The Analyst

Assessment of detection methods in trace analysis by means of a statistically based in-house validation concept

Bernd Jülicher*[a], Petra Gowik[a] and Steffen Uhlig[b]

[a] *EU Reference Laboratory for Residues of Veterinary Drugs, Division of Animal Drug Registration, Residue Control and Feed Additives, Federal Institute for Health Protection of Consumers and Verterinary Medicine, Diedersdorfer Weg 1, D-12277 Berlin, Germany*
[b] *Institute for Statistics and Econometrics, Free University Berlin, Garystrasse 21, D-14195 Berlin, Germany*

A matrix-considering in-house validation concept for analytical methods is presented which takes into account the uncertainty due to matrix- and time-induced deviations. It is based on a variance component model for univariate quantitative measurement data that can be adapted to both screening and confirmation methods and to both zero-tolerance and threshold decisions. The model allows the calculation of critical concentrations for given α-errors and the calculation of the corresponding power function to evaluate the performance of an analytical method. The model is applied to a real-life validation experiment.

Keywords: *Validation; calibration; variance component model; power function; uncertainty; trace analysis; residue analysis; limit of detection; critical concentration*

In view of the globalization of markets, the reliability of analytical methods used for regulatory purposes (*e.g.*, food control, environmental monitoring) has become increasingly important. To protect consumers from intolerable health hazards and to prevent market distortions, the competent authorities have to guarantee the comparable, reliable control of established threshold levels. One approach to ensure the analytical quality is to recommend or even to require the use of precisely described methods as the Codex Alimentarius Commission[1] and the US Food Safety and Inspection Service[2] do, whereas another approach is to establish minimum quality criteria for the different analytical techniques as the European Union[3] does in the field of veterinary drug residue control.

A widely accepted prerequisite for the application or even recommendation of an analytical method is its validation.[1,4–8] Although virtually all procedures for the validation of methods are based on the determination of specific performance characteristics, *e.g.*, the limits of 'detection and quantification',[7–13] there is no commonly accepted validation procedure for their assessment. Even within the concepts that are based on the construction of a calibration curve,[7,8,12] there is no consensus about the choice of calibration samples and the number of replicates.

Most validation procedures include the performance of a collaborative study[1,4,6–8,12–16] to evaluate the reproducibility of the method, which means that enormous efforts have to be made. Consequently, the availability of fully validated methods for regulatory purposes is limited.

In order to gain flexibility and to reduce time and costs, an in-house validation procedure for the processing of quantitative data is presented here that takes into account matrix- and time-induced deviations based on a statistical variance component model. By applying this model to a real-life validation experiment, the practicability of the in-house validation procedure is demonstrated. The reliability of the examined method is evaluated by calculating the critical concentrations for given α- and β-errors.

Requirements for validation procedures in trace analysis

A basic requirement for analytical methods used in trace analysis and especially in residue control is their ability to distinguish between upwards and downwards deviations from a threshold level in accordance with a defined error probability. Therefore, the method should show as low a variance as possible so that these decisions can be made with the appropriate sharpness and security. At the beginning of the validation, preliminary examinations have to be carried out to determine the substances detectable by the method, the present and potential future matrix populations of the laboratory and the measurement range to which the validation will apply. Any change of the matrix resulting, for instance, from different origins of the samples or different freshness or storage conditions, could have an unexpected influence on the analysis and therefore reduce the evidential value of the results. Hence it has to be defined in advance under which conditions the method will be considered to be valid. Most validation procedures[1,5–11,13–16] assume that the distribution of the measurement values of the matrix of the unknown sample equals the distribution of the measurement values of the matrix of the calibration samples. Hence a huge amount of laboratory work and costs would be unavoidable if one of these procedures was to be repeated for each matrix being considered. These considerations suggest an investigation of the matrix variability, which so far has not been undertaken.

Additional variability caused by time-related effects such as measurement series deviations due to instrument instabilities can occur. This additional variability must also be taken into account when performance characteristics are determined. Usually, attempts are made to compensate for this by the construction of a calibration curve within each measurement series and by the addition of an internal standard if available. However, even when internal standards are used, time effects may be observed. Therefore, the validity of a validation calibration curve should be guaranteed for a period to be defined by the validation procedure itself.

