Estimating uncertainties of analytical results using information from the validation process

Alicia Maroto, Jordi Riu, Ricard Boqué and F. Xavier Rius

Department of Analytical and Organic Chemistry. Rovira i Virgili University.
Pl. Imperial Tàrraco, 1. 43005-Tarragona. Spain.

19 Pages, 2 Tables and 5 Figures

Author for correspondence:

F. Xavier Rius
Department of Analytical and Organic Chemistry,
Universitat Rovira i Virgili
Pl. Imperial Tàrraco, 1
E - 43005 TARRAGONA
SPAIN

Tel.: +34 - 77 - 559562
Fax.: +34 - 77 - 559563
e-mail: rius@quimica.urv.es

ABSTRACT

A new approach for calculating uncertainties of analytical results based on the information from the validation process is proposed. This approach complements the existing approaches proposed to date and can be applied to any validated analytical method. The precision estimates generated during the process of assessment of the accuracy take into account the uncertainties of preprocessing steps and analytical measurement steps as long as the different factors that influence these steps are representatively varied in the whole validation process. Since the accuracy of an analytical method should be always assessed before applying it to future working samples, little extra work needs to be done to estimate the final uncertainty. Other sources of uncertainty not previously considered (e.g. uncertainty associated to sampling, to differences between the test-sample and the working sample, etc.) are subsequently included and mathematically combined with the uncertainty arising from the assessment of the accuracy to provide the overall uncertainty. These ideas are illustrated with a case study of the determination of copper in contaminated land.

INTRODUCTION

These days analysts are increasingly being urged to ensure that their results are both reliable and comparable. The estimation of uncertainty is therefore one of the main focuses of interest in the field of measurement chemistry [1]. The main reason for this is that the analytical result is a piece of information which is used for various purposes. Important decisions are often based on these results, so both the traceability and the degree of confidence (i.e. the uncertainty), given as parameters that define the quality of these results, must be demonstrated.

The International Vocabulary of basic and general terms in Metrology, VIM, states in its latest definition that uncertainty is "a parameter, associated with the result of a measurement, that characterises the dispersion of the values that could reasonably be attributed to the measurand" [2]. This definition must be interpreted if a link between traceability and uncertainty is to be established. If traceability has not been assessed against a high quality reference in metrological terms, the correctness of all the possible systematic effects of the measurements cannot be guaranteed, and it is therefore impossible to ensure that the confidence interval comprises the true value. So, if the quantitative analytical result is expressed as Value 1 $\pm$ Uncertainty, every analyst should verify that Value 1 is an unbiased value which is traceable, whenever possible, to an internationally recognised reference.

The main approaches for calculating uncertainty which have been proposed to date are the ISO approach (commonly known as "*bottom-up*") and the Analytical Methods Committee approach (commonly known as "*top-down*"). The ISO approach was originally proposed for quantifying uncertainty in physical measurements [3] and was subsequently adapted by EURACHEM [4] for chemical measurements. It is based on identifying, quantifying and combining all the sources of uncertainty of the measurement. Although the ISO approach improves knowledge of the measurement procedure, the difficulty of applying it in chemical measurements hampers its widespread use. On the other hand, the approach proposed by the Analytical Methods Committee [5] is based on information gathered from interlaboratory exercises. In this approach the laboratory is seen from a higher level (i.e. as a member of the population of laboratories). Consequently, systematic and random errors within individual laboratories become random errors when they are considered from this "higher level". The "*top-down*" approach takes into account factors of the measurement that are intrinsically

impossible to know; however, the lack of interlaboratory information for many measurands in a diversity of matrices and the fact that a collaborative trial carried out in the past may not be related to the present conditions in the laboratory often means that it is difficult to use.

To complement the existing approaches for calculating uncertainties, we propose a new procedure that takes advantage of the precision estimates generated in assessing the accuracy (trueness and precision) of the analytical methods. Since the accuracy of all the methods should always be assessed before applying them to future working samples, this approach means that little extra work needs to be done when calculating the uncertainty of the final result. Following the error propagation procedure, we combine the uncertainty arising from the assessment of accuracy with other components not previously considered, such as the uncertainty arising from sampling or lack of homogeneity, to obtain the overall uncertainty.

ASSESSMENT OF ACCURACY

The accuracy of the measurement methods should always be assessed before applying them to future working samples. Verifying accuracy is one of the basic steps in the process of achieving traceability. The international concept of traceability [2] and its interpretation in chemical analysis [6-10] can be applied to chemical measurement processes by comparing the measurement method statistically to an appropriate reference. The measurement method, which will be applied to future samples to be analysed in the same conditions, is therefore compared for precision and trueness to the reference results.

Figure 1 compares the results obtained with the measurement method to two different types of references: a reference measurement method and a certified reference material (CRM). However, other references, whose metrological importance is shown in Figure 2, could also be used.

Figure 1a compares the accuracy of the tested method B against a reference method A. This is a common situation in many laboratories, where an alternative method has been developed to substitute a reference method in routine analysis. One of the most advanced guidelines for comparing two methods can be found in ISO 5725-6 [11]. However, it is not applicable in a single laboratory because it is based on interlaboratory information. Therefore, a recently developed procedure

will be used [12]. This is based on validating a method by generating different precision estimates in a laboratory. If the accuracy of method B has to be assessed against reference method A in a single laboratory, the experimental design should take into account all the possible factors of variation. The experimenter must decide how many factors are to be varied representatively and how many analyses are to be carried out. The number of factors depends on the resources (number of analysts, instruments, time, etc.) required by the measurement process and the experimenter's knowledge of the method. Moreover, the number of analyses to be made should be chosen by considering the minimum bias and the minimum ratio of precision measures that the analyst wishes to detect and the α probability of falsely concluding the presence of bias and the β probability of falsely concluding the absence of bias. The sample to be analysed (i.e. the test sample), should be homogeneous, stable and as representative as possible of the samples to be subsequently analysed.

The experimental design proposed to vary the chosen factors representatively is a fully-nested design [13]. As stated in [13], the factors affected mainly by systematic effects must be in the highest ranks, while the factors affected mainly by random effects must be in the lowest ranks. For instance, if the factors studied are the repeatability and the run on which the measurement is performed, the runs will be in a higher rank than the repeatability because the repeatability is less affected by systematic errors**.** Moreover, in this case nesting the run-effect in the replicates is not possible. A fully-nested design which takes into account the variation of the runs and the replicates is shown in Figure 3.

The proposed experimental design provides precision estimates of the different factors studied. Following the proposed two-factor fully-nested design, the precision estimates generated are the repeatability standard deviation, s_r, the between-run standard deviation, s_{run}, the run-different intermediate standard deviation, $s_{I(run)}$, and the standard deviation of the means, $s_{\bar{x}}$. The typical analysis of variance (ANOVA) table and the calculation of the different variance terms is shown in Table 1. All the important sources of variation of the measurement result (e.g. the analyst, the calibration, the day, etc.) have been varied between each run. The run-different intermediate standard deviation can therefore be considered as an estimate of the [time + analyst + calibration + …]-different intermediate precision [13]. As a result, the run-different intermediate standard deviation includes all the important factors of variation of the measurement result. However, it must be pointed out that the estimation of the between-run standard deviation may not be

as good as the one that would have been obtained if a fully-nested design had been carried out by including all the important sources of variation like time, analyst, calibration, etc. as factors in the experimental design. However, in this latter case, the cost in terms of work increases considerably.

The precision estimates and the grand means obtained for both methods are then used to assess the accuracy of method B. It is important to note that the different runs within which the results of both methods have been obtained need not (necessarily) be the same (i.e. laboratory A obtained the results in 10 consecutive days while laboratory B did it in 20 days, because they analysed a sample every two days).

