# **Can *in situ* geochemical measurements be *more* fit-for-purpose than those made *ex situ*?**

Michael H. Ramseya\* and Katy A. Boona,

aSchool of Life Sciences, University of Sussex, Brighton, BN1 9QG UK

\* Corresponding author(m.h.ramsey@sussex.ac.uk; Tel: +44 (0)1273 678085; Fax: +44 (0)1273 678937)

**Abstract**

# It is argued that the selection of the most appropriate geochemical measurement technique should be based upon the fitness of its measurement results for any specified purpose, regardless of whether the measurement are made *in situ* or *ex situ*. Using this approach, *in situ* measurements made in the field are shown to have some definite advantages over those made *ex situ* in a laboratory. A case study is used to show that there are cases where *in situ* measurements can be more fit-for-purpose than their *ex situ* equivalents. This is primarily because the uncertainty of both types of measurement is usually limited by the uncertainty arising from the field sampling process. That uncertainty is mainly caused by small-scale heterogeneity (in space or time) in the analyte concentration within the environmental material (e.g. soil, water or air).

**Introduction**

## The traditional view of geochemical measurements that are made *in situ,* that is without disturbing the test material from its original location, is that they are substantially less reliable than laboratory-based measurements made *ex situ*. *In situ* measurement techniques are now increasingly available but their application is limited by this largely subjective perception that the measurements they produce are not fit-for-purpose[[1]](#footnote-2) (FFP) (Thompson and Ramsey, 1995). However, there are now quantitative methods available to judge the fitness-for-purpose (FnFP) of measurements, whatever that purpose is (Ramsey et al, 1992; Thompson and Ramsey, 1995) and it is possible therefore to judge the FnFP of *in situ* measurements objectively and compare this against the FnFP of traditional laboratory *ex situ* measurements. The purpose of geochemical measurements vary widely from the scientific investigation of geochemical processes, to regulatory decision as to whether concentrations exceed a regulatory threshold, or the financial decision as to whether remediation is economically feasible. In the case of pure science, FFP requires the uncertainty of the measurements to obscure the geochemical variation and allow a hypothesis to be tested with sufficient statistical power (Ramsey et al, 1992). For the situations where there are potential financial consequences arising from measurement uncertainty, it is possible to judge FFP as being where there is a minimal expectation of financial loss from the measurement process and the consequences (Thompson and Fearn, 1996).

## Examples of *in situ* measurements include the determination of elements, such as arsenic, in top soil by just resting a portable x-ray fluorescence (XRF) spectrometer on the surface of the fresh soil after a turf is cut. It could also include the measurement of the activity concentration of a radionuclide, such as137Cs, using a γ-ray spectrometer held some distance (e.g. 25 cm) above the surface of a soil. Similarly is could be the measurement of the concentration of a gas, such as CH4 using a miniature non-dispersive IR sensor, in the ambient air around a landfill site.

## The unacceptability of *in situ* (and on site) measurements for some purposes is certainly explicit in the procedures of some regulators. For example the UK Environment Agency has stated ‘for regulatory purposes, we require laboratories to be accredited to the current version of the European and international standard, ISO/IEC 17025 for this MCERTS performance standard’ (EA, 2006). As *in situ* measurement methods do not take place in laboratories and are not open to accreditation, this implies that *in situ* measurements are not acceptable for regulatory purposes. It is understandable that regulators, and scientists, would want reliability in measurements, but this statement seems to overlook that all measurement results are wrong to some extent, even if they are made by *ex situ* methods, whether or not they are accredited. This is because all measurement results have uncertainty and are never exactly equal to the true value that is being sought in the material being characterised (i.e., the sampling target).

This paper argues that *in situ* measurement results, although they often have higher levels of uncertainty, can be fit for some purposes. Moreover, their ability to measure *in situ* and instantaneously can mean that they can be more fit-for-purpose than traditional lab-based *ex situ* measurements, even if the latter are from an accredited method. Measurements from a case study will be used to support this argument, and to clarify (i) the identity of the true value that is to be estimated, (ii) the sampling target to which that true value applies, and (iii) the strengths and weaknesses of *in situ* measurements.

