17th IMEKO TC 10 and EUROLAB Virtual Conference
"Global Trends in Testing, Diagnostics & Inspection for 2030"
October 20-22, 2020.

Measurement Uncertainty from Sampling: Implication for Testing, Diagnostics and Inspection

Ramsey, Michael H[1]

[1] *Chair of Eurachem Working Group on Uncertainty from Sampling, School of Life Sciences, University of Sussex, Falmer, Brighton, BN1 9QG, UK (m.h.ramsey@sussex.ac.uk, 00441273480559)*

***Abstract* – Measurement uncertainty (MU) has become recognised as the primary metric of data quality for quantitative chemical measurements made in the laboratory (i.e. *ex situ*). Furthermore, the primary sampling of the material under consideration, is now generally considered to be the first step in the measurement process. It follows that the MU that arises from sampling (UfS) needs to be included in the overall estimate of MU. This principle also applies when measurements are made *in situ*, without removal of a physical sample, as is the case for most hand-held and locally positioned measurement devices. A worked example with Portable XRF is used to explain how MU and UfS can be estimated for *in situ* measurements. It is argued that this role of UfS within MU is equally applicable to quantitative measurements made for Testing, Diagnostics and Inspection in general, and will improve the reliability of compliance decision made on whatever material is under consideration.**

Keywords – Measurement Uncertainty, Sampling, Testing, Guidance

I. INTRODUCTION

Sampling and analysis are both integral and essential parts of the whole Measurement Cycle (Fig. 1), which describes the interaction between the analyst and the client that drives the measurement and testing process , as previously discussed by Barwick [1].

It has become generally accepted that the measurement process (Fig. 2), within this cycle, begins when the primary sample is taken from the sampling target [2]. This is different from the common assumption that the process begins either when the laboratory sample enters the door of the laboratory, or even just when an analytical instrument gives a measurement value. The sampling target is defined as 'the portion of material, at a particular time, that the sample is intended to represent', so for example, it may be a batch of product or an area of land.

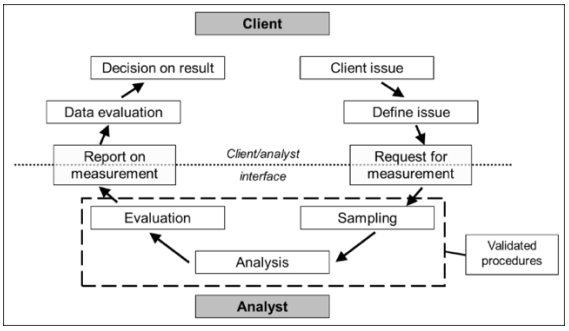


Fig. 1. The Measurement Cycle showing the integral role of sampling [1]

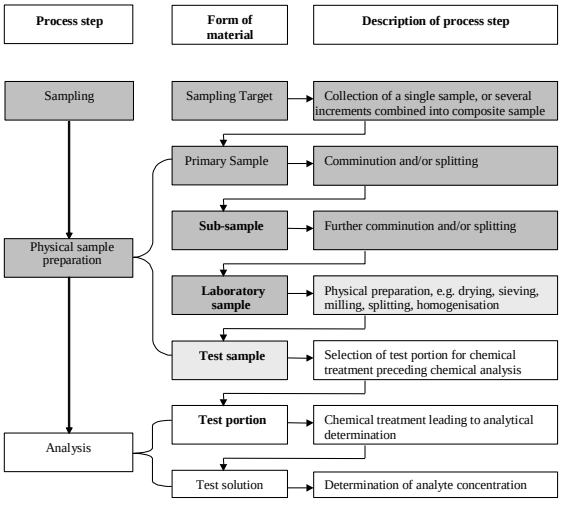


Fig. 2: Sampling as the first step in a typical measurement process [2]

A consequence of this model of the measurement process, is that the uncertainty of the measurement value (MU) must include the contribution that arises from the sampling process (UfS). The UfS is mainly caused by the small-scale heterogeneity of the analyte within the sampling target. The person resonsible for taking the measurement therefore needs to consider the quality of the

primary sampling, as well as the quality of the instrumental analysis. The primary metric for judging the quality of a measurement value is its uncertainty (MU) and thereby, whether the measurement value is fit for its stated purpose (FFP). Consequently, it is essential to estimate UfS in order to make a realistic estimate of MU. This is more realistic than assuming that a sample taken by a correct protocol is 'representative' by definition, and that the sampling process is not, therefore, a source of MU.