Before comparing the two methods, the consistency of the individual results should be checked by graphical analysis and/or tests for outliers [14]. Moreover, the normality of the measurement results should always be checked.

To assess the precision of method B, all the intermediate precision estimates must be compared statistically with those of the reference method by performing an *F*-test. For instance, when the runs and the repeatability are the factors varied, the repeatability variance and the run-different intermediate variance of both methods must be compared. Both methods should be as precise as possible because the probability of falsely concluding the trueness of method B against method A increases when the precision of both methods diminishes.

Trueness can be assessed by checking whether the difference between the grand mean for reference method A, $\overline{\overline{x}}_A$, and the grand mean of method B, $\overline{\overline{x}}_B$, is not statistically significant:

$$\left|\overline{\overline{x}}_A - \overline{\overline{x}}_B\right| \leq t_{\alpha/2} \cdot s_d \qquad (1)$$

where s_d depends on the variance of the means of both methods. Therefore, an *F*-test is performed to compare the variances of the means of both methods, $s^2_{\bar{x}_A}$ and $s^2_{\bar{x}_B}$. If the difference between these variances is not significant, s_d is given by equation 2a:

$$s_d = \sqrt{\left(\frac{(p_A - 1)\cdot {s_{\bar{X}_A}}^2 + (p_B - 1)\cdot {s_{\bar{X}_B}}^2}{p_A + p_B - 2}\right)\left(\frac{1}{p_A} + \frac{1}{p_B}\right)} \qquad (2a)$$

If the difference between the variances of the means is statistically significant, s_d is given by equation 2b:

$$s_d = \sqrt{\frac{{s_{\bar{X}_A}}^2}{p_A} + \frac{{s_{\bar{X}_B}}^2}{p_B}} \qquad (2b)$$

If the fully-nested design of figure 3 is followed, p corresponds to the number of runs on which the analysis is carried out. In this case, the variance of the means, $s_{\bar{x}}^2$ is calculated as:

$$s_{\bar{x}}^2 = \frac{\sum_{i=1}^{p}\left(\bar{x}_i - \bar{\bar{x}}\right)^2}{p - 1} \qquad (3)$$

where $\bar{x}_i$ is the mean of the replicate measurements carried out in the run *i*.

If the difference between the grand mean of reference method A and the grand mean of method B is not significant, the trueness of method B against method A is demonstrated, but only when the factors considered in the experimental design are varied. Therefore, method B is traceable to method A in the conditions of variation studied.

The accuracy of a method can be checked by means other than just a reference method, as Figure 2 shows. For instance, in Figure 1b the process of comparing method B against a CRM is shown. In this case, the CRM is analysed with method B by varying all the factors that can affect the measurement method representatively. Finally, the results obtained, $\bar{\bar{x}}_B$ and $s_{\bar{X}_B}$ are compared statistically to the certified reference values, $\bar{\bar{x}}_A$ and $s_{\bar{X}_A}$. If the reference used in assessing accuracy has a low level of traceability, higher levels can be achieved by linking this reference to a recognised interlaboratory exercise, to a primary method, to a CRM or, in selected cases, to the SI units. This latter path is still not established in many practical instances, however. Moreover, as the number of steps in the traceability chain increases, the measurement result becomes less

precise due to the addition of uncertainty terms and the propagation of errors. It is always advisable, therefore, to trace the procedure to the highest available reference. All future analyses using tested method B in conditions of routine work will be traceable to the reference used in assessing accuracy as long as the working samples are similar to the one used in the internal validation, and the quality assurance conditions are implemented in the laboratory effectively.

CALCULATION OF UNCERTAINTY

While verifying the accuracy of the measurement, the analyst generates information about different intermediate precision estimates. Since these estimates have been obtained by varying the factors that affect the measurement result representatively, they take into account the uncertainties of different steps in the analytical process. For instance, uncertainties due to the separation and preconcentration of the analyte of interest, matrix effects, environmental conditions, instruments or operators are considered as long as the factors that influence them are varied representatively. However, there may be other sources of variation which have not been considered in the assessment of accuracy. These sources may be related to global factors or local factors.

The *global factors* are inherent to the measurement process. Therefore, these sources of uncertainty must always be considered in the overall uncertainty. They may be related to uncertainties not taken into account while assessing the accuracy or to the bias of the reference (estimated by linking this reference to another one with higher levels of traceability). For instance, if the reagent's suppliers had not been representatively varied while assessing accuracy, possible future changes in the supplier would introduce a new source of uncertainty that should subsequently be considered in the overall uncertainty of all the samples to be analysed in the future. The uncertainty arising from laboratory bias or from method bias should also be taken into account when the higher level of traceability achieved corresponds to a reference method carried out in the same laboratory as the method to be tested. When the higher level is related to a CRM, the source of uncertainty that should be considered is the uncertainty arising from the bias of this reference material. Finally, all the sources of uncertainty arising from the different links of the established chain of traceability must also be considered a component of uncertainty in all the samples to be analysed in the future.

The *local factors* are characteristic of the sample to be analysed. These factors therefore vary according to the type of sample analysed. One of the main sources of uncertainty related to local factors are the differences in sampling and pretreatment procedures between the test sample and the working sample. Other possible sources of uncertainty are the differences in the measurand value or in the matrix of the test sample and the working sample.

All the different uncertainty terms previously mentioned must be calculated or estimated and mathematically combined to obtain the overall uncertainty value of the analytical result. This process is illustrated in Figure 4. In this scheme, the uncertainty components have been grouped into four terms:

1- Uncertainty arising from the analytical procedure

The uncertainty which arises from the experimental variation when applying the analytical procedure to obtain future measurement results is associated to the different values that the run bias and the random error of method B can take on within the laboratory. If every sample is analysed in p_s different runs and, in each run n_s replicates are carried out, an estimation of the uncertainty of the grand mean, $\bar{\bar{x}}_s$, of these measurements is given by the between-run variance, the repeatability variance and the number of runs and replicates carried out within each run.

$$u_{proc} = \sqrt{\frac{s_{run}^2}{p_s} + \frac{s_r^2}{n_s}} \qquad (4)$$

However, the result of a future working sample is seldom obtained as the grand mean of measurement results performed in p_s different runs and with n_s replicates in each run. Normally, the result is obtained with only one analysis or at most as the mean of three replicates. In this case, information about the between-run variance and repeatability variance cannot be obtained because of the small number of replicates and runs carried out. However, the between-run variance and repeatability variance obtained in assessing the accuracy of method B can be used as estimates of the s_{run}^2 and s_r^2 related to the working sample. This can be done whenever the working sample is similar to the test sample in the assessment of accuracy and the quality assurance conditions are effectively implemented in the

laboratory. Therefore, the uncertainty of the procedure can be calculated with equation 4.

