composition of the target is thus $x \pm u$, and that is the information that the end user should use to support a decision about the target.

But the end user also needs good value for money. Sampling and analysis both involve costs, and quite generally the cost increases as the uncertainty gets smaller. Furthermore, it is important to recognise that additional costs may be incurred by the end user as the outcome of the occasional incorrect decision. Incorrect decisions become more likely, and in some instances more costly as well, as the uncertainty increases. Add these two cost functions together, and the long term costs show a minimum at a particular level of uncertainty u_f (Figure 1). That minimum cost is the ideal and has been used as a definition for the otherwise vague idea of fitness for purpose.

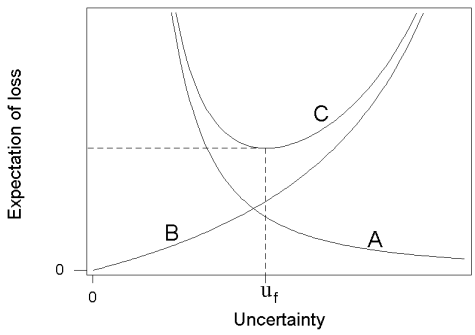

Figure 1. Schematic diagram of costs versus uncertainty. Line A shows the cost of measurement, line B the cost of incorrect decision, and line C the sum of the two costs. The uncertainty u_f at the minimum cost is regarded as fit for purpose.

* For simplicity here, we regard the 'sampling' as including all operations up to and including the preparation of the 'test sample', that is, the laboratory sample in a form ready for the analyst to measure out test portions for analysis. 'Analysis' covers all subsequent operations including taking the test portion. Other divisions of the sequence of operations are possible.

Background to the lettuce study

Leafy green vegetables provide a major source of nitrate to the consumer. EU Regulation 1822/2005 sets a maximum level for nitrate in lettuce and spinach of 4500 mg kg^{-1} and requires that member states carry out appropriate monitoring. Recommended methods of analysis and sampling are available for the determination of nitrate, but information on the combined uncertainty attached to the results was lacking before this study.

Preliminary investigation of uncertainty

The sampling targets in the study were individual bays of lettuce in a large greenhouse, each containing up to 12 000 heads. The original procedure specified that primary samples should comprise ten heads taken in a predetermined pattern from each bay. For the validation study, the sampling was duplicated in a randomised way. The primary samples were processed in the laboratory by splitting the heads and blending them. The test sample was thus a puree. Two test portions were taken from the test sample for duplicate analysis by HPLC. No bias was identified in the analytical results. The design of the experiment is shown in Figure 2, and the results in Box 1.

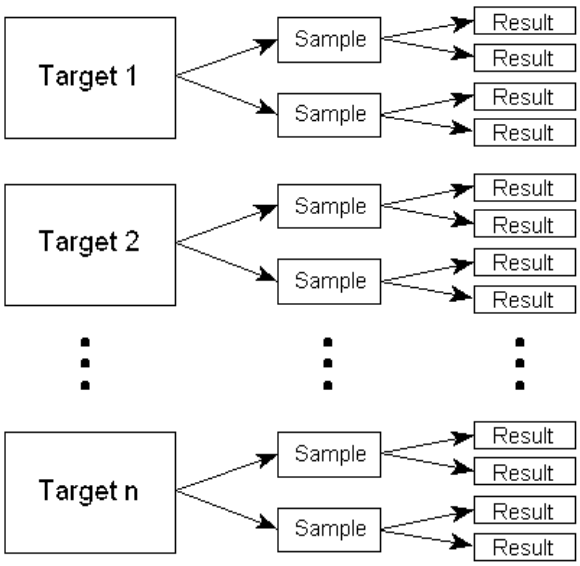

Figure 2. Experimental design for estimating variance components. In the example the targets were eight bays of lettuce.

Box 1. The validation results in ppm NO_3 (mass fraction × 10^6) were as follows.

Mean result:	4408
Analytical repeatability SD u_S :	168
Sampling SD u_A :	319
Combined standard uncertainty:	360.5

Theory part 1: modifying the uncertainties and its effect on measurement costs

A common method of sampling (as in the lettuce study) is to collect a number of small 'increments' from random positions within the target and combine them to form the primary sample (a so-called aggregate or composite sample). Suppose the recommended sampling procedure specifies that *n* increments are taken giving rise to an uncertainty u_S at a cost l_S. We could modify the uncertainty obtained by adjusting the number of increments. If we wanted to reduce the uncertainty to half of its previous value, that is, to $u_S/2$, sampling theory shows that we would need to combine 4*n* increments. The cost[†] of sampling would therefore increase fourfold to $4l_S$. Generally, the cost L_S for sampling the target with any specified uncertainty w_S would be

$$\text{Eq 1} \qquad L_S = l_S\left(\frac{u_S}{w_S}\right)^2 .$$

We can postulate an analogous general inverse square relationship for the cost L_A of analysis[‡], when a validated analytical method provides an uncertainty u_A at a cost l_A. This gives

$$\text{Eq 2} \qquad L_A = l_A\left(\frac{u_A}{w_A}\right)^2 .$$

From Eqs 1and 2, we have the combined uncertainty of

$$\text{Eq 3} \qquad u = \sqrt{w_S^2 + w_A^2}$$

at a total cost for the measurement of

$$\text{Eq 4} \qquad L_m = L_S + L_A = l_S\left(\frac{u_S}{w_S}\right)^2 + l_A\left(\frac{u_A}{w_A}\right)^2 .$$

