

Arpad Ambrus

Reliability of measurements of pesticide residues in food

Received: 23 October 2002
Accepted: 5 August 2003
Published online: 14 April 2004
© Springer-Verlag 2004

A. Ambrus (✉)
FAO/IAEA Agriculture
and Biotechnology Laboratory,
IAEA, Vienna, Austria
e-mail: a.ambrus@iaea.org
Tel.: +43-1-260028395
Fax: +43-1-260028222

Abstract This paper accounts for the major sources of errors associated with pesticide residue analysis and illustrates their magnitude based on the currently available information. The sampling, sample processing and analysis may significantly influence the uncertainty and accuracy of analytical data. Their combined effects should be considered in deciding on the reliability of the results. In the case of plant material, the average random sampling (coefficient of variation, $CV=28-40\%$) and sample processing (CV up to 100%) errors are significant components of the combined uncertainty of the results. The average relative uncertainty of the analytical phase alone is about $17-25\%$ in the usual $0.01-10$ mg/kg concentration range. The major con-

tributor to this error can be the gas-liquid chromatography (GLC) or high-performance liquid chromatography (HPLC) analysis especially close to the lowest calibrated level. The expectable minimum of the combined relative standard uncertainty of the pesticide residue analytical results is in the range of $33-49\%$ depending on the sample size. The gross and systematic errors may be much larger than the random error. Special attention is required to obtain representative random samples and to eliminate the loss of residues during sample preparation and processing.

Keywords Measurement uncertainty · Measurement reliability · Pesticide residue analysis

Introduction

Analytical measurements are made to obtain information on some properties of the sampled object. The interpretation of the results and making correct decisions require information on the accuracy and precision of the measurements. In most measurements we can distinguish three types of errors: *gross* or *spurious*, random, and systematic [1]. *Gross* errors, which occur occasionally even in the very best laboratories, are best described as errors from which there is no recovery—the experiment simply has to be started again from scratch to give a valid result. Elementary examples might include a sample that has been contaminated by poor handling, accidentally discarding an extract, failure of a piece of equipment due to an unexpected loss of electrical power, accidental use

of the wrong reagent (analytical standard) or the correct reagent at the wrong concentration, and so on. In some cases *gross* errors can occur without the experimenter being aware of them at the time: they are only revealed by further experiments, or by comparisons with results from other laboratories. *Random* errors are present in all measurements and cause replicate results to fall on either side of the mean (average) value. Even in very simple analyses (e.g. acid/base titration) [2] they usually arise from a combination of experimental steps, each of which is not perfectly repeatable from one replicate to the next. Random errors give rise to results both above and below the true value, and thus sum to zero over many measurements. The random error of a measurement cannot be compensated for, but its effects can be reduced by increasing the number of observations. *Systematic* errors

also occur in most experiments, or at least in some stages of most experiments, but their effects are quite different. A particular systematic error always affects a series of measurements in the same way, e.g. all the data in a sample have values which are too high, or too low. Some simple examples are: a balance providing incorrect reading of weight; a partly decomposed analytical standard taken with the stated purity, the partition coefficient of a pesticide residue between the organic phase and the sample, etc. The sum of all the systematic errors in an experiment (note that some may be positive and others negative) is called the *bias* in the experiment. Since they do *not* sum to zero over a large number of measurements, it is clear that individual systematic errors cannot be detected directly by replicate analyses. The greatest problem with such errors is that they may be very large indeed, and yet go undetected if appropriate precautions are not taken. In practice, it is clear that systematic errors in an analysis can only be identified if the technique is applied to a reference material containing a known amount of the measurand. However, only if the reference material matches identically in terms of analyte, matrix and concentration does it meet the ideal conditions for determining the performance of the method. Note, that when using carefully homogenized reference materials, the uncertainty of sample processing cannot be estimated and the efficiency of extraction may not be correctly assessed, though they can have a critical effect on the results.

The uncertainty of the measurements is mainly due to some random effects. The separate contributors to uncertainty are called uncertainty components. There are two basically different approaches for estimating the uncertainty of analytical measurements: the “top-down” approach introduced by the Analytical Methods Committee [3] and recommended for practical use, for instance, by the Nordic Committee on Food Analysis [4], and the “bottom-up” or error budget introduced by ISO [5] and elaborated by EURACHEM [2].

The top-down method is primarily based on the results of inter-laboratory proficiency tests, collaborative trials and internal quality control data, and thus it may take into account the between-laboratories variability of the results. It provides the most reliable estimate of the expectable performance of a method and/or laboratory, and provides the ground for judging the equivalency of results obtained in different laboratories. It also eliminates the problems encountered when the bottom-up approach is used, which are related to: the nature of chemical measurements, considering unknowable interactions and interferences, overlooking important variables and double counting others, and adjusting for uncontrollable “Type B” components [6].

In spite of the above-mentioned deficiencies, the bottom-up approach has its own merit especially in identifying and quantifying the uncertainty of individual components or steps of the determination. By taking into

consideration the contribution of the individual procedures or steps to the overall uncertainty of the results, the analytical procedures and testing strategy can be optimized to be fit for the purpose of the analysis with minimum cost.

The determination of pesticide residues in the usual concentration range of 0.0001–100 mg/kg is a complex task. The activities involved may be divided into field/external operations (collection, preparation, packing/shipping of samples) and laboratory operations (sample preparation, sample processing and analysis). Some of them may consist of several distinct steps. Some of the operations (e.g. sample preparation, storage of samples) may be carried out at more than one place, similarly the storage conditions affect the residue levels during the whole procedure. The main steps of the procedures and the major sources of errors are illustrated in Table 1. The definition of terms used in this paper are summarized in Table 2.

The appropriate reference materials for pesticide residue analysis are rarely available. Therefore, the accuracy and bias of the analytical method are usually checked with recovery studies, which can only provide information on the performance of the method from the point of fortification. The individual recoveries are affected by both random and systematic errors. The sum of systematic errors is usually indicated by the average recovery, while the standard deviation of individual recovery values reflects the uncertainty of the measurements.

In the past, analysts chose the most convenient approach and claimed responsibility for the analysis of the samples delivered to the laboratory, and reported the uncertainty of the results based on the laboratory measurements only. Disproportionally large amounts of effort, time and money were invested in the characterization of the performance of analytical procedures, while very little attention was given to study the effect of the equally important sampling, sample preparation/processing and the stability of analytes during the latter procedures and storage of samples. However, such practice is not correct and may lead to wrong decisions.

The objectives of this paper are:

- To summarize the methods of estimating the uncertainty of individual steps of the determination of pesticide residues which can be considered as the major components contributing to the combined uncertainty of the pesticide residue data
- To review the currently available information on the major sources and magnitude of random and systematic errors involved in the procedure of obtaining pesticide residue data from collections of samples to reporting of final results
- To illustrate the use of uncertainty information for the optimization of the procedure and interpretation of the results.

Table 1 Main steps of the determination of pesticide residues and the major sources of error

Sources of systematic error	Step	Sources of random error
Field/external operation Selection of sampling position	Sampling (uncertainty component: S_S)	Large variation of residue concentration in/on treated objects. Number of primary samples taken (sample size)
Wrong labelling Decomposition of analytes Laboratory operations	Shipping and storage	
The portion of sample to be analysed (analytical sample) selected incorrectly	Sample preparation	The analytical sample comes in contact and gets contaminated with other portions of the sample Rinsing, brushing is performed to various extent
Decomposition of analyte during sample processing	Sample processing, S_{Sp}	Inhomogeneity of the analyte in natural units of the analytical sample Inhomogeneity of the analyte in the ground/chopped analytical sample Variation of temperature during the homogenization process Texture (maturity) of analytical sample affecting the efficiency of homogenization process
Incomplete recovery of analyte	Analysis, S_A Extraction/clean-up	Variation of the composition (e.g. water, fat and sugar content) of sample materials taken from a commodity Temperature of sample/solvent matrix
Interference of co-extracted materials	Quantitative determination	Variation of nominal volume of devices within the permitted tolerance intervals Precision of balances
Unknown purity of reference analytical standard Biased weight/volume measurements Personal bias in reading analogue instruments/equipment		Varying chromatographic conditions (matrix effect, system inertness, detector response, etc.) Changing of laboratory environmental conditions during analysis Variation of conversion rate of derivatization reactions

Gross or spurious errors cannot be predicted; hence they are not discussed here.

Estimation of the standard uncertainty of the major uncertainty components of pesticide residue data

Analytical results are usually considered normally distributed about the mean because the variations are derived from the cumulative effects of small errors [7]. Hill and von Holst showed [8] that the frequency distribution of results produced by multiplication of normally distributed populations deviate from normal distribution. The deviation from the normal distribution increases with the level of dispersion of the measurements. However, the analysis of a number of large proficiency and control databases indicated [9] that the distribution of the recovery data, after removal of outliers, was fairly normal with a small inclination toward the high side. Since the procedures developed for the measurement and expression of precision work in practice for most unimodal distributions [10], the small deviation from normality does not affect their applicability for assessing the combined uncertainty of pesticide residue analysis.