In the most common case i.e. n_s=1 and p_s=1, equation 4 becomes:

$$u_{\text{proc}} = \sqrt{s_{\text{run}}^2 + s_{\text{r}}^2} = \sqrt{\frac{MS_{\text{RUN}} - MS_{\text{E}}}{n} + MS_{\text{E}}} = \sqrt{\frac{MS_{\text{RUN}}}{n} + \left(1 - \frac{1}{n}\right) \cdot MS_{\text{E}}} \qquad (5)$$

and its associated number of effective degrees of freedom, ν_{proc}, are estimated with the Welch-Satterthwaite approach [16]

$$\nu_{\text{proc}} = \frac{u_{\text{proc}}^4}{\dfrac{\left(MS_{\text{RUN}} / n\right)^2}{p-1} + \dfrac{\left(\left(1 - \frac{1}{n}\right) MS_{\text{E}}\right)^2}{p \cdot (n-1)}} \qquad (6)$$

2- Uncertainty arising from the assessment of accuracy

In assessing accuracy, what is checked is whether the laboratory and/or the method bias is statistically significant. The type of bias checked depends on the metrological quality of the reference used. If the bias is not statistically significant, the method is traceable to the reference but there still remains a source of uncertainty in the estimation of this bias. If the reference used is a reference method A, the estimated bias of method B is calculated from the difference between the grand means obtained from both methods when analysing the same test sample:

$$\hat{\delta}_{\text{method}} = \bar{\bar{x}}_{\text{B}} - \bar{\bar{x}}_{\text{A}} \qquad (7)$$

The standard uncertainty associated to the estimated method bias is given by s_d.

$$u_{\text{assessment}} = s_d \qquad (8)$$

This standard deviation is estimated from the variance of the method to be tested and from the variance of the reference (if available) used in the assessment of accuracy. If the difference between variances is statistically not significant, s_d is

calculated using equation 2a with p_A+p_B-2 degrees of freedom. If the difference between variances is statistically significant, s_d is estimated by equation 2b. Since in this case s_d is a compound variance, the degrees of freedom are obtained by applying the Welch-Satterthwaite approach:

$$\nu_{\text{assessment}} = \frac{s_d^4}{\frac{\left(s_{\bar{x}_A}^2 / p_A\right)^2}{p_A - 1} + \frac{\left(s_{\bar{x}_B}^2 / p_B\right)^2}{p_B - 1}} \tag{9}$$

3- Uncertainty arising from preprocessing steps

This component is necessary when the sources of uncertainty arise from subsampling and from preprocessing steps applied to future working samples that have not been considered in the process of the assessment of accuracy. Field sampling uncertainty is not considered here because the uncertainty estimated normally corresponds to the sample that has reached the laboratory once the field sampling has been performed. However, the component of uncertainty due to field sampling is usually very high and exceeds that from other steps, so it should be taken into account whenever the laboratory has the responsibility for sampling. To estimate this uncertainty is rather complex but it can be done by collaborative trials in which all the factors that can affect sampling (protocol, sampler, replicate, etc.) are varied. This subject has been extensively treated by Ramsey *et al*. [15]

The uncertainty arising from subsampling i.e. from the heterogeneity of the material, has not been considered when assessing the accuracy. In this process, the test sample or reference material must be homogeneous by definition. However, future working samples are not expected to be homogeneous, and so the uncertainty arising from the heterogeneity of the sample once in the laboratory must be included in this term. The uncertainty arising from other preprocessing steps (filtration, weighing, drying, etc.) not carried out on the test sample used in assessing the accuracy must also be included in this term. The uncertainty arising from heterogeneity and/or preprocessing steps can be estimated from a sample with the same characteristics as the working samples to be analysed subsequently. The subsampling and/or the preprocessing steps must be carried out by representatively changing all the factors that can affect them (e.g. operator, material, time, etc.). Then the analytical results of each portion of the sample

subsampled and/or pretreated are obtained under repeatability conditions. If only one individual measurement is taken, the contribution of the analytical measurement repeatability can be subtracted by using the information of the repeatability variance from the assessment of the accuracy,

$$s^2_{\text{preproc.steps}} = s^2_{\text{I(preproc.steps)}} - s^2_{\text{r}} \quad (10)$$

where $s^2_{\text{I(preproc.steps)}}$ is the preprocessing step-different intermediate variance and $s^2_{\text{preproc.steps}}$ is the between-preprocessing steps variance.

However, if each portion of the sample subsampled and/or pretreated is analysed at least twice under repeatability conditions, the analysis of variance (ANOVA) can be applied to separate the repeatability variance of the analytical measurement from the variance arising from the subsampling and/or the preprocessing steps, $s^2_{\text{preproc.steps}}$. Like in the assessment of accuracy, the experimental design adopted in this case is a fully-nested design.

Once the variance of the between-preprocessing steps has been obtained, equation 11 estimates the standard uncertainty due to the preprocessing steps of future working samples with the same characteristics as the sample analysed:

$$u_{\text{preproc.steps}} = s_{\text{preproc.steps}} \quad (11)$$

where the degrees of freedom depend on the number of portions of the sample subsampled and/or pretreated.

4- Uncertainty arising from other sources

The last term in the equation in Figure 4 contains all the other components of uncertainty that have not yet been considered. Most of these components can be considered as “type B” uncertainties [3,4] e.g. associated to differences between the working sample and the test sample, or to the bias of the reference (estimated by linking the reference to other references which have higher levels of traceability). Computation of some of these terms is not straightforward and they should therefore be evaluated with the experience and knowledge of the analyst, bibliography or any other source of reliable information in mind.

The combined standard uncertainty of this term is estimated as the square root of the sum of the variances of the different components:

$$u_{\text{other terms}} = \sqrt{\sum_i s_i^2} \tag{12}$$

Since $u_{\text{other terms}}$ is an estimate of a compound variance, the degrees of freedom, $\nu_{\text{other terms}}$, associated to this standard combined uncertainty are estimated using the Welch-Satterthwaite approach:

$$\nu_{\text{other terms}} = \frac{\left(u^2_{\text{other terms}}\right)^2}{\sum_i \frac{\left(s_i^2\right)^2}{\nu_i}} \tag{13}$$

where ν_i are the degrees of freedom associated to each individual variance term, s_i^2. In many cases, however, there is no information about the degrees of freedom of the components related to this last term. In such a case, a coverage factor, *k*, can be used instead of the two-sided *t*-tabulated value [3].

5- Estimating overall uncertainty

Once the variances of each term and their associated degrees of freedom have been estimated, the overall standard uncertainty, *u*, is estimated by combining these variances.

$$u = \sqrt{u^2_{\text{proc}} + u^2_{\text{assessment}} + u^2_{\text{preproc.steps}} + u^2_{\text{other terms}}} \tag{14}$$

The overall expanded uncertainty is estimated by multiplying the overall standard uncertainty by the two-sided t_{eff}-tabulated value, at a given α probability and ν_{eff} degrees of freedom.

$$U = t_{\alpha/2,\text{eff}} \cdot \sqrt{u^2_{\text{proc}} + u^2_{\text{assessment}} + u^2_{\text{preproc.steps}} + u^2_{\text{other terms}}} \tag{15}$$

where the effective degrees of freedom, ν_{eff}, are estimated using the Welch-Satterthwaite approach [16,17]. The overall expanded uncertainty is a confidence

interval within which the conventional true value is considered to lie with selected probabilities of error.

Although the overall uncertainty has been estimated from precision information from test samples (u_{proc} and $u_{assessment}$ in assessing accuracy and $u_{preproc.steps}$ from a sample similar to the working samples to be analysed), the uncertainty calculated with equation 15 is a reliable estimate of the uncertainty related to future working samples. The only requirement is that the samples used to estimate precision have similar matrix and level of analyte concentration as the future working samples. The uncertainty arising from possible differences in the analyte concentration and/or in the matrix type between the test sample and the working sample should be considered and included, whenever possible, as type B uncertainties. Further analyses of working samples using the same procedure, whenever this is under statistical control, would provide similar uncertainty values. Therefore, just as traceability establishes a link between the measurement method and the results corresponding to the analysed samples, the estimated uncertainty can be associated to the measured quantities of similar samples in the future.

A CASE STUDY: DETERMINING COPPER IN CONTAMINATED LANDS.

A tested (candidate) method B, whose accuracy against a reference method A had previously been demonstrated, is used to determine copper content in contaminated lands. Method B treats the sample with a mixture of nitric and perchloric acids, dissolves the analyte in diluted hydrochloric acid and then analyses it by ICP-AES. The reference method A is based on treating the sample with the same acids but the copper content of the final acidic solution is determined by AAS.

The main sources of uncertainty arise from the experimental variation, the estimation of method bias, the estimation of laboratory bias (estimated by linking the reference method A to a CRM) and the lack of homogeneity of the sample to be analysed.

