# Systematic errors in analytical measurement results

**Author/Contributor:**
Hibbert, D. Brynn

# Systematic Errors in Analytical Measurement Results

D Brynn Hibbert
School of Chemistry
University of New South Wales
Sydney, NSW 2052
Australia

Ph: +61 2 9385 4713
Fax: +61 2 9385 6141
e-mail: b.hibbert@unsw.edu.au

**Abstract**

Definitions of the concepts of bias and recovery are discussed and approaches to dealing with them described. The Guide to Uncertainty in Measurement (GUM) recommends correction for all significant bias, but it is also possible to expand measurement uncertainty to take account of uncorrected bias. Run, laboratory and method bias are subsumed in the bias of a particular measurement result, but can be useful as concepts used in method validation. Multivariate calibration brings its own biases that must be quantified and minimised.

## 1 Introduction

The concepts of 'bias' (and 'recovery') are important aspects of the understanding of a measurement result in analytical chemistry. This paper will discuss the present definitions and will review different approaches to dealing with systematic effects. In addition to the metrological debate, field laboratories need to be able to estimate and, if necessary correct for, systematic effects. Example of present practice will be given. The concept of bias of a measurement result is best understood in terms of a measurement model that recognises systematic and random components of error.

$$\hat{x} = x + \delta + \varepsilon \tag{1}$$

The true value of a measurand, $x$, is estimated by $\hat{x}$ which differs from it by a systematic component, the bias $\delta$ and a random component $\varepsilon$. The random error is considered to be Normally distributed with expectation zero and standard deviation $\sigma$. Therefore a large number of measurements will have a mean of $(x + \delta)$ as shown in Fig. 1. A single measurement result cannot distinguish between systematic and random error, but several measurements combined with knowledge about the characteristics of the method can allow calculation of an interval about that contains the true value with a certain level of confidence.

$$x = \hat{x} \pm U \tag{2}$$

$U$ is known as the expanded uncertainty [1] and is obtained from considerations of all aspects of the uncertainty of the measurement result. The so-called GUM approach (after Guide to the Expression of Uncertainty in Measurement [1]) was first published in 1993 and has been the basis of the recommended methods for characterising a measurement result. A requirement of the international standard for testing laboratories (ISO/IEC 17025 [2]) is that an appropriate measurement uncertainty be estimated for the results. One reason that traditional concepts of systematic and random error have been subsumed into the uncertainty approach is that depending on the information used, one kind of error can be turned into another, and so there is no general definition of these terms and the measuring system must be described very carefully. This discussion will be expanded below.
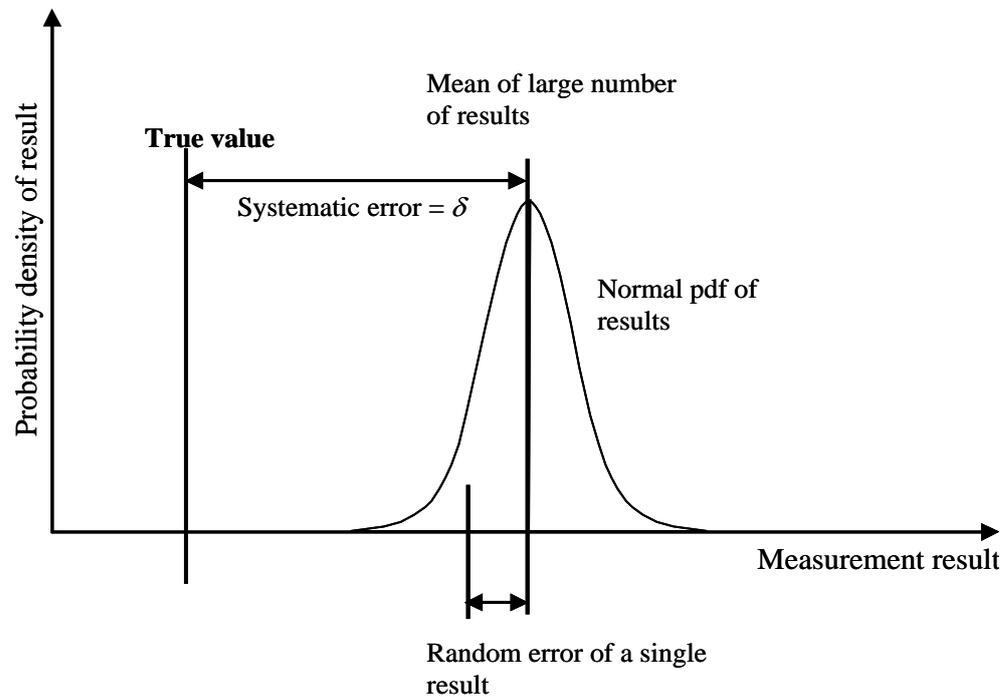
**Fig. 1. Schematic showing the relationship between random and systematic error in a measurement result**