For accredited laboratories, the international standard ISO/IEC 17025:2017 [3] now makes it clear that UfS should be included in an estimate of MU, unless it is explicitly excluded. In particular, Section 7.6.1 states "Laboratories shall identify the contributions to measurement uncertainty. When evaluating measurement uncertainty all contributions that are of significance *including those arising from sampling*, shall be taken into account using appropriate methods of analysis". Also, Clause 7.6.3 requires that "A laboratory performing testing shall evaluate measurement uncertainty. Where the test method precludes rigorous evaluation of measurement uncertainty, an estimation shall be made based on an understanding of the theoretical principles or practical experience of the performance of the method." International Laboratory Accreditation Cooperation (ILAC) Accreditation Committee did conclude (at a recent meeting) that 7.6.3. is valid for sampling because sampling in the context of ISO/IEC 17025:2017 is an activity associated with subsequent testing performed by a laboratory [4].

II. ESTIMATION OF MU FROM SAMPLING (UfS)

Guidance on how to estimate UfS has recently been published by Eurachem, in collaboration with Cooperation on International Traceability in Analytical Chemistry (CITAC), EUROLAB, Nordtest and Royal Society of Chemistry/Analytical Methods Committee [2]. The two main types of estimation methods described are either based upon modelling, or upon empirical measurements. The 'Duplicate Method' is the empirical approach that is most likely to be widely applicable to UfS estimation in testing, diagnostics and inspection.

In the Duplicate Method, duplicated samples are taken on a small proportion of the sampling targets (e.g. 10 % of the sampled batches of product, but at least 8. [5]), and duplicate measurements are usually made on both of these samples in a balanced design (Fig 3).

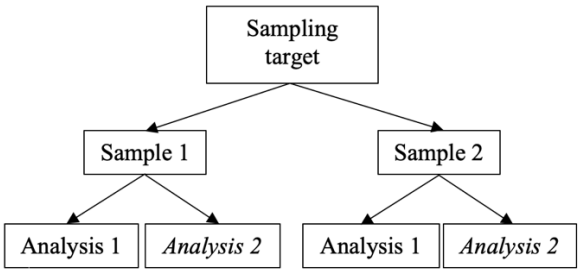


Fig. 3: Balanced experimental design for the estimation of UfS. A simplified design dispenses with the duplicate analyses (Analysis 2). This still enables the estimation of MU, but not directly of its two components ([2], pages 18 & 98)

The statistical procedure called Analysis of Variance (ANOVA) is then generally used to quantify the three components of the total variance, s^2_{total} , where *s* is the standard deviation, as shown by

$$s^2_{total} = s^2_{between-target} + s^2_{sampling} + s^2_{analytical} \quad (1)$$

The standard measurement uncertainty (u) arises from a combination of the sampling and analytical sources:

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

The generally more useful expanded relative measurement uncertainty with 95 % confidence (U'_{meas} or more generally U'), for a measurement value (x) is given by

$$U' = 100 \frac{2 s_{meas}}{x} \ \% \quad (3)$$

Dispensing with the duplicate analyses (Analysis 2 in Fig. 3) in the balanced design saves time and money and still enables the estimation of MU. It does not, however, give *direct* estimates of the two component uncertainties (e.g. $s_{sampling}$, $s_{analytical}$) to enable the identification of the dominant source of MU. However, UfS can be calculated subsequently, using an external estimate of $s_{analytical}$ in Equation (2).

The duplicate method only estimates the random components of the MU, in the form of repeatability. This will therefore tend to underestimate $U'_{\text{analytical}}$, because it ignores further sources of variability included in rigorous external estimates [6]. The remaining systematic components of the MU can be quantified as analytical bias and sampling bias. Analytical bias for *ex situ* measurements is usually determined by the measurement of certified reference materials (CRMs) with composition similar to that of the test material. The sampling bias is harder to estimate, and has often specifically been excluded from the MU estimate, but it can be estimated

using results of Sampling Proficiency Testing ([2], p 17).