According to the general principles of propagation of random error [2], which do not assume a normal distri-

bution, the combined standard uncertainty of measured residues (S_{Res}) may be expressed as

$$S_{Res} = \sqrt{S_S^2 + S_L^2} \quad (1)$$

where S_S is the uncertainty of sampling and S_L is the combined uncertainty of the laboratory phase including sample processing (S_p) and analysis (A). The sample preparation step (such as gentle rinsing or brushing to remove soil, or taking outer loose leaves from cabbages) cannot usually be validated and its contribution to the uncertainty of the results cannot be estimated. Standard operating procedures (SOPs) must provide sufficiently detailed information to carry them out in a consistent manner.

If the whole sample is analysed the mean residue remains the same and the equation can be written as:

$$CV_{Res} = \sqrt{CV_S^2 + CV_L^2} \quad (2)$$

The CV_L may be calculated as:

$$CV_L = \sqrt{CV_{Sp}^2 + CV_A^2} \quad (3)$$

The CV_L can be conveniently determined from the residues measured in duplicate portions of samples containing field-incurred residues. Reference materials are not suitable for this purpose as they are thoroughly homoge-

Table 2 Glossary of terms used in the text

Analyte	The chemical species of which the concentration (or mass) is to be determined, e.g. a pesticide or a metabolite, breakdown product, or derivative of a pesticide.
Analytical sample	The laboratory sample after removal of any parts that are not to be analysed, e.g. bones, adhering soil. It may or may not be comminuted and mixed before withdrawing test portions.
Batch, (analysis)	For extraction, clean-up and similar processes, a batch is a series of samples dealt with by an analyst (or team of analysts) in parallel, usually in 1 day and should incorporate at least one recovery determination. For the determination system, a batch is a series undertaken without a significant time break and which incorporates all relevant calibration determinations (also referred to as an “analysis sequence”, a “chromatography sequence”, etc.). With formats such as 96-well plates, a plate or group of plates may form a batch. A determination batch may incorporate more than one extraction batch, or part of one. In the latter case, determination of recovery is incorporated into one of the determination batches.
Bias	The difference between the mean measured value and the true value.
Blank	(i) Material (a sample, or a portion or extract of a sample) known not to contain detectable levels of the analyte(s) sought. Also known as a matrix blank. (ii) A complete analysis conducted using the solvents and reagents only, in the absence of any sample (water may be substituted for the sample, to make the analysis realistic). Also known as a reagent blank or procedural blank.
Calibration	Determination of the relationship between the observed signal (response produced by the detection system) and known quantities of the analyte.
Certified reference material (CRM)	See reference material.
Combined standard uncertainty	For a measurement result, y , the total uncertainty, $u_{c(y)}$ is an estimated standard deviation equal to the positive square root of the total variance obtained by combining all uncertainty components using the law of propagation of uncertainty (error propagation law)
Comminution	The process of disintegrating a solid sample.
Consignment	A consignment may consist of one or more lots.
Contamination	Unintended introduction of the analyte into a sample, extract, internal standard solution, etc. by any route and at any stage during sampling or analysis.
Residue definition	The definition of a residue is that combination of the pesticide and its metabolites, derivatives and related compounds to which the maximum residue limit (MRL) applies or which is used for dietary exposure assessment.
Determination system	Any system used to detect and determine the concentration or mass of the analyte. For example, gas chromatography- flame photometry detection (GC-FPD), liquid chromatography- mass spectrometry (LC-MS/MS), LC with post-column derivatization, enzyme-linked immuno-sorbent assay (ELISA), thin layer chromatography (TLC) with densitometry, or bioassay.
ECD	Electron-capture detector.
ELISA	Enzyme-linked immuno-sorbent assay.
GLC	Gas chromatography (gas-liquid chromatography).
Half-life of pesticide residues	The time required for decreasing the residue concentration to half of its initial value.
Internal quality control (IQC)	Measurement and recording requirements intended to demonstrate the performance of the analytical method in routine practice. Synonymous with the terms: analytical quality control (AQC) and performance verification. Concurrent AQC data are those generated during analysis of the batch in which the particular sample is included.
Laboratory sample	The sample sent to and received by the laboratory.
Level	In this document, refers to concentration (e.g. mg/kg, $\mu\text{g}/\text{ml}$) or quantity (e.g. ng, pg).
Limit of detection (LOD)	The minimum concentration or mass of the analyte that can be detected with acceptable certainty, though not quantifiable with acceptable precision.
Limit of quantitation (quantification) (LOQ)	The minimum concentration or mass of the analyte that can be quantified with acceptable precision. Should apply to the complete analytical method.
Lot	A quantity of a food material delivered at one time and known, or presumed, by the sampling officer to have uniform characteristics such as origin, producer, variety, packer, type of packing, markings, consignor, etc.
Lowest calibrated level (LCL)	The lowest concentration (or mass) of analyte with which the determination system is successfully calibrated, throughout the analysis batch. It is the level below which there is no experimental evidence to demonstrate that residues will have been detected and calibrated satisfactorily. It will normally correspond to the reporting limit.
Matrix effect	An influence of one or more undetected components from the sample on the measurement of the analyte concentration or mass. The response of some determination systems (e.g. GC, LC-MS, ELISA) to certain analytes may be affected by the presence of co-extractives from the sample (matrix).
Matrix-matched calibration	Calibration intended to compensate for matrix effects and acceptable interference, if present.
Method	A sequence of analytical procedures, from receipt of a sample through to the calculation of results.
Method validation	The process of characterizing the performance to be expected of a method in terms of its scope, specificity, accuracy (bias), sensitivity, repeatability and reproducibility.
MRL	Maximum residue limit
NPD	Nitrogen-phosphorus detector
Performance verification	See internal quality control (IQC)
Procedural blank	See blank

Table 2 (continued)

“Pure” standard	A relatively pure sample of the solid/liquid analyte (or internal standard), of known purity. Usually >90% purity, except for certain technical pesticides.
Reagent blank	See blank
Recovery (of analyte through an analytical method)	The proportion of analyte remaining at the point of the final determination, following its addition (usually to a blank sample) immediately prior to extraction. Usually expressed as a percentage. The mean recovery provides a measure of part of the bias in results. It may or may not be the most significant source of bias in any particular case.
Reference material	Material characterized with respect to its notionally homogeneous content of analyte. Certified reference materials (CRMs) are normally characterized in a number of laboratories, for concentration and homogeneity of distribution of analyte.
Repeatability	The precision (standard deviation) of measurement of an analyte (usually obtained from recovery or analysis of reference materials), obtained using the same method on the same sample(s) in a single laboratory over a short period of time, during which differences in the materials and equipment used and/or the analysts involved will not occur.
Representative analyte	An analyte used to assess probable analytical performance in respect of other analytes notionally sought in the analysis. Acceptable data for a representative analyte are assumed to show that performance is satisfactory for the represented analytes. Representative analytes must include those for which the worst performance is expected.
Representative matrix	Sample material or an extract of a commodity used as an indicator of method performance, or for matrix-matched calibration, in the analysis of broadly similar commodities. Similarity is usually determined according to the content of water, acids, sugars, lipids, secondary plant metabolites, etc., physical characteristics or matrix effects.
Reproducibility	The precision (standard deviation) of measurement of an analyte (usually by means of recovery or analysis of reference materials), obtained using the same method in a number of laboratories, by different analysts, or over a period in which differences in the materials and equipment will occur. Within-laboratory reproducibility is that produced in a single laboratory under these conditions.
Response	The absolute or relative signal output from the detector when presented with the analyte.
RSD	Relative standard deviation (coefficient of variation-CV).
Sample preparation	The first of two processes which may be required to convert the laboratory sample into the analytical sample. The removal of parts that are not to be analysed, if required.
Sample processing	The second of two processes which may be required to convert the laboratory sample into the test sample. The process of homogenization, comminution, mixing, etc., if required, to prepare a statistically well-mixed material from the analytical sample.
Spike or spiking	Addition of analyte for the purposes of recovery determination or standard addition.
Standard	A general term which may refer to a “pure” standard, stock standard, working standard or calibration standard.
Standard uncertainty	Expressed as the standard deviation of an uncertainty component.
Stock standard	The most concentrated solution (or solid dilution, etc.) of the “pure” standard or internal standard, from which aliquots are used to prepare working standards or calibration standards.
Supervised trials	Supervised trials for estimating maximum residue levels are scientific studies in which pesticides are applied to crops or animals according to specified conditions intended to reflect commercial practice, after which harvested crops or tissues of slaughtered animals are analysed for pesticide residues.
Test portion	A representative sub-sample of the analytical sample, i.e. the portion which is to be analysed. Alternative expression: analytical portion
Uncertainty	A parameter associated with the result of a measurement, that characterises the dispersion of the values that could reasonably be attributed to the measurand [2]
Unit (as part of sample)	A single fruit, vegetable, animal, cereal grain, can, etc. For example, an apple, a T-bone steak, a grain of wheat, a can of tomato soup.
Validation	See method validation
Violative residue	A residue which exceeds the MRL or is unlawful for any other reason.
Working standard	A general term used to describe dilutions produced from the stock standard, which are used, for example, to spike for recovery determination or to prepare calibration standards.