The uncertainty arising from the experimental variation caused by applying the analytical procedure to obtain future measurement results and the uncertainty from estimating the method bias are calculated from the information generated in validating (assessing the accuracy of) method B against method A. Method B was

validated with a test sample which is very similar to the contaminated land samples (i.e. the level of the measurand and the matrix are similar in both samples), so most of the sources of uncertainty arising from analytical measurement and pretreatment steps can be estimated with the information from the assessment of the accuracy. A two-factor fully-nested design was followed for both methods. The factors representatively varied were the replicate and the run. Between each run, factors such as the analyst, the calibration and the day were varied. Therefore, information about the repeatability standard deviation and the between-run standard deviation was obtained.

1- Assessment of accuracy

The accuracy of method B was assessed against method A with a homogenous and stable test sample. The measurements were carried out following a two-factor fully-nested design. From the requirements, two measurements were made in repeatability conditions on ten different runs. The results, together with the grand means, $\overline{\overline{x}}_A$ and $\overline{\overline{x}}_B$, the variance of the means, $s^2_{\overline{x}_A}$ and $s^2_{\overline{x}_B}$, the between-run variance, $s_{run,A}^2$ and $s_{run,B}^2$, and the repeatability variance, $s_{r,A}^2$ and $s_{r,B}^2$, for both methods, are shown in Table 2. A graphical presentation of the data from Table 2 for methods A and B is shown in Figure 5. It can be seen that for both methods all the individual results are within the confidence interval defined by twice the run-different intermediate standard deviation of the analytical method.

1.1- *Comparison of precision*

After checking the absence of outlying results and the normality of the data, the repeatability variance and the run-different intermediate variance of each method were compared by *F*-tests [11,12]. There was no evidence that both methods have different repeatabilities and different run-different intermediate precision at 95% significance.

1.2- *Comparison of trueness*

The variance of the means of both methods must also be compared by means of an *F*-test to find out which expression must be used for calculating s_d.

$$F_{cal} = \frac{s^2_{\bar{x}_A}}{s^2_{\bar{x}_B}} = 184 \quad (16)$$

This value was compared to the two-sided tabulated value at probability α=0.05 and $\nu_A=\nu_B$=9 degrees of freedom. The tabulated *F*-value is 4.03. As $F_{cal} \leq F(0.975, 9,9)$, the difference between the variance of the means of both methods is not significant. Therefore, equation 2a was used to calculate s_d and a value of 1.04 was obtained.

The trueness of method B was checked by calculating the following statistic which is based on equation 1:

$$t_{cal} = \frac{\left|\bar{\bar{x}}_A - \bar{\bar{x}}_B\right|}{s_d} = 142 \quad (17)$$

Finally, the t_{cal} was compared to the two-sided *t*-value, at a probability α=0.05 and ν=18 degrees of freedom. The *t*-tabulated value is 2.10. Since $t_{cal}<t(0.975,18)$, the difference between the two grand means was found to be not significant. Therefore, since the accuracy of method B has been assessed against method A, method B is traceable to reference method A.

2- Estimating uncertainty

2.1- *Uncertainty arising from the analytical procedure*

The standard uncertainty arising from the experimental variation of future measurements is given by equation 4. Since each portion of the working sample (i.e. the contaminated land), was analysed only in one run and with one analysis (i.e. p_s=1 and n_s=1), the standard uncertainty is obtained by combining the repeatability variance and the between-run variance obtained in assessing the accuracy for method B. The standard uncertainty was 2.14 $\mu g \cdot g^{-1}$ and its associated degrees of freedom (estimated from equation 6) were 13.

2.2- Uncertainty arising from the assessment of accuracy

Since the traceability of method B was assessed against method A, the method bias was not found to be statistically significant. The standard uncertainty arising from the assessment of accuracy was estimated with s_d. From equation 2a, the value for $u_{\text{assessment}}$ was 1.04 $\mu g \cdot g^{-1}$ with 18 degrees of freedom.

2.3- *Uncertainty arising from preprocessing steps*

Since the samples of soil are not homogeneous and the lack of homogeneity was not considered in the assessment of accuracy, an uncertainty term related to the heterogeneity of the sample was included in the overall uncertainty. This term was estimated from the information generated previously with a sample similar to the working sample. An estimate of the between-preprocessing steps standard deviation, $s_{\text{preproc.steps}}$, was obtained by taking n_{ps}=10 different portions of the sample. The different portions subsampled were analysed under repeatability conditions (i.e. the sample pretreatments and measurement were performed in the same run). The standard deviation of the results was $s_{\text{I(preproc.steps)}}$=4.7 $\mu g \cdot g^{-1}$ with 9 degrees of freedom. Since the standard deviation obtained is much larger than the repeatability of the analytical method, the repeatability variance can be neglected when it is compared to the preprocessing step-different intermediate variance.

2.4- *Uncertainty arising from other sources*

To obtain a higher level of traceability, method A was traced to a CRM which had similar properties to the test sample and the contaminated land. This provided an estimate of the laboratory bias of method A.

The copper content of the CRM was analysed with method A. As in assessing the accuracy of method B, the experimental design was a two-factor fully-nested design. The factors varied in the fully-nested design were the replicate and the run. Between each run, factors such as the analyst, the day, the calibration, etc. were also varied. The grand mean, $\bar{\bar{x}}_{\text{A}}$=30.46 $\mu g \cdot g^{-1}$, and the variance of the means, $s^2_{\bar{x}_{\text{A}}}$=2.74, were used to assess the traceability of method A to the CRM. The certified values from the results of ten laboratories were $\bar{\bar{x}}_{\text{CRM}}$=31.4 $\mu g \cdot g^{-1}$ and $s^2_{\bar{x}_{\text{CRM}}}$ = 5.2. The uncertainty, $u_{\text{other tems}}$, related to this higher link in traceability was estimated as 0.85 $\mu g \cdot g^{-1}$, with 20 degrees of freedom.

2.5- Estimating the overall uncertainty

The overall uncertainty was estimated by combining the uncertainties arising from the analytical procedure, the assessment of accuracy, the preprocessing steps of the sample and higher levels of traceability and then applying equation 15:

$$U = t_{\text{eff},\alpha/2}\sqrt{u^2_{\text{proc}} + u^2_{\text{assessment}} + u^2_{\text{preproc.steps}} + u^2_{\text{other terms}}} = 1144\ \mu g \cdot g^{-1}$$

where t_{eff} is the two-sided tabulated *t*-value for α=0.05 and ν_{eff}= 14.

Nevertheless, this estimated uncertainty does not take into account every component. For instance, because of a lack of references for a higher metrological level than the CRM, the uncertainty arising from the bias of the CRM has not been considered in the term which corresponds to other sources. Moreover, sources of uncertainty arising from some preprocessing steps or from some analytical measurement steps, influenced by factors of variation not considered in the assessment of accuracy, may have been overlooked.

DISCUSSION

We have presented a new way of calculating the uncertainty in analytical results as a complement to the two main approaches proposed to date: the ISO approach and the Analytical Methods Committee approach. The new proposal takes advantage of the information generated in assessing the accuracy of a given analytical procedure. The advantage of evaluating the uncertainty with this new approach is that the effort made to check the accuracy can also be used to calculate the uncertainty. Therefore, not a great deal of extra work needs to be done. The main benefits of this approach are its conceptual simplicity, its low cost and its universal application.

Compared to the ISO approach, the new proposal has a more holistic character. It means that it is no longer necessary to know the various individual steps into which the analytical method can be broken down because the analytical procedure is taken as a whole.

