Measurement uncertainty: Approaches to the evaluation of uncertainties associated with recovery†

Vicki J. Barwick and Stephen L. R. Ellison

Laboratory of the Government Chemist, Queens Road, Teddington, Middlesex, UK TW11 0LY

Received 9th March 1999, Accepted 14th May 1999

A number of approaches for evaluating recovery and its contribution to uncertainty budgets for analytical methods are considered in detail. The recovery, *R*, for a particular sample is considered as comprising three elements, $\bar{R}_m$, R_s and R_{rep}. These relate to the recovery for the method; the effect of sample matrix and/or analyte concentration on recovery; and how well the behaviour of spiked samples represents that of test samples. The uncertainty associated with *R*, *u(R)*, will have contributions from $u(\bar{R}_m)$, $u(R_s)$ and $u(R_{rep})$. The evaluation of these components depends on the method scope and the availability, or otherwise, of representative certified reference materials. Procedures for evaluating these parameters are considered and illustrated with worked examples. Techniques discussed include the use of certified reference materials and spiking studies, and the use of extraction profiling to predict recoveries. All the approaches discussed evaluate the recovery and its uncertainty for the analytical method as a whole. It is concluded that this is a useful approach as it reduces the amount of experimental work required. In addition, most of the required data are frequently available from method validation studies.

Introduction

In recent years, the subject of the evaluation of measurement uncertainty in analytical chemistry has generated a significant level of interest and discussion.[1–6] It is generally acknowledged that the fitness for purpose of an analytical method cannot be assessed without some estimate of the measurement uncertainty to compare with the confidence required. Well characterised and controlled uncertainties are also fundamental to the implementation of traceability as a means of ensuring comparability of results; large uncertainties imply poor comparability. The *Guide to the Expression of Uncertainty in Measurement* (GUM) published by ISO[7] establishes general rules for evaluating and expressing uncertainty for a wide range of measurements. The guide was interpreted for analytical chemistry by Eurachem in 1995.[8] The approach described in the GUM requires the identification of all possible sources of uncertainty associated with the procedure; the estimation of their magnitude from either experimental or published data; and the combination of these individual uncertainties to give standard and expanded uncertainties for the procedure as a whole. Some applications of this approach to analytical chemistry have been published.[9,10] However, the GUM principles are significantly different from the methods currently used in analytical chemistry for estimating uncertainty[11–13] which generally make use of 'whole method' performance parameters, such as precision and recovery, obtained during in-house method validation studies or during method development and collaborative study.[14–16] In earlier papers we have illustrated the use of precision and recovery data in uncertainty estimates for a range of analytical techniques.[17–19] Though this establishes the principle of applying validation data to uncertainty estimation, the approach relies on estimation of uncertainties associated with recovery, including those associated with matrix change or incomplete extraction. Though recovery itself is routinely estimated during method validation, there is no general approach to the estimation of the uncertainty associated with recovery. This paper accordingly describes and illustrates some approaches to the evaluation of recovery and its uncertainty.

Theoretical basis

General approach

In this paper, recovery *R* is defined as the ratio c_{obs}/c_{ref} of observed concentration c_{obs} to a reference value c_{ref} for the particular material tested. If known, *R* could be used to correct an observation to an appropriate reference scale. Were such a correction made, it is clear that any uncertainty in *R* will contribute to uncertainties in the declared result.

R is, however, not usually obtained or considered obtainable for test samples. It is instead estimated indirectly, for example by experiments on related reference materials with a certified concentration, by comparison with an alternative definitive method, or by observing the amount of an added spike recovered from a sample matrix. In practice, measures are usually taken to ensure that the recovery is likely to be reasonably close to unity, and the assumption then made that $R = 1$. The main uncertainties associated with recovery arise from this assumption. To quantify the uncertainty, it is necessary to consider the degree to which a particular sample matrix under test is represented by the reference material employed and, where relevant, the extent to which spiking provides a representation of native analyte behaviour.

To treat these uncertainties explicitly, it is useful to consider the recovery *R* for a particular sample as comprising three components:

(i) $\bar{R}_M$ (so denoted because it is usually estimated as a mean of several determinations) is an estimate of the recovery obtained from, for example, the analysis of a CRM or a spiked sample. $\bar{R}_m$ may be considered as a 'reference' recovery, or more generally a 'method recovery' since it would normally be expected to apply to all determinations using the method, at least in a particular laboratory. The uncertainty in $\bar{R}_m$ is composed of the uncertainty in the nominal value (*e.g.*, the uncertainty in the certified value of a reference material) and the

uncertainty in the observed value (*e.g.*, the standard deviation of the mean of replicate analyses).

(ii) R_s is a correction factor to take account of differences in the recovery for a particular sample compared to the recovery observed for the material used to estimate $\bar{R}_m$.

(iii) R_{rep} is a correction factor to take account of the fact that a spiked sample may behave differently to a real sample with incurred analyte.

These three elements are combined multiplicatively to give an estimate of the recovery for a particular sample, *i.e*, $R = \bar{R}_m \times R_s \times R_{rep}$. It therefore follows that the uncertainty in R, $u(R)$, will have contributions from and $u(\bar{R}_m)$, $u(R_s)$ and $u(R_{rep})$.

How each of these components and their uncertainties are evaluated will depend on the method scope and the availability of reference materials. In the simplest case the method scope covers a single matrix type and analyte concentration for which a representative CRM is available. However, the situation is often more complex than this. The method scope may cover a range of matrices and/or analyte concentrations and there may be no suitable CRM available. The approaches we suggest for estimating recovery and the associated uncertainty for a range of situations are summarised in Table 1. The remainder of the paper first discusses the different strategies and presents the relevant calculations, then presents experimental illustrations of selected approaches.

Estimating $\bar{R}_m$ and $u(\bar{R}_m)$ using a representative CRM

If a representative reference material is available, $\bar{R}_m$ is estimated by comparing the mean of replicate analyses of the CRM with the certified value:

$$\bar{R}_m = \frac{\bar{C}_{obs}}{C_{CRM}} \quad (1)$$

where $\bar{C}_{obs}$ is the mean of the results from the replicate analysis of the CRM and C_{CRM} is the certified value for the CRM. The uncertainty associated with the estimate of, $\bar{R}_m$, $u(\bar{R}_m)$, is estimated by:

Table 1 Summary of suggested methods for evaluating recovery

	Recovery component		
Method scope/availability of CRMs	$\bar{R}_m$	R_s	R_{rep}
Single matrix and analyte concentration. Representative CRM available.	Determine $\bar{R}_m$ and $u(\bar{R}_m)$ from replicate analysis of the CRM.	Not applicable[a]	Not applicable[b]
Single matrix and analyte concentration. No representative CRM available.	Determine $\bar{R}_m$ by: Analysis of a representative matrix spiked at a representative concentration; or Comparison of result obtained for a typical sample with results obtained from a standard procedure; or Changing the extraction system (*e.g.*, using a stronger solvent) to see if any more of the analyte can be recovered; or Analysing a 'worst case' CRM. If a CRM is available which has a matrix that is known to be more difficult to extract the analyte from compared to the sample, it can be assumed that the recovery for the sample will be no worse than the recovery observed for the CRM.	Not applicable[a]	Evaluate how representative the spike is of the native material. Possible approaches include: Monitoring the extraction of spiked and native analytes with time; Comparing spiked recovery with the recovery from a non representative CRM.
Multiple matrices and/or analyte concentrations. One representative CRM available.	Determine $\bar{R}_m$ and $u\bar{R}_m$ from replicate analysis of the CRM.	Determine the recovery for a range of representative sample matrices spiked at representative concentrations. $u(R_s)$ is calculated from the spread of the recovery estimates.	Not applicable[b]
Multiple matrices and/or analyte concentrations. No representative CRM available.	Determine $\bar{R}_m$ and $u(\bar{R}_m)$ from analysis of representative sample matrices spiked at representative concentrations.	Estimate $u(R_s)$ from the data obtained in the calculation of $\bar{R}_m$.	Evaluate how representative the spike is of the native material. Possible approaches include: Monitoring the extraction of spiked and native analytes with time; Comparing spiked recovery with the recovery from a non representative CRM.