# **Uncertainty of measurements**

## One key concept that gives a new perspective on *in situ* measurements, and is central to supporting the argument in this paper, is uncertainty of measurement (Umeas). A clear definition of measurement uncertainty from 1993 is *‘an estimate attached to a test result which characterises the range of values within which the true value is asserted to lie’* (ISO, 1993)*.* This concept of measurement uncertainty moves the emphasis away from the properties of an analytical method (e.g. precision and bias), to deal with the key property of the measurement result itself. This uncertaintyincludes both random and systematic effects, unlike either of the terms precision or bias, where they are separated. It also includes contributions from the whole measurement process beginning at the moment of primary sampling, through the sample preparation and into the chemical analysis. Uncertainty is often expressed in its expanded form (using two standard deviations for 95.4 % confidence) and relative to the concentration value as a percentage (U%). The total expanded relative uncertainty of the measurement (Umeas) has contribution from the sampling (Usamp) and the chemical analysis (Uanal). Estimates of the random component of the total uncertainty, including contribution from these individual sources, can be quantified with an appropriate experimental design, such as duplicated sampling at some proportion of sampling targets selected at random (e.g. 10 %), followed by a statistical technique such as analysis of variance (ANOVA) (Ramsey and Ellison, 2007). This ‘duplicate method’ gives and average value for the uncertainty across the selected site, and is equally applicable to estimate uncertainty in both *in situ* and *ex situ* measurement results. Many applications to a range of different geochemical media have shown that the measurement uncertainty arising from the sampling process is usually the limiting factor for both *in situ* and *ex situ* measurements (Boon et al, 2008; EA, 2006). One source of this uncertainty is the interplay between ambiguity in the sampling protocol and small-scale contaminant heterogeneity. This heterogeneity is that which is found within the sampling target, which might be a single location in a soil survey. Because sampling often contributes more than 80 % to the total measurement uncertainty, the analytical contribution is typically less than 20 % of the total variance (Boon et al, 2007). This means that the analytical contribution from both *in situ* & *ex situ* techniques is much less important overall, than that arising from the sampling process. *In situ* and *ex situ* measurements can therefore have similar overall uncertainty, and be effectively equally reliable. Uncertainty of measurement thereby provides anew perspective on the role of *in situ* measurements*.* It also provides a tool to decide whether measurements (*in situ* or *ex situ*) are sufficiently reliable and fit-for-purpose (FFP), as will be discussed below.

# **A clearer definition of the true value**

# The definition of measurement uncertainty from 1993 shows the central role of the true value of the analyte concentration in the sampling target. A more recent definition of measurement uncertainty is a ‘*non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used* (JCGM 200, 2008). In place of the true value, this definition uses the concept of the measurand, which is itself defined as the *‘quantity**intended to be measured’* (Ramsey and Ellison, 2007), and is approximately equivalent in meaning to the true value.

# The correct identification of the true value (or measurand) is however crucial to appreciating the benefits of *in situ* measurements. Taking an example of arsenic in soil, let us assume that the true concentration (mass fraction) of arsenic in a field soil is 50 mg/kg. However a field soil contains stones, plant roots, small animals and water that are usually excluded from the sample in the preparation processes that are carried out in a laboratory prior to *ex situ* analysis. A laboratory may well therefore provide an estimate of the arsenic concentration of ~100 mg/kg, on dry weight basis, for the <2 mm fraction (dried, sieved and ground), which is specified in most definitions of a soil (MAFF, 1986). Nevertheless, for exposure assessment, the arsenic concentration that plants/animals are being exposed to is closer to 50 mg/kg. In this case therefore, the *in situ* technique, which measures the concentration of As in the field soil (containing the stones, roots and water which may contain arsenci), is therefore providing a better estimate of the true value that the plants and animals are being exposed to. An even clearer example is the case of soil from shot gun firing ranges. The lead shot generally have unweathered diameters above 2mm, and are therefore physically excluded from the soils used in *ex situ* measurements. In contrast, an *in situ* measurement would retain the lead shot in the material being analysed.

### Additionally the arsenic concentration will also vary spatially within the soil, usually being higher in the fine particles (e.g. clays) than in the larger particles (e.g. quartz). *In situ* measurement techniques with high spatial resolution (e.g. PXRF down to 0.1 m, and microprobe XRF down to 0.001m)) can then be used to quantify this *in situ* heterogeneity (Taylor et al, 2005), and to measure the arsenic concentration in specific parts of the soil, such as the particles in contact with the plant roots. Furthermore, volatile analytes such as arsenic can be lost by the processes of grinding, homogenisation and chemical digestion that are required for most *ex situ* analysis. Overall, therefore, there are substantial advantages in measuring some analytes (e.g. volatile or bioaccessible) in fresh soil using an *in situ* technique, even though this material is wet, course gained and more heterogeneous. The courser grain size of this material will give a higher contribution to the uncertainty from the random effects, but will avoid the sample preparation that can contribute greater systematic effects. The environmental interpretation can therefore be much more reliable with the *in situ* measurement, if the uncertainty is fully quantified. This is because *in situ* measurements (e.g. by PXRF) can be more accurate (i.e. closer to the true value) than *ex situ* measurements.

# A similar argument can be made for the measurement of activity concentration for the assessment of radioactively contaminated land (and hence dose to a person standing on the land). Let us assume that the true value (or measurand) is a particular activity concentration of 137Cs in Bq/kg in a specific sampling target, such as a defined volume of soil (e.g. 0.03m3 over a depth of 0-10 cm). One option is to measure the counts per second directly using an *in situ* field measurement device (e.g. a γ-ray sodium iodide detector). A suitable mathematical model (e.g. ISOCS (Venkataraman et al, 2005)) can then be used to estimate the activity concentration in that particular volume of soil. The alternative *ex situ* option is to take a physical soil sample (of necessarily more limited dimensions, such as 8 x 8 x 10 cm (0.0005 m3) with a coring device, for transportation to the laboratory. The activity concentration of the 137Cs can then be measured *ex situ* in that different volume of soil. If the ultimate objective is to use the measured activity to estimate dose to a person standing on the soil surface (e.g. where the *in situ* detector was positioned) then the *in situ* measurement might well be closer to the true value as it automatically includes all the soil contamination that might be contributing to the dose.