The definitions given are based on Refs. [2, 24, 55, 57].

nized. If the relative difference of the residues measured in replicate portions is $R_{\Delta i} = 2(R_{i1} - R_{i2})/(R_{i1} + R_{i2})$, then

$$CV_L = \sqrt{\frac{\sum_{i=1}^n R_{\Delta i}^2}{2n}} \quad (4)$$

Assuming that only random error affects the duplicate measurements, their average must be zero, thus the degree of freedom is equal to n , the number of measurement pairs

[11]. The CV_A is determined from the results of recovery studies performed with spiked analytical portions. The CV_{Sp} can be determined from CV_L and CV_A after the rearrangement of Eq. 3. If the whole sample (e.g. crop unit) is analysed the uncertainty of sample processing is zero.

The relative uncertainty of analytical results corrected for the average recovery, CV_{Acor} , may be calculated from the CV_A values of the procedure.

$$CV_{\text{Acor}} = \sqrt{CV_{\bar{Q}}^2 + \frac{CV_A^2}{n}} \quad (5)$$

where $CV_{\bar{Q}} = CV_A/\sqrt{n}$ is the relative uncertainty of the average recovery, \bar{Q} .

The analytical phase may include, for instance, the extraction (Extr), clean-up (Cln), evaporation (Evap), derivatization (Der) and instrumental determination (Ch). If the relative uncertainties of individual steps were quantified, the combined relative uncertainty of the analytical phase could be expressed as:

$$CV_A = \sqrt{CV_{\text{Extr}}^2 + CV_{\text{Cln}}^2 + CV_{\text{Evap}}^2 + \dots + CV_{\text{Ch}}^2} \quad (6)$$

The uncertainty of chromatographic analysis, S_{Ch} , can be conveniently quantified, while the determination of the contributions of other steps is time consuming, or may only be properly done by applying radio-labelled compounds.

The uncertainty of the predicted analyte concentration (S_{Ch}) is calculated as [2]

$$S_{\text{Ch}} = \sqrt{S_{x_0}^2 + S_{\text{AS}}^2} \quad (7)$$

where S_{x_0} is the standard deviation of the analyte concentration calculated from the calibration data (see Eqs. 8, 9), and S_{AS} is the combined uncertainty of the analyte concentrations in the standard solutions.

The standard deviation of x_0 can be calculated [1] either from ordinary linear regression (OLR)

$$s_{s_0} = \frac{s_{y/x}}{b} \left\{ \frac{1}{m} + \frac{1}{nk} + \frac{(y_0 - \bar{y})^2}{b^2 \sum_i (x_i - \bar{x})^2} \right\}^{1/2} \quad (8)$$

or from weighted linear regression (WLR)

$$s_{s_{0w}} = \frac{s_{(y/x)w}}{b} \left\{ \frac{1}{mw_0} + \frac{1}{nk} + \frac{(y_0 - \bar{y}_w)^2}{b^2 \left(\sum_i w_i x_i^2 - n\bar{x}_w^2 \right)} \right\}^{1/2} \quad (9)$$

where w_0 and w_i are the weights corresponding to the signal, y_0 , of predicted analyte concentration and the i th standard concentration x_i , n is the number of calibration points; k and m are the number of replicate standard and sample injections, respectively. The estimated S_{x_0} has $nk-2$ degrees of freedom.

The relative uncertainty of the predicted concentration is:

$$CV_{x_0} = \frac{S_{x_0}}{x_0} \quad (10)$$

For preparing a working standard solution in three steps the analyte concentration, C_{AS} , is calculated as:

$$C_{\text{AS}} = \frac{wPV_{p1}V_{p2}}{V_{f1}V_{f2}V_{f3}} \quad (11)$$

where w is the mass of analytical standard, P is its purity and V_{f1} , V_{f2} , V_{f3} are the volumes of the volumetric flasks and V_{p1} , V_{p2} are the volumes of pipetted amounts. The combined uncertainty is calculated [2] from the relative uncertainties (e.g. purity of analytical standard: CV_p) of the steps involved:

$$CV_{\text{AS}} = \sqrt{CV_{\text{wht}}^2 + CV_p^2 + CV_{V1}^2 + CV_{V2}^2 + CV_{V3}^2 + CV_{p1}^2 + CV_{p2}^2} \quad (12)$$

One of the basic conditions for the application of the linear regression is that the error in the reference materials used for calibration should be zero or negligible compared to that of the response, S_y . Therefore, the uncertainty of the preparation of the standard solutions should be estimated and the assumption of $S_y \gg S_{\text{AS}}$ should be verified. As the numerical value of S_{AS} changes proportionally with the magnitude of the concentration, the precondition of the application of linear regression could be easily fulfilled by expressing the concentration or amount of the reference material in different units.

In order to compare the errors of y and x they must be expressed in equivalent scales [12]. If the linear relationship is expressed as:

$$y_i = a + bx_i \quad (13)$$

and

$$bx_i = x'_i \quad (14)$$

then

$$y_i = a + x'_i \quad (15)$$

Equation 15 is a linear relationship between x' and y with a slope of 1. If the random error of y and x are σ_y and σ_{AS} , respectively, to satisfy the preconditions for linear regression the following inequalities should stand:

$$|b|\sigma_{\text{AS}} \ll \sigma_y \text{ and } |b|\sigma_{\text{AS}}/\sigma_y \ll 1 \quad (16)$$

The situation is illustrated with an example: the calibration was carried out with standard solutions ranging between 9.19 pg/ μ l and 1103 pg/ μ l. The signal-injected concentration relationship and the parameters of the linear regression ($y = a + bx$) are summarized in Table 3. Assuming 1% uncertainty of standard solution, for the 36.77 pg/ μ l and 0.03677 ng/ μ l concentrations the S_{AS} is equal to 0.368 and 0.000368 ng/ μ l, respectively. As the S_y is unchanged, $66.74=3337 \times 0.02$, the S_y/S_{AS} ratios are 1.81×10^2 and $1.81 \times 10^5 \gg 1$, respectively. Thus expressing the analyte concentrations in ng/ μ l, the precondition of the application of linear regression is fully satisfied. However, if we perform the calculation with the values transformed according to Eq. 15, we obtain $99.5 \times 0.368 =$

Table 3 Parameters of linear regression depending on the concentration scale of analytical standard

Response ^a	385.3	3337	20006	40669	59065	82929	109285	^b	^a
pg/ μ l ^b	9.192	36.768	183.84	367.68	551.52	827.28	1103.04	99.5	1318
ng/ μ l	0.009192	0.036768	0.18384	0.36768	0.55152	0.82728	1.10304	99519	1318
$x_i' = 99.5 \times x$	914.60	3658.42	18292.08	36584.16	54876.24	82314.36	109752.5	1.000	1318

^a The injections were made with an auto-sampler with an average relative 2% random error of the signal

^b Injected concentrations. The estimated relative random error of 36.77 pg/ μ l analytical standard was 1%.

99519 \times 0.000368=36.61, which is not much smaller than 66.74. Thus the inequality described in Eq. 16 does not stand, and the precondition for applying linear regression is not fulfilled. Note that the method of calculation of linear regression (OLR or WLR) does not affect the above considerations as the uncertainty of the analytical standards, S_{AS} , and the uncertainty of the replicate injections, S_y , are independent from the regression calculation.

The confidence intervals for the concentration of the analyte (x_0) predicted from linear regression equation can be calculated as:

$$CL_{x_0} = x_0 \pm t_{2a, nk-2} \times s_{x_0} \quad (17)$$

where S_{x_0} is the standard deviation calculated with Eq. 8 or 9.

Where the combined uncertainty, $S_{c(y)}$, is calculated from the linear combination of the variances of its components, according to the Welch–Statterthwaite formula the degree of freedom of the estimated uncertainty is:

$$v_{\text{eff}} = \frac{S_{c(y)}^4}{\sum_{i=1}^N \frac{S_{i(y)}^4}{v_i}} \quad (18)$$

with $v_{\text{eff}} \leq \sum_{i=1}^N v_i$. The $S_{c(y)} = u_{c(y)}$ values may be replaced with $S_{c(y)}/y$ values where the combined uncertainty is calculated from the relative standard deviations [5].

Evaluation and reporting of the uncertainty of measurements

The acceptable values for the within-laboratory repeatability and reproducibility at different residue levels have been given in the *Guidelines for single-laboratory validation of methods for trace-level organic chemicals* [13], and these values were adopted by the Codex Committee on pesticide residues [14]. The performance of the method should be acceptable if the relative standard uncertainty of recoveries obtained during the validation or performance verification of a method are lower or not significantly higher than the reference values, which can be obtained from relevant guidelines, or from collaborative studies or proficiency tests.

The standard deviation obtained in the laboratory (S) can be considered similar to the established one [15] (σ) if

$$\frac{S^2}{\sigma^2} \leq \frac{\chi_{1-\alpha, v}}{v} \text{ or } F_{0.95, v, \infty} \quad (19)$$

α is usually selected at 0.05 level. The $\chi_{1-\alpha, v}/v$ is equal to the critical value of the F -test at $F_{0.95, v, \infty}$.