[a] As the method scope covers only a single matrix type and analyte concentration the estimate of $\bar{R}_m$ and its uncertainty can be based on the analysis of a sample which is truly representative of real samples. There is therefore no need to include a correction factor to take account of differences in recovery for a particular sample, compared to the sample used in the estimation of $\bar{R}_m$ (*i.e.*, R_s is assumed to equal 1 with negligible uncertainty). [b] If $\bar{R}_m$ is estimated from the analysis of a *representative* certified reference material it can be assumed that this will behave in a similar manner to incurred analyte in a real sample. A correction factor to take account of the fact that the recovery of the analyte from the material used to estimate $\bar{R}_m$ may not be representative of the recovery from a real sample is therefore unnecessary (*i.e.*, R_{rep} is implicitly assumed to equal 1 with negligible uncertainty).

$$u(\bar{R}_m) = \bar{R}_m \times \sqrt{\left(\frac{s_{obs}^2}{n \times \bar{C}_{obs}^2}\right) + \left(\frac{u(C_{CRM})}{C_{CRM}}\right)^2} \quad (2)$$

where s_{obs} is the standard deviation of the results from the replicate analyses of the CRM, n is the number of replicates and $u(C_{CRM})$ is the standard uncertainty in the certified value for the CRM. The contribution of $\bar{R}_m$ and its uncertainty to the combined uncertainty for the method depends on whether the recovery is significantly different from 1, and if so, whether or not a correction is made. This is discussed in detail later. If the reference material is only approximately representative of a typical sample, additional sources of uncertainty may need to be considered. These include effects of matrix or interferences in test samples which may differ from those in the CRM.

Estimating $\bar{R}_m$ and $u(\bar{R}_m)$ from spiking studies

In the absence of a suitable reference material, recovery is frequently estimated through spiking studies, *i.e.*, the addition of the analyte to a previously studied material. The spiked sample is prepared in such a way as to represent as closely as possible a natural sample with incurred analyte. A number of options are available. In the simplest case, a bulk sample of a suitable sample matrix known to be free from the analyte of interest is spiked with an appropriate concentration of the analyte. The bulk spiked sample is then analysed in replicate. $\bar{R}_m$ is given by:

$$\bar{R}_m = \frac{\bar{C}_{obs}}{C_{spike}} \quad (3)$$

where $\bar{C}_{obs}$ is the mean of the replicate analyses of the spiked sample and C_{spike} is the nominal concentration of analyte in the spiked sample. The uncertainty is estimated by:

$$u(\bar{R}_m) = \bar{R}_m \times \sqrt{\left(\frac{s_{obs}^2}{n \times \bar{C}_{obs}^2}\right) + \left(\frac{u(C_{spike})}{C_{spike}}\right)^2} \quad (4)$$

where s_{obs} is the standard deviation of the results from the replicate analyses of the spiked sample, n is the number of replicates and $u(C_{spike})$ is the standard uncertainty in the concentration of the spiked sample.

If no blank sample matrix is available, a bulk spiked sample can be prepared from a matrix which contains the analyte. The spiked sample is then analysed in replicate. $\bar{R}_m$ is given by:

$$\bar{R}_m = \frac{\bar{C}_{obs} - \hat{C}_{native}}{C_{spike}} \quad (5)$$

where $\hat{C}_{native}$ is the *observed* concentration of the analyte in the unspiked sample. Note that since we are concerned only with the difference between the spiked and unspiked observations, $\hat{C}_{native}$ does not have to represent the 'true' value of the concentration of the analyte in the unspiked matrix; eqn. (5) represents the change in observation divided by the change in concentration. The uncertainty is estimated by:

$$u(\bar{R}_m) = R_m \times \sqrt{\frac{s_{obs}^2/n + s_{native}^2}{(\bar{C}_{obs} - \hat{C}_{native})^2} + \left(\frac{u(C_{spike})}{C_{spike}}\right)^2} \quad (6)$$

where s_{native} is the standard deviation of the mean of the results of repeat analyses of the unspiked matrix.

If it is impractical to prepare a homogeneous bulk spiked sample for sub-sampling, then individual spiked samples can be prepared. If the spiked samples are prepared from approximately the same weight of a blank sample matrix, and the same weight of the spike is added to each sample, the recovery is given by:

$$\bar{R}_m = \frac{\bar{m}_{obs}}{m_{spike}} \quad (7)$$

where $\bar{m}_{obs}$ is the mean weight of the spike recovered from the samples and m_{spike} is the weight of the spike added to each sample. $u(\bar{R}_m)$ is therefore estimated by:

$$u(\bar{R}_m) = \bar{R}_m \times \sqrt{\frac{s_{m_{obs}}^2}{n \times \bar{m}_{obs}^2} + \left(\frac{u(m_{spike})}{m_{spike}}\right)^2} \quad (8)$$

where $s_{m_{obs}}$ is the standard deviation of the results obtained from the spiked samples, n is the number of spiked samples analysed and $u(m_{spike})$ is the uncertainty in the amount of spike added to each sample.

If the spiked samples are prepared from a sample matrix which contains the analyte the situation is somewhat more complex. The recovery for each sample, $R_{m(i)}$, is given by:

$$R_{m(i)} = \frac{C_{obs(i)} - \hat{C}_{native}}{C_{spike(i)}} \quad (9)$$

where $C_{obs(i)}$ is the concentration of the analyte observed for sample i, $\hat{C}_{native}$ the observed response for the unspiked sample as before [eqn. (5)], and $C_{spike(i)}$ is the concentration of the spike added to sample i. The mean recovery, $\bar{R}_m$, is given by:

$$\bar{R}_m = \frac{1}{n}\sum_{i=1}^{n} \frac{C_{obs(i)} - \hat{C}_{native}}{C_{spike(i)}} \quad (10)$$

Therefore:

$$\bar{R}_m = \frac{1}{n}\left[\sum_{i=1}^{n} \frac{C_{obs(i)}}{C_{spike+(i)}} - \hat{C}_{native}\sum_{i=1}^{n} \frac{1}{C_{spike(i)}}\right] \quad (11)$$

The uncertainty is calculated using the expression of the form:[7]

$$u[y(p,q,\ldots)] = \sqrt{\left(\frac{\partial y}{\partial p}\right)^2 [u(p)]^2 + \left(\frac{\partial y}{\partial q}\right)^2 [u(q)]^2 + \ldots} \quad (12)$$

Differentiating eqn. (11) gives:

$$\frac{\partial \bar{R}_m}{\partial C_{obs(i)}} = \frac{1}{n} \times \frac{1}{C_{spike(i)}} \quad \frac{\partial \bar{R}_m}{\partial \hat{C}_{native}} = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{C_{spike(i)}} \quad \frac{\partial \bar{R}_m}{\partial C_{spike(i)}}$$

$$= \frac{1}{n} \times \frac{(C_{obs(i)} - \hat{C}_{native})}{C_{spike(i)}^2}$$

$$u(\bar{R}_m)^2 = \sum_{i=1}^{n}\left[\frac{1}{n} \times \frac{1}{C_{spike(i)}} \times u(C_{obs(i)})\right]^2 + \left[\frac{1}{n}\sum_{i=1}^{n} \frac{1}{C_{spike(i)}}\right]^2 \times u(\hat{C}_{native})^2 + \sum_{i=1}^{n}\left[\frac{1}{n} \times \frac{(C_{obs(i)} - \hat{C}_{native})}{C_{spike(i)}^2} \times u(C_{spike(i)})\right]^2 \quad (13)$$