# **Strengths and weaknesses of *in situ* measurements**

## There are many other strengths of *in situ* measurement techniques. Probably most important for many users is that they provide virtually immediate estimates of contaminant concentration, which can enable immediate decisions to be made. Secondly, the generally lower cost of each *in situ* rather than *ex situ* measurement, means that higher sampling densities are possible in the field. For many applications this higher sampling density means that a more reliable overall assessment can be made of the spatial distribution of a contaminant. Taken in conjunction with the immediacy of the results, this also means that iterative sampling designs become possible. For example a follow-up survey to delineate the extent of ‘hot spots’ of contamination in soil can be made during one site visit, but this can also have a potential draw-back of a perceived lack of objectivity (as described below).

## Thirdly, the test material for *in situ* measurements is fresh, with very little need for sample preparation or storage. This results in no contamination during handling, no change in composition whist the material during transported to the laboratory (e.g. loss of dissolved O2 in water), and no losses of the analyte during drying, grinding or homogenisation (e.g. As, Hg, VOCs, pesticides). This can reduce the high level of measurement uncertainty that can arise from such causes (Lyn et al, 2003).

## A generally less well appreciated advantage of *in situ* measurements is that the structure of the *in situ* heterogeneity of the test material is preserved. This contrasts with *ex situ* analysis where the test material is usually homogenised. *In situ* techniques can therefore be used to quantify this *in situ* heterogeneity (Boon et al, 2007). Once quantified, this new information on the *in situ* heterogeneity of an analyte or contaminant can be used for many purposes. This includes the recreation of intermediate levels of field heterogeneity in pot trials to estimate its effects on the uptake of heavy metals by plants (Thomas el at, 2008). This can provide a much more realistic environment in which to quantify this uptake by plants than either the traditional homogeneous spatial distribution of contaminant, or the binary ‘checker board’ soil treatments that have been employed to simulate extreme levels of heterogeneity (Haines, 2002).

Incidentally the *in situ* measurement option has a further advantage in that it avoids the costs of disposal and transportation of the wastes that are incurred when an *ex situ* method is used, which is particularly advantageous for radioactive contaminants.

## One potential weakness of *in situ* measurements, as discussed above, is that they are potentially less reliable than *ex situ* measurements. It is more difficult to apply analytical quality control (AQC) and to establish clear traceability for the measurements back, for example, to a certified reference material in the field situation, rather than in the lab where this is now routine practice. Measurement times used *in situ* tend to be shorter than those used in the laboratory. For example for gamma ray counting of 137Cs, 10 minutes per location would be considered a long count time if the survey was to measure 100 locations over two working days (~17 hours in total). However, in the lab an *ex situ* count time of 12 hours would be feasible. Because this is 72 time longer than the *in situ* count time, it would be expected to reduce the detection limit by a factor of ~8 (the square root of 72). Procedures for training the operators of the *in situ* techniques are also less well developed than they are for laboratory methods. Furthermore, this current lack of sufficient AQC in the field can make it harder to detect insufficient training. Nonetheless, many of the AQC procedures developed in laboratories can be applied to *in situ* measurements. For example, certified reference materials (CRMs) can be analysed for the estimation of analytical bias in the field, as happens routinely in the laboratory. However, the mismatch between the physical state of the test and reference materials can become more of an issue for materials in their original field state. Proficiency tests (PTs) have proved very effective in improving the quality of lab measurements, and now the field-based equivalents have been developed and have proven to be effective in either sampling PTs (Squire and Ramsey, 2001), or as part of a technology verification programme (EPA, 1995).

## Limited sample size is an issue with some *in situ* techniques, such as PXRF where the depth of measurement for some elements is as little as 1 mm, which gives a mass of the sample (or test portion) of around 1 g. Such a small sample mass would be considered insufficient in traditional sampling theory, but can be considered acceptable for some purposes, as long as the uncertainty is quantified. This apparent limitation can be overcome for soils by making the ‘*in situ’* measurements on core samples that, although removed from the ground, can reveal the original *in situ* variation in analyte concentration over a much greater depth. As discussed above, this property can also be considered as an advantage in quantifying the heterogeneity of the analyte concentration within the sampling target. A major weakness of many *in situ* techniques is that the detection limits may not be low enough to quantify some analytes (or contaminants) at their background concentrations. For example the typical detection limit for arsenic in soil is around 20 mg/kg (for a 60 sec reading on a Niton XLT 700), which is above the typical background concentration of 7 mg/kg. Similarly for ambient air analysis, the current detection limit for methane using a small non-dispersive IR sensor is around 50 ppmv, which is well above the typical background concentration of 2 ppmv.