If we want to judge the acceptability of the results of replicate measurements made in one laboratory from replicate analytical portions, or from random samples taken independently from the same lot, the critical difference

$$CD = f \times S_L = f \times CV_L \times R_{\text{ave}} \text{ or } CD = f \times S_{\text{Res}} = f \times CV_{\text{Res}} \times R_{\text{ave}} \quad (20)$$

should be calculated taking the CV values from Eqs. 2, 3, and the corresponding factor (f), which depends on the number of replicate measurements and the selected probability, from appropriate statistical tables for Studentized extreme range for individual observations [16]. Where the degree of freedom of the calculated standard deviation is ≥ 19 , the use $f=2.8$ for duplicate measurements is usually acceptable [17].

Where two independent composite samples taken from a single lot are analysed in two laboratories, the results obtained can be considered equivalent if their difference is $\leq CD_{\text{ResR}}$ calculated with Eq. 20 inserting CV_{ResR} from Eq. 21 and the appropriate factor.

$$CV_{\text{ResR}} = \sqrt{CV_S^2 + 0.5(CV_{\text{Sp1}}^2 + CV_{\text{Sp2}}^2) + 0.5(CV_{\text{A1}}^2 + CV_{\text{A2}}^2)} \quad (21)$$

Where CV_S is the uncertainty of sampling, CV_{Sp1} and CV_{Sp2} are the relative uncertainties of sample processing in the laboratories and CV_{A1} and CV_{A2} are the within-laboratory reproducibilities of the analytical method.

If the reproducibility of the analytical procedure between two laboratories is calculated from the relative standard deviation (RSD) of analysis obtained from proficiency or collaborative studies, the $RSD_{\text{R}} = CV_{\text{AR}}$ includes the between-laboratories uncertainty of the analysis and it can be inserted in Eq. 21 instead of $0.5(CV_{\text{A1}}^2 + CV_{\text{A2}}^2)$.

Further guidance for deciding on the acceptability of replicate measurements in various situations, and reporting the results can be found in ISO Standard 5725-6 [17].

The estimates for the repeatability and reproducibility standard deviations and the corresponding variances, based on a single set of measurements, vary from study to study. McClure and Lee [18] suggested the calculation and reporting of their confidence intervals to indicate their closeness to the true value. The upper and lower limits of the variances are calculated with effective degrees of freedom, v_{eff} , and appropriate values taken from the tabulated χ^2 -distribution

$$v_{\text{eff}}S^2/\chi_{1-\alpha, v}^2 \leq \sigma^2 \leq v_{\text{eff}}S^2/\chi_{\alpha, v}^2 \quad (22)$$

where $1-\alpha$ is the probability that σ^2 lies between the calculated intervals. The authors give practical examples for the calculation of confidence intervals for the repeatability, reproducibility standard deviations and their ratios.

According to the EURACHEM/CITAC Guide the expanded uncertainty should be used in reporting the results of measurements. It can be calculated as

$$U_p = t_{2\alpha, v_{\text{eff}}} \times U_c \quad (23)$$

and the result can be reported as $Y=y \pm U_p$.

The EURACHEM/CITAC Guide recommends the use of a coverage factor of 2 instead of the $t_{2\alpha, v_{\text{eff}}}$. This approach is appropriate if the combined uncertainty of the measurement, U_c , is estimated based on a large (>20) data set.

Contribution of the major steps of residue analysis to the combined uncertainty and trueness of the results

Currently the main emphasis is placed on the identification and quantification of random errors and lesser attention is given to the sources of systematic errors, which can grossly affect the accuracy and trueness of the results. Detailed analysis of the sources of errors can be found in a number of publications [19–21]. The most important factors are summarized below.

Sampling

The distribution of pesticide residues in crop units or primary samples within a treated area is strongly skewed positive [22]. The within-field variation, CV, of pesticide residues, for instance, in 16 sets of apple data each consisting of 100–320 apples ranged from 39 to 119%. The logarithmic transformation of the residues measured in unit samples resulted in approximately normal distribution in about half of the 76 data sets tested [23].

The variability of residues in composite samples, $s_{\bar{x}}$, depends on the sample size (the number of primary

samples, n , making up the composite sample) and the spread of residues (s) in primary samples:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} \quad (24)$$

The Codex sampling procedure specifies the minimum size of samples ($n=10$ for small and medium size crops and 5 for large crops) for determining pesticide residues in plant material [24]. The composite random samples drawn from over 8,800 unit crop residue data confirmed the preliminary conclusions [25, 26] that samples of size 5 are approaching normal distribution, size 10 are close to normal and for practical purposes can be considered normal, and size >25 are normally distributed [23]. The uncertainty did not depend on the crop or pesticide residue. The average relative standard uncertainties for sample size 10 and 5 were 28 and 40%, respectively. They are somewhat lower than those estimated previously based on limited data [25]. The analysis of replicate composite samples taken from the treated areas supports the values obtained with drawing random composite samples from residues in crop units [27].

The relative uncertainty of 28% can be used as typical sampling uncertainty of small- and medium-size crops with minimum sample size of 10. The estimated uncertainty of 40% for large crops, such as cabbage, cucumber, mango and papaya, was estimated from the residue data obtained for medium size crops assuming a sample size of 5. Therefore it can be used as a preliminary estimate, as residue data on cucumbers were only available for large crops. The average CV_S value was calculated from 76 data sets, thus the corresponding standard deviation, S_S , has a degree of freedom of 75, which can be considered infinite, except when the effective degree of freedom is calculated with Eq. 18.

The change of the pesticide residue concentration in/on the treated objects is usually illustrated with decline curves obtained by plotting the residue concentration as the function of the time elapsed between the treatment with the pesticide and the sampling. The uncertainty of sampling may substantially affect the shape of decline curves, and the estimated half-life of residues or the expected average residue concentration at a given time after the treatment with the pesticide, and the width of their confidence intervals. For instance, the average residue at 10 days after the application of pesticide was calculated from the decline curves obtained from 7 sets of composite samples of size 24 taken independently from the treated field [27]. At day 10 the calculated average residues, with their confidence intervals in brackets, ranged from 0.025 (0.008–0.08) to 0.041 (0.012–0.141) mg/kg. The estimated time for decreasing the initial residue concentrations to half varied similarly. Since usually a single decline study is carried out in an experimental plot, the uncertainty of the estimated values is not recognized, and

cannot be taken into account. Precaution is required when conclusions are drawn from such results.

The field-to-field variation of average residues is also very large. Evaluation of the field trials reported to the FAO/ World Health Organization (WHO) Joint Meeting on Pesticide Residues revealed that the residue distribution between-fields is strongly skewed positive with an apparent *CV* of the residues ranging from 30 to 150% in large composite samples [22]. Note that the between-fields residue distribution is far from normal, and normal statistics cannot be used for estimating residue ranges, but the calculated *CV* can be used to illustrate the spread of residues. Thus, if samples are taken from mixed lots the sampling uncertainty is much larger (around 70–90% on an average) than that corresponding to a single lot. Further, in a mixed consignment a lot containing residues above the maximum residue limit (MRL) can easily remain unobserved. Consequently, sampling of mixed lots should be avoided as far as practically possible.

In addition to the inevitable uncertainty of random sampling, very large systematic or gross errors can be caused by, for instance, non-representative samples, degradation of analyte during the shipping or storage of samples, and contamination or mixing up samples.

Most of these errors remain unnoticed by the analysts receiving the samples, and decisions are made based on invalid results.

The Codex and many national MRLs, for pesticides in animal products, except milk and eggs, apply to the primary sample (e.g. a tissue or organ of a single animal). The distribution of residues in such samples is far from normal. Therefore, the so-called distribution-free statistics should be used for deciding on the required number of samples (*n*) to be taken for estimation of a specified percentage of the population (β_p) with a given probability (β_t), or detection of the violation of an MRL [24]. Their relationship is described by:

$$1 - \beta_t = (\beta_p)^n \text{ or } n = \log(1 - \beta_t) / \log(\beta_p) \quad (25)$$

Sample preparation

The sample preparation should be clearly distinguished from sample processing. The sample preparation, involving the separation of the portion of the laboratory sample to produce the analytical sample, can be a major source of systematic and uncontrollable gross error. The Codex MRLs are based on the specified portion of the commodity [28]. The description of the commodity to which the MRLs apply may be different in some national legislations. Since the residues are not equally distributed within a commodity, the results of the analysis can be substantially different in most cases if, for instance, the portions analysed are: cabbage and lettuce with or without wrapper leaves; whole citrus fruit or only the pulp; washed or unwashed potato; peach or plum with or without

stone. Therefore, the analyst should select the portion of the commodity to be analysed according to the relevant protocol or the purpose of the work and describe the procedure applied when reporting the results.

Sample processing

Analysts usually assume that their sample processing procedure results in analytical portions which are representative of the analytical sample, even if they cut up the plant material into small pieces with a knife or pair of scissors. A number of recent studies [29–31] reported on the estimation of uncertainty of analytical procedures did not consider the contribution of sample processing either.

Collaborative studies and proficiency tests are carried out with material tested for homogeneity. CRMs are also carefully homogenized. The recovery studies are usually performed with test portions spiked shortly before extraction. Such studies cannot reveal any information on the efficiency of sample processing.