Another crucial point in validation is the determination of the blank value. For several analytical techniques, *e.g.*, mass spectrometry, it is problematic to determine. Establishing a virtual blank value by means of linear extrapolation of the calibration curve[6–8] is questionable, because adsorption or contamination effects, especially in the region of the blank value, may result in a pronounced non-linear course of the calibration curve. It seems more reasonable to adapt the model to the purpose of analysis.

Finally, the reliability of the analysis should be sufficient for court proceedings, *i.e.*, the underlying concept should adequately reflect the true conditions. On the one hand, possible compensation claims resulting from false-positive decisions (α-error) have to be excluded to the greatest extent possible. On the other hand, consumer safety has to be guaranteed, *i.e.*, virtually all contaminated samples should be discovered. The analytical strategy used so far to comply with these requirements is a combination of screening and confirmatory methods.[17] One requirement for a successful screening method is the minimisation of the false negative rate (β-error); for a satisfactory confirmatory method, it is the minimisation of both the β- and the α-error. Consequently, the intended use of a method, *e.g.*, screening or confirmation, influences the validation criteria. Definition of the maximum tolerated error probabilities for the analytical methods applied is a prerequisite for comparability of the analytical results. The different analytical communities, *e.g.*, environmental monitoring, residue control, food control, pharmaceutical or forensic chemistry, should agree on error probabilities for common use in their respective fields.

Statistical model

For the validation of an analytical method, the determination of the method-specific parameters has to be based on a statistical model that reflects the existing analytical conditions, *e.g.*, system instabilities, matrix variability and the distribution of measurement values.

In the model presented (see also Appendix), it is assumed that the fundamental relationship between measurement Y, concentration x_i and matrix j is given by

$$Y_{ij} = \mu(x_i) + \tilde{a}_j + \tilde{b}_j x + \varepsilon_{ij}, \quad (1)$$

where $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_I$ represent the spike concentrations, $i = 1, \ldots, I$ denote the spike levels and $j = 1, \ldots, J$ denote the matrix of the calibration sample. Y may represent the measured raw values (height or area of peaks) and also estimated concentration values obtained using a standard calibration curve established within the measurement series. $\mu(x)$ is called the overall calibration curve. It is assumed to be linear within the calibration range, *i.e.*, $\mu(x) = a + bx$ within the interval $[x_1; x_I]$. The definition of the matrix levels j depends on the calibration experiment; j could, for instance, represent an animal or a species, or different parts of an animal such as muscle or liver. Additionally, each j could represent another measurement series. A clear definition of the different j levels is essential for the validation procedure. The measurement error is denoted by ε_{ij}, which represents the variability of Y_{ij} when matrix j is fixed. The interpretation of ε_{ij} is closely related to the experiment and to the definition of j. In either case, ε_{ij} is a random variable with zero mean. Its variance can depend on concentration x_i and on the level j, *i.e.*, on compartments, species or instrument conditions. In the model presented, the variance of ε_{ij} is assumed to be constant, $\mathrm{Var}(\varepsilon_{ij}) = \sigma_0^2$, $\tilde{a}_j$ and $\tilde{b}j$ are random effects (possibly correlated) with zero mean and variances σ_a^2 and σ_b^2, respectively. Additionally, all random variables, ε_{ij}, $\tilde{a}_j$ and $\tilde{b}j$, are assumed to have a normal distribution.

If estimates $\hat{a}_j$ and $\hat{b}_j$ of the parameters $a_j = a + \tilde{a}_j$ and $b_j = b + \tilde{b}_j$ are available, the concentration x of the unknown sample with signal Y can be estimated from $x = (Y - \hat{a}_j)/\hat{b}_j$. However, the matrix-specific calibration curve normally fails for practical reasons such as the availability of sample material. Therefore, in the model presented, the overall calibration curve $\mu(x)$ is used, where $\mu(x)$ determines the relationship between x and Y on average over all matrices of the defined population. The contaminant concentration x of the unknown sample with signal y will be estimated using $x = (y - \hat{a})/\hat{b}$, where $\hat{a}$ and $\hat{b}$ denote the estimated coefficients of the overall calibration curve.