The main disadvantages of the approach proposed by the Analytical Methods Committee (i.e. the lack of interlaboratory information for the specific type of matrix required and the fact that the interlaboratory study may not be related to the conditions of the laboratory where the result was obtained) are also overcome by this new approach. However, if the interlaboratory information is available and is related to the present conditions of the laboratory, the “top-down” approach considers components of uncertainty, such as laboratory bias or changes in supplier, that must sometimes be taken into account in the new approach as “Type B” uncertainties.

This new approach also has some drawbacks. The main restrictions are related to constraints in validating the accuracy of measurement methods. Since in practice the accuracy of the analytical methods can only be assessed by a limited number of references placed at high levels in the traceability hierarchy, the analyst has often to resort to references placed at low levels in the traceability hierarchy. The analyst should be aware that the lower the level the reference is placed at, the more terms need to be added to the uncertainty budget in order to include the uncertainties arising from systematic effects. For instance, if the accuracy of a method is assessed against a reference method carried out in the same laboratory and not linked to references which have higher levels of traceability (i.e. not linked to a CRM or to a collaborative trial) uncertainty terms such as laboratory bias and reference method bias should be included in the overall uncertainty. Moreover, the analyst must be aware that the factors chosen to be varied in the assessment of accuracy may not include some of the sources of variation of the measurement method. Depending on the number of factors chosen, therefore, other terms of uncertainty which were not previously taken into account in assessing the accuracy must be included. However, the extra work needed to evaluate them is not comparable to the work needed to estimate the uncertainty with the ISO approach. This makes this approach useful in routine analysis or if no information about the different sources of uncertainty is available.

More work still needs to be done to evaluate the component of uncertainty arising from differences in the level of the measurand and the type of matrix between the test sample and the work sample. Future studies must focus on developing detailed procedures and guidelines which describe the experimental steps and statistics required to test the accuracy of measurement methods and their traceability to different references of high metrological quality.

REFERENCES

[1] CEN / CENELEC (1989) European Standard EN 45001: General Criteria for the operation of testing laboratories: Brussels, Belgium

[2] BIPM,IEC,IFCC,ISO,IUPAC,IUPAP,OIML, *International Vocabulary of basic and general terms in Metrology, VIM,* ISO, Geneva, 1993.

[3] BIPM,IEC,IFCC,ISO,IUPAC,IUPAP,OIML, *Guide to the expression of uncertainty in measurement,* ISO, Geneva, 1993.

[4] EURACHEM, Q*uantifying uncertainty in analytical measurements,* EURACHEM Secretariat, P.O. Box 46, Teddington, Middlesex, TW11 0LY, UK, 1995.

[5] Analytical Methods Committee, *Analyst,* 1995, **120**, 2303-2308.

[6] De Bièvre, P., Kaarls, R., Peiser, H.S., Rasberry, S.D., and Reed, W.P., *Accred. Qual. Assur* , 1997, **2**, 168-179.

[7] De Bièvre, P., Kaarls, R., Peiser, H.S., Rasberry, S.D., and Reed, W.P., *Accred. Qual. Assur.* , 1997, **2**, 270-274.

[8] Valcárcel, M., and Ríos, A., *Analyst,* 1995, **120,** 2291-2297.

[9] Thompson, M., *Analyst* , 1996, **126**. 285-288.

[10] King, B., *Analyst,* 1997, **122**.197-204.

[11] International Organization for Standardization, *ISO 5725-6. Accuracy (trueness and precision) of measurement methods and results.* ISO. Geneva, 1994.

[12] Massart, D.L., and Smeyers-Verbeke, J., *Robust process analytical methods for industrial practice and their validation*. Report Project SMT4-CT95-2031, 1996.

[13] International Organization for Standardization, *ISO 5725-3. Accuracy (trueness and precision) of measurement methods and results.* ISO. Geneva, 1994.

[14] Barnett, V., and Lewis, T., *Outliers in Statistical Data*, 3rd Ed. Wiley, Chichester, 1994.

[15] Ramsey, M. , Argyraki, A. , Thompson, M. , *Analyst*, 1995, **120**, 2309-2312.

[16] Satterthwaite, F.E., Psychometrika, 1941, **6**, 309-316.

[17] Dietrich C.F, *Uncertainty, Calibration and Probability. The statistic of Scientific and Industrial Measurement*, 2nd Ed. Adam Hilger, Bristol, 1991.

Table 1

Table 1a. ANOVA table for a two-factor fully-nested design.

Source	Mean Squares	Degrees of freedom	Expected mean square
run	MS_{RUN}	p-1	$n\sigma_{run}^2+\sigma_r^2$
residual	MS_E	$p\cdot(n$-1)	σ_r^2
total		$p\cdot n$-1	

Table 1b. Calculation of variances for a two-factor fully-nested design.

Variance	Expression	Degrees of freedom
repeatability variance s_r^2	MS_E	$p\cdot(n$-1)
between-run variance s^2_{run}	$\frac{MS_{RUN}-MS_E}{n}$	
run-different intermediate variance s_I^2	$s_r^2+s_{run}^2$	
variance of the means $s_{\bar{x}}^2$	$\frac{MS_{RUN}}{n}$	p-1

Table 2. Analytical results obtained for method A and method B in the assessment of accuracy by carrying out a two-factor fully-nested design. The copper content is expressed as $\mu g \cdot g^{-1}$.

Method A			Method B		
Run	Mean	Replicate	Run	Mean	Replicate
1	26.15	27.2	1	29.55	30.1
		25.1			29.0
2	29.15	28.3	2	31.40	32.6
		30.0			30.2
3	24.75	26.5	3	29.45	28.9
		23.0			30.0
4	24.65	24.1	4	27.75	29.1
		25.2			26.4
5	31.20	32.1	5	26.60	27.2
		30.3			26.0
6	30.25	31.5	6	30.60	31.2
		29.0			30.0
7	31.80	32.6	7	31.65	32.5
		31.0			30.8
8	28.90	29.5	8	31.05	32.1
		28.3			30.0
9	29.75	30.2	9	27.35	28.2
		29.3			26.5
10	26.15	27.3	10	32.10	33.0
		25.0			31.2
$\overline{\overline{x}}_A = 28.27$ $s^2_{\overline{x}_A} = 6.98$ $s_{r,A}^2 = 2.02$ $s_{run,A}^2 = 5.97$			$\overline{\overline{x}}_B = 29.75$ $s^2_{\overline{x}_B} = 3.79$ $s_{r,B}^2 = 1.59$ $s_{run,B}^2 = 2.99$		

FIGURE CAPTIONS

Figure 1. Plan showing the assessment of the accuracy of a tested method B. a) Comparison against a reference method A. b) Comparison with a certified reference material, CRM.

Figure 2. Hierarchy of some relevant references in chemical analysis and their relationship to the classical metrological pyramid.

Figure 3. Two-factor fully-nested design. The factors studied are the runs and the replicates. *p* is the number of runs on which the measurement is carried out and *n* is the number of replicates performed every run. $\bar{x}_i$ is the mean of the *j* replicate measurements performed on run *i*. The grand mean, $\bar{\bar{x}}$, is calculated as the mean of the mean values obtained on the different runs.

Figure 4. Scheme showing the process for calculating the overall uncertainty of chemical measurement results using the information from the validation of the accuracy.

Figure 5. Graphical presentation of the analytical results obtained for method A and method B in the assessment of accuracy.