Systematic studies on the uncertainty of sample processing [32–34] revealed, however, that the random error of sample processing can be very high and may contribute substantially to the combined uncertainty of the results. When the analytical sample is statistically well mixed, a sampling constant, K_s , can be defined:

$$K_s = mCV^2 \quad (26)$$

The K_s is the weight of a single increment that must be withdrawn from a well-mixed material to hold the relative sampling uncertainty, *CV%*, to 1% at the 68% confidence level [35]. It provides the relationship between the mass of analytical portion and the RSDs of the residues, expressed in percentage, being in the analytical portions, and indicates that the square of relative uncertainty of sample processing is inversely proportional to the mass of analytical sample. The sampling constant may be influenced by the equipment used for chopping/homogenizing the sample and the sample matrix, but it is independent from the analyte. Commodities with a soft texture, such as oranges, can be relatively easily homogenized with a sampling constant ≤ 1 kg. The homogenization of commodities with hard peel and soft flesh, such as tomatoes, can be much more difficult and the average sampling constant may be as high as 18 kg with confidence intervals of about 11 and 44 kg [34]. Taking the upper confidence limit of sample processing (44 kg) and 30- and 5-g analytical portions, we obtain about 38 and 94% for uncertainty of sample processing, respectively. In another study, the uncertainty of sample processing of sweet peppers was found to be negligible [36]. These findings highlight the importance of testing the uncertainty of sample processing for each piece of equipment and type of commodity, especially where small sample portions are used for extraction. It should always be verified that

Table 4 Predicted between-laboratory reproducibility in pesticide residue analysis based on the Horwitz equation:RSD, relative standard deviation

Concentration (mg/kg)	RSD%	Examples of occurrence of pesticide residues
100	8	Raw agriculture commodities after post-harvest treatment
10	11	
1	16	Raw agriculture commodities, meat, milk
0.1	23	
0.01	32	Surface water
0.001	45	
0.0001	64	Drinking water

statistically well-mixed analytical samples can be prepared which allows the withdrawal of small (2–10 g) test portions for extraction.

The efficiency of sample processing can be substantially improved by applying dry ice to the sample during chopping or grinding. In the presence of dry ice even frozen alfalfa, wheat straw or maize stalk can be ground to a powder.

The efficiency of sample processing can be checked with the reanalysis of replicate analytical portions containing field-incurred residues. The relative uncertainty can be calculated with Eqs. 4, 3.

During the sample processing the analytes may evaporate and they are subject to hydrolysis and enzymatic reactions. Systematic studies [37, 38] revealed substantial decrease (40–70%) of the concentration of several analytes such as captan, captafol, folpet, chlorothalonil and dichlofluanid during sample processing at ambient temperature, with extremes such as captan in lettuce (96%) and chlorothalonil in onion (100%).

Cryogenic processing (processing of deep-frozen samples in the presence of dry ice) significantly improved the stability of residues, and the vast majority (94 of 96) of pesticides were stable. The losses of several pesticides [bitertanol (95%), heptenophos (50%), isophenphos (40%) and tolylfluanid (48%)] reported to occur at ambient temperature during processing of apples did not occur during cryogenic processing [39].

The loss of residues during sample processing may remain unnoticed, as the usual recovery studies incorporate the loss of analyte only from the point of fortification of analytical portions. Therefore, the stability of residues during sample processing has to be specifically studied during the validation of the analytical procedure. The chopping/mincing of deep-frozen samples is difficult and requires very robust equipment. A less demanding intermediate solution, involving cooling down the fresh sample quickly with dry ice during the processing and continuing the homogenization until free-flowing powder is obtained, may be applicable in many cases. Its efficiency should be tested with compounds found to be unstable during processing at ambient temperature.

Analysis

Horwitz and co-workers [40] analysed the results of over 7,000 collaborative studies and defined the relationship between the analyte concentration and the inter-laboratory reproducibility standard deviation. The equation for the calculation of the predicted relative standard deviation, PRSD, can be written in various equivalent forms:

$$\text{PRSD}\% = 2^{(1-0.5 \log C)} = 2C^{(-0.1505)} \quad (27)$$

Where C is the analyte concentration expressed in decimal fraction (1 mg/kg=0.000001).

These equations describe the so-called Horwitz curve, and have been widely used for the assessment of the performance of the laboratories in many proficiency testing programmes and provided the basis for the acceptance criteria of Codex Alimentarius Commission establishing food quality standards [41]. They are applicable to all concentrations, methods and analytes. The predicted values for the typical pesticide concentration range are given in Table 4.

The relationship between the actual and predicted CV is described by the HORRAT=RSD/PRSD value. It can be expected to be about 0.5–0.7 for within-laboratory studies and about 1 for between-laboratory studies [42], although values in the interval from about 0.5–2 times the expected value may be acceptable [43].

Large proficiency and quality control databases of pesticide residues (>100,000 measurements in 14 data sets from the State of California, US Department of Agriculture, US FDA, Standards Council of Canada and FAPAS, UK) were statistically evaluated [44]. The USDA acceptability limits of 50–150% were used for individual recoveries and data sets consisting of ≥ 8 results were considered. The analysis indicated 94% average recovery for pesticide residues being present in fruits, vegetables and milk at an average concentration of 0.1 mg/kg with an overall average RSD of 17%. The average apparent HORRAT value was 0.8. These indicate the general applicability of Eq. 27.

The analysis of variance revealed that none of the major factors (analyte, commodity, laboratory and concentration) alone has much of an effect on the overall variability. The analyte factor was responsible for about 10% of the attributable variance and 7% could be ac-

counted for by the commodity-analyte interaction. The other factors had a minor effect. About 60–70% of the variance could not be associated with any particular factor or combination of factors. The authors concluded that most of the variability of pesticide residue analysis is “random” in the sense of being inherent and not assignable to specific factor fluctuations. Further, because the pesticide-commodity interaction term is small, it is reasonable to make comparisons between various pesticides and commodities. This important finding supports the concept of using “typical” representative commodities and pesticides for characterizing the performance of multi-residue analytical methods [13].

The results of 61 proficiency tests from 5 countries involving 24 matrices and 869 RSD values were analysed to determine the applicability of Eq. 27 for a much narrower concentration and analyte range [45]. It was found that the between-laboratory RSD for fruits, vegetables and cereal grains remained practically constant at an average of 25% over the entire concentration range of 0.001–10 mg/kg, while for fatty foods the Horwitz equation remained valid. The authors identified a number of factors that may be responsible for the practically constant RSD. Those considered most important are: proficiency tests are often conducted at pesticide concentrations near the limit of quantification, but pesticides which are difficult to analyse are included at higher concentrations thus the precision parameters reported reflect the additional problems related to their analysis; the between-laboratory variation of the efficiency of extraction of pesticides of widely differing polarity from matrices of substantially different composition; pesticide residues occurring in plant materials are usually less stable than those present in animal products; the matrix effect of extracts of fruits, vegetables and grains on the response of pesticides may be more variable than the matrix effect of extracts of animal products. It should be noted, however, that the estimated RSD of 25% is very close to the 23% calculated from Eq. 27 for the mean (0.1 mg/kg) of the examined residue range (0.001–10 mg/kg).

The statistical analysis [46] of over 2,900 recovery values obtained in fruit and vegetable samples with 105 pesticide residues ranging from 0.001 to 2 mg/kg in 3 laboratories over 4 years resulted in average within-laboratory long-term reproducibility RSD values in the range of 15–30%, which are in accord with the above findings.

In addition to the assessment of the overall performance of the methods, it is worth considering the sources and magnitude of individual errors affecting their uncertainty and trueness.

Extraction and clean-up

When an MRL is established the residue components included in the limit are defined. As the MRL should be

expressed as simply as possible to facilitate regulatory control, the residue definition for MRL often includes only the parent compound or a single residue component. In contrast, the residue definition for the purpose of dietary risk assessment must include all residue components of toxicological significance, which can be the parent compound, metabolites or degradation products. Consequently, there can be a significant difference in the reported results based on the measurement of a single residue component or several components. One of the often-overlooked sources of systematic error is the selection of residue components (parent compound and some metabolites or degradation products) to be determined. To check compliance with MRLs or determining residues for dietary risk assessment, all residues included in the residue definitions should be recovered as completely as possible. However, residues not included in the residue definition should not be taken into account in the reported results.

Another major source of bias may also be the efficiency of extraction. The recovery studies carried out with portions of chopped samples fortified before extraction may not reveal realistic information on the efficiency of extraction of field-incurred residues, which can be most conveniently determined with radio-labelled compounds applied in a similar manner as in the practical use of the pesticide. Where the extraction efficiency is not available from radio-labelled studies, comparison of residue levels obtained from field-incurred residues with different solvents and extraction procedures can provide information on the efficiency of extraction. Among the most widely used extraction solvents, ethyl acetate and acetone gave similar results, but ethyl acetate proved to be more efficient for highly polar residues [47].