The Eurachem UfS Guide [2] describes six worked examples of UfS estimation using several different approaches, applied to quantitative laboratory measurements made *ex situ* on a wide range of analytes present in many different materials including food, feed, water and soil.

III. UfS ESTIMATION FOR TESTING, DIAGNOSTICS AND INSPECTION

The 2nd Edition of the Eurachem UfS Guide in 2019 was extended from the 1st Edition in 2007 to include a discussion of uncertainty estimation in a wider range of situations beyond those of traditional *ex situ* measurements [2, p37]. *In situ* measurements are those where an analytical instrument is placed on the surface of the test material in the sampling target, without removing a physical sample. *In situ* measurements are often made with either hand-held measurement devices, or when unmanned sensors are locally positioned (e.g. in a process stream). Users of *in situ* measurement instruments are, therefore, often unaware that a 'sample' has still effectively been taken, and that a measurement value may therefore represent only a small proportion of the sampling target. Consequently, the heterogeneity of the analyte concentration within the sampling target will be a primary cause the UfS, as has already been discussed for traditional *ex situ* lab measurements. The UfS can be even higher for *in situ* applications, because the material measured is not homogenized as would be the case for a removed sample for *ex situ* measurement in the laboratory. This uncertainty becomes evident when the testing instrument is repositioned in nominally the same location on the sampling target, but gives quite a different measurement value. This will be evident in the Case Study discussed below.

IV. CASE STUDY: ESTIMATION OF UfS AND MU FOR MEASUREMENTS MADE *IN SITU*

The random components of the uncertainty of *in situ* measurements are usually estimated as repeatability using the Duplicate Method, which has been widely employed for *ex situ* measurements, as already explained. The equivalent of the 'duplicate samples' are taken by placing the *in situ* measurement device twice, reflecting two independent interpretations of the measurement protocol. In the Case Study, a hand-held portable x-ray fluorescence spectometer (PXRF) (Fig. 4) was used to measure Pb concentration in the topsoil at the site of a medieval Pb smelter at Wirksworth, Derbyshire, UK [7].

The instrument was placed on the soil at two estimates of the sampling location for a particular sampling target, separated by a distance representing the spatial uncertainty of the survey technique. In this study, the spatial uncertainty was ± 2m, so the duplicate sample was located 2m away from the first in a randomly chosen direction. These two sampling points are both equally likely interpretations of the protocol, given that particular surveying technology. The duplicate *in situ* readings will reflect the effect on the uncertainty of the small-scale spatial heterogeneity in the analyte concentration at that location (i.e. within each sampling target). Duplicate samples taken from at least eight such sampling targets, selected at random across the investigation site, will reflect the typical measurement uncertainty caused by heterogeneity. In this study for research purposes, 24 sampling targets had duplicates samples, but this could have been reduced to 8 for a routine investigation. When sampling in the temporal domain, for example for river waters, the investigator should take the duplicate samples with a time-lapse that similarly reflects the temporal ambiguity in the sampling protocol.

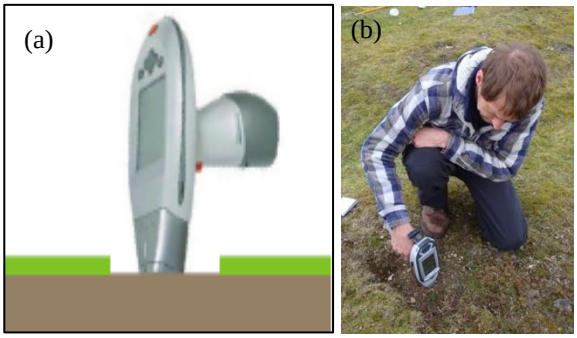


Fig. 4. (a) Schematic diagram and (b) photo of a PXRF being placed on the surface of soil to make an in situ measurement of Pb concentration, as in the Case Study