Calibration experiment

The model was applied to a quantitative method used in veterinary drug residue control. In order to assess the influence of different operational conditions (freshness, storage conditions) and different matrices (muscles of calves, cows, pigs and turkeys from different feeding conditions with different fat contents), 26 calibration runs were performed, each at four concentration levels, $x = 0.3$, 0.6, 0.9 and 1.2 μg kg^{-1} chloramphenicol (CAP), in muscle. The operating conditions and the matrices were chosen randomly. In the European Union the use of CAP in food-producing animals is banned and therefore the validation had to be carried out around the zero concentration range.

The measuring results were obtained by means of a sample preparation consisting of several liquid- and solid-phase extraction steps followed by a derivatisation step to prepare the sample for GC–MS analysis. Deuterated chloramphenicol-d_5 was used as internal standard.

Complying with the prerequisite of an unambiguous identification of an analyte, a measuring result was only considered acceptable if the identification criteria for GC–MS had been fulfilled, *e.g.*, the presence of four characteristic fragments (diagnostic ions) of the analyte within given margins at the correct retention time.[3] Fragmentation took place in a negative chemical ionisation source. The reactant gas was ammonia. Quantitative evaluation was carried out by using the internal standard according to the peak-area mode customary in gas chromatography. For this purpose, the area ratio of the most intense ion of the analyte and of the internal standard, respectively, were used.

Statistical analysis of the calibration experiment

For each calibration run j the corresponding calibration function $\hat{a}_j + \hat{b}_j x$ and the residual standard deviation s_j were calculated. Table 1 gives the results.

The measurement data and the calibration functions $\hat{a}_j + \hat{b}_j x$ of the 26 calibration runs are presented graphically in Fig. 1. There was no indication that the measurement data were not a random sample from a normal population. The influence of different matrices and different operating conditions on the constant $\hat{a}_j$ and the slope $\hat{b}_j$ was examined by several t-tests (Fig. 2). No significant effect of different matrices or operating conditions was detected (details omitted). If there were no effects resulting from different matrices, operating conditions or calibration runs, the theoretical (unknown) calibration function of run j at concentration x would equal the theoretical overall calibration function:

$$Y_{ij} = a_j + b_j x_i + \varepsilon_{ij} = a + bx_i + \varepsilon_{ij}$$

where ε_{ij} denotes the random error. Applying this restricted model, the $(1 - \alpha)$ prediction interval for measurement values at concentration x could be calculated:

$$\hat{a} + \hat{b}x \pm s_0 t_{52,1-\alpha/2} \sqrt{1 + \frac{1}{26}(1, x) M^{-1} \begin{pmatrix} 1 \\ x \end{pmatrix}}$$

where

$$M = \begin{pmatrix} 4 & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^2 \end{pmatrix} = \begin{pmatrix} 4 & 3 \\ 3 & 2.7 \end{pmatrix}$$

denotes the information matrix for $(\hat{a}_j, \hat{b}_j)$; $t_{52.1-\alpha/2}$ denotes the $1 - \alpha/2$-quantile of the t-distribution with 52 degrees of freedom (which are derived from the 52 degrees of freedom of s_0^2). The resulting 98% prediction interval is shown in Fig. 3. Ten out of the 4×26 measurement values are not in the interval, which is far more than 2%. It can be concluded that

there are additional sources of error which model (1) does not take into account.