The contribution of various steps of the residue analytical methods to the overall variability of the results were systematically studied by applying ^{14}C -labelled chlorpyrifos as an internal standard and a well-established and widely used multi-residue method based on ethyl acetate extraction, SX-3 GPC clean-up and GC-electron-capture detector (ECD) and nitrogen-phosphorous detector (NPD) determination [38]. As the liquid scintillation detection of labelled chlorpyrifos could be carried out with a relative uncertainty of about 1–1.5% at any stage of the determination process, the uncertainty of the individual procedures could be precisely quantified. The overall uncertainty of the method, CV_A , calculated from the GC determination of chlorpyrifos, was 17.2%. The uncertainty of the extraction and clean-up steps were 5.5 and 5.6%, respectively. Further reproducibility studies, performed by different analysts in the IAEA Agrochemical Unit on different occasions using ^{14}C -labelled chlorpyrifos and lindane at 0.05–1 mg/kg level, confirmed that the ethyl acetate extraction can be carried out with a relative repeatability standard deviation of ≤ 2.5 –5%.

Table 5 Illustration of the uncertainty components of a diluted analytical standard solution

	Weight (g)	Relative uncertainty		
		Volumetric	4-digit balance	5-digit balance
Purity of analytical standard		0.00058	0.00058	0.00058
Weighting of analytical standard	0.0254	0.00124	0.00664	0.00124
Dilution to 25 ml in an A-grade volumetric flask	17.21334	0.00121	9.798×10^{-6}	1.837×10^{-6}
Taking 100 μ l	0.07199	0.00657	0.0023	0.00044
Dilution to 25 ml in an A-grade volumetric flask	17.20969	0.00121	9.798×10^{-6}	1.837×10^{-6}
Taking 100 μ l	0.07125	0.00657	0.0023	0.00044
Dilution to 25 ml in an A-grade volumetric flask	17.19284	0.00121	0.00121	0.00121
Combined uncertainty ($\pm 7^\circ\text{C}$ temperature range)		0.00963	0.00744	0.00193
Combined uncertainty ($\pm 2^\circ\text{C}$ temperature range)		0.00942	0.00738	0.001709
Combined uncertainty ^a		0.01418	0.02561	0.004978
Calculated concentration		1.624×10^{-8} ^b		1.757×10^{-8} ^c

^a Calculated with $\pm 7^\circ\text{C}$ temperature range for method (b) (dilutions of 1 ml and 10 μ l aliquots instead of 2×100 μ l).

^b Concentration of standard solution calculated with the nominal volume of volumetric glassware.

^c Concentration of standard solution calculated with the weights except the last dilution.

GLC and HPLC determination

In an extensive inter-laboratory comparison study the between-laboratories RSD of GLC determination of 18 organophosphorus pesticides, fenpropimorph and iprodione in ethyl acetate extracts of apple, carrot, tomato, orange and wheat with 47 GC instruments alone ranged between 16 and 56% [45]. The GC response was also shown to be highly variable within one laboratory due to the matrix effect of the sample extract [48–51]. The uncertainty estimation of routine instrumental analysis [52] indicated that the GLC analysis may amount to the 75–95% of the CV_A at the lowest calibrated level.

Since the contribution of GLC determination to the combined uncertainty of the analysis is very high, the factors (see Eq. 7) affecting its accuracy and precision were studied in more detail.

Diluted standard solutions were prepared by different methods to assess the error in their concentration:

- 25.4 mg of analytical standard ($99.9 \pm 0.1\% = 0.999 \pm 0.001$) was weighed with a 5-digit analytical balance (linearity ± 0.03 mg, repeatability SD ± 0.02 mg). The stock solution was prepared by dissolving the standard in an A-grade 25-ml (± 0.04 ml) volumetric flask. 100 μ l of stock solution was transferred with a Hamilton syringe (± 1 μ l) to a 25-ml volumetric flask in order to obtain the intermediate solution. The working solution was made by taking 100 μ l of intermediate solution with a Hamilton syringe and diluting it to 25 ml in a volumetric flask.
- 25.4 mg of analytical standard was weighed with a 4-digit analytical balance (linearity ± 0.2 mg, repeatability SD ± 0.03 mg) and diluted to 25 ml. 1 ml of stock solution was pipetted (± 0.007 ml) into a 25-ml volumetric flask, made up to mark, then 10 μ l of intermediate solution (± 0.1 μ l) was taken with a Hamilton syringe and diluted to 25 ml.

bility SD ± 0.03 mg) and diluted to 25 ml. 1 ml of stock solution was pipetted (± 0.007 ml) into a 25-ml volumetric flask, made up to mark, then 10 μ l of intermediate solution (± 0.1 μ l) was taken with a Hamilton syringe and diluted to 25 ml.

The uncertainties were calculated assuming $\pm 7^\circ\text{C}$ change of temperature during the day, which can occur if the laboratory is not air-conditioned, and $\pm 2^\circ\text{C}$ for an ideally air-conditioned laboratory. In addition, the uncertainties were calculated for the case where the standard solution was prepared based on weighing without the last dilution to 25 ml.

The results are summarized in Table 5. The combined uncertainty of the diluted solution made by volumetric dilution is about 0.96% with method (a) and 1.4% with method (b) (the limiting factor (1%) is pipetting of 1 ml). The larger temperature range only had a marginal effect on the uncertainty. Preparing the solutions based on weighing with a 5-digit balance significantly improved the precision of the standard solution (0.2%), but a 4-digit balance provides only a slight improvement (0.7%) because weighing of 25 mg of material has an uncertainty of about 0.6%. Weighing small amounts of analytical standards was considered to be the major source of the combined uncertainty in other studies [31] as well. Another notable finding is the 7.5% difference, which is >7.5 times higher than the estimated uncertainty of the solution, in the concentrations calculated from the nominal volumes of the A-grade glassware and from the results of weighing. Since the series of standard solutions for the calibration are prepared with different glassware, the deviation from the nominal value may be positive or neg-

ative and may significantly increase the uncertainty of the calibration solutions.

Considering the repeatability of injections with modern auto-samplers (<2%), the uncertainty of analytical standards used for calibration should be around 0.3–0.5% to satisfy the preconditions of linear regression (Eq. 16). Therefore the standard solution should be prepared based on weight measurements. Where the intermediate solutions of the components of the mixtures of analytical standards are prepared according to method (a) and the mixture is prepared in the last step, the number of compounds in the mixture does not affect the combined uncertainty of the mixture.

The uncertainty of predicted analyte concentration based on multi-point calibration was studied using 68 GC-ECD, GC-NPD and HPLC- ultra violet (UV) calibration data sets obtained during daily work in 8 laboratories. The data sets consisted of calibration points from 3×2 to 8×3. WLR and OLR methods were applied for all data sets [53]. It was found that, regardless of the actual concentration range of the calibration, the relative random error (calculated with Eqs. 8, 9, 10) at lowest calibrated level (LCL) ranged in the case of OLR between 3 and 110%, typically 20 and 50%, and for WLR between 1 and 18%, typically 3 and 9%. At or above 1/3 of the calibrated range the CV_{x_0} ranged between 1 and 7%, and there was no significant difference between the estimates obtained with WLR or OLR. Similarly, no difference was found in the uncertainty of the predicted concentration at the upper end of the calibrated range estimated with OLR or WLR when pesticide residues in apples were determined with ECD [52].

The method of standard addition did not provide the expected advantages in compensating the matrix effect, mainly because the calibration line obtained with matrix-matched standard solutions rarely passes zero coordinates [54]. Consequently, when analytical standards are added to the sample extract the calculated analyte concentration in the sample is biased proportionally to the intercept, which would be obtained if the analytical standards were given to the blank sample extract.

Combined uncertainty of the results

For the estimation of the uncertainty of the residue data the uncertainties of sampling, sample processing and analysis have to be taken into consideration. As shown, their expectable relative uncertainties range between 28 and 40%, not-significant –94% and 7–30%, respectively. Table 6 gives examples, based on the observed or likely uncertainties, for the dependence of combined uncertainty on the uncertainties of individual steps. The results indicate that any of the three main components can be the major contributing factor to the combined uncertainty,

Table 6 Effect of uncertainty of sampling sample processing and analysis on the combined uncertainty of residue data

CV_S	CV_{Sp}	CV_A	CV_{Res}	CV_L
0.28	0.94	0.25	1.01	0.97
0.28	0.38	0.25	0.53	0.45
0.28	0.1	0.25	0.39	0.27
0.28	0.1	0.2	0.36	0.22
0.28	0.1	0.15	0.33	0.18
0.28	0.1	0.1	0.31	0.14
0.28	0.05	0.25	0.38	0.25
0.28	0.05	0.15	0.32	0.16
0.28	0.05	0.1	0.30	0.11
0.28	0.03	0.1	0.30	0.10
0.28	0.05	0.05	0.29	0.07
0.4	0.1	0.25	0.48	0.27
0.4	0.1	0.15	0.44	0.18
0.4	0.1	0.1	0.42	0.14

which cannot be smaller than any of its components (Eqs. 2, 3).

The uncertainty of sample processing can be very high (around 100%) especially when small test portions (2–5 g) are withdrawn for extraction. The resulted combined uncertainty would likely be unacceptable for most purposes. On the other hand, it can be seen that decreasing the uncertainty of sample processing from 10 to 5% decreases the combined uncertainty by only 1%. Taking into account the technical difficulties of obtaining $CV_{Sp} \leq 5\%$ for many sample types and the possible consequences of the extended and intensive homogenization process on the stability of residues, one may conclude that a CV_{Sp} of $\leq 10\%$ would be an optimal target value in pesticide residue analysis.