## 2  Definitions

In the forthcoming third edition of the International Vocabulary of Basic and General Terms in Metrology (VIM3, [3]) measurement bias (synonym bias) will be defined as

**2.19**
**measurement bias**
**bias**
**systematic measurement error** or its estimate, with respect to a **reference quantity value**

where bold terms are also defined in the VIM. Systematic measurement error is defined as

**2.18**
**systematic error of measurement**
**systematic error**
component of **measurement error** that in replicate **measurements** remains constant or varies in a predictable manner
NOTES
1 — The **reference quantity value** for a systematic measurement error is a **true quantity value**, or a **measured quantity value** of a **measurement standard** of negligible **measurement uncertainty**, or a **conventional quantity value**.
2 — Systematic measurement error, and its causes, can be known or unknown. A **correction** can be applied to compensate for a known systematic measurement error.
3 — Systematic measurement error equals the difference of measurement error and **random measurement error**.

The depiction in Fig. 1 is consistent with these definitions. The definitions also imply that bias can be estimated by measurement of a reference quantity value, and then subsequently corrected for.

## 3  Components of the systematic error of a measurement result

In the definitions given above there is no distinction made among different sources of systematic error. However identification of the source of systematic error can impact

on its estimation and treatment. The headings below represent sources that have been proposed as worthy of individual attention. They may overlap, and it must be remembered that for a single measurement result there is but one, unknowable, 'measurement error' – the difference between $x$ and $\xi$ in Eq. 1.

### 3.1 Sampling bias

When the measurand is a quantity of a larger whole, sampling error can be a major systematic effect, and will not be treated in the same way as for effects in the laboratory procedures. A goal of a sampling protocol is often to randomise effects that can be then treated statistically [4]. Ramsey has pointed out that the traditional approach may remove sampling bias by definition, but not in practice – i.e. an assertion of random sampling does not guarantee the desired result [5]. 'Analytical bias' (Ramsey's term for systematic effects arising during the laboratory measurement, to distinguish the effect from 'sampling bias') is usually estimated by measurement of a reference material. By analogy sampling bias can be estimated by use of a Reference Sampling Target (RST). The RST is synthesized to have a known concentration of analyte [6], or it is a routine sample that has been selected for the purpose and its quantity value established by an interlaboratory study [7]. The certified value may also be specified for its spatial extent. The second method suggested by Ramsey, which is designed to randomize sampling bias, is to use multiple sampling protocols, again in a collaborative study [5]. Each different bias will contribute to the sampling variance in an assumed random fashion. Therefore a realistic uncertainty due to sampling can be estimated and a decision regarding whether a measurement will be fit for purpose made. Ramsey calls this 'appropriate' sampling.

### 3.2 Calibration bias

Calibration bias for direct reading instruments (where the indication of the measuring instrument is expressed in the same quantity as the standard) has been identified by Cuadros-Rodriguez et al [8] in contrast to factors employed in indirect calibration. An example is a direct-reading balance, for which the bias, measured by weighing a calibrated mass, is subtracted from subsequent measurements as a correction. The correction can be an absolute value or a relative correction factor that multiplies the uncorrected result.