The measurement uncertainty estimated using the duplicate method alone does not include the systematic component arising from any bias in the chemical analysis or the field sampling. The bias from the chemical analysis alone can be estimated by measurements made on matrix-matched CRMs and can easily be included in the estimate of the measurement uncertainty ([2] Example A2, p 50). However, unlike most test materials in the real world, the CRMs are homogeneous, fine grained, and dry. To overcome this mis-match, the approach most often adopted to estimate systematic sampling effects for *in situ* measurements, is to compare them against *ex situ* measurements made for the same analyte on the same sampling targets. Technically, as well as a matching analyte, there is also a need to match the value of the 'measurand', which is effectively the true value that is being estimated (e.g. total Pb concentration in the dry soil). For this Case Study, the take-away samples were extracted at the locations where PXRF measurement had been made, and then analysed by ICP-AES (traceable to CRMs) after drying, sieving, grinding and acid digestion in a remote laboratory (i.e. *ex situ*). The application of a simplified balanced design (Fig 1, without *Analysis 2*) for 24 sampling targets gave the measured values of Pb

concentration (Table 1, Columns S1Pb & S2Pb).

Table 1. Measurements of Pb concentration at 24 sampling targets in the Case Study. Columns S1Pb & S2Pb are duplicated in situ PXRF measurements (i.e. samples S1 and S2) used for the estimation of the random component of MU. Ex situ measurements made by ICP-AES on prepared samples from these same targets (ICP Target Av.) were used to estimate the systematic component of MU (i.e. bias) against the average (Av.) of the PXRF values (PXRF Target Av). Raw data from [8]

	PXRF in situ			ICP ex situ
Target	**S1Pb**	**S2Pb**	**Target Av.**	**Target Av.**
Number	mg/kg	mg/kg	mg/kg	mg/kg
1	1005	1633	1319	7340
2	4631	3723	4177	8815
3	1415	2264	1840	1522
4	865	1350	1108	1290
5	2899	2216	2558	9340
6	721	1758	1240	3080
7	2122	1014	1568	4180
8	1321	1043	1182	1926
9	3348	3904	3626	3670
10	11543	5570	8557	6718
11	2904	2833	2869	5630
12	2617	2762	2690	3630
13	976	786	881	6880
14	6127	3874	5001	9370
15	331	576	454	1522
16	12878	8948	10913	21877
17	3246	4332	3789	5230
18	9006	6098	7552	18784
19	1936	1989	1963	2800
20	5811	6289	6050	10584
21	4611	2880	3746	7316
22	1326	1442	1384	2235
23	1215	2713	1964	3860
24	2070	2305	2188	5210

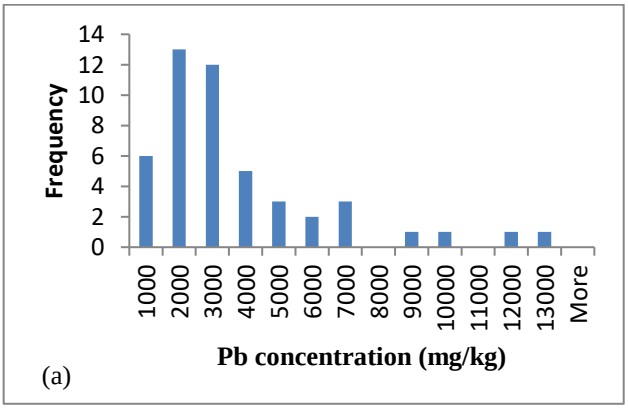
When these values were processed by robust analysis of variance program RANOVA2 [10], they gave an estimate of MU (U'_{meas}) of 55 %. The use of such robust statistical methods makes the calculations much less susceptible to the effects of a small proportion of outlying values (i.e. < 10 %), which are evident in some duplicates (e.g. Targets 10 & 18). Additional *ex situ* PXRF measurements were made in the laboratory on prepared versions of removed samples from the same 24 targets. These conformed to the fully balanced experimental design (Fig. 3), i.e. with duplicated analyses. The $U'_{analysis}$ was estimated as repeatability by ANOVA to be 3 %. Assuming that the instrumental performance of the PXRF was similar when used *ex situ* and *in situ*, that would indicate that the value for the $U'_{sampling}$ used *in situ* was 54.9 % ($\sqrt{(55^2 - 3^2)}$, using Equation 2). The limitations of using the repeatability to estimate $U'_{analysis}$, dicussed above, therefore have little effect on the overall estimate of MU. More genrally, these findings illustrate that although the PXRF instrument reports an uncertainty (U') of around 3 %, the actual value of the MU is much higher at around 55 %, when UfS is included.