For a more detailed analysis of the error, the scatter of the 26 calibration functions has to be investigated. These calibration functions can be written

$$\hat{a}_j + \hat{b}_j x = a_j + b_j x + \text{estimation error}$$

where $\hat{a}_j + \hat{b}_j x$ denotes the observed (estimated) calibration function and $a_j + b_j x$ denotes the unknown, theoretical calibration function at calibration run j. The variance of the observed calibration function is the sum of the variance of the theoretical calibration function and the variance of the estimation error, *i.e.*,

$$\text{Var}(\hat{a}_j + \hat{b}_j x) = \text{Var}(a_j + b_j x) + \text{Var}(\text{estimation error})$$

where j is assumed to be randomly chosen. The estimation error refers to the estimation of the calibration function at concentration x for one calibration run. Its variance can be computed as

$$\text{Var}(\text{estimation error}) = \sigma_0^2(1, x)M^{-1}\begin{pmatrix}1\\x\end{pmatrix}$$

where σ_0^2 denotes the variance of the error ε_{ij}. Therefore, the variance of the theoretical calibration function $a_j + b_j x$ can be estimated:

$$\text{Var}(a_j + b_j x) = s^2_{\hat{a}_j + \hat{b}_j x} - s_0^2(1, x)M^{-1}\begin{pmatrix}1\\x\end{pmatrix}$$

where $s^2_{\hat{a}_j + \hat{b}_j x}$ denotes the empirical variance of the estimated calibration functions and $s_0^2 = \frac{1}{26}\sum s_j^2$ denotes the residual variance. Hence an overall calibration function can be estimated as

$$\mu(x) = \hat{a} + \hat{b}x = \frac{1}{26}\sum_j \hat{a}_j + \hat{b}_j x$$

The respective variances were calculated at the limits, $x = 0.3$ μg kg^{-1} and $x = 1.2$ μg kg^{-1}, and at the centre of the calibration range, $x = 0.75$ μg kg^{-1}. The results are given in Table 2.

At concentration $x = 0.3$ μg kg^{-1}, the empirical variance $s^2_{\hat{a}_j + \hat{b}_j x}$ is slightly lower than the variance of the estimation error. Since variances are non-negative, the variance of the calibration function at $x = 0.3$ μg kg^{-1} can be estimated, $\text{Var}(a_j + b_j \times 0.3) \approx 0$, *i.e.*, there is no indication that the calibration function at the lower limit of the calibration range is dependent on the run j. In contrast to this result, at concentration $x = 0.75$ μg kg^{-1}, the empirical variance is significantly higher than the variance of the estimation error, *i.e.*, $\text{Var}(a_j + b_j \times 0.75) > 0$. This result is in accordance with the observed scatter of the calibration functions (Fig. 1). Apparently the dispersion becomes larger at higher concentrations. This is just an empirical result—analytical implications will not be discussed in this paper. Other analytical methods, matrices and substances may yield different results.

Based on model (1), the $(1 - \alpha)$ prediction interval for the measurement values can be recalculated. For this purpose, the variance of the theoretical calibration function $a_j + b_j x$ has to be taken into account. Then the $(1 - \alpha)$ prediction interval for the measurement values can be computed as

$$\hat{a} + \hat{b}x \pm t_{25,1-\alpha/2}\sqrt{1 + s_0^2(1, x)M^{-1}\begin{pmatrix}1\\x\end{pmatrix} + \text{Var}(a_j + b_j x)}$$

The resulting 98% prediction interval is shown in Fig. 4. This interval covers the empirical distribution of the measurement values very well. This is a clear indication that the general model proposed in this paper is adequate to deal with matrix- and time-induced deviations.

Computation of the critical concentrations for banned substances

Based on the assumption that the calibration function can be extrapolated linearly, the critical concentration CC_α could be determined as illustrated in Fig. 5(*a*). As discussed above, this assumption is, however, questionable. For confirmatory purposes it may be more appropriate to assume a worst case scenario, as shown in Fig. 5(*b*).