Indirect calibration is required for the majority of analytical measurements, and involves the establishment of the relation between the indication of the measuring instrument (counts, voltages, currents, peak areas etc) and values of the quantity being measured. If the calibration equation holds for the measuring system, is linear in the coefficients, and the random components of error are known to be constant or proportional to the quantity value, then there are algebraic solutions for the coefficients and uncertainties of quantity values estimates [9]. A bias arises when the calibration equation does not fit the quantity value-response relation; for example when a straight line is forced through data that curves at higher concentrations. The potential error in misusing straight line calibrations has been demonstrated by Mulholland and Hibbert [10], and by using an inappropriate function by Kirkup and Mulholland [11]. It is important therefore that in establishing the calibration the adequacy of the mathematical model is demonstrated.

### 3.3 Recovery

Although incomplete recovery contributes systematic errors to a measurement result and so is included in this discussion, the actions that give rise to the need to consider recovery perhaps should be seen as a part of the overall measurement procedure. Indeed one definition of the term concerns the physical separation of an analyte from a matrix, as opposed to our ability to measure a quantity in the course of an analysis. Confusion among the different terminology and interpretations of the concept 'recovery' still permeates the analytical community. For a flavour of the problems in defining this term see the concise paper of Dybkaer [12] which was written in response to two IUPAC recommendations [13, 14]. Here we shall adopt the Dybkaer terminology, although there appears to be no great dispute about the intent of the definitions of the concept. If in a chemical analysis it is necessary to apply a procedure to the sample to be analysed that separates or derivatises or otherwise changes the sample before presentation to the measuring instrument, there is a possibility that not all of the analyte will be measured by the instrument. The ratio of the measured amount (Dybkaer : the initially estimated quantity) and the amount actually present in the sample (Dybkaer: the actual amount) is called the recovered quantity ratio (colloquially the 'recovery') [12]. The problem arises when the recovery has to be estimated, in order to apply a recovery factor to subsequent measurement results. For a routine measurement the actual amount is perforce unknown, and so a separate procedure is needed to establish the recovery factor. If a well characterised matrix-matched certified reference material is available, the recovery factor is simply the mean of a suitably large number of measurement results of this material divided by its certified quantity value (see Eq. 3). A statistical test for a null hypothesis of $R = 1$ can be performed to decide if any correction is to be applied to routine results, and then an uncertainty component must be included when the combined standard uncertainty is calculated for the corrected or uncorrected field measurement result.

$$R = \frac{\sum_{i=1}^{i=p} C(\text{ref})_{\text{meas}}}{pC(\text{ref})_{\text{cert}}} = \frac{\overline{C}(\text{ref})_{\text{meas}}}{C(\text{ref})_{\text{cert}}} \tag{3}$$

$$\frac{u(R)}{R} = \sqrt{\left(\frac{u(C(\text{ref})_{\text{cert}})}{C(\text{ref})_{\text{cert}}}\right)^2 + \frac{1}{p}\left(\frac{u(\overline{C}(\text{ref})_{\text{meas}})}{\overline{C}(\text{ref})_{\text{meas}}}\right)^2} \tag{4}$$

where $C(\text{ref})_{\text{meas}}$ is the initially estimated quantity that is measured $p$ times to give the mean, and $C(\text{ref})_{\text{cert}}$ the actual amount which is usually certified in the documentation accompanying the reference material. The standard uncertainty of the estimate of the quantity value of the reference material is $u(\overline{C}(\text{ref})_{\text{meas}})$ and the standard uncertainty of the reference value itself is $u(C(\text{ref})_{\text{cert}})$. In the majority of cases there is no matrix matched reference material, and so the commutability of any material used must be considered. Thus, if a blank matrix material is spiked with neat analyte the question arises as to what extent the analyte is taken up in the matrix in the same manner as a field sample. Perhaps the spike will be recovered more easily? In principle recovery is estimated in the same way, by replicate measurements of the spiked material. The uncertainty of the recovery factor must now include an estimate of the error arising from lack of commutability. This component is difficult to estimate and might be of a magnitude that makes it difficult to observe a significant recovery. Under this regime, equations for the recovery and its uncertainty become

$$R = \frac{\sum_{i=1}^{i=p} C(\text{ref})_{\text{meas}}}{p\,C(\text{ref})_{\text{grav}}} \times f = \frac{\overline{C}(\text{ref})_{\text{meas}}}{C(\text{ref})_{\text{grav}}} \times f \tag{5}$$

$$\frac{u(R)}{R} = \sqrt{\left(\frac{u(C(\text{ref})_{\text{grav}})}{C(\text{ref})_{\text{grav}}}\right)^2 + \frac{1}{p}\left(\frac{u(\overline{C}(\text{ref})_{\text{meas}})}{\overline{C}(\text{ref})_{\text{meas}}}\right)^2 + u(f)^2} \tag{6}$$

$C(\text{ref})_{\text{grav}}$ is the quantity value of the analyte in a gravimetrically prepared spike, and $f$ appears formally in the equation for $R$ as a correction for the lack of commutability with a value of 1.