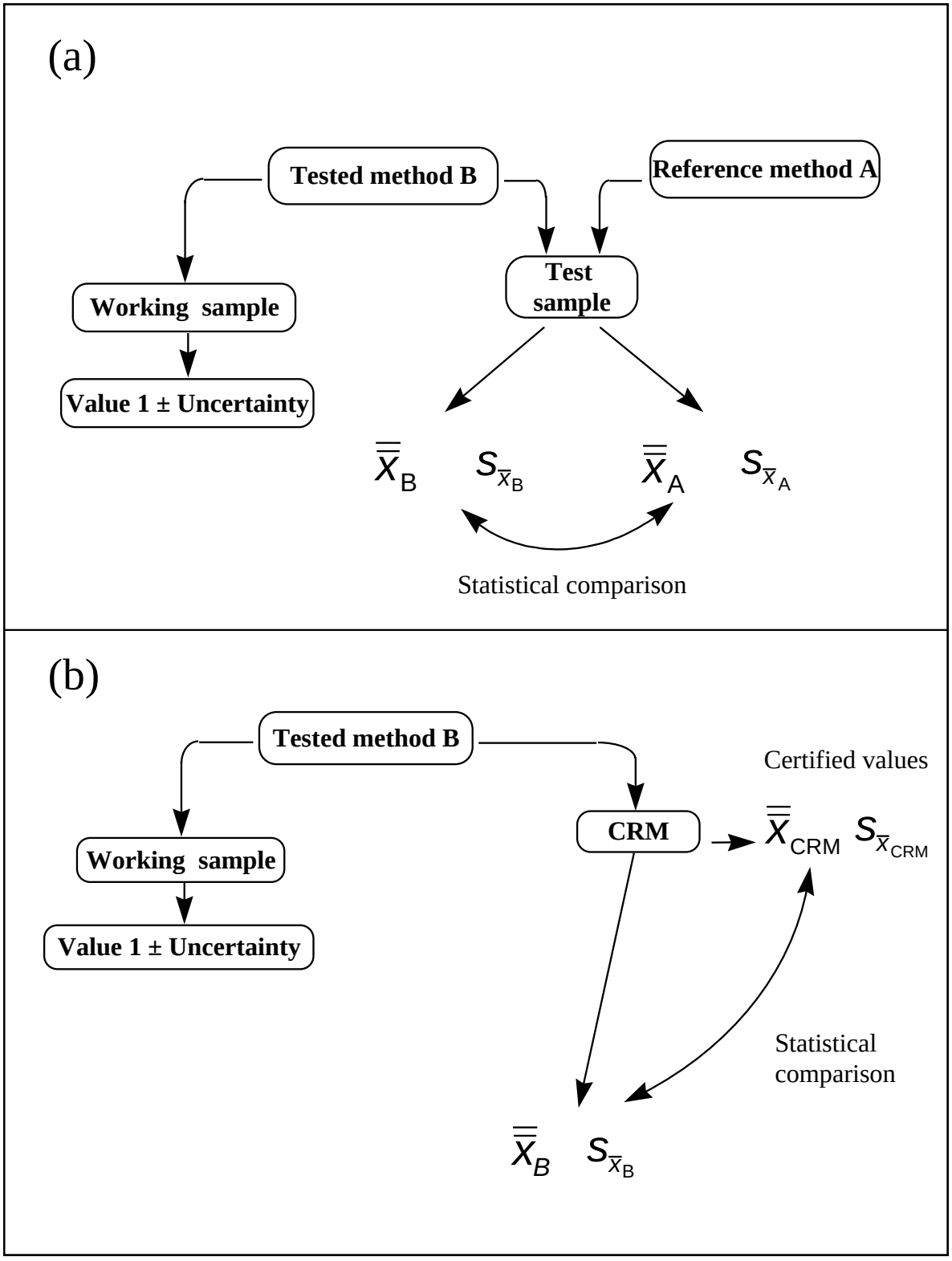

(a)
Tested method B
Reference method A
Working sample
Test sample
Value 1 ± Uncertainty
$\overline{\overline{x}}_B$
$s_{\overline{x}_B}$
$\overline{\overline{x}}_A$
$s_{\overline{x}_A}$
Statistical comparison
(b)
Tested method B
Certified values
CRM
$\overline{\overline{x}}_{CRM}$
$s_{\overline{x}_{CRM}}$
Working sample
Value 1 ± Uncertainty
Statistical comparison
$\overline{\overline{x}}_B$
$s_{\overline{x}_B}$

Figure 1

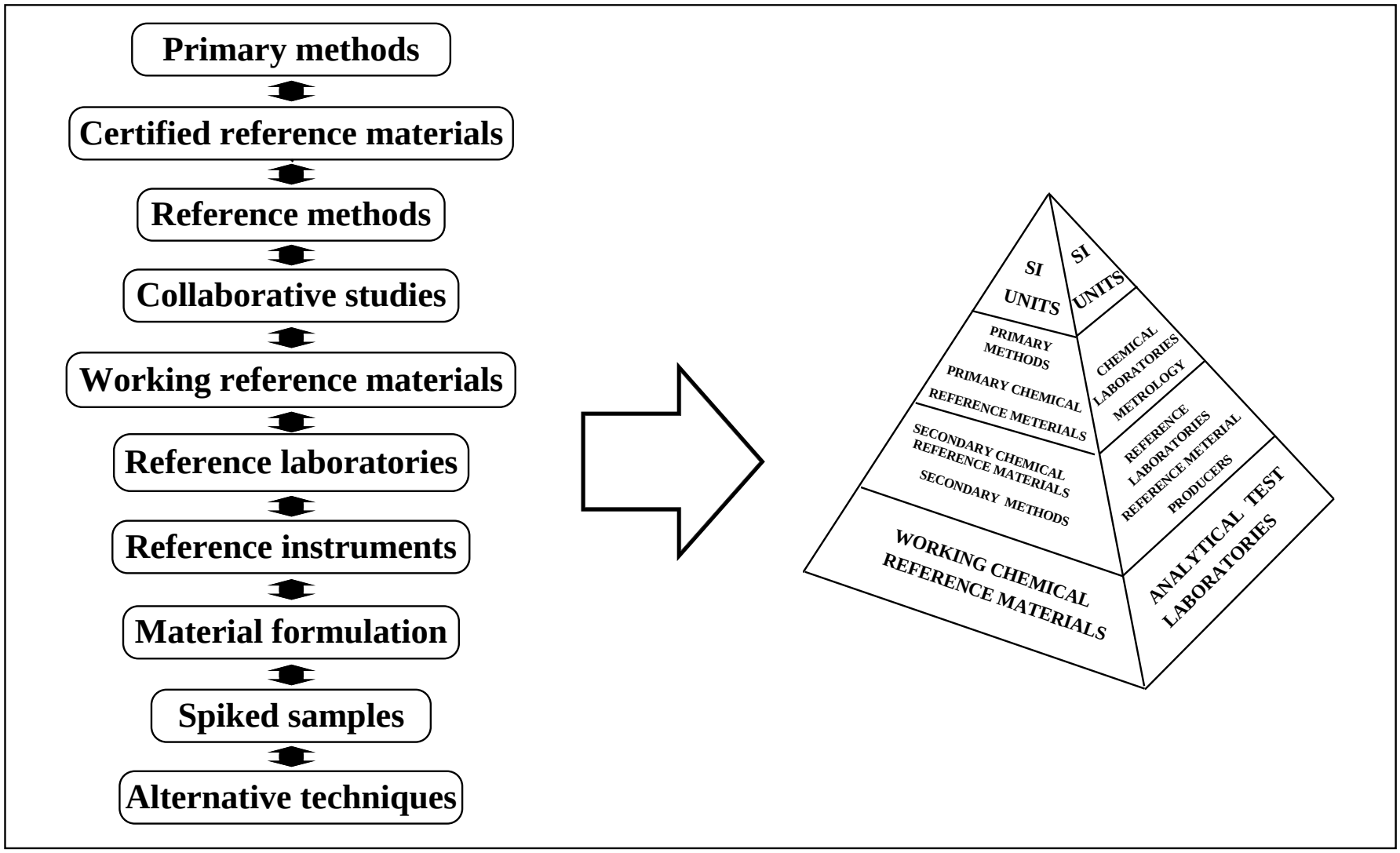

Primary methods
Certified reference materials
Reference methods
Collaborative studies
Working reference materials
Reference laboratories
Reference instruments
Material formulation
Spiked samples
Alternative techniques
SI UNITS
SI UNITS
PRIMARY METHODS
PRIMARY CHEMICAL REFERENCE METERIALS
SECONDARY CHEMICAL REFERENCE MATERIALS
SECONDARY METHODS
WORKING CHEMICAL REFERENCE MATERIALS
CHEMICAL LABORATORIES METROLOGY
REFERENCE LABORATORIES REFERENCE METERIAL PRODUCERS
ANALYTICAL TEST LABORATORIES