For any required combined uncertainty u at least cost, we need to find the minimal value of L_m in Eq 4 that is consistent with the constraint of Eq 3. This requires the use of a mathematical tool called the Lagrange multiplier. The operation of the Lagrange multiplier is illustrated in Figure 3. After some algebra (and only minor brain damage) we find that the minimal measurement cost is

[†] This ignores any fixed 'overhead' costs, such as that of travelling to the target.

[‡] This is not an exact analogy with sampling, as the replicated analytical results would have to be obtained under reproducibility conditions to approximate uncertainty.

Eq 5 $$L_m = \left(u_S\sqrt{l_S} + u_A\sqrt{l_A}\right)^2 \Big/ u^2 ,$$

This minimal cost is obtained when

Eq 6 $$\frac{w_S}{w_A} = \left(\frac{l_S u_S^2}{l_A u_A^2}\right)^{1/4} .$$

Eq 6 is in itself a useful result. The fourth root tells us that the optimum value for w_S/w_A will be close to unity unless the combination of the costs and the uncertainties of the standard methods are wildly discrepant. For example, if $l_S u_S^2 / l_A u_A^2 = 100$, then w_S/w_A is still only 3.2.

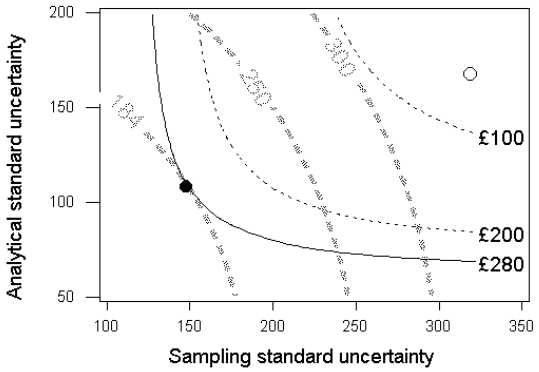

Figure 3. Contour lines of cost (black) and combined uncertainty (grey). The two lines showing the optimal values share a common tangent at the optimum point (solid circle). The open circle shows the uncertainty from the initial 'validation', where costs are low but combined uncertainty unsuitably high.

Theory part 2: optimising the combined uncertainty

Eq 5 shows how the cost of a measurement depends on *u*, but we also need to consider the cost of mistaken decisions. Here we shall consider only the cost of false rejection—when the end user decides to discard (or rework) the target, at a loss of L_f when it is actually satisfactory. The average loss per target is L_f multiplied by the probability of false rejection.

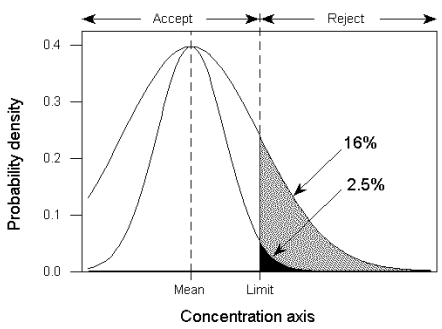

Figure 4. Probability of false rejection with a particular mean concentration and rejection limit, with different degrees of uncertainty. The shaded areas (grey or black) show the probabilities of rejection.

When rejection is based on a legal or other limit, we can calculate this probability by assuming a frequency distribution of the results. A normal distribution is suitable in the present case. Figure 4 shows an instance where the concentration of analyte in the target is below the limit but because of uncertainty there is a substantial probability of obtaining a result above the limit. We see that the probability of rejection is increased when the uncertainty is increased. These probabilities can be calculated for a given distribution.

The total average cost for the measurement and the subsequent decision is therefore $L = L_m + L_f$.