### 3.4 Analytical biases

It is possible to conceive of bias terms arising from different aspects of the analytical process. A rather poor method that consistently gave low results would not be susceptible to more quality control at the laboratory level. Uncorrected the method would be unlikely to give results that could be compared with those from a different method. The Analytical Methods Committee of the Royal Society of Chemistry [15] in 1995, and then later Thompson [16] split the bias term of Eq. 1 into contributions from the method, laboratory and run, which together form a 'ladder of errors' [16]. Fig. 2 shows the relationships among these concepts.
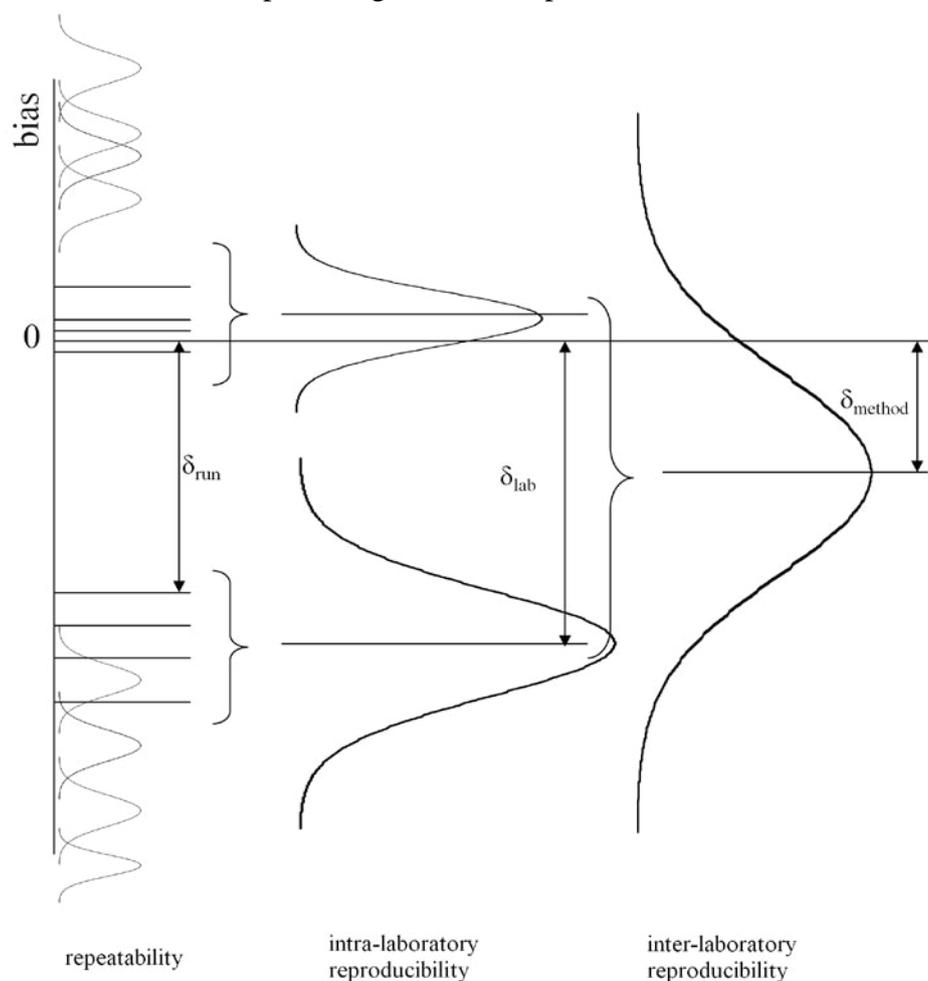