Under certain experimental conditions, eqn. (13) can be simplified. Firstly, if $u(C_{spike(i)}) << u(C_{obs(i)})$ and $u(C_{native})$ the expression becomes:

$$u(\bar{R}_m) = \frac{1}{n}\sqrt{\sum_{i=1}^{n}\left(\frac{u(C_{obs(i)})}{C_{spike(i)}}\right)^2 + \left(\sum_{i=1}^{n} \frac{1}{C_{spike(i)}}\right)^2 u(\hat{C}_{native})^2} \quad (14)$$

This is often the case, as spiking is generally achieved by adding an aliquot of a solution or a known weight of the analyte. The uncertainties associated with such operations are usually small compared to the uncertainties associated with the observation of the amount of the analyte in a sample [*i.e.*, $u(C_{\mathrm{obs}(i)})$ and $u(\hat{C}_{\mathrm{native}})$]). Furthermore, if the standard deviation of the $C_{\mathrm{spike}(i)}$ values is small compared to the mean of the $C_{\mathrm{spike}(i)}$ values, $\bar{C}_{\mathrm{spike}}$ can be used in the calculation. Eqn. (14) therefore simplifies further to:

$$u(\bar{R}_{\mathrm{m}}) = \frac{1}{n} \times \frac{1}{\bar{C}_{\mathrm{spike}}} \sqrt{\sum_{i=1}^{n} u(C_{\mathrm{obs}(i)})^2 + n^2 \times u(\hat{C}_{\mathrm{native}})^2} \quad (15)$$

This is likely to be the case in recovery studies at a single level using similar quantities of the sample matrix in the preparation of each spiked sample. Finally, if the estimates of $u(C_{\mathrm{obs}(i)})$ are all similar, the mean can be used. This leads to:

$$u(\bar{R}_{\mathrm{m}}) = \frac{1}{\bar{C}_{\mathrm{spike}}} \sqrt{\frac{\bar{u}(C_{\mathrm{obs}(i)})^2}{n} + u(\hat{C}_{\mathrm{native}})^2} \quad (16)$$

Again, this is likely to be the case when each sample is spiked at the same concentration so that all the $C_{\mathrm{obs}(i)}$ values are of similar magnitude.

Estimating $\bar{R}_{\mathrm{m}}$ and $u(\bar{R}_{\mathrm{m}})$ by comparison with a standard method

An alternative approach to estimating $\bar{R}_{\mathrm{m}}$ is by comparison with the results obtained from a standard method of known uncertainty. A representative sample is analysed, in replicate, using both the method under evaluation and the standard method. $\bar{R}_{\mathrm{m}}$ is given by:

$$\bar{R}_{\mathrm{m}} = \frac{\bar{C}_{\mathrm{method}}}{\bar{C}_{\mathrm{standard}}} \quad (17)$$

where $\bar{C}_{\mathrm{method}}$ is the mean of the results obtained using the method under consideration and $\bar{C}_{\mathrm{standard}}$ is the mean of the results obtained using the standard method. The uncertainty in the recovery, $u(\bar{R}_{\mathrm{m}})$, is therefore estimated by:

$$u(\bar{R}_{\mathrm{m}}) = \bar{R}_{\mathrm{m}} \times \sqrt{\left(\frac{s^2_{\mathrm{method}}}{n \times \bar{C}^2_{\mathrm{method}}}\right) + \left(\frac{u(C_{\mathrm{standard}})}{\bar{C}_{\mathrm{standard}}}\right)^2} \quad (18)$$

where s_{method} is the standard deviation of the results obtained using the method, n is the number of replicates and $u(C_{\mathrm{standard}})$ is the standard uncertainty associated with the standard method.

Alternative approaches to estimating $\bar{R}_{\mathrm{m}}$ and $u(\bar{R}_{\mathrm{m}})$

In the absence of appropriate CRMs or standard methods, and if preparing spiked samples is impractical, alternative methods of investigating the recovery are required. However, such techniques generally require an element of judgement on the part of the analyst and can often only be used as an initial indication of the uncertainty associated with method recovery. If the results of such a study indicate that the uncertainties associated with recovery are a significant contribution to the uncertainty budget, further investigation will be required to obtain a better estimate. The main techniques available include repeated extraction experiments, monitoring the progress of extraction with time, and analysis of 'worst case' materials. These approaches are discussed in turn below.

Repeated extraction. Samples are re-extracted either under the same experimental conditions, or preferably with a more vigorous extraction system (*e.g.*, a more polar extraction solvent). The amount of analyte extracted under the normal application of the method is compared with the total amount extracted (amount extracted initially plus the amount extracted by subsequent re-extractions). $\bar{R}_{\mathrm{m}}$ is the ratio of these estimates. If re-extraction was achieved using the same conditions as the initial extraction, the difference between the true recovery and the assumed value of 1 is known to be at least $1 - \bar{R}_{\mathrm{m}}$. The difference could be greater, as repeated extractions under the same experimental conditions may not quantitatively recover all of the analyte from the sample. In such cases we estimate the uncertainty, $u(\bar{R}_{\mathrm{m}})$, associated with the assumed value of $\bar{R}_{\mathrm{m}} = 1$ (*i.e.*, perfect recovery) as $(1 - \bar{R}_{\mathrm{m}})$.

If repeat extractions were carried out using a more vigorous extraction system, there is greater confidence associated with the observed difference between $\bar{R}_{\mathrm{m}}$ and the assumed value of 1. This is because it is more likely that the repeat extractions will have quantitatively extracted the remainder of the analyte from the sample, thus giving greater confidence in the estimate of $\bar{R}_{\mathrm{m}}$. In such cases we estimate $u(\bar{R}_{\mathrm{m}})$ as, $(1 - \bar{R}_{\mathrm{m}})/k$ where k is the coverage factor which will be used to calculate the expanded uncertainty.

Monitoring extraction with time. For some methods, it may be possible to build up an extraction profile for the method and use it to predict how close the extraction is to completion. A procedure for doing this in supercritical fluid extraction (SFE) has been described by Bartle *et al.*[20] The prediction relies on the extraction profile following an approximately exponential form after an initial rapid extraction. If the extraction is carried out for at least as long as the initial non-exponential period to obtain a mass of extracted analyte, m_1, followed by extraction over two subsequent equal time periods to obtain masses of analyte m_2 and m_3, then m_0, the total mass of the analyte in the sample, is given by:

$$m_0 = m_1 + \frac{m_2^2}{m_2 - m_3} \quad (19)$$

To estimate $\bar{R}_{\mathrm{m}}$ the mass extracted during the normal application of the method is compared with the predicted total mass m_0. The uncertainty associated with $\bar{R}_{\mathrm{m}}$ will have contributions from the uncertainty associated with the observed mass (standard deviation of the mean of n observations) and the uncertainty associated with the prediction of m_0.