Increasing the sample size (Eq. 24), which is limited by the capacity and efficiency of sample processing equipment, may reduce the sampling uncertainty. If a sample of size 25 would be taken and processed, the sampling uncertainty would be reduced to 18%. As the MRLs refer to the residues in samples of specified size, the effect of increased sample size on the combined uncertainty must be considered in interpreting the results [19].

The information on the relative contribution of various components to the combined uncertainty may be very useful in optimizing the experimental design for a study or an analytical procedure. The number and size of samples to be taken, the mass of analytical portion and the number of replicate samples to be analysed can be better decided if the uncertainties of the procedures are taken into account. Attempting to reduce the uncertainty of individual steps may not be worth doing beyond a certain level, as the overall uncertainty will not be improved significantly.

Incorporation of the uncertainties of sampling and sample processing may substantially increase the uncertainty of the reported results compared to the uncertainty values based on the analysis alone. Taking into consid-

eration the typical uncertainties of sampling (28, 40%) and within-laboratory reproducibility of analysis (15–25%), and a realistic 10% for sample processing, we obtain 33–39% and 42–48% combined uncertainty for medium- and large-size crops, respectively. This uncertainty should be taken into account when one decides on the acceptability of the results. For instance, if the mean residue in a composite sample of size 10 is 1 mg/kg, the difference between the residues in 2 replicate analytical portions (assuming $v_A=14$, $v_{Sp}=6$ [33] the v_{Ceff} is 20 and 18 for $CV_L=18$ and 27%, respectively) can be up to $2.95 \times 0.18 \times 1 \text{ mg/kg} = 0.53 \text{ mg/kg}$ and $2.971 \times 0.27 \times 1 \text{ mg/kg} = 0.8 \text{ mg/kg}$ (Eqs. 3, 20). Similarly, where two independent samples were taken from a lot and analysed in two different laboratories ($CV_{Res}=0.39$, $v_{Ceff}=61$) the acceptable difference between the results obtained would be $2.829 \times 0.39 \times 1 \text{ mg/kg} = 1.1 \text{ mg/kg}$.

Consideration of the bias of the measurements

The average recoveries reported, indicating the sum of systematic errors (bias) of the measurements, usually range between 80–100%, but occasionally lower and to a lesser extent higher values might occur. Average recoveries between 70–110% are generally accepted, though it is recognized that lower recoveries may be unavoidable and acceptable if multi-residue methods are applied [21]. When the average recovery is statistically significantly different, based on the *t*-test [55], from 100%, the results should generally be corrected for the average recovery [2, 56]. Since the uncertainty of the mean recovery, $CV_{\bar{Q}} = CV_A / \sqrt{n}$, affects the uncertainty of the corrected results (Eq. 5), it is important that the mean recovery is determined from a sufficiently large number, *n*, of replicate analyses incorporating the whole range of matrices and the expected concentration levels of analyte. If the mean recovery is determined from ≥ 15 measurements, the contribution of the uncertainty of recovery to the combined uncertainty becomes negligible ($CV_{Acor} \leq 1.03 CV_A$). On the other hand, if corrections would be made with a single procedural recovery, the uncertainty of the corrected result would be $1.41 CV_A$. Therefore, such correction should never be done. One may argue that if the spiked sample for determining the recovery rate is analysed along with the unknown samples in the same sequence, it is more likely that a low recovery rate coincides with low results of the unknown sample, thereby leading to a correlation between the results. The experimental data, however, do not support this argument. The recovery values obtained for performance verification usually symmetrically fluctuate around the mean recovery, which indicate that the measured values are subject to random variation. If the procedural recovery performed with an analytical batch is within the expected range, based on the

mean recovery and within-laboratory reproducibility of the method, we have all reason to assume that the method was applied with expected performance, and the results obtained are independent from each other. It means that the variability of the results is random and not serially correlated. Consequently, it is unlikely that within one analytical batch all the residues will be low or high. Further, the dispersion of the data can be mainly influenced by the so-called Type-B effects [6] depending on the nature of individual samples which are different within one analytical batch and can be quite different from the one used for spiking. Therefore, the correct approach is to use the typical recovery established from the method validation and the long-term performance verification (within-laboratory reproducibility studies) for correction of the measured residue values, if necessary.

Correction for recovery is not generally accepted in pesticide residue analysis. The major argument for not correcting the results is based on the fact that the legal MRLs are established taking into account the results of supervised field trials, which were not corrected for recovery. Using uncorrected residue data for estimating MRLs is generally valid because the supervised field trial samples are analysed by specific individual methods optimized for the analysis of the given pesticide residue-commodity matrix, and the recovery of the analyte with these methods is usually above 90%, which is not significantly different from 100%. However, if a given analyte is determined with a multi-residue method providing an average recovery significantly lower than 100%, the analyst introduces a systematic error if the uncorrected residue concentration is reported. In order to avoid any ambiguity in reporting results of supervised field trials, when a correction is necessary, the analyst should give the uncorrected as well as the corrected value, and the reason for and the method of the correction [57].

Another problem is providing data for the estimation of exposure to pesticide residues. In this case the best estimate of the actual residue level should always be given, that is the residues measured should be corrected for the mean recovery, if that is significantly different from 100%.

Conclusions

The accuracy and uncertainty of the results of pesticide residue analysis are influenced by the performance of sampling, sample preparation, sample processing and the analysis.

The inevitable random variation of residues in composite samples depends on the size of the sample. When the Codex sampling procedure is applied, the typical uncertainty of sampling of small- and medium-size crops from a single lot is about 28%, while for large crops about 40% can be expected. Note that the sampling uncertainty

can be much larger (130–150%) if the commodity sampled composed of several lots. Further information is required on the residue distribution in natural units of leafy and root vegetables before the sampling uncertainty, and the overall combined uncertainty of residue data can be estimated in such crops.

Since the MRL refers to the residues of tissues of individual animals, and the variation of residues in different muscle and fat tissues is relatively small compared to plants, the sampling has negligible effect on the uncertainty of the residue data obtained for animal products. However, a much larger number of samples is required to verify the compliance of a lot with the MRLs, than in case of plant commodities from which composite samples are taken.

In addition to random error, various systematic and gross errors may occur. Most of these errors, made during sampling and sample preparation, remain unnoticed by the analysts receiving the samples, therefore application of clearly written sampling instructions applied by trained sampling officers are the basic preconditions for obtaining reliable results on the residue content of the sampled commodity.

Sample preparation and processing may also introduce substantial errors which are difficult to monitor and quantify, and strict adherence to agreed sufficiently detailed protocols standard operating procedures (SOPs) are necessary to keep them at the minimum. A large portion of the residues may be lost during sample processing. Comminuting deep-frozen samples in the presence of dry ice eliminated or reduced the loss significantly for the majority of pesticides tested.

The uncertainty of the analysis is the result of the combination of a number of factors. The quantification of their individual contribution to the combined uncertainty is very often difficult and time consuming. Therefore, the most realistic estimate of the uncertainty of the analytical phase can be obtained from the proficiency tests, in combination with internal quality control check results, which include the between-laboratory reproducibility component as well. Recent evaluations of large data sets of proficiency and internal quality control tests indicate that the average between-laboratories reproducibility RSD (CV_A) can be expected to be between 17 and 25%. It is emphasized, however, that these values do not include the uncertainty of sampling and sample processing.

The analyte and its interaction with concentration and laboratory contributed up to only 10% of the total variance. For plant matrices the analyte concentration does not seem to affect the relative uncertainty in the 0.001–10 mg/kg range, provided that the analyte concentration is above the level of 2–3 limit of quantification (LOQ). These findings support the establishment of typical per-

formance parameters (average recovery, repeatability and reproducibility CV_{Atyp}) of multi-residue procedures based on the validation of the method with representative analytes and commodities.

As the estimation of standard deviation based on a few samples is very imprecise, the method performance parameters obtained during method validation must be confirmed and if necessary refined with the results of internal quality control tests performed to verify the daily performance of the method. The typical method performance parameters established based on collaborative studies or proficiency tests cannot be directly used in individual laboratories, unless the laboratory first demonstrates similar or better analytical reproducibility values.

The reanalysis of replicate analytical portions is a very powerful internal quality control check, as it provides information on the reproducibility of the whole procedure including sample processing, therefore it should be included in each analytical batch.

The major source of uncertainty of analysis is the GLC and HPLC determination that may contribute up to 75–95% of the variability at the lower third of the calibrated range. Quantification of residues at the lowest calibrated level should be avoided. The analytical standard solutions should preferably be prepared based on weighing with a 5-digit balance to keep their uncertainty at minimum. The extraction and clean-up procedures used in multi-residue methods may be performed relatively reproducibly, but attention is required for selection of appropriate residue components and assuring the efficiency of extraction.

Average recoveries of 80–100% were reported for a wide range of pesticides. In general the 70–110% average recovery range is acceptable. The laboratory should report the residue values as accurately as possible. It may require the correction of the measured values with the average recovery if it is significantly different from 100%, but never with a single recovery value obtained in the analytical batch.