## One major criticism of *in situ* measurements has been that they encouraging judgemental and subjective sampling and measurement taking. For example, recent guidance on the use of rapid measurement tools on site has stated that *‘The compulsion to chase contamination should be resisted’* (Environment Agency, 2009). It has been considered for many years that *ex-situ* lab-based measurements are therefore more compatible with systematic, and therefore non-judgemental, sampling designs. However, the recent development of many low cost *in situ* measurement devices (e.g. sensors), means that is now possible to set up non-judgemental sampling designs (such as the regular grid) for *in situ* measurements using a sensor network. One example of this would be the recent deployment of a network of sensors in an urban environment to measure air quality (Blythe et al, 2008). By this means, *in situ* measurements can therefore, be considered just as objective as *ex situ* measurements. Furthermore, the continuing presence of *in situ* sensors in such a network enables contamination events to be delineated in the temporal domain as well as in the usual spatial dimension. This capability is not really feasible with the *ex situ* lab analysis approach, so in that case *in situ* measurements will potentially be more fit for that purpose than those made using an *ex situ* approach.

##

# **Comparing the accuracy and fitness-for-purpose of *in situ* & *ex situ* measurements**

The accuracy of a measurement result has recently been defined as ‘closeness of agreement between a measuredquantity valueand a truequantity valueof a measurand’ (ISO, 1993).Using these terms, it is usually assumed that the *ex situ* lab measurement is effectively the true value, and that the *in situ* measurement should be corrected to agree with it *.* However, it is also the case that both *ex situ* and *in situ* measurements are incorrect to some extent compared with the true value (i.e. both have uncertainty). The *ex situ* value may be more reproducible and traceable, and be easier to quality control, but the *in situ* value might well be more accurate (i.e. closer to the true field concentration), in representing the real exposure experienced by a living organisms for example.

## Some criteria for judging thefitness-for-purpose (FnFP) of measurement results include economic considerations, so on a simplistic level the lower cost of *in situ* measurements can lead to a greater number of samples/measurements being taken within an overall budget. This can therefore increase reliability of an investigation, for example in the detection and delineation of hot spots of contamination in either space or time. A more detailed approach incorporates the uncertainty of the measurements, using an economic loss function to assess FnFP (Thompson and Fearn, 1996). One initial application of this approach to both *in situ* and *ex situ* measurements showed that the *in situ* measurements weremore FFP than *ex situ,* primarilydue to higher sampling density achievable with the former. *In situ* measurements generated a lower overall cost than the lab-based *ex-situ* measurements, despite having higher uncertainty on the individual analyses (Taylor et al, 2004).

# **Case study to compare the uncertainty and fitness of *in situ* and *ex situ* measurements**

## The site chosen to investigate the relative fitness of *in situ* and *ex situ* measurements was a 34 hectare golf course in NW England (Fig. 1). It had beenbuilt upon a disposal site for calcium sulphate waste that was generated from the Leblanc soda manufacture process, and was thereby contaminated with arsenic. Measurements of the arsenic concentration in the topsoil were made *in situ* using a PXRF (Niton XLt 700) and compared against *ex situ* measurements made on 20 matching top soil samples (1000 g from 0-100 mm), dried, ground, digested, and analysed in a laboratory using an MCERTS accredited hydride generation-AAS method. The aim of the investigation was to refine the delineation of the arsenic contaminated area, based on a previous low density survey (Barnes, Pers Comms), classified against a site specific action level (threshold value) of 59 mg As/kg dry weight (DW).

Figure 1

## The random component of the measurement uncertainty was estimated using the duplicate method for both *in situ* and *ex situ* methods (Environment Agency, 2006). The systematic component was estimated using measurements on a certified reference materials (CRM, NIST 2711 and LGC 6144). The ‘bias’ between the measurements made on the samples by the two methods at each site was modelled using a functional relationship estimated by maximum likelihood (FREML) (Ellison, 2003), which allows for uncertainty on both axes (unlike least squares regression). Further details of the site, and the application of some other intermediary ‘on site’ methods using the PXRF, after various degrees of preparation of the samples, are reported elsewhere.

## **Results of Case Study**

## The random contributions to the expanded relative measurement uncertainty is high for both the *in situ* and the *ex situ* measurements (Umeas > 100 %, Table 1). By expressing the uncertainty in a relative form, the value is expected to stay reasonably constant for concentrations well above (i.e. > 10 times) the analytical detection limit (Ramsey and Ellison, 2010). Analytical uncertainty for the *ex situ* measurements (Uanal =2.9 %) is a small proportion (~7 %) of that for measurements made *in situ* (Uanal =40 %). This suggests that the proximity of the PXRF measurements to the detection limit (20 mg As/kg) is increasing the contribution of the analysis to the uncertainty at this site. However both uncertainty values are dominated by the sampling uncertainty, which contributes more than 93 % of the measurement variance in both cases. This is likely to arise from the small-scale heterogeneity of the spatial distribution of the arsenic contamination in the soil.

Table 1

## The systematic component of the uncertainty arises partially from the analytical bias of the methods. The estimate of analytical bias for PXRF using the CRMs was -25%, which isunusually large for this technique. The CRM was analysed as a dry, fine homogeneous powder which is quite unlike the field soils analyses that were analysed *in situ*. When the ‘bias’ between the PXRF and Lab measurement techniques was estimated using the concentration estimates for all of the sample materials rather than CRMs (using the functional relationship shown in Fig. 2), the bias was estimated as -43 ± 14 %. The scatter in this relationship, and the large error bars, show that single measurements can deviate substantially from the general trend, probably due to small-scale heterogeneity in the spatial distribution of the arsenic.