Analysis of a worst case CRM. If a CRM is available which has a matrix known to provide an extreme example (*i.e.*, more difficult to extract the analyte from than test samples), the recovery observed for the CRM can provide a worst case estimate on which to base the recovery for real samples. The recovery observed from replicate analyses of the CRM is denoted R_{CRM}. Since the CRM matrix is known to be more difficult to extract the analyte from, it is reasonable to assume that recoveries for test samples are more likely to be closer to 1 than to R_{CRM}. It is therefore appropriate to consider R_{CRM} as representing the lower limit of a triangular distribution.[7] As a first estimate, $\bar{R}_{\mathrm{m}}$ is assumed to equal 1, with an uncertainty, $u(\bar{R}_{\mathrm{m}})$, of:

$$u(\bar{R}_{\mathrm{m}}) = \frac{1 - R_{\mathrm{CRM}}}{\sqrt{6}} \quad (20)$$

Note that if there is no evidence to suggest where in the range $1 - R_{\mathrm{CRM}}$ the recovery for test samples is likely to lie, a rectangular distribution should be assumed. $u(\bar{R}_{\mathrm{m}})$ is then estimated by $(1 - \bar{R}_{\mathrm{m}})/\sqrt{3}$.

Evaluating the contribution of $\bar{R}_m$ to $u(R)$ when significance tests are used

Recovery is often tested for evidence of significant difference from 1.0 (100%). In such circumstances, the contribution of $\bar{R}_m$ and its uncertainty to the overall uncertainty for the method will depend on whether it is found to be significantly different from 1, and if so, whether or not a correction is made. General rules for calculating uncertainty estimates for these different circumstances have been discussed in detail elsewhere.[21] Here, we summarise them for an estimate of $\bar{R}_m$ and its uncertainty, assuming that significance is checked by comparison of a statistic $t = |\bar{R}_m - 1|/u(\bar{R}_m)'$ with a critical value t_{crit}. Identical principles hold for estimation of uncertainties associated with any other recovery component subjected to a significance test. Three cases arise:

1. $\bar{R}_m$, taking into account $u(\bar{R}_m)'$, is not significantly different from 1 so no correction to the final result is applied. Eqn. (21) applies:†

$$u(\bar{R}_m) = \frac{t_{crit} \times u(\bar{R}_m)'}{1.96} \quad (21)$$

where $u(\bar{R}_m)$ is the required standard uncertainty associated with the estimate (1.0) of $\bar{R}_m$.

2. $\bar{R}_m$, taking into account $u(\bar{R}_m)'$, is significantly different from 1 and a correction to the final result is applied. $u(\bar{R}_m)$ is again given by eqn. (21).

3. $\bar{R}_m$, taking into account $u(\bar{R}_m)$, is significantly different from 1 but a correction to the final result is not applied. The standard uncertainty is increased to ensure that the range quoted will include the true value using eqn. (22):

$$u(\bar{R}_m) = \sqrt{\left(\frac{1-\bar{R}_m}{k}\right)^2 + u(\bar{R}_m)'^2} \quad (22)$$

where k is the coverage factor that will be used in the calculation of the expanded uncertainty.

A fourth special case applies to empirical methods. An empirical method is a standardised method agreed upon for the purposes of comparative measurement within a particular field of application; the measurand is accordingly defined by the method. In such cases the recovery is arbitrarily defined as unity and the uncertainty associated with it as zero. Note, however, that a particular laboratory implementation of the method will generally need verification, and where a reference material is used to check the local performance, the above considerations will apply.

Estimating R_s and $u(R_s)$ from spiking studies

Where the method scope covers a range of sample matrices and/or analyte concentrations, an additional uncertainty term is required to take account of differences in the recovery of a particular sample type, compared to the material used to estimate $\bar{R}_m$. This can be evaluated by analysing a representative range of spiked samples, covering typical matrices and analyte concentrations, in replicate. The mean recovery for each sample type is calculated. R_s is normally implicitly assumed to be equal to 1. However, there will be an uncertainty associated with this assumption, which appears in the spread of mean recoveries observed for the different spiked samples (strictly, from the between-matrix component of variance, but in practice, the dispersion of mean values usually provides a reasonable estimate). The uncertainty, $u(R_s)$, is therefore the standard deviation of the mean recoveries for each sample type. Where R_s differs significantly from 1.0, an additional allowance should be made as in eqn. (22).

Estimating R_{rep} and $u(R_{rep})$

R_{rep} is generally assumed to equal one, indicating that the recovery from a spiked sample perfectly represents the recovery observed for incurred analyte. The uncertainty $u(R_{rep})$ is a measure of the uncertainty associated with that assumption, *i.e.*, how different R_{rep} might be from the assumed value of 1. The complexity of evaluating how well a spike represents the behaviour of native material varies from matrix to matrix and with the method being studied. In some cases it can be argued that a spike is a good representation of a real sample, for example in liquid samples where the analyte is simply dissolved in the matrix. In addition, if the method involves total dissolution or destruction of the matrix, for example by ashing, there may be no reason to believe that a spike would behave any differently from the incurred analyte. However, problems arise for more complex matrices and where the method involves extraction rather than total destruction or dissolution. Possible approaches to investigating the performance of spiked *versus* real samples include monitoring the extraction of spiked and native analytes with time, and comparison of spiked recovery with the recovery from a less representative CRM. However, these may not be appropriate in all cases. If the analyst cannot obtain any experimental evidence on the appropriateness of spiking, then judgements and/or assumptions have to be made. Ideally, R_{rep} should be evaluated by the analysis of a reference material (even if it is not directly comparable to the test samples) and by comparing the recovery obtained with those observed from the spiking studies. The uncertainty $u(R_{rep})$ is then estimated as:

$$u(R_{rep}) = \sqrt{\left(\frac{1-R_{rep}}{k}\right)^2 + (u(R_{rep})')^2} \quad (23)$$

where k is the coverage factor which will be used to calculate the expanded uncertainty and $u(R_{rep})'$ is the uncertainty associated with the estimate of R_{rep}.

The most straightforward approach is to spike the CRM and compare the recovery observed with that observed from the analysis of the unspiked reference material. In such cases R_{rep} is given by:

$$R_{rep} = \frac{\bar{C}_{obs(spike)} - \bar{C}_{obs(CRM)}}{C_{spike}} \times \frac{C_{CRM}}{\bar{C}_{obs(CRM)}} \quad (24)$$

where $\bar{C}_{obs(spike)}$ is the mean concentration observed from replicate analyses of the spiked CRM, $\bar{C}_{obs(CRM)}$ is the mean concentration observed from replicate analyses of the unspiked CRM, C_{spike} is the concentration of the spike added and C_{CRM} is the certified concentration of the reference material. The uncertainty, $u(R_{rep})'$, is obtained by differentiating eqn. (24) and applying eqn. (12):

$$u(R_{rep})' = R_{rep} \times \sqrt{\left(\frac{u(\bar{C}_{obs(spike)})}{\bar{C}_{obs(spike)} - \bar{C}_{obs(CRM)}}\right)^2 + \left(\frac{u(\bar{C}_{obs(CRM)}) \times u(\bar{C}_{obs(spike)})}{\bar{C}_{obs(CRM)} \times (\bar{C}_{obs(CRM)} - \bar{C}_{obs(spike)})}\right)^2 + \left(\frac{u(C_{CRM})}{C_{CRM}}\right)^2 + \left(\frac{u(C_{spike})}{C_{spike}}\right)^2} \quad (25)$$

† Note that the standard uncertainty is actually given by $u(\bar{R}_m)$. The multiplication by $t_{crit}/1.96$ effectively increases the estimate slightly to allow for small numbers of degrees of freedom, which would otherwise need to be considered in forming the combined expanded uncertainty.[7]