When the Codex standard sampling procedure is applied for small and medium, and large crops the overall combined relative uncertainty of residue data can be expected in the best case to be in the range of 33–39 and 42–48%, respectively. These values are much larger than those usually reported by laboratories for their analyses, but they realistically reflect the practical situation, and should be accepted and taken into consideration when decisions are made based on the results.

Acknowledgements The useful comments of Ms. Britt Maestroni and Dr. Eurof Walters are gratefully acknowledged.

References

- Miller JN, Miller JC (2000) *Statistics and chemometrics for analytical chemistry*, 4th edn. Pearson Education Limited, Harlow, UK
- EURACHEM/CITAC (2000) *Guide quantifying uncertainty in analytical measurements*, 2nd edn. <http://www.measurementuncertainty.org>
- Analytical Methods Committee (1995) *Analyst* 120:2303–2308
- Nordic Committee on Food Analysis (NMKL) (1997) *Estimation and expression of measurement uncertainty in chemical analysis*, NMKL Procedure No. 5
- ISO-GUM (1995) *Guide to the expression of uncertainty of measurement*. International Standard Organisation (ISO), Geneva
- Horwitz W (1998) *J AOAC Int* 81:785–794
- Thompson M, Howarth RJ (1980) *Analyst* 105:1188–1195
- Hill ARC, von Holst C (2001) *Analyst* 126:2044–2052
- Horwitz W, Jackson T, Chirtel SJ (2001) *J AOAC Int* 84:919–935
- ISO-5725-3 (1994) *Accuracy (trueness and precision) of measurement methods and results. Part 3. Intermediate measures of the precision of a standard measurement method*. ISO, Geneva
- Youden WJ, Steiner EH (1975) *Statistical manual of AOAC*. AOAC, Washington
- Mandel J (1984) *J Qual Techn* 16:1–14
- Alder L, Ambrus A, Hill ARC, Holland PT, Lantos J, Lee SM, MacNeil JD, O'Rangers J, van Zoonen P (2000) *Guidelines for single-laboratory validation of analytical methods for trace-level concentrations of organic chemicals*. In: Fajgelj A, Ambrus A (eds) *Principles and practices of method validation*. Royal Society of Chemistry, Cambridge, pp 179–252
- Report of the 35th Session of codex committee on pesticide residues (2003) <ftp://ftp.fao.org/codex/alnorm03/A10324Ae.pdf>
- ISO-5725-4 (1994) *Accuracy (trueness and precision) of measurement methods and results. Part 4. Basic methods for the determination of the trueness of a standard measurement method*. ISO, Geneva
- Lentner C (ed) (1982) *Geigy Scientific Tables vol 2*, 8th revised edn. Ciba-Geigy Ltd, Basle
- ISO-5725-6 (1994) *Accuracy (trueness and precision) of measurement methods and results. Part 6. Use in practice of accuracy values*. ISO, Geneva
- McClure F, Lee JK (2001) *J AOAC Int* 84:940–946
- Ambrus A (2000) *Ital J Food Sci* 3:259–278
- Hill ARC, Reynolds SL (1999) *Analyst* 124:953–958
- Hill ARC (2000) *Quality control procedures for pesticide residues analysis: guidelines for monitoring pesticide residues in the European Union*, 2nd edn. Document No. SANCO/3103/2000
- Ambrus A (2000) *Food Addit Contam* 17:519–537
- Ambrus A., Soboleva E (2002) *Contribution of sampling to the variability of residue data*. 4th European pesticide residue workshop, Rome
- Codex Secretariat. *Recommended method of sampling for the determination of pesticide residues for compliance with MRLs*, ftp://ftp.fao.org/codex/standard/en/cxg_033e.pdf
- Ambrus A (1996) *J Environ Sci Health* B31:435–442
- Ambrus A (1999) *Quality of residue data*. In: Brooks GT, Roberts TR (eds) *Pesticide chemistry and bioscience*. Royal Society of Chemistry, Cambridge, pp 339–350
- Ambrus A, Lantos J (2002) *Evaluation of the studies on decline of pesticide residues*. *J Agric Food Chem* 50:4846–4851
- FAO (1993) *Portion of commodities to which codex maximum residue limits apply and which is analyzed*. In *Joint FAO/WHO Food Standards Programme Codex Alimentarius V. 2 Pesticide residues in food*, 2nd edn. Food and Agriculture Organization (FAO), Rome, pp 391–404
- Bettencourt da Silva RJN, Lino MJ, Santos JR, Camoes MFGFC (2000) *Analyst* 125:1459–1464
- Bettencourt da Silva RJN, Lino MJ, Ribeiro A, Santos JR (2001) *Analyst* 126:743–746
- Cuadros-Rodríguez L, Hernández Torres ME, Almansa López E, Egea González FJ, Arrebola Liébanas FJ, Martínez Vidal JL (2002) *Analytic Chimia Acta* 454:297–314
- Ambrus A, Solymosné ME, Korsós I, (1996) *J Environ Sci Health* B31:443–450
- Maestroni B, Ghods A, El-Bidaoui M, Rathor N, Jarju OP, Ton T, Ambrus A (2000) *Testing the efficiency and uncertainty of sample processing using ¹⁴C-labelled Chlorpyrifos, Part I*. In: Fajgelj A, Ambrus A (eds) *Principles of method validation*. Royal Society of Chemistry, Cambridge, pp 49–58
- Maestroni B, Ghods A, El-Bidaoui M, Rathor N, Ton T, Ambrus A (2000) *Testing the efficiency and uncertainty of sample processing using ¹⁴C-labelled Chlorpyrifos, Part II*. In: Fajgelj A, Ambrus A (eds) *Principles of method validation*. Royal Society of Chemistry, Cambridge, pp 59–74
- Wallace D, Kratochvil B (1987) *Anal Chem* 59:226–232
- Bettencourt da Silva RJN, Lino MJ, Santos JR (2002) *Accred Qual Assur* 7:195–201
- Hill ARC, Harris CA, Warburton AG (2000) *Effects of sample processing on pesticide residues in fruit and vegetables*. In: Fajgelj A, Ambrus A (eds) *Principles of method validation*. Royal Society of Chemistry, Cambridge pp 41–48
- El-Bidaoui M, Jarju OP, Maestroni M, Phakaeiw Y, Ambrus A (2000) *Testing the effect of sample processing and storage on stability of residues*. In: Fajgelj A, Ambrus A (eds) *Principles of method validation*. Royal Society of Chemistry, Cambridge, pp 75–88
- Fussel RJ, Jackson Addie K, Reynolds SL, Wilson MF (2002) *J Agric Food Chem* 50:441–448
- Horwitz W, Kamps LR, Boyer KW (1980) *J Assoc Off Anal Chem* 63:1344–1354
- Codex Alimentarius Volume 3 (1995) *Residues of veterinary drugs in foods*. FAO, Rome
- Horwitz W, Britton P, Chirtel SJ (1988) *J AOAC Int* 81:1257–1265
- Horwitz W (2000) *The potential use of quality control data to validate pesticide residue method performance*. In: Fajgelj A, Ambrus A (eds) *Principles and practices of method validation*. Royal Society of Chemistry, Cambridge, pp 1–8
- Horwitz W, Jackson T, Chirtel SJ (2001) *J AOAC Int* 84:919–935
- Alder L, Korth W, Patey AL, van der Schee HA, Schoeneweis S (2001) *J AOAC Int* 84:1569–1578
- Ambrus A, Dobi D, Kadenczki L (2002) *Utilization of internal quality control data for establishing performance characteristics of analytical methods*, Poster 6a.21, 10th IPUAC International Congress on the Chemistry of Crop Protection, Basel
- Andersson A, Palshedden H (1991) *Fresenius J Anal Chem* 339:365–367
- Erney AM, Gillespie AM, Gilvydis DM, Poole CF (1993) *J Chromatogr* 638:57–63
- Erney A.M., Pawlowski T, Poole CF (1997) *J High Resolut Chromatogr* 20:375–378

-
50. Hajslova J, Holadova K, Kocurek V, Pouska J, Godula M, Cuhra P, Kempny M (1998) *J Chromatogr A* 800:283–295
 51. Soboleva E, Rathor N, Mageto A, Ambrus A (2000) Estimation of significance of matrix-induced chromatographic effects. In: Fajgelj A, Ambrus A (eds) *Principles of method validation*. Royal Society of Chemistry, Cambridge, pp 138–156
 52. Bettencourt da Silva RJN, Santos JR, Camoes MFGFC (2002) *Analyst* 127:957–963
 53. Ambrus A, Jarju PO, Soboleva E (2002) Uncertainty of analyte concentration predicted by GLC and HPLC analysis, Poster 6a.22, 10th IPUAC International Congress on the Chemistry of Crop Protection, Basel
 54. El-Bidaoui M, Ambrus A, Pouliquen I (2002) Alternative methods for matrix-matched calibration, Poster 6a.20, 10th IPUAC International Congress on the Chemistry of Crop Protection, Basel
 55. Willetts P, Wood R (1998) *Accred Qual Assur* 3:231–236
 56. Thompson M, Ellison S, Fajgelj A, Willetts P, Wood R (1999) *Pure Appl Chem* 71:337–348
 57. FAO (2002) *FAO Plant Production and Protection Paper* 170