Figure 2

The concentration estimates made using the PXRF are on the course-grained *in situ* soils, that contained stones, biota, air and water, and are based upon a soil depth penetration estimated to be ~2 mm for As. This compares with the *ex situ* measurements that were made on dried, fine grained homogenised soil powders from which the stones (diameter >2 mm), coarse biota, air and water had all been removed, and are based upon a soil depth of ~100 mm. This raises the question of whether this value can really be called ‘bias’ in terms of the formal definition.

A recent definition of *bias* is an ‘estimate of a systematic measurement error’ (Ramsey and Ellison, 2007). To clarify this *measurement error* is defined as ‘measured quantity value minus a reference quantity value‘ (Ramsey and Ellison, 2007), and *systematic measurement error* is defined as the ‘component of measurement error that in replicate measurements remains constant or varies in a predictable manner’ (Ramsey and Ellison, 2007). Using these definitions, and knowing that the reference quantity value is equal to the true quality value in this case (note 1 in section 2.17 (Ramsey and Ellison, 2007)), then if the measurand (true value) is the arsenic concentration (mass ratio) in the fresh *in situ* soil, then it would appear that the difference between the *in situ* and *ex situ* measurement results is equal to ‘bias’ in the strict sense. However, when this difference is considered as bias, then the *in situ* measurement could arguably be taken as a better proxy for the true value than the *ex situ* measurement that is traditionally recommended15.

In this general case, the correction of the *in situ* to agree with the *ex situ* measurements is questionable, and will reduce the accuracy of the *in situ* measurement values. However in the particular example, the true value is specified in the site specific action level using the dry rather than the fresh weight. The correction is therefore justified in terms of the moisture content of the soil, but not when considering the other parts of the field soils that have been excluded for lab analysis, such as particles with a diameter greater than 2mm.

**How much U is acceptable? Are either set of measurement Fit-For-Purpose?**

# The optimal level of uncertainty was estimated for both the *in situ* and the *ex situ* measurements using the OCLI method (Ramsey et al, 2002), using the input data give in Table 2. This method calculates the overall cost of a site investigation (expressed as an expectation of loss, E(L)), at an average concentration at those locations within the site that are at risk of misclassification. The cost is based upon not just the cost of taking the measurements (sampling and chemical analysis), but also on the costs arising from the consequences of misclassifying the land due to the measurement uncertainty. This misclassification can be of two types. In the false positive misclassification the land is erroneously classified as contaminated when it is not, and may therefore be needlessly remediated. For the false negative case, the land is erroneously classified as uncontaminated and contamination left in place, leading to costs such as those from subsequent delays in site development, human health effects, or subsequent litigation.

Table2

# **Results of optimisation of uncertainty for PXRF versus Lab As**For the false negative situation (Fig. 3b & 3d), neither the *ex situ* nor the *in situ* measurements are considered fit-for-purpose (FFP). This is because both sets of measurements have actual uncertainty values which give predicted costs E(L) that are much higher than the minimum possible for *this* purpose (shown as the optimal values). However, the cost at the current actual uncertainty is much lower for *in situ* (£211K) than for *ex situ* (£421K). The lower cost (E(L)) for *in situ* measurements therefore means that they are more FFP than *ex situ* measurements.

The optimal uncertainty is low for both measurement options (5.2 mg/kg for *in situ*, and 3.8 mg/kg for *ex situ*). Achieving this optimal uncertainty is predicted to give potential savings of approximately £4m (for *in situ*) for the whole site. This is calculated by multiplying the reduced expectation of loss for each sampling location (£210K) by the total number of locations (i.e. 20).

Achieving these potential savings requires a reduction in the measurement uncertainty at each location. This can only be achieved by reducing the uncertainty arising from the sampling, as this contributes more than 93 % of the total measurement uncertainty. Ideally this reduction would be by a factor of 4.6 (i.e. 23.8 /5.2), and sampling theory predicts that this would require the taking of 20-fold composite samples at each location, creating a sample mass of around 20 kg. As an intermediate improvement, even 5-fold composite samples would be predicted to reduce the sampling uncertainty by a factor of 2.2 (√5) from 20 to 9 mg/kg, and reduce costs (E(L)) by 56 %.

For the false positive situation (Fig 3a & 3c), both measurements have slightly lower U than is needed for *this* purpose. There is a low predicted cost (E(L) = £40 per location) by either technique, at both the actual and the optimal U. There is therefore no need to modify either method for this purpose as they both produce measurements that are already marginally less uncertain than required for this purpose.

There is however a need to decide on whether to correct for the ‘bias’ in the *in situ* measurements to make them comparable with the *ex situ*. This depends on how the regulatory threshold is expressed. The site specific action level (threshold value) was set at 59 mg As/kg expressed on a dry weight basis. It can be argued therefore, that the *in situ* measurements should be corrected to the *ex situ* values that are reported on that same basis15. To investigate the effect that this would have, all the in situ measurements were corrected (by dividing by 0.574), and the uncertainty from the bias (14%) added to the uncertainty budget of the corrected measurement values (to give a very similar value of 154.7%). This produced broadly similar conclusions, except that cost (E(L)) for the PRXRF measurements (£373K) was still lower than that for the *ex situ* measurements but only by 11%. In future perhaps the threshold value could be set on a fresh weight basis (as is the case for some fresh foods), and for the whole soil rather than just the <2 mm fraction, in which case the *ex situ* measurements might be better corrected to agree with the *in situ* values.