In this case, we are only interested in the dispersion of results obtained for the mean values $\overline{C}_{obs(spike)}$ and $\overline{C}_{obs(CRM)}$. The corresponding uncertainties are therefore estimated by the standard deviation of the mean of the observed concentrations in each case. Note that the above equation holds if the spiking study was based on the replicate analysis of a single spiked sample of the CRM. If the study was based on the analysis of a number of individual portions of the CRM (of similar weight), all spiked at a similar concentration, eqns. (24) and (25) are modified slightly:

$$R_{rep} = \frac{\overline{C}_{obs(spike)} - \overline{C}_{obs(CRM)}}{\overline{C}_{spike}} \times \frac{C_{CRM}}{\overline{C}_{obs(CRM)}} \quad (26)$$

$$u(R_{rep})' = R_{rep} \times \sqrt{\left(\frac{\overline{u}(C_{obs(spike)})}{\overline{C}_{obs(spike)} - \overline{C}_{obs(CRM)}}\right)^2 + \left(\frac{u(\overline{C}_{obs(CRM)}) \times u(\overline{C}_{obs(spike)})}{\overline{C}_{obs(CRM)} \times (\overline{C}_{obs(CRM)} - \overline{C}_{obs(spike)})}\right)^2 + \left(\frac{u(C_{CRM})}{C_{CRM}}\right)^2 + \left(\frac{\overline{u}(C_{spike})}{C_{spike}}\right)^2} \quad (27)$$

where $\overline{u}(C_{obs(spike)})$ is the average of the uncertainties associated with each of the $C_{obs(spike)}$ values divided by the square root of the number of determinations of $C_{obs(spike)}$, $\overline{C}_{spike}$ is the average of the concentrations of the spike added to each sample and $\overline{u}(C_{spike})$ is the average of the uncertainties associated with each of the C_{spike} values. This approach is illustrated in the Experimental section.

If there is no CRM available then the analyst will have to make a judgement based on the information available.

Calculating *R* and *u*(*R*)

The recovery for a particular sample, R, is given by $R = \overline{R}_m \times R_s \times R_{rep}$. However, since R_s and R_{rep} are generally assumed to equal 1, $R = \overline{R}_m$. The value of $\overline{R}_m$ used depends on whether or not it is significantly different from 1, and if so, whether a correction to the result for a particular sample is applied. The uncertainty associated with R, $u(R)$ is estimated by:

$$u(R) = R \times \sqrt{\left(\frac{u(\overline{R}_m)}{\overline{R}_m}\right)^2 + \left(\frac{u(R_s)}{R_s}\right)^2 + \left(\frac{u(R_{rep})}{R_{rep}}\right)^2} \quad (28)$$

However, if $R_s = R_{rep} = 1$, eqn. (28) simplifies to:

$$u(R) = \overline{R}_m \times \sqrt{\left(\frac{u(\overline{R}_m)}{R_m}\right)^2 + u(R_s)^2 + u(R_{rep})^2} \quad (29)$$

Experimental

Polycyclic aromatic hydrocarbon (PAH) extraction studies

The study was based on the analysis of a coal carbonisation site soil reference material (RM), LGC RM 6138 (Office of Reference Materials, LGC (Teddington), Middlesex, UK). The 16 analytes are listed in Table 2, together with the relevant reference values and their associated uncertainties. 20 portions of the material were analysed using the method outlined below, nine of which were spiked with a solution of the 16 target PAHs. The spiking solution had a nominal concentration of 200 μg m^{-1} and was prepared from a 2000 μg ml^{-1} stock solution (Supelco, Bellefonte, PA, USA). The actual (stated) concentrations of each of the analytes in the spiking solution are presented in Table 2. 0.5 ml of the spike was added to approximately 10 g reference material before the addition of the deuterated surrogate recovery standards. The spiked samples were otherwise treated in exactly the same way as the unspiked samples.

The method studied is used for the determination of PAHs in soils samples.[22] In normal use, air dried soil samples are soxhlet extracted with dichloromethane for six hours. Prior to extraction, 0.5 ml of a solution of deuterated PAHs (naphthalene-d_8, acenaphthene-d_{10}, phenanthrene-d_{10}, chrysene-d_{12}, perylene-d_{12}) with a nominal concentration of 80 μg l^{-1}, are added as surrogate recovery standards. The PAHs are quantified by capillary gas chromatography-mass spectrometry (GC-MS). An internal standard, 4-terphenyl-d_{14}, is added to the sample and standard solutions prior to analysis by GC-MS. Calibration

Table 2 Reference and spiked analyte concentrations in RM LGC 6138—Coal Carbonisation Site Soil l

	CRM certified values		Spiking levels			
Analyte	Reference value, C_{CRM} mg kg^{-1} [a]	Uncertainty/ mg kg^{-1}	Stock solution concentration/ μg ml^{-1} [c]	Standard deviation/ μg ml^{-1} [c]	Spiking solution concentration/ μg ml^{-1}	Amount of analyte added to sample/mg
Naphthalene	32	4.0	2054	9.4	205	0.103
Acenaphthylene	7.0	1.4	1998	1.2	200	0.100
Acenaphthene	6.4	0.8	1992	2.4	199	0.100
Fluorene	15.3	1.4	1983	10.2	198	0.099
Phenanthrene	114	7.0	1974	18.9	197	0.099
Anthracene	22	3.0	1985	7.1	199	0.099
Fluoranthene	118	8.0	1990	9.2	199	0.100
Pyrene	103	6.0	1991	1.5	199	0.100
Benzo[*a*]anthracene	42	4.0	1987	7.5	199	0.099
Chrysene	44	5.0	1983	11.8	198	0.099
Benzo[*b*]fluoranthene	42	6.0	1987	8.1	199	0.099
Benzo[*k*]fluoranthene	21	5.0	1991	5.4	199	0.100
Benzo[*a*]pyrene	36	5.0	1989	0.8	199	0.099
Indeno[1,2,3-*cd*]pyrene	25	2.0	1992	14	199	0.100
Dibenz[*a*,*h*]anthracene	7.6	4.0	1998	3.8	200	0.100
Benzo[*g*,*h*,*i*]perylene	28	4.0	1993	8.4	199	0.100

[a] RM 6138. Robust mean value (median) of the results, on a dry sample weight basis. [b] RM 6138. The uncertainty quoted is the half width of the 95% confidence interval based on the robust standard deviation of the results. [c] Values quoted by supplier (Supelco).

curves are obtained for each PAH from standards prepared from a stock solution with a nominal concentration of 2000 μg ml^{-1}.

Results and discussion

Investigation of spiked *versus* native recoveries

The aim of this study was to investigate how representative spiked recoveries are of extraction of the native analyte and to determine the uncertainty associated with estimating recovery through spiking.

Results. The values of $\bar{R}_m$ calculated from replicate analysis of the RM using eqn. (1) are presented in Table 3, together with the corresponding estimates of $\bar{R}_m$ obtained from the analysis of the spiked samples calculated using eqn. (10). In the latter case, C_{native} was taken as the mean concentration observed during the analysis of the unspiked reference material. *t*-tests[23] were performed to compare the two estimates of recovery obtained for each analyte; the results are also included in Table 3. The results for acenaphthylene, acenaphthene, fluorene, anthracene, benzo[*b*]fluoranthene, indeno[1.2.3-*cd*]pyrene and dibenz[*a*,*h*]-anthracene indicated a significant difference (at the 95% confidence level) between the recovery estimates obtained from the analysis of the RM and those obtained from the analysis of spike samples. In the case of acenaphthylene, acenaphthene and fluorene the difference is due to the unusually high (in the case of acenaphthylene) or low estimates of recovery obtained from the analysis of the RM.