**Fig. 2. Systematic errors for a method with a large inherent method bias showing the relationship between method, intra-laboratory and run bias as defined by Thompson [16].**

O'Donnell and Hibbert have considered bias from the VIM definition and thus discuss different biases in relation to the conditions under which the measurement result is obtained [17]. They emphasise that only a single systematic error can be measured in a given set of experiments, whether done under repeatability conditions (run bias), intra-laboratory reproducibility conditions (laboratory bias) or as part of an inter-laboratory method validation study (method bias). How lower level biases become random components as the analysis moves up the hierarchy are shown in Fig. 3. Although the usages of the terms run bias, laboratory bias and method bias differ, it can be seen how they are related from Figs. 2 and 3.
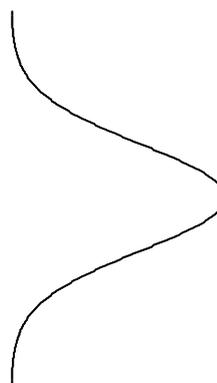


**Fig. 3. A hierarchy of experimental conditions showing how run bias is randomised in intra-laboratory studies, and laboratory bias is randomised in inter-laboratory method validation studies.**

## 4 Bias of empirical methods

An empirical method, one for which the measurand is defined by the method, has zero bias by definition, when used exactly according to the prescribed method. An example is the measurement of chemical oxygen demand by ISO 6060, in which dichromate is reduced by oxidisable material in the sample.

## 5 Treatment of bias

Authoritative guides on the treatment of measurement results and the estimation of measurement uncertainty agree that all significant bias should be estimated and corrected for [1, 18]. Section 3.2.4 of the GUM reads "It is assumed that the result of a measurement has been corrected for all recognised significant systematic effects and that every effort has been made to identify such effects". The example given is of a clear case in physical measurement in which the finite impedance of a voltmeter gives rise to a bias, the form of which is well known and related to the measurable impedance. In chemical measurement, as has been shown, the nature of bias can be less obvious and arise at different stages of an analysis. Because of the greater complexity of systematic errors in chemistry, and perhaps because a particular value

of the bias will be unique to each measurement, some sectors have tended to avoid explicit correction for bias and instead have used an estimate of the magnitude of the bias to augment the measurement uncertainty. O'Donnell and Hibbert have reviewed such methods in comparison with the GUM-recommended approach by Monte Carlo simulation [17]. Correction of a measurement result with inclusion of the uncertainty of the correction in the combined standard uncertainty is demonstrated to be the best approach. If a correction is not to be done then the method known as $SUMU_{\text{Max}}$ [19] in which the absolute value of the estimate of run bias ($\delta_{\text{run}}$ as shown in Fig. 3) is added to the expanded uncertainty gives the best estimate, albeit a consistent overestimate.

$$U(C_{\text{test}}) = ku_{\text{c}}(C_{\text{test}}) + |\delta_{\text{run}}| \tag{7}$$

$u_{\text{c}}$ is the combined standard uncertainty and $k$ is the coverage factor which is determined by the required level of confidence and number of degrees of freedom of $u_{\text{c}}$.

The definition of run bias as the difference between the average of results obtained under repeatability conditions and the true value leads to a straight forward approach to correction for bias in batch analyses and estimation of measurement uncertainty. In a routine batch measurement if suitable matrix reference materials are available then run bias can be estimated and if significant a correction applied. O'Donnell and Hibbert argue that if this is done, then the only components of the measurement uncertainty are the repeatability precision that pertains to the run conditions under which the measurement and bias estimation are being made, and the uncertainty of the correction [20]. For a bias correction this becomes

$$\delta_{\text{run}} = \frac{\sum_{1}^{p} C(\text{CRM})_{\text{meas}}}{p} - C(\text{CRM})_{\text{cert}} = \overline{C}(\text{CRM})_{\text{meas}} - C(\text{CRM})_{\text{cert}} \tag{8}$$

$$\overline{C}(\text{sample}) = \frac{\sum_{1}^{n} C(\text{sample})_{\text{meas}}}{n} + \delta \tag{9}$$