The uncertainty of each measurement value (x) can be expressed as being between lower and upper confidence limits (LCL and UCL), usually calculated as x-U' and x+U' respectively. One refinement of this approach discussed in the recent Guide ([2]. p21,48), when the MU is large (>20-30 %), is to express MU as an uncertainty factor (^{F}U).

$$^{F}U = \exp(2s_{G,meas}) \quad (4)$$

Where, $s_{G,meas}$ is the the equivalent of s_{meas} calculated using ANOVA on the natural logarithms of the measurement values ($s_{G,meas}$ *is* 0.308 for Case Study).

In that case, the LCL and UCL are calculated as $x/\,^{F}U$ and $x*^{F}U$, respectively. The large MU is often due to the frequency distribution being log-normal (i.e. positively skewed) rather than normal (i.e. Gaussian). For this Case Study the distribution of the PXRF measurements (Table 1), is log-normal (Fig. 5).



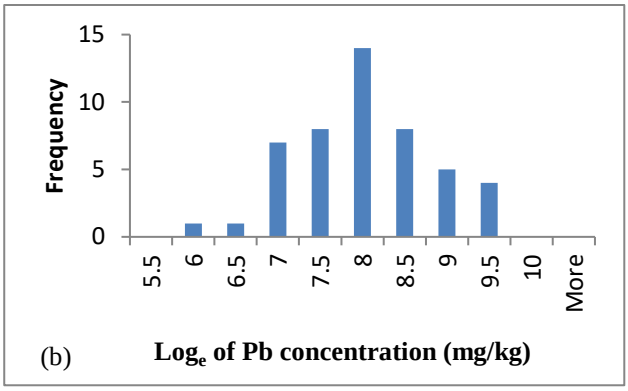


Fig. 5. Histograms of Pb concentrations measured by PXRF in Case Study showing (a) positive skew of original frequency distribution, and (b) near-normal distribution when measurements are log-transformed

For this Case Study, expressing MU as *U'* for a typical single measurement value of 3000 mg/kg, the CLs would therefore be 1350 and 4650 mg/kg (i.e. 3000 +/- 55 %). However, because the *U',* is over 30 %, a more reliable estimate would be to use ^{F}U, which is calculated (by the RANOVA2 program) to be 1.85. This gives the LCL and UCL for this same typical measurement value as 1622 and 5550 mg/kg (i.e. 3000 */ 1.85). This confidence interval is asymmetric (like Fig. 5a), with the measured value being much closer to the LCL than to the UCL.

The systematic component of the MU from the analytical bias was then estimated by comparing the average value of the *in situ* PXRF measurements, against the *ex situ* ICP-AES measurement (Table 1, Columns PXRF Target Av. and ICP Target Av. respectively). This relationship was modelled as a function of concentration using FREML (functional relationship estimation by maximum likelihood [10, 11]). In FREML the uncertainty of both variables is properly taken into account. It is also possible to use ordinary least-squares regression for this purpose, but this can only allow for uncertainty in the y axis (e.g. the PXRF) and ignores the uncertainty for the x-axis (e.g. the ICP-AES). Assuming no lack-of-fit, the slope coefficient of the linear model ($b_{(1)}$) gives the rotational component of the bias, and the intercept coefficient ($b_{(0)}$) gives the translational component. The FREML linear relationship for the Pb concentration [Pb] in this Case Study is shown in Fig. 6. The equation describing the relationship, showing both coefficients and their standard errors (in parantheses) is

$$[\text{Pb}]_{\text{in situ}} = 0.60\ (\pm\ 0.09) \times [\text{Pb}]_{\text{ex situ}} - 120\ (\pm\ 288) \qquad (5)$$

The estimated rotational bias of the *in situ* PXRF measurements, compared against the *ex situ* ICP measurements, calculated from the slope coefficient, is - 40 % (± 9 %) (i.e. 100 x (1 - 0.60). No translational bias was detected, as the intercept coefficient (-120 mk/kg (± 288)) was not statistically different from zero. Possible causes of this measurement bias were identified as soil moisture, material >2mm, surface roughness in the PXRF 'undisturbed sample', and the difference in depth between the undisturbed sample for *in situ* PXRF (~1mm) and the removed *ex situ* field sample for ICP-AES (150 mm) [7].