Using the approach discussed previously, R_{rep} and $u(R_{rep})$ were calculated for each of the analytes by applying eqns. (26) and (27) respectively. The relevant data, and the resulting estimates of R_{rep} and $u(R_{rep})$ are presented in Table 4. The estimates of $u(C_{obs(spike)})$ and $u(C_{obs(CRM)})$ were based on the observed relative standard deviations obtained from the replicate analysis of RM 6138 (see Table 3). The values for $u(C_{spike})$ were calculated using the data given in Table 2, as described in the following section. Note that the values of $u(R_{rep})$ form a

Table 3 Estimates of recovery obtained from the replicate analysis of RM 6138

	Native recovery[a]			Spike recovery[b]		
Analyte	$\bar{C}_{obs}$/ mg kg^{-1}	s_{obs}/ mg kg^{-1}	$\bar{R}_m$	$\bar{R}_m$	$s(R_m)$	*t*-test result p,(v)[c]
Naphthalene	29.5	2.1	0.920	0.876	0.13	0.33 (18)
Acenaphthylene	12.3	1.7	1.915	0.863	0.063	8.3×10^{-8} (11)
Acenaphthene	3.30	0.25	0.471	0.825	0.043	8.3×10^{-14} (18)
Fluorene	8.86	0.63	0.579	0.837	0.048	1.5×10^{-10} (18)
Phenanthrene	101	5.7	0.884	1.041	0.29	0.15 (8)
Anthracene	23.6	2.0	1.070	0.969	0.099	0.027 (18)
Fluoranthene	108	7.0	0.920	0.959	0.38	0.77 (8)
Pyrene	89.3	6.2	0.867	0.955	0.29	0.40 (9)
Benzo[*a*]anthracene	37.8	2.9	0.899	0.884	0.15	0.79 (11)
Chrysene	40.9	3.1	0.929	0.878	0.15	0.38 (11)
Benzo[*b*]fluoranthene	30.9	2.6	0.737	0.824	0.11	0.043 (18)
Benzo[*k*]fluoranthene	15.9	1.0	0.758	0.765	0.19	0.91 (9)
Benzo[*a*]pyrene	30.4	2.7	0.845	0.825	0.13	0.68 (18)
Indeno[1.2.3-*cd*]pyrene	25.2	2.3	0.972	0.688	0.19	8.9×10^{-6} (18)
Dibenz[*a*,*h*]anthracene	5.63	0.41	0.740	0.644	0.057	0.0011 (18)
Benzo[*g*,*h*,*i*]perylene	24.0	2.6	0.856	0.818	0.10	0.38 (18)

[a] Using eqn. (1), with 11 determinations for all analytes. [b] Estimates of recovery obtained from the analysis of samples of RM 6138 spiked with approximately 10 mg kg^{-1} PAHs , calculated using eqn. (10) with 9 determinations for all analytes. [c] *p*-value for 2-tailed *t*-test (Null hypothesis: equal recoveries; alternative hypothesis: unequal recoveries). Where an *F*-test showed significant variance difference, an unequal variance *t*-test was applied. The figure in parentheses is the number of freedom (equal variance test) or effective degrees of freedom (unequal variance test) for the test statistic.

Table 4 Estimates of R_{rep} and $u(R_{rep})$ obtained from the comparison of the recovery of spiked and native PAHs[a]

Analyte	$\bar{C}_{obs(spike)}$ (mg kg^{-1})	$\bar{C}_{obs(CRM)}$ (mg kg^{-1})	$\bar{C}_{spike}$ (mg kg^{-1})	R_{rep}	$\bar{u}(C_{obs(spike)})$ (mg kg^{-1})	$\bar{u}(C_{obs(CRM)})$ (mg kg^{-1})	$\bar{u}(C_{spike})$ (mg kg^{-1})	$u(R_{rep})'$	$\lvert 1 - R_{rep}\rvert/k$	$u(R_{rep})$
Naphthalene	38.39	29.46	10.20	0.952	0.896	0.630	0.0877	0.143	0.0241	0.145
Acenaphthylene	20.82	12.26	9.92	0.450	0.972	0.526	0.0695	0.0753	0.275	0.285
Acenaphthene	11.46	3.30	9.89	1.75	0.306	0.076	0.0703	0.199	0.375	0.424
Fluorene	17.10	8.86	9.85	1.45	0.399	0.190	0.0857	0.117	0.222	0.251
Phenanthrene	110.97	100.78	9.81	1.18	2.219	1.170	0.117	0.338	0.0882	0.349
Anthracene	33.10	23.55	9.86	0.906	0.883	0.598	0.0779	0.132	0.0478	0.140
Fluoranthene	117.98	108.51	9.88	1.04	2.360	2.100	0.0830	0.363	0.0208	0.364
Pyrene	98.77	89.34	9.89	1.10	2.305	1.866	0.0692	0.362	0.0494	0.365
Benzo[*a*]anthracene	46.49	37.77	9.87	0.983	1.240	0.873	0.0790	0.191	0.00889	0.191
Chrysene	49.50	40.87	9.85	0.945	1.155	0.921	0.0906	0.184	0.0279	0.186
Benzo[*b*]fluoranthene	39.06	30.94	9.87	1.12	1.042	0.792	0.0799	0.215	0.0590	0.223
Benzo[*k*]fluoranthene	23.47	15.91	9.89	1.01	0.548	0.315	0.0742	0.156	0.00439	0.156
Benzo[*a*]pyrene	38.58	30.43	9.88	0.976	1.157	0.828	0.0692	0.200	0.0120	0.200
Indeno[1.2.3-*cd*]pyrene	32.07	25.27	9.89	0.708	0.962	0.701	0.0980	0.147	0.146	0.208
Dibenz[*a*,*h*]anthracene	12.02	5.63	9.92	0.870	0.280	0.122	0.0724	0.128	0.0648	0.144
Benzo[*g*,*h*,*i*]perylene	32.06	23.96	9.90	0.956	1.069	0.772	0.0812	0.189	0.0225	0.190

[a] The study involved the analysis of 9 individual spiked samples which gave results $C_{obs(spike(i))}$ each with a standard deviation $s_{obs(spike(i))}$. $\bar{C}_{obs(spike)}$ is the mean of the $C_{obs(spike(i))}$ values and $\bar{u}(C_{obs(spike)})$ is the mean of the $s_{obs(spike(i))}$ values divided by $\sqrt{9}$. $\bar{C}_{obs(CRM)}$ is the mean of the results of 11 analyses of the RM and $u(\bar{C}_{obs(CRM)})$ is the standard deviation of the mean.

major part of the estimated uncertainty, and are typically much larger than the reference uncertainties.

Calculation of $\bar{R}_m$ and $u(\bar{R}_m)$ from spiking studies

This study illustrates the determination of $\bar{R}_m$ and $u(\bar{R}_m)$ from the analysis of portions of a sample matrix containing the analyte, spiked at an appropriate concentration of the analyte. Approximately 10 g samples of soil reference material LGC RM 6138 were spiked with 0.5 ml of a 205 μg ml^{-1} solution of naphthalene in dichloromethane. Previous analyses of the reference material had a mean of 29.5 mg kg^{-1} with a standard deviation of the mean of 0.63 mg kg^{-1} (n = 11). These values correspond to C_{native} and $u(C_{native})$ in eqns. (9) and (13) respectively. The uncertainty in the concentration of the spike added has contributions from the uncertainties associated with the concentration of the stock solution and the volumetric glassware used to prepare and add the spiking solution. The uncertainty in the concentration of the stock solution was quoted by the supplier (Supelco, Bellefonte, PA, USA) as 0.005 as a relative standard deviation. Based on previous experience in our laboratory, the combined uncertainties associated with the volumetric glassware were estimated as 0.007 (as a relative standard deviation). Combining these elements using root sum of squares gives an uncertainty in the concentration of the spike added of 0.0086 as a relative standard deviation. This value was therefore used to calculate $u(C_{spike(i)})$ in eqn. (13). Note that there is also a contribution from the weight of sample taken, however previous work has shown that such uncertainties are generally insignificant. The additional uncertainty has therefore not been included. Based on the results of earlier studies of the method, precision was estimated as 0.07 as a relative standard deviation. This value was used to calculate the $u(C_{obs(i)})$ values in eqn. (13). The relevant results are presented in Table 5.