$$u_{\text{c}}\left(\overline{C}(\text{sample})\right) = \sqrt{\left(\frac{u_r\left(\overline{C}(\text{sample})\right)}{\sqrt{n}}\right)^2 + \left(\frac{u_r\left(\overline{C}(\text{CRM})_{\text{meas}}\right)}{\sqrt{p}}\right)^2 + u\left(C(\text{CRM})_{\text{cert}}\right)^2} \tag{10}$$

where a certified reference material (CRM), appropriately matrix matched, is measured $p$ times and the test sample is measured $n$ times, all under repeatability conditions with standard uncertainty $u_{\text{r}}$. The correction in Eq. (9) is applied if the value of $\delta$ is significantly different from zero by a Student-t test

$$t = \frac{\delta_{\text{run}}}{u(\delta)} \tag{11}$$

where the standard uncertainty of the estimate of the bias, is the combination of the measurement and certification terms in Eq. 10

$$u(\delta) = \sqrt{\left(\frac{u_r\left(\overline{C}(\text{CRM})_{\text{meas}}\right)}{\sqrt{p}}\right)^2 + u\left(C(\text{CRM})_{\text{cert}}\right)^2} \tag{12}$$

If it can be demonstrated that the measurements are all made under repeatability conditions, for example using the tests in ISO 5725 [21] then the appropriate uncertainty ($u_{\text{r}}$) is the repeatability from quality control data with essentially infinite degrees of freedom.

An example of the analysis of the concentration of creatinine in urine by a spectrophotometric method in a commercial analyser has been given by O'Donnell and Hibbert [20].

## 6   Some examples of bias in practice

Multivariate methods of calibration cause concern about bias relative to more traditional methods. Sugar content of cane was measured by principal components regression of mid-infrared data. A bias for sucrose is reported as 0.041 g / 100 mL which is claimed to be better than that for direct polarimetry [22]. Other infrared applications for which bias in transferring calibration from one instrument to another is important include analysis of red grapes [23], the properties of wood in the presence of blue stain [24], and the analysis of petrochemicals [25]. The more general question of robustness of multivariate calibration using near infrared is discussed by Zeaiter [26].

An example of what might be called sampling bias has been identified in geophysics with the analysis of rainwater. Ayers et al showed that biological effects on the ionic composition of rainwater are not restricted to the previously reported pH [27]. Ammonium, potassium, nitrate, sulfate, methanesulfonate, and phosphate ions are also removed by biological processes, but remain in the rainwater in biomass. The implication is  that most previous rainwater composition studies based on ionic analyses will have systematically underestimated nutrient deposition. The analysis of bias in geochemical analysis is of concern in a work by de Castilho, who reports Monte Carlo simulations to test statistical methods for detecting analytical bias [28]. Charlet and Marchal report the use of certified reference materials in making metrologically sound measurements of heavy metals in groundwater [29]. They underline the role of a matrix CRM in estimating bias in method validation.