Figure 3

**Probabilistic mapping of contamination**

Values of measurement uncertainty on each measurement result (e.g. *in situ*) can be used to make a probabilistic interpretation of the spatial distribution of the contamination (Ramsey and Argyraki, 1997). In this case the uncertainty estimate of 154% was applied to each *in situ* arsenic measurement to classify each location as being in one of four categories. These are either

1. under the threshold (site specific assessment criteria of 59 mg As/kg), (< 23 mg/kg)
2. probably under the threshold, but potentially a false negative, (23 – 59 mg/kg)
3. probably over the threshold, but potentially a false positive, (59 – ∞ mg/kg) or
4. definitely over the threshold, all at the 95% confidence interval. (> ∞ mg/kg)

The site map can then be used to show how the sampling locations fall into these four categories (Fig. 4). Because the estimated value of Umeas is over 100%, it is not mathematically possible to have any values that are in the top category (definitely over the threshold). The previous preliminary delineation of the ‘hot spots’ of arsenic contamination are broadly confirmed, as shown to be slightly conservative.

**Figure 4**

**Comparison of *in situ* and on site measurements**

## Some of the arguments made for *in situ* measurements can also be made for measurements that are made ‘on site’. On site measurements are often made on the site of the investigation, perhaps in a field lab, but with the test material removed from its original location, and often prepared in some way. On site results are also generally considered to be of lower quality than *ex situ* lab measurements, and prone to the encouragement of judgemental rather than systematic sampling strategies. They are subject to many of the same sources of uncertainty as *in situ* measurements, and often dominated by the contribution from the sampling process, so benefit from the integrated approach to the measurement process. The key differences between on site and *in situ* measurements are that (1) they don’t give information on the *in situ* heterogeneity of the material under study (2) the sample mass is not limited by the geometry of the measurement devise, so (3) larger masses of material can be taken in the sample and prepared in similar ways to those used for ex situ analysis (drying, mixing, disaggregation, sieving, grinding and splitting). A detailed investigation of the fitness of measurements made on site (but not *in situ*), and a comparison against *ex situ* measurements for a wide range of analytes (including organic compounds) in soils at different types of sites, is given elsewhere (Boon and Ramsey,).

# **Conclusions**

## The overall conclusion is that *in situ* measurements can be more fit-for-purpose (FFP) than *ex situ* lab measurements, in some circumstances. This extends the finding of an earlier comparison when the *in situ* measurements also produced a lower cost (expectation of loss), but only because they were applied at ahigher sampling density (Lyn et al, 2003). It is clear that all of the measurements made in this current study, *ex situ* as well as *in situ,* as in all others, are ‘wrong’ to some degree because they all have uncertainty of measurement. By definition this uncertainty estimate should include the true value of the contaminant concentration. If this true value (or measurand in metrological terminology) is specified as the field concentration, not just that in the sieved, dried and ground lab sample, then this has implication for the fitness of *in situ* measurements. Some analytes may be lost in storage, drying, sieving, mixing or splitting (e.g. As, Hg, Total Petrol Hydrocarbons). In these cases ‘correcting’ *in situ* field results to agree with *ex situ* lab data can make the measurements less accurate. The results of the case study, as with most previously reported studies, show that the main source of uncertainty is usually the sampling procedure, especially as it interacts with *in situ* heterogeneity of the material in the sampling target.

## The choice between *ex situ* and *in situ* measurement techniquesshould not be made on their uncertainty values alone, even when these are quantified and explicitly stated. The objectives of the measurement strategy also need to be considered. If the objective is to assess the temporal variability of the contaminant concentration then *in situ* measurement strategies have a particular advantage if the sensor can be left in place over that time period. Consideration should also be given to the relative cost of *in situ* versus *ex situ*, and how this will affect the number of samples that can be taken and therefore how representative the measurements are of whole sampling target (e.g. site). These effects apply at any spatial scale, so they are equally applicable to microprobe measurements where *in situ* measurements are the norm, but are also often compared with bulk analysis by traditional *ex situ* techniques.

##

## Measurements that are fit-for-purpose can be either those that either minimize the overall cost (i.e. expectation of loss), or allow scientific hypothesis to be tested, with sufficient statistical power, at minimum cost. *In situ* measurements should be assessed for FnFP using the same procedures applied to *ex situ* techniques. It is clear from this study that there should be no prior assumption that *in situ* measurements are inferior, but rather the fitness of all measurements should be judged objectively.