Using eqn. (10), $\bar{R}_m$ was calculated as 0.876. Applying eqn. (13), $u(\bar{R}_m)$ was calculated as 0.1074. However in this case, simplifications can be applied, as discussed previously. Firstly, the estimates of $u(C_{spike(i)})$ are much smaller than $u(C_{obs(i)})$ and $u(C_{native})$; typically 0.09 mg kg^{-1} compared to 2.7 mg kg^{-1} and 0.63 mg kg^{-1} respectively. Eqn. (14) can therefore be applied. This gives an estimate for $u(\bar{R}_m)$ of 0.1074. In addition, since the same amount of spiking solution was added to each sample, and the weights of each sample were similar, the standard deviation of the $C_{spike(i)}$ values (0.072 mg kg^{-1}) is small compared to the mean of the $C_{spike(i)}$ values (10.20 mg kg^{-1}). The mean can therefore be used and eqn. (15) applied. This gives an estimate for $u(\bar{R}_m)$ of 0.1074. Finally, the estimates of the uncertainty associated with $C_{obs(i)}$ are all similar (standard deviation of $u(C_{obs(i)})$ = 0.092) so eqn. (16) can be applied. This leads to an estimate for $u(\bar{R}_m)$ of 0.1073. This example illustrates that when the above assumptions hold a relatively simple calculation can be used to obtain an estimate of $u(\bar{R}_m)$.

Table 5 Data from the analysis of samples of RM 6138 spiked with naphthalene

Sample no.	Weight/g	$C_{obs(i)}$/ mg kg^{-1}	$C_{spike(i)}$ / mg kg^{-1}	$u(C_{obs(i)})$/ mg kg^{-1}	$u(C_{spike(i)})$/ mg kg^{-1}
1	10.02	37.21	10.25	2.605	0.0881
2	10.01	39.90	10.26	2.793	0.0882
3	10.04	40.16	10.23	2.811	0.0880
4	10.15	39.29	10.12	2.750	0.0870
5	10.20	36.36	10.07	2.545	0.0866
6	10.10	39.27	10.17	2.749	0.0874
7	9.98	38.12	10.29	2.668	0.0885
8	10.07	37.51	10.20	2.626	0.0877
9	10.03	37.73	10.24	2.641	0.0881

Estimation of $\bar{R}_m$ and $u(\bar{R}_m)$ from extraction monitoring studies

The aim of this study was to determine whether it would be possible to predict the total concentration of the analytes in the sample using data obtained after the usual 6 h extraction period. An estimate of the method recovery could then be obtained by comparing the amount extracted after 6 h with the predicted total amount present in the sample.

Two portions of RM 6138 were prepared for analysis as described previously. The samples were extracted for a total of 14 h (two 7 h periods on consecutive days). After 1, 2, 4 and 6 soxhlet cycles, and hourly intervals thereafter, small aliquots of the extraction solvent were removed for analysis. After 7 h the extraction was halted and left overnight. The extraction was then restarted and aliquots removed at hourly intervals for a further 7 h. For comparison, a further two samples were extracted for 14 h, after which time an aliquot of the extraction solvent was removed and submitted for analysis by GC-MS.

Based on the concentrations observed after extracting the samples for 4, 5 and 6 h, eqn. (19) was used to calculate the total analyte concentration in the sample. The results are summarised in Table 6. The results obtained after 14 h extraction are also included for comparison.

The results indicate that in some cases the predicted total concentration in the sample, m_0, is similar to the concentration observed after 14 h extraction (see m_0/C_{TOTAL} column in Table 6). This assumes that after 14 h all of the PAHs present have been extracted from the sample, thus providing a reasonable estimate of the total amount of the analytes present. However, the predictions are variable, as can be seen from the difference in the results obtained for samples 1 and 2. The differences were significantly greater than those observed for the duplicate 14 h extractions. Table 7 compares the concentration observed for each sample after 6 h extraction with the predicted total concentrations. In theory, this could be used to obtain an estimate of $\bar{R}_m$. However, the uncertainties associated with such an estimate are likely to be large. The expression for estimating the uncertainty in m_0, $u(m_0)$, was obtained by differentiating eqn. (19) to give:

$$u(m_0) = \sqrt{\left[\begin{array}{l} u(m_1)^2 + \left(\left(\frac{m_2^2}{(m_2 - m_3)^2}\right) \times u(m_3)\right)^2 \\ + \left(\left(\frac{2 \times m_2}{(m_2 - m_3)} - \frac{m_2^2}{(m_2 - m_3)^2}\right) \times u(m_2)\right)^2 \end{array}\right]} \quad (30)$$

Note the terms $\frac{1}{m_2 - m_3}$ which lead to a very large estimate of $u(m_0)$ if $m_2 \approx m_3$.

Based on previous studies of the method precision, the uncertainty associated with each of the experimental observations used to calculate m_1, m_2 and m_3 was estimated as 0.07 (as a relative standard deviation). The uncertainties in m_2 and m_3 are calculated by taking the root sum of squares of the uncertainties (as standard deviations) of each of the values used in their calculation. $u(m_0)$ was calculated for each analyte and compared with m_0 to give an indication of the relative uncertainty in each case. The results are presented in Table 8. The results indicate that m_0 could be used as a rough estimate of the total amount of analyte in the sample, although in some cases the uncertainties are rather large. For example, in the case of benzo[*a*]anthracene in sample 1 and benzo[*k*]fluoranthene in sample 2, the associated uncertainties are too large for practical

Table 6 Predicted concentrations of PAHs in RM 6138

Analyte	m_0/ng ml^{-1}[a] Sample 1	Sample 2	Total extracted after 14 h (C_{TOTAL}/ng ml^{-1})[a]	m_0/C_{TOTAL} Sample 1	Sample 2
Naphthalene	2931.31	2660.2	3027.8	0.968	0.879
Acenaphthylene	1180.35	970.6	1257.5	0.939	0.772
Fluorene	607.4	456.8	647.8	0.938	0.705
Phenanthrene	10062.4	8864.2	10374.4	0.970	0.854
Anthracene	2780.8	2446.8	3039.0	0.915	0.805
Fluoranthene	10976.6	9901.8	11384.9	0.964	0.870
Pyrene	8980.7	8161.9	9320.3	0.964	0.876
Benzo[*a*]anthracene	1553.1	3304.4	4033.1	0.385	0.819
Chrysene	4111.5	3767.4	4354.8	0.944	0.865
Benzo[*b*]fluoranthene	3253.2	8143.8	3500.7	0.929	2.326
Benzo[*k*]fluoranthene	1265.4	1007.8	1667.8	0.759	0.604
Benzo[*a*]pyrene	3125.5	2632.1	3349.7	0.933	0.786
Indeno[1.2.3-*cd*]pyrene	2755.3	2280.4	2990.0	0.922	0.763
Dibenz[*a,h*]anthracene	674.8	671.9	725.9	0.930	0.926
Benzo[*g,h,i*]perylene	2269.8	2107.6	2530.7	0.897	0.833

[a] All concentrations corrected to 10 g sample.