How to treat this systematic component of MU for *in situ* measurements is an issue that still needs further discussion, by users of *in situ* measurements in general, to reach a consensus, as has recently been identified [12]. In brief, one option is to 'correct' the *in situ* measurements ($[\text{Pb}]_{\text{PXRF, corr}}$) to agree with the *ex situ* values, by applying a rearrangement of the bias model (omitting the non-significant intercept for the Case Study).

$$[Pb]_{PXRF,corr} = \frac{[Pb]_{PXRF,raw} - b(0)}{b(1)} = \frac{[Pb]_{PXRF,raw}}{0.60} \qquad (6)$$

The uncertainty of the correction (s'_{bias}, 0.09, as <0.2) can be combined into $s_{G,meas}$ using an approximation [9]

$$s_{G,meas} = \sqrt{s^2_{G,meas} + (s'_{bias})^2} \qquad (7)$$

The expanded uncertainty factor ^{F}U can then be calculated using Equation (4), as 1.88 for this case study.

The second possible option is not to correct, but to add the entire bias, and its uncertainty, to the MU.

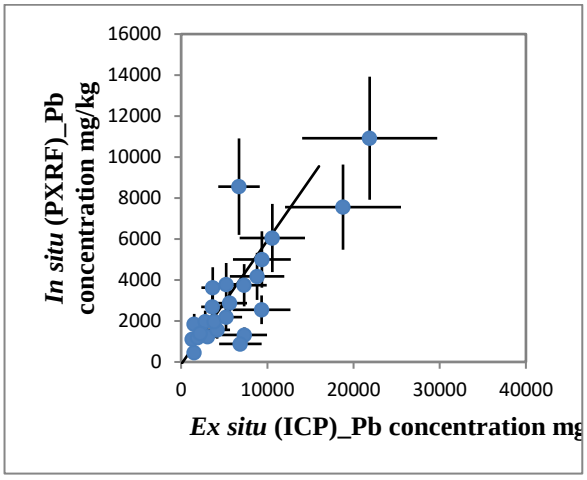


Fig. 6. Systematic component of uncertainty, or 'measurement bias', between in situ and ex situ measurements of Pb concentration at 24 matching locations in Case Study, estimated using a linear FREML model (Equation 5). Error bars are 1s

The two procedures described above where applied to 24 sampling targets for research purposes, but they could be implemented with fewer, down to a minimum of 8 sampling targets for the Duplicate Method. This would save money, but would result in somewhat less reliable estimates of MU[5].

V. BENEFITS OF KNOWING UNCERTAINTY

The Case Study makes clear that realistic estimates of the uncertainty of measurement values require that the UfS is included. This applies whether the measurements are made *in situ* without disturbing the test material, or *ex situ* on a removed sample in a remote laboratory. This conclusion should therefore be equally applicable to quantitative measurements that are made for the purposes of Testing, Diagnostics and Inspection in general.

One advantage of having a reliable estimate of MU (that includes UfS) comes in the assessment of regulatory compliance. Eurachem has also published detailed guidance on compliance assessment [13]. At the most basic level, when the MU is appreciable there is a chance that although a measurement value is below a regulatory threshold, the true value of the analyte concentration may exceed the threshold. The risk of such a 'false negative' classification (and also a false positive), can be avoided if the MU is known, and allowed for in the decision rule. If the MU is used in this way but underestimated, for example by ignoring UfS, then there is still an appreciable risk of an unsuspected false classification (negative or positive). The misclassification of products/targets is even more likely if UfS is not included in the MU estimation, because it is typically a far larger component of the overall MU than that arising from the instrumental analysis alone. For example, at the time of the case study, there was one regulatory threshold for Pb in soil of 2000 mg/kg. For example, the first PXRF measurement value on Target 1 is 1005 mg/kg (Table 1). If this value is corrected for measurement 'bias' using Equation (6) it gives 1675 mg/kg. The confidence interval using an MU based upon the U'_{analysis} of 3 % indicates that the true value lies between 1625 and 1725 mg/kg, which would exclude a false positive classification. By including the UfS, and bias correction in the MU estimate (^{F}U = 1.88), we get the wider confidence interval of 891 to 3149 mg/kg, that does indicate the possibility that the true value of Pb concentration at this target may indeed be over the threshold of 2000 mg/kg.