Table 1 Results of the calibration experiment

Run No.	0.3 μg kg^{-1}	0.6 μg kg^{-1}	0.9 μg kg^{-1}	1.2 μg kg^{-1}	$\hat{a}_j$	$\hat{b}_j$	s_j
1	0.36	0.66	1.06	1.31	0.035	1.08	0.0433
2	0.28	0.62	0.90	1.24	−0.030	1.05	0.0190
3	0.31	0.65	0.92	1.09	0.090	0.87	0.0603
4	0.31	0.61	0.88	1.12	0.055	0.90	0.0212
5	0.30	0.66	0.95	1.26	0.00	1.06	0.0227
6	0.35	0.72	1.05	1.27	0.075	1.03	0.0542
7	0.31	0.64	0.99	1.31	−0.025	1.12	0.0087
8	0.32	0.67	0.87	1.26	0.025	1.01	0.0556
9	0.26	0.58	0.92	1.20	−0.050	1.05	0.0190
10	0.31	0.64	0.84	1.19	0.035	0.95	0.0448
11	0.33	0.62	0.90	1.19	0.045	0.95	0.0032
12	0.33	0.69	0.86	1.21	0.070	0.94	0.0586
13	0.28	0.60	0.89	1.17	−0.005	0.99	0.0145
14	0.32	0.67	0.93	1.17	0.070	0.94	0.0404
15	0.36	0.50	0.88	1.13	0.045	0.90	0.0702
16	0.35	0.69	1.03	1.30	0.045	1.06	0.0271
17	0.28	0.57	0.91	1.24	−0.055	1.07	0.0170
18	0.32	0.58	0.87	1.14	0.040	0.92	0.0087
19	0.32	0.69	0.88	1.25	0.040	0.99	0.0569
20	0.32	0.70	1.02	1.19	0.075	0.98	0.0756
21	0.31	0.60	0.75	1.14	0.040	0.88	0.0697
22	0.35	0.66	1.10	1.27	0.045	1.07	0.0803
23	0.29	0.62	0.92	1.25	−0.025	1.06	0.0095
24	0.31	0.73	0.91	1.30	0.025	1.05	0.0719
25	0.37	0.74	1.04	1.45	0.015	1.18	0.0318
26	0.29	0.56	0.89	1.28	−0.070	1.10	0.0424

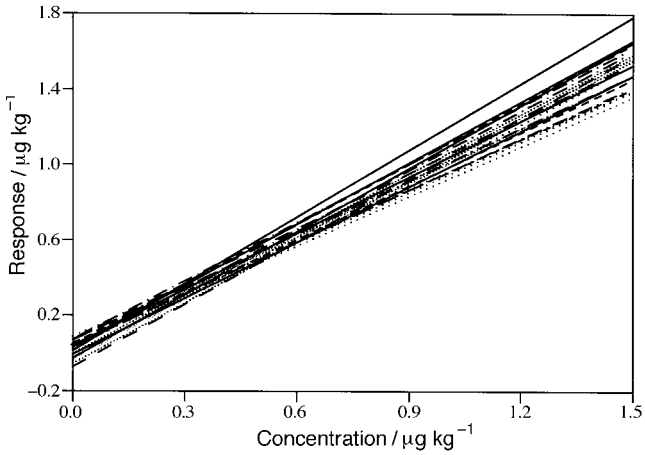

Fig. 1 Calibration functions of the calibration experiment.

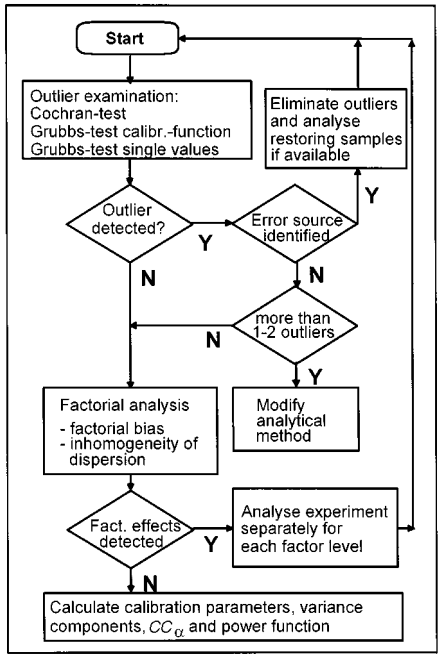

Fig. 2 Flow chart of the preliminary analysis of the data of the calibration experiment.