**References**

1. ISO, *Guide to the Expression of Uncertainty in Measurement*. 1995, International Organisation for Standardisation: Geneva.
2. ISO/IEC, *General requirements for the competence of calibration and testing laboratories*, in *17025*, ISO, Editor. 2005, International Organization for Standardization: Geneva.
3. Joint Committee for Guides in Metrology, *International vocabulary of basic and general terms in metrology, 3rd edition*, I. BIPM, IFCC, ISO, IUPAC, IUPAP, OIML, Editor. 2007, International Organization for Standardization: Geneva, Switzerland.
4. ISO, *Statistical aspects of sampling from bulk materials -- Part 1: General principles* in *11648-1*. 2003, International Organization for Standardization: Geneva.
5. Ramsey, M.H., *Appropriate rather than representative sampling, based on acceptable levels of uncertainty*. Accreditation and Quality Assurance, 2002. **7**(7): p. 274-280.
6. Ramsey, M.H., S. Squire, and M.J. Gardner, *Synthetic reference sampling target for the estimation of measurement uncertainty*. Analyst, 1999. **124**(11): p. 1701-1706.
7. Argyraki, A., M.H. Ramsey, and M. Thompson, *Proficiency testing in sampling: Pilot study on contaminated land*. Analyst, 1995. **120**(12): p. 2799-2803.
8. Cuadros-Rodriguez, L., et al., *Calibration in chemical measurement processes: I. A metrological approach*. Trends in Analytical Chemistry, 2001. **20**(4): p. 195-206.
9. Hibbert, D.B., *The uncertainty of a result from a linear calibration*. Analyst, 2006. **131**(12): p. 1273-1278.
10. Mulholland, M. and D.B. Hibbert, *Linearity and the limitations of least squares calibration*. Journal of Chromatography A, 1997. **762**(1-2): p. 73-82.
11. Kirkup, L. and M. Mulholland, *Comparison of linear and non-linear equations for univariate calibration*. Journal of Chromatography A, 2004. **1029**(1-2): p. 1-11.
12. Dybkær, R., *Recovery revisited*. Accreditation and Quality Assurance, 2005. **10**(6): p. 302-303.
13. Thompson, M., et al., *Harmonized guidelines for the use of recovery information in analytical measurement*. Pure and Applied Chemistry, 1999. **71**(2): p. 337 - 348.
14. Burns, D.T., K. Danzer, and A. Townshend, *Use of the terms "Recovery" and "Apparent Recovery" in Analytical Procedures*. Pure and Applied Chemistry, 2002. **74**(11): p. 2201-2205.
15. Analytical Methods Committee, *Uncertainty of Measurement: Implications of Its Use in Analytical Science*. The Analyst, 1995. **120**: p. 2303-2308.
16. Thompson, M., *Towards a unified model of errors in analytical measurement*. The Analyst, 2000. **125**(11): p. 2020-2025.
17. O'Donnell, G.E. and D.B. Hibbert, *Treatment of bias in estimating measurement uncertainty*. The Analyst, 2005. **130**: p. 721 - 729.
18. EURACHEM/CITAC, *EURACHEM/CITAC Guide - Quantifying Uncertainty in Analytical Measurement*. 2nd ed, ed. M.R. S.L.R. Ellison, A. Williams. 2000: EURACHEM / CITAC.
19. Maroto, A., et al., *Should non-significant bias be included in the uncertainty budget?* Accreditation and Quality Assurance, 2002. **7**(3): p. 90-94.
20. O'Donnell, G. and D.B. Hibbert, *The estimation of uncertainty of measurement from in-house method validation and analytical run data*. submitted to Accreditation and Quality Assurance, 2007. **12**.
21. ISO, *Accuracy (trueness and precision) of measurement methods and results - Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method* in *5725-2*. 1994, International Organization for Standardization: Geneva.
22. Cadet, F., *Measurement of sugar content by multidimensional analysis and mid-infrared spectroscopy*. Talanta, 1999. **48**(4): p. 867-875.

23. Cozzolino, D., et al., *Effect of both homogenisation and storage on the spectra of red grapes and on the measurement of total anthocyanins, total soluble solids and pH by visual near infrared spectroscopy.* Journal of near Infrared Spectroscopy, 2005. **13**(4): p. 213-223.

24. Via, B.K., et al., *Prediction of wood mechanical and chemical properties in the presence and absence of blue stain using two near infrared instruments.* Journal of near Infrared Spectroscopy, 2005. **13**(4): p. 201-212.

25. Hoffmann, U. and N. Zanier-Szydlowski, *Portability of near infrared spectroscopic calibrations for petrochemical parameters.* Journal of near Infrared Spectroscopy, 1999. **7**(1): p. 33-45.

26. Zeaiter, M., et al., *Robustness of models developed by multivariate calibration. Part I: The assessment of robustness.* Trends in Analytical Chemistry, 2004. **23**(2): p. 157-170.

27. Ayers, G.P., R.W. Gillett, and P.W. Selleck, *A universal bias in inorganic rainwater chemical composition data.* Geophysical Research Letters, 2003. **30**(13): p. 1715-1715.

28. de Castilho, M.V., *A comparison of statistical techniques for detecting analytical bias in geoanalysis.* Geostandards and Geoanalytical Research, 2004. **28**(2): p. 277-290.

29. Charlet, P. and A. Marschal, *Improvement in the traceability of environmental analysis by the relevant use of certified pure solutions and a matrix certified reference material.* Trends in Analytical Chemistry, 2004. **23**(3): p. 178-184.