VI. CONCLUSIONS

Procedures are well developed and widely applied to estimate the uncertainty of quantitative measurements, including the often-dominant contribution from primary sampling (UfS). Eurachem Guidance is available [2] to explain how these procedures can be applied to a wide range of analytes in many different products and natural systems. One example has been given here to show how these procedures can be applied to measurement made *in situ*, without removal of a sample, using a hand-held device. These procedures are applicable to quantitative measurement made more generally for testing, diagnostics and inspection, whether they are made *in situ*, or in the laboratory.

Ignoring MU, or excluding UfS from its estimation, can cause misclassification of the product (i.e. sampling target) and consequent financial losses. When MU is estimated rigorously, including UfS, it becomes possible to make much more reliable decisions on the compliance of the product. This approach will therefore have important implications for testing, diagnostics and inspection in general.

REFERENCES

[1] **Barwick V.J.**: Ensuring the quality of analytical measurements – current support and future challenges. 2020, This volume**.**

[2] **Ramsey M.H., Ellison S. L. R., Rostron P.(**eds.): Eurachem/EUROLAB/ CITAC/Nordtest/ AMC Guide: *Measurement uncertainty arising from sampling: a guide to methods and approach*, Second Edition, 2019, Eurachem. http://www.eurachem.org/index.php/publications/guides/musamp

[3] **ISO/IEC 17025** – General requirements for the competence of testing and calibration laboratories. ISO, Switzerland, 2017.

[4] **Oehlenschlaeger E.** ILAC Guidance on contribution to measurement uncertainty arising from sampling and testing. Presentation made at *Eurachem Workshop 'Uncertainty from sampling and analysis for accredited laboratories'* BAM, Berlin, November 2019 https://www.eurachem.org/index.php/events/completed/277-wks-mu2019

[5] **Lyn, J.A., Ramsey, M.H., Coad, S., Damant, A.P., Wood, R. and Boon K.A.**: The duplicate method of uncertainty estimation: are eight targets enough? *Analyst* 2007, Vol. 132, pp. 1147-1152 *(*DOI: 10.1039/b702691a).

[6] **Ellison S. L. R.; Williams A.** (eds): Eurachem/CITAC Guide: *Quantifying Uncertainty in Analytical Measurement*, 3rd edition, 2012, ISBN 978-0-948926-30-3. Available from http://www.eurchem.org/

[7] **Argyraki, A., Ramsey, M.H. and Potts P.J.:** Evaluation of portable XRF for in-situ measurements of lead on contaminated land. *Analyst*, 1997,Vol. 122, pp. 743-749.

[8] **Argyraki A,:** Estimation of measurement uncertainty in the sampling of contaminated land. PhD Thesis, Imperial College, University of London, 1997.

[9] **Ramsey M. H. and Ellison S. L. R.**: Combined uncertainty factor for sampling and analysis. *Accreditation and Quality Assurance*, 2017, Vol. 22, No. 4, pp. 187-189 DOI 10.1007/s00769-017-1271-y

[10] **AMC Software:** downloaded gratis from: https://www.rsc.org/Membership/Networking/InterestGroups/Analytical/AMC/Software/

[11] **Analytical Methods Committee**: Fitting a linear functional relationship to data with error on both variables, Technical Brief No.10, Royal Society of Chemistry, London, 2020. http://www.rsc.org/Membership/Networking/InterestGroups/Analytical/AMC/TechnicalBriefs.asp

[12] **Ramsey M.H.:** Challenges for the estimation of uncertainty of measurements made *in situ.* Accreditation and Quality Assurance: Journal for Quality, Comparability and Reliability in Chemical Measurement. 2020. DOI: 10.1007/s00769-020-01446-4

[13] **Ellison S.L.R. and Williams A.:(**Eds) *Eurachem/CITAC guide: Use of uncertainty information in compliance assessment.* 2007. From www.eurachem.org.