The overall judgement as to whether to select an *in situ*, or an *ex situ* measurement strategy will depend primarily on whether the measurement results they produce are expected to be fit-for the specific purpose of the investigation, be it scientifically, regulatory or financially motivated. Once FnFP is established, then the other factors discussed here can be used to choose between these two options. For example the immediacy of the *in situ* results may prove decisive if rapid decisions have to be made, particularly in a remote location (e.g. whether to take more samples in an initial survey). As *in situ* measurement technologies improve in performance, such as detection limit, the key requirement is to have objective criteria with which to make the judgement between using either an *in situ* or an *ex situ* strategy.

# **Acknowledgements**

## The authors would like to acknowledge the contributions from Jacqui Thomas and Peter Rostron at University of Sussex, Bob Barnes and Steve Phipps at the Environment Agency, and funding from TSB (Technology Strategy Board) and DTI Technology Programme, Project No TP/5/CON/6/I/H0065B.

**References**

Barnes B. Environmental Agency, personal communication of unpublished data

Blythe PT, Neasham,J, Sharif,B, Watson,P,Bell,MC, Edwards S, Suresh,V, Wagner, J and Bryan H., 2008. An environmental sensor system for pervasively monitoring road networks, IET - Road Transport Info. and Control 2008. Manchester: IEEE, 1-6.

Boon K.A., Ramsey M.H. McKenna S., and Yeo M., 2008. The use of measurement uncertainty to assess the reliability of on-site field test kits for the investigation of contaminated land. Proceedings of ConSoil 2008 (10th International UFZ-Deltares/TNO Conference on Soil-Water Systems), Milan, Italy, 3-6 June 2008 (ISBN: 978-3-00-024598-5). Theme C, 64-73

Boon K.A., Ramsey M.H. and Taylor P.D., 2007 Estimating and optimising measurement uncertainty in environmental monitoring: an example using six contrasting contaminated land investigations. Geostandards and Geoanalytical Research, 31, 237-249

Boon, K.A. and Ramsey, M.H. Judging the fitness of on-site measurements by their uncertainty, including the contributions from sampling and systematic effects. Environmental Science and Technology (submitted)

Ellison S.L.R., 2003. Excel add-in: amc\_freml.xla (version 0.1) - Linear Functional relationship estimation based on RSC AMC Technical Brief No.10 (http://www.rsc.org/images/brief10\_tcm18-25920.pdf ). Royal Society of Chemistry Analytical Methods Committee. Available at: http://www.rsc.org/Membership/Networking/InterestGroups/Analytical/AMC/Software/FREML.asp (Accessed 11/01/11)

Environment Agency, 2006. MCERTS Performance Standard for Laboratories Undertaking Chemical Testing of Soil, Version 3. Environment Agency, Preston, UK.

Environment Agency, 2009. Framework for the use of rapid measurement techniques (RMT) in the risk management of land contamination ISBN 978-1-84432-982-3 (page 18)

EPA, 1995. Superfund Innovative Technology Evaluation Program, Final Demonstration Plan for the Evaluation of Field-Portable X-Ray Fluorescence Technologies. US Environmental Protection Agency, Environmental Monitoring Systems Laboratory, Las Vegas, Nevada 89193

Finnamore J., Denton B. and Nathanail C. P., 2000. Land contamination - management of financial risk. Construction Industry Research and Information Association (CIRIA

Haines B., 2002. Zincophilic root foraging in Thlaspi caerulescens. New Phytologist 155: 363–372.

ISO, 1993. 3534-1 Statistics – Vocabulary and Symbols, International Organization for Standardization, Geneva

JCGM 200, 2008. International Vocabulary of Metrology – Basic and General Concepts and Associated Terms. VIM, 3rd edition). Joint Committee for Guides in Metrology.

Lyn, J.A., Ramsey, M.H., Fussel, R.J. and Wood, R., 2003. Measurement uncertainty from physical sample preparation: estimation including systematic error Analyst, 2003, 128 (11), 1391 – 1398

MAFF (Ministry of Agriculture, Fisheries & Food), 1986. The analysis of agricultural materials: A manual of the analytical methods used by the agricultural development and advisory service (3rd Edition). London: Ministry of Agriculture, Fisheries & Food. Her Majesty's Stationary Office.

Ramsey, M.H. and Argyraki A., 1997. Estimation of measurement uncertainty from field sampling: implications for the classification of contaminated land. Science of the Total Environment, 198, 243 – 257

Ramsey M.H., and Ellison S. L. R.,(eds.), 2007. Eurachem/EUROLAB/ CITAC/Nordtest/ AMC Guide: Measurement uncertainty arising from sampling: a guide to methods and approaches Eurachem ISBN 978 0 948926 26 6. (http://www.eurachem.org/guides/UfS\_2007.pdf)

Ramsey M.H., and Ellison S. L. R., 2010. Reply to Comments on EURACHEM/CITAC Guide “Measurement uncertainty arising from sampling” Accreditation and Quality Assurance: Journal for Quality, Comparability and Reliability in Chemical Measurement, 15:533-535

Ramsey M.H., Taylor P. D. and Lee J.C., 2002. Optimized contaminated land investigation at minimum overall cost to achieve fitness-for-purpose, Journal of Environmental Monitoring, 4, 5, 809 – 814

Ramsey, M.H., Thompson M. and Hale M., 1992. Objective Evaluation of Precision Requirements for Geochemical Analysis using Robust Analysis of Variance. Journal of Geochemical Exploration, 44, 23-36

