Notes on sampling in chemical measurement

Sampling in context

The topic of uncertainty from sampling for chemical measurements has long been neglected but in many other application sectors, sampling uncertainty is a substantial or even dominant term in the uncertainty budget and therefore it is very important and hard to be ignored.

It is a fact that the end-user of analytical results needs to know the combined uncertainty (analytical and sampling) to make valid decisions about the sampling “target”. These decisions can be the commercial value of a batch or lot of a material, or whether material conforms with a legal or contractual specification.

Nearly all analysis is preceded by sampling, the process of taking a small portion (the sample) from the much larger amount (the target), the composition of which is of interest.

Normally such sample is small enough to be sent to the laboratory for further physical preparation such as drying and grinding before analysis. In the laboratory, we start the analytical process by first taking a subsample from the sample received and preparing it as a test sample for chemical treatment preceding to measurement.

Of course, taking a sample is pointless unless it reasonably approximates the average composition of the target; such a sample is said to be “representative”.

We know obtaining representative samples is sometimes very difficult, especially when the target is very large and heterogeneous; a shipload of mineral ore for example, and parts of the target may be difficult to access. Hence, sampling introduces error into the final test result. Heterogeneity in the target can make samples from the same target differ greatly in composition.

However, it is quite a standard practice for analytical chemists to accept the samples drawn by others without further queries on how these samples have been drawn and whether the sampling procedure applied to the sampling target (statistically called the *population*) is appropriate. They assume the sampling uncertainty is negligible which is actually not true.

Laboratory analysts should realize that any residual heterogeneity in the laboratory sample received gives rise to an uncertainty that is attributed to the analytical variation. If such a sampling uncertainty is significant, it should form an uncertainty budget and be included in the evaluation of analytical uncertainty.

Today, the revised ISO/IEC 17025:2017 accreditation standards regards sampling as an integral part of the measurement process and includes its contribution to the combined uncertainty. This new requirement has put accredited laboratories on notice that sampling uncertainty can no longer be neglected.

Sampling uncertainty in context

All sampling targets are actually or potentially heterogeneous: the chemical composition can vary from point to point in the material. This implies that replicate test samples from a single target will differ in composition from each other and from the target.

Such variation gives rise to uncertainty from sampling u_s, that is additional to and independent of the u_a, derived from purely analytical activities. The combined measurement uncertainty on the composition of the target is thus:

$$u = \sqrt{u_s^2 + u_a^2}$$

This combined uncertainty is relevant to the needs of the end-user of the data who is required to make rational decisions about the target (not about the laboratory sample).

Indeed, sampling uncertainty can make a substantial contribution to the combined uncertainty in the environmental studies and in the examination of raw materials such as food or mineral ores.

So, to create a conceptual framework for sampling uncertainty, we must consider ideas like precision and bias applied to sampling, and carry out operations such as the validation of methods and quality control, like what have been done in analytical works. But, there are differences in the way that sampling uncertainty can be tackled. This is because sampling uncertainty is partly the outcome of the heterogeneity of the material under test as well as the process of collecting it.

A working value of u_s will have to be a robust average that is typical of the material as a whole. This potential variation in the degree of heterogeneity implies that quality control is especially important in sampling. An analytical result on a sample from a target with an anomalously high value of u_s could be unfit for purpose, even though the validated sampling protocol was scrupulously followed.

Statistical tools for sampling uncertainty

Randomization is the keyword in taking samples from the lot (target) and analyzing them in the laboratory in random order.

Secondly a properly designed sampling protocols allows the analysis of variance (ANOVA) be applied to the measurement results to give useful estimates of the sampling standard uncertainty expressed as standard deviation. ANOVA is used to estimate sampling and analytical variance.

We can assume from the start that the samples can be different form each other, therefore we are not really interested in a significance test.