Table 7 Comparison of the concentration of analyte extracted after 6 h with the predicted total concentration

Analyte	Concentration after 6 h extraction ($C_{6\,h}$/ng ml^{-1}) Sample 1	Sample 2	$C_{6\,h}/m_0$ Sample 1	Sample2
Naphthalene	3039.0	2406.1	1.04	0.90
Acenaphthylene	1192.7	767.9	1.01	0.79
Fluorene	626.8	418.0	1.03	0.92
Phenanthrene	10263.8	8077.0	1.02	0.91
Anthracene	2924.4	2220.5	1.05	0.91
Fluoranthene	11318.2	8957.6	1.03	0.90
Pyrene	9326.5	7214.6	1.04	0.88
Benzo[*a*]anthracene	3962.7	2963.3	2.55	0.90
Chrysene	4378.4	3226.0	1.06	0.86
Benzo[*b*]fluoranthene	3266.8	2553.4	1.00	0.31
Benzo[*k*]fluoranthene	1286.7	1146.8	1.02	1.14
Benzo[*a*]pyrene	3159.7	2348.5	1.01	0.89
Indeno[1.2.3-*cd*]pyrene	2769.4	1982.4	1.01	0.87
Dibenz[*a,h*]anthracene	680.3	580.0	1.01	0.86
Benzo[*g,h,i*]perylene	2235.0	1693.5	0.98	0.80

Table 8 Uncertainties associated with m_0

Analyte	Sample 1 $u(m_0)$/ng ml^{-1}	$u(m_0)/m_0$	Sample 2 $u(m_0)$/ng ml^{-1}	$u(m_0)/m_0$
Naphthalene	299.2	0.10	606.0	0.23
Acenaphthylene	164.8	0.14	604.9	0.63
Fluorene	38.2	0.11	76.2	0.17
Phenanthrene	1 217.5	0.12	1 838.9	0.21
Anthracene	225.2	0.08	582.3	0.24
Fluoranthene	1 093.7	0.10	2 353.7	0.24
Pyrene	926.8	0.10	3 079.2	0.38
Benzo[*a*]anthracene	235 884	151.9	840.7	0.25
Chrysene	326.8	0.08	2 570.4	0.68
Benzo[*b*]fluoranthene	388.5	0.12	382 747	47.0
Benzo[*k*]fluoranthene	192.0	0.15	77.6	0.08
Benzo[*a*]pyrene	467.7	0.15	751.7	0.29
Indeno[1.2.3-*cd*]pyrene	350.4	0.13	998.9	0.44
Dibenz[*a,h*]anthracene	90.3	0.13	323.2	0.48
Benzo[*g,h,i*]perylene	215.8	0.10	2 097.7	1.00

application. This arises when $m_2 - m_3$ is small compared to m_2. Note, however, that the present calculation is crude and based on only three data points from a continuous sequence. It should therefore be possible to improve the estimates and uncertainties significantly using improved modelling and curve fitting methods.

Conclusions

This paper has considered various approaches for estimating analytical bias (measured as recovery) and its associated uncertainty. It is useful to consider the recovery, R, for a particular sample as being composed of three components; $\bar{R}_m$, R_s and R_{rep} respectively representing a reference recovery, specific sample correction and allowance for imperfect representativeness in spiking studies. $\bar{R}_m$ and its uncertainty $u(\bar{R}_m)$ can be estimated by existing methods. The simplest and most effective involves the analysis of a relevant certified reference material. Other techniques studied (re-extraction, extraction modelling, analysis of 'worst case' CRMs) currently appear to lead to larger (sometimes impractically large) associated uncertainties, but can provide an initial estimate of $\bar{R}_m$.

Where a method covers many matrices and a restricted number of CRMs is available, R_s and its uncertainty can be assessed from studies on various different matrices; it is particularly important to study a representative range. $u(R_s)$ tends to be larger than $u(\bar{R}_m)$.

In the absence of relevant reference materials, analysis of spiked samples is common, but uncertainty calculations may be intricate. Treatment of the data can be substantially simplified by appropriate approximations. However, estimating the necessary additional term R_{rep} and its uncertainty is one of the more problematic aspects of a recovery study and leads to very substantial uncertainties.

There are, therefore, methods available for characterisation of uncertainty arising from recovery, but only direct application of relevant certified reference materials or reference methods currently provides a completely general method of characterising recovery well (that is, with small uncertainty) for a particular sample type. Other general methods, particularly modelling approaches, currently lead to large uncertainties, but do appear capable of improvement. This has important implications for international efforts to improve comparability *via* traceability to national and international standards. Large uncertainties in nominally traceable measurements imply poor comparability, and there are large uncertainties associated with matrix change and analyte recovery. Given the practical impossibility of producing relevant reference materials for all matrices and analytes, it is important that these effects are characterised well to minimise uncertainties associated with the inevitable use of imperfectly matched reference materials. In

other words, general implementation of traceability through reference materials will require improvements in methodology for characterising recovery and matrix effects if traceability is to be a generally useful means of ensuring comparability in analytical chemistry.

Acknowledgement

Production of this paper was supported under contract with the Department of Trade and Industry as part of the National Measurement System Valid Analytical Measurement Programme.

References

1 M. Thompson, *Analyst*, 1995, **120**, 117N.
2 W. Horwitz and A. Albert, *Analyst*, 1997, **122**, 615.
3 S. L. R. Ellison, W. Wegsheider and A. Williams, *Anal. Chem.*, 1997, **69**, 607A.
4 J. S. Kane, *Analyst*, 1997, **122**, 1283.
5 S. L. R. Ellison and A. Williams, *Accred. Qual. Assur*., 1998, **3**, 6.
6 S. L. R. Ellison, *Accred. Qual. Assur.*, 1998, **3**, 95.
7 ISO, *Guide to the Expression of Uncertainty in Measurement*, International Standards Organisation, Geneva, 1993.
8 EURACHEM GUIDE: *Quantifying Uncertainty in Analytical Measurement*, Laboratory of the Government Chemist, London, 1995.
9 M. Pueyo, J. Obiols and E. Vilalta, *Anal. Commun.*, 1996, **33**, 205.
10 A. Williams, *Anal. Proc*., 1993, **30**, 248.
11 Analytical Methods Committee, *Analyst*, 1995, **120**, 2303.
12 S. L. R. Ellison, *Accred. Qual. Assur.* 1998, **3,** 95
13 S. L. R. Ellison and A. Williams, *Accred. Qual. Assur.*, 1998, **3**, 6.
14 W. Horwitz, *Pure Appl. Chem.*, 1988, **60**, 885.
15 AOAC recommendation, *J. Assoc. Off. Anal. Chem.*, 1989, **72**, 694.
16 *ISO 5725:1994, Accuracy (Trueness and Precision) of Measurement Methods and Results*, International Standards Organisation, Geneva, 1995.
17 S. L. R. Ellison and V. J. Barwick, *Accred. Qual. Assur.*, 1998, **3**, 101.
18 S. L. R. Ellison and V. J. Barwick, *Analyst*, 1998, **123**, 1387.
19 V. J. Barwick and S. L. R. Ellison, *Anal. Commun.*, 1998, **35** 377
20 K. D. Bartle, A. A. Clifford, S. B. Hawthorne, J. J. Langenfeld, D. J. Miller and R. Robinson, *J. Supercrit. Fluids*, 1990, **3**, 143.
21 S. L. R. Ellison and A. Williams, in *The use of recovery factors in trace analysis*, ed. M. Parkany, Royal Society of Chemistry, Cambridge, UK, 1996.
22 *Soil Quality—Determination of polycyclic aromatic hydrocarbons (PAH)*, Draft British Standard, 1998.
23 T. J. Farrant, *Practical Statistics for the Analytical Scientist, A Bench Guide*, Royal Society of Chemistry, Cambridge, UK, 1997.

Paper 9/01845J