Squire, S and Ramsey, M.H., 2001. Inter-organisational sampling trials for the uncertainty estimation of landfill gas measurements. Journal of Environmental Monitoring, 3, 3, 288 ­ 294

Taylor P.D., Ramsey M.H. and Potts P.J., 2004. Balancing measurement uncertainty against financial benefits: a comparison of in situ and ex situ analysis of contaminated land. Environmental Science and Technology 38, 6824-6831

Taylor P.D., Ramsey M.H. and Potts, P.J., 2005. Spatial contaminant heterogeneity: quantification with scale of measurement at contrasting sites. Journal of Environmental Monitoring. 7, 1364 – 1370

Thomas J.Y., Ramsey M.H., John E.A, and Barnes R., 2008. Quantification of in situ heterogeneity of contaminants in soil: a fundamental prerequisite to understanding processes controlling plant uptake. Proceedings of ConSoil 2008 (10th International UFZ-Deltares/TNO Conference on Soil-Water Systems), Milan, Italy, 3-6 June 2008 (ISBN: 978-3-00-024598-5). Theme C, 101-106

Thompson M and Fearn T., 1996. What exactly is fitness for purpose in analytical measurements? The Analyst 121, 275-278

Thompson M. and Ramsey M.H., 1995. Quality concepts and practices applied to sampling – an exploratory study. Analyst, 120, 261–270

Venkataraman R., Bronson F., Atrashkevich V., Field M., Young B.M., 2005. Improved detector response characterization method in ISOCS and LabSOCS. Journal of Radioanalytical and Nuclear Chemistry, 264, 1, 213 -219


## **Figure 1. Schematic map of test site, on a former golf course, showing the locations of the 20 sampling targets where both *in situ* measurement were made and samples taken for *ex situ* measurements in the laboratory. The areas marked in red are where the As concentration was formerly considered to be above the site specific action level of 59 mg/kg, based upon results of a previous survey.**

1:1


## **Figure 2. Estimation of ‘bias’ between *in situ* and *ex situ* measurements of arsenic concentration in the actual samples, allowing for the uncertainty in both sets of measurements. The model reveals a significant rotational ‘bias’ of -43** ± **14 % in the *in situ* measurements compared with the *ex situ* measurements, but a no significant translational ‘bias’.**

**False Positive**

**False Negative**

***Ex situ***

***In situ***

**Figure 3. Optimisation of the measurement uncertainty *of* arsenic for (a) *ex situ* false positive scenario and (b) *ex situ* false negative scenario, (c) *in situ* false positive scenario and (d) *in situ* false negative scenario. These show in cases (b) and (d) that the actual uncertainty ( ♦ ) is well above the optimal value ( ♦ ) that would produce the minimum overall cost (expressed as expectation of loss). The overall situation is very similar for both the *ex situ* and the *in situ* measurements, except that the cost for the *ex situ* analysis is approximately twice as high.**

**Figure 4. Map of the site showing probabilistic classification of whether the arsenic concentration exceeds the site specific assessment criteria (threshold) value of 59 mg/kg**

## **Table 1. Estimates of the expanded relative measurement uncertainty, and its sampling and analytical components, for the *ex situ* and *in situ* measurements of arsenic in the case study.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Robust mean****(mg/kg)** | **Usamp (%)**  | **Uanal (%)**  | **Umeas (%)**  |
| ***Ex situ* (Lab)**  | 100.8 | 109.8 | 2.9 | 109.8 |
| ***In situ* (PXRF)**  | 30.9 | 148.7 | 40.2 | 154.1 |

**Table 2. Input data for the optimisation of the uncertainty of both *in situ* and *ex situ* (lab-based) measurements using the OCLI method.**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Lab** ***ex situ*** | **PXRF *in situ*** | **Explanation**  |
| **Costs (£)**  |
| Sampling (Lsamp)  | 25 | Cost per sample  |
| Analysis (Lanal)  | 20 | 15 | Cost per sample.  |
| **Consequential Costs (C, in £)**  |
| False positive  | 1,360,000 | Potential cost per sample , e.g. unnecessary remediation. (top 1m dig & dump)  |
| False negative  | 1,000,000 | Potential cost per sample of litigation if contamination found after development (£20m for whole site (Finnamore et al, 2000)) |
| **Uncertainty (mg/kg)**  |
| Sampling (ssamp)  | 55.33 | 19.60 | From Robust ANOVA on measurements produced from balanced design. |
| Analytical (sanal)  | 1.48 | 13.56 |
| Measurement (smeas)  | 55.35 | 23.84 |
| **Threshold value (T) (mg/kg)**  | 59 | Site-specific action level for As. |
| **Contaminant concentration (cm) (mg/kg)**  |
| False positive  | 326.97 | 211.83 | Average of all locations “Probably over the action level”. |
| False negative  | 47.97 | 39.82 | Average of all locations “Possibly over the action level”. |

1. Fitness for purpose is defined as ‘the degree to which data produced by a measurement process enables a user to make technically and administratively correct decisions for a stated purpose’ (Thompson and Ramsey, 1995) [↑](#footnote-ref-2)