Figure 2

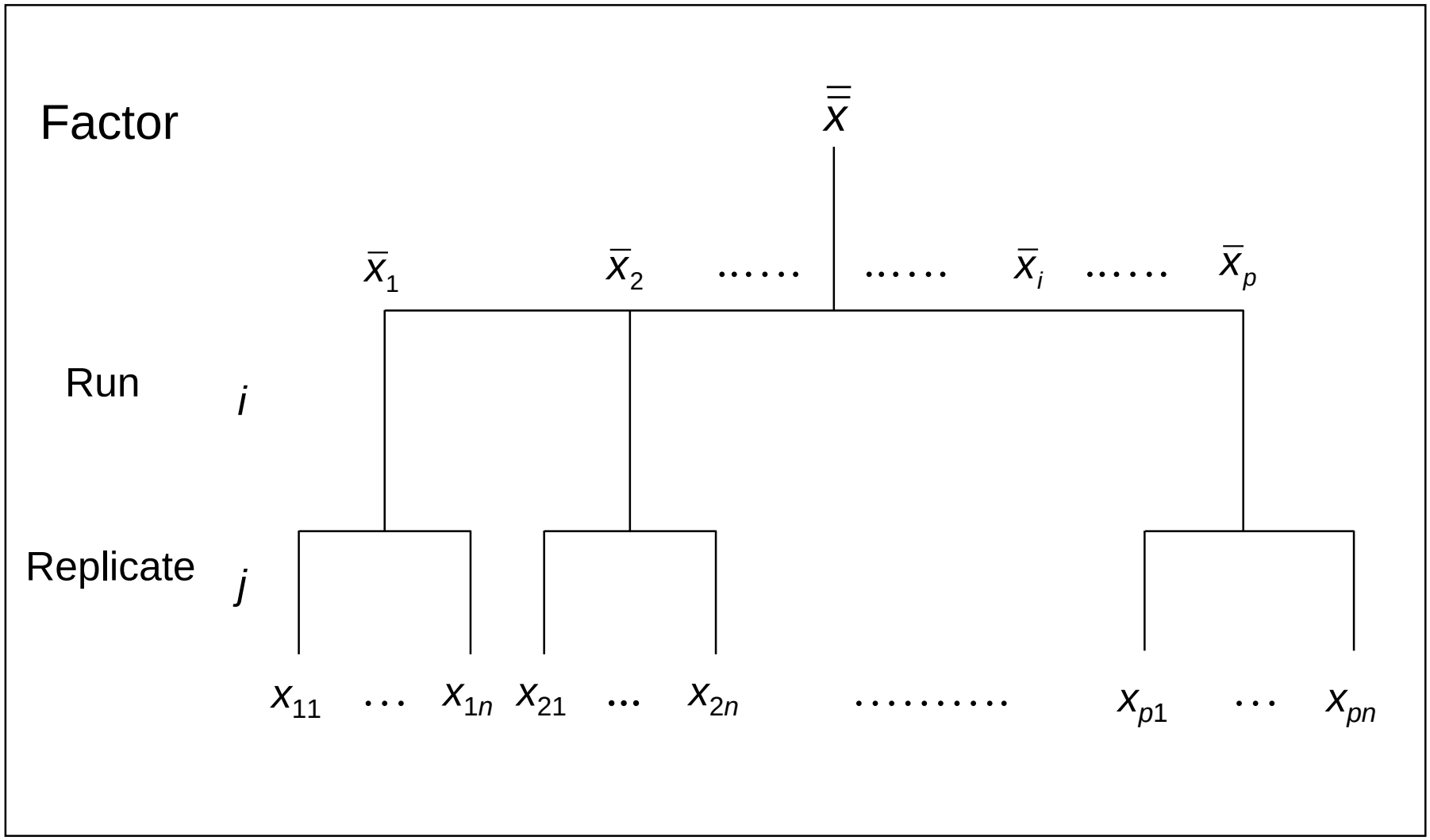

Factor
$\bar{\bar{x}}$
$\bar{x}_1$
$\bar{x}_2$
.....
.....
$\bar{x}_i$
.....
$\bar{x}_p$
Run
i
Replicate
j
x_{11} ... x_{1n} x_{21} ... x_{2n} ..... x_{p1} ... x_{pn}