Back to the example

Recommended methods of sampling and analysis were put through a validation procedure involving duplicate sampling from eight targets (bays of lettuce) and duplicate analysis of each sample (Figure 2 and Box 1). Variance components calculated by ANOVA gave a sampling uncertainty of $u_S = 319$ at a cost of l_S = £40, and an analytical uncertainty $u_A = 168$, coincidentally at the same cost of l_A = £40. The cost of the measurement was therefore £80, and the combined uncertainty was $\sqrt{319^2 + 168^2} = 361$. We see from Eq 6 that in this instance the optimal ratio of the uncertainty components is

$$\frac{w_S}{w_A} = \left(\frac{40 \times 319^2}{40 \times 168^2}\right)^{1/4} = 1.38 .$$

Finding the fit-for-purpose uncertainty

For the present study, only false non-compliance is considered. A non-compliance, false or otherwise, would result in the rejection of a batch of lettuce with a loss to the producer of £5280. (False compliance could also result in a loss, but is more considerably more difficult to assess and is omitted here.)

For a nitrate concentration of 4408 ppm and a acceptable maximum limit of 4780 ppm, we need to find the minimum value of

$$L = ((319 + 168)\sqrt{40})^2 / u^2 + 5280\, p(X > 4780) ,$$

with the probability $p(X > 4780)$ calculated when X is a random result from a normal distribution with mean 4408 and standard deviation *u*. (There are commands in statistical software, including Excel, that can be used to calculate this probability.) A simple way to find the minimum is to evaluate *L* at closely-spaced values of *u* and plot the results with the points joined. Figure 5 shows the outcome with a minimum average cost per batch of £394 at a combined uncertainty of 184 ppm. With the original sampling and analytical methods giving an uncertainty of 319 ppm, the grower would face an average loss per batch of £740.

As we know that the required combined uncertainty $u = \sqrt{w_S^2 + w_A^2} = 184$, and $w_S / w_A = 1.38$, solving the two simultaneous equations gives $w_S = 148$, $w_A = 108$. Thus the sampling uncertainty provided by the original method would ideally need to be reduced by 54%. This could be achieved simply by taking more increments per sample. The corresponding reduction for the analytical uncertainty would be 36%. From Eq 5 we see that the optimised combined uncertainty would impose a cost for the complete measurement of $\left(319 \times \sqrt{40} + 168 \times \sqrt{40}\right)^2 / 184^2 = £280$. This is a large increase over the cost of the original measurement. There are other fit-for-purpose combinations of uncertainties, as can be seen in Figure 1, but all would cost even more. Whatever the eventual choice, in the long term the grower could save money by spending more on the measurement, thereby reducing the proportion of incorrectly condemned batches of lettuce.

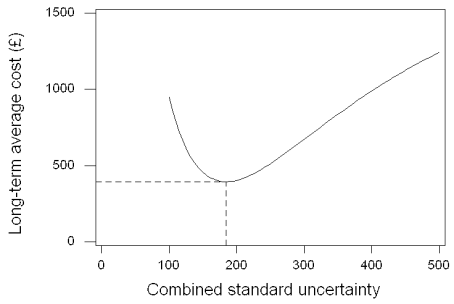

Figure 5. The long term average cost of an accept/reject decision as a function of combined uncertainty.

Assessment of the optimisation method

All right—we admit it! The lettuce example, although real, was simplified to demonstrate the principle of uncertainty optimisation in action. Look at the following potential complications that were not considered here.

- The simple functions (Eqs 1, 2) giving the costs of sampling and analysis in relation to uncertainty may be wrong in some circumstances. For example, if the sampler has to travel to the other end of the country to collect a sample comprising *n* increments, it would probably cost no more to collect an alternative comprising 4*n* increments.
- It may be impossible to modify the original analytical method to obtain the indicated uncertainty. We may have to accept the closest available method.
- We have taken no account of the potential costs of a false acceptance of a batch. This is more difficult to do because it involves the perhaps small probability of detection by external agencies, and the probably very large but hard-to-assess financial penalty contingent on such detection.
- The optimisation was carried out at a single concentration of the analyte (the mean result from the preliminary validation), but we would obviously get a different solution at other concentrations present in different batches. Moreover, the combined uncertainty may vary with concentration. The optimisation should have taken these extra sources of variation into account.

All of the above difficulties can be overcome, although that may require the services of an expert. The principle, however, is easy to understand and appropriate for a wide range of analytical applications.

We seem to have the tools to determine the long-term optimum uncertainty for a particular application**.** Having this knowledge is clearly worthwhile financially. However, whether or not to adopt the optimal test conditions is a commercial decision—for whatever reason, the end user might prefer to minimise short-term losses.

Further reading

Measurement uncertainty arising from sampling—a guide to methods and approaches (2007) 102pp. The Guide, written under the Chairmanship of Prof M H Ramsey, is the joint production of Eurachem, CITAC, Eurolab, Nordtest and the Analytical Methods Committee. It contains chapters on fundamental concepts, estimation of sampling uncertainty, and management issues. Six practical examples are examined in detail. Download gratis from the Eurachem website www.eurachem.org/guides

This Technical Brief was produced for the Analytical Methods Committee by the Subcommittee for Uncertainty from Sampling under the chairmanship of Prof M H Ramsey.