Figure 3

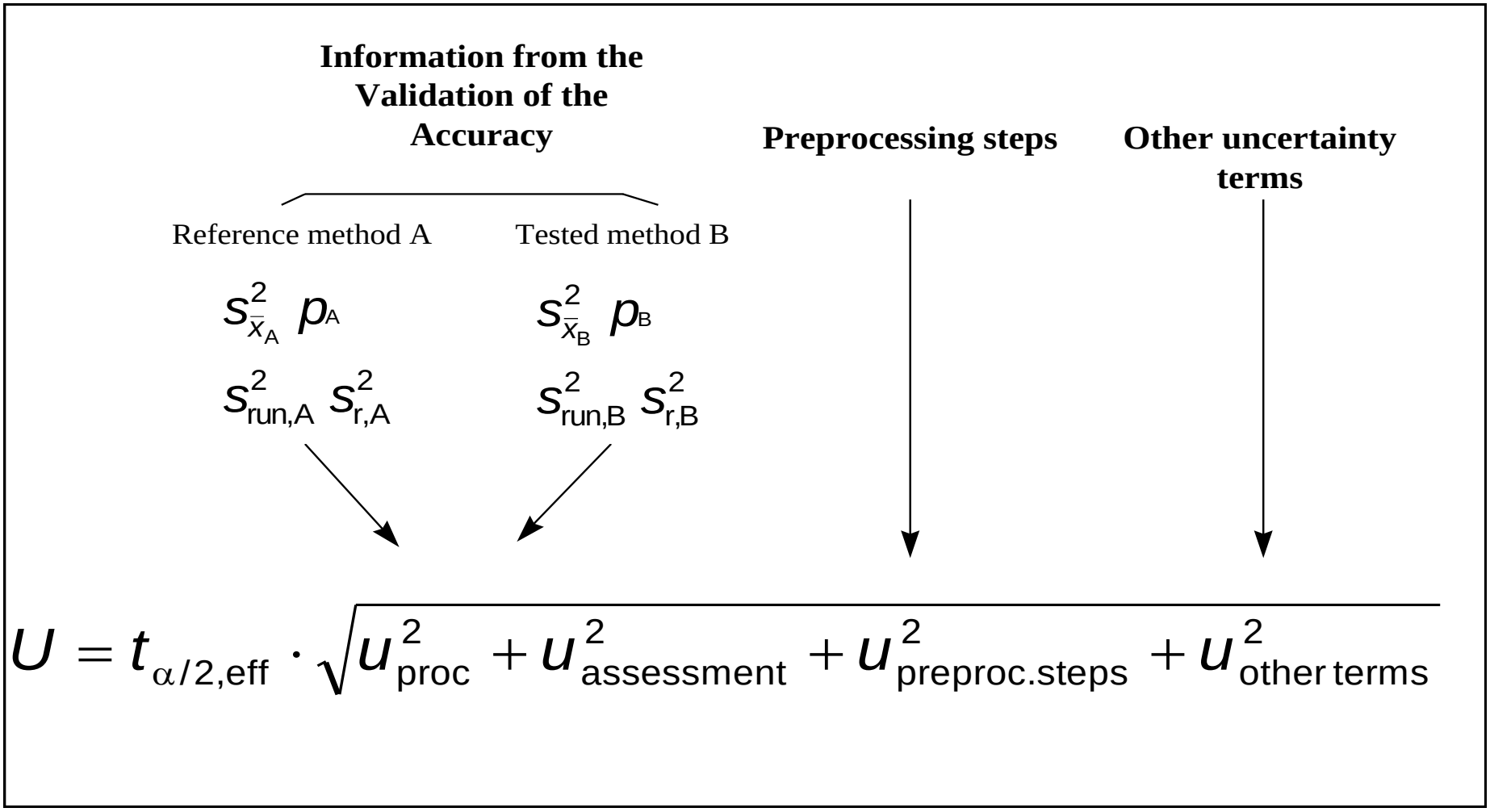
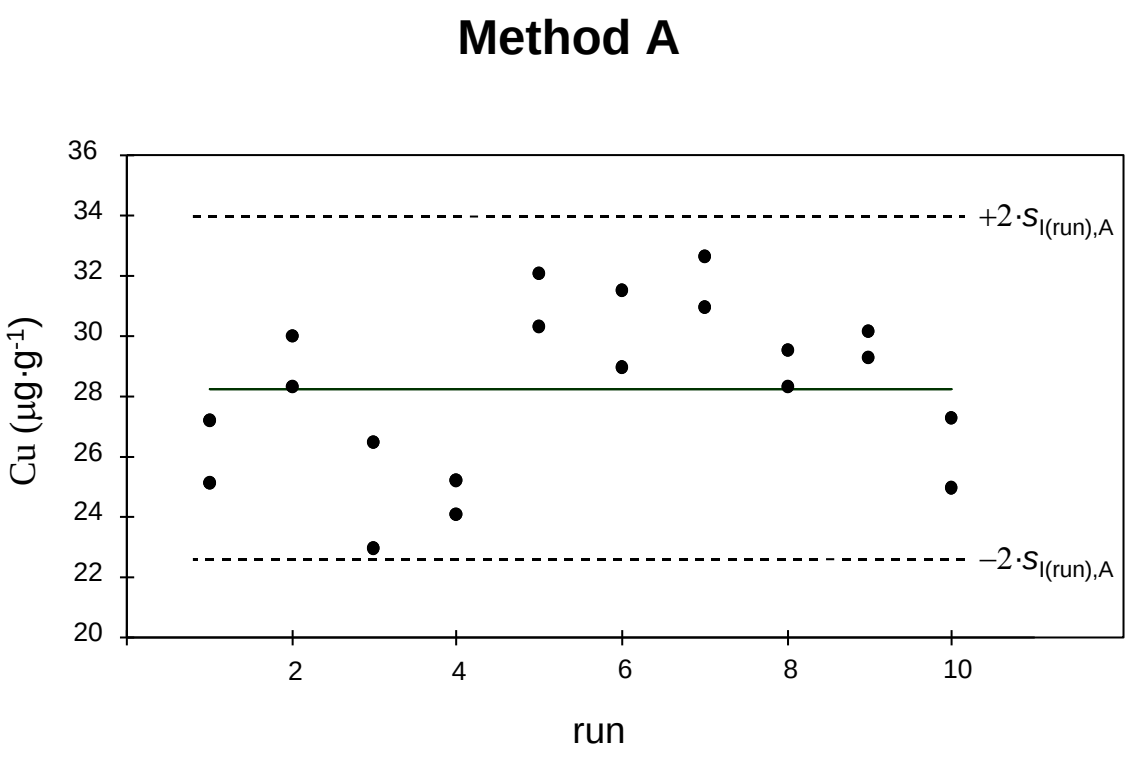
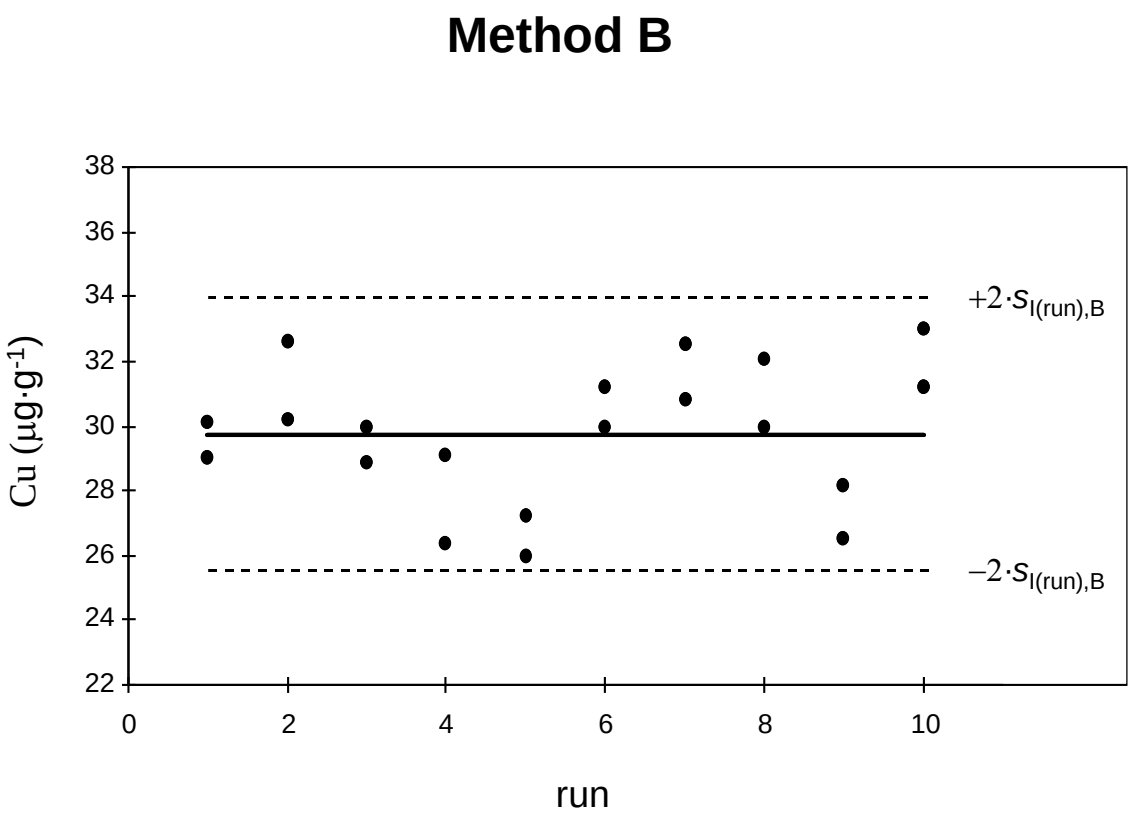

Information from the Validation of the Accuracy
Preprocessing steps
Other uncertainty terms
Reference method A
Tested method B
$s^2_{\bar{x}_A}$ p_A
$s^2_{\bar{x}_B}$ p_B
$s^2_{run,A}$ $s^2_{r,A}$
$s^2_{run,B}$ $s^2_{r,B}$
$U = t_{\alpha/2,eff} \cdot \sqrt{u^2_{proc} + u^2_{assessment} + u^2_{preproc.steps} + u^2_{other\ terms}}$

Figure 4

Figure 5


Method A
36
34
32
30
28
26
24
22
20
Cu (µg·g-1)
+2·sI(run),A
–2·sI(run),A
2
4
6
8
10
run


Method B
38
36
34
32
30
28
26
24
22
Cu (µg·g-1)
+2·sI(run),B
–2·sI(run),B
0
2
4
6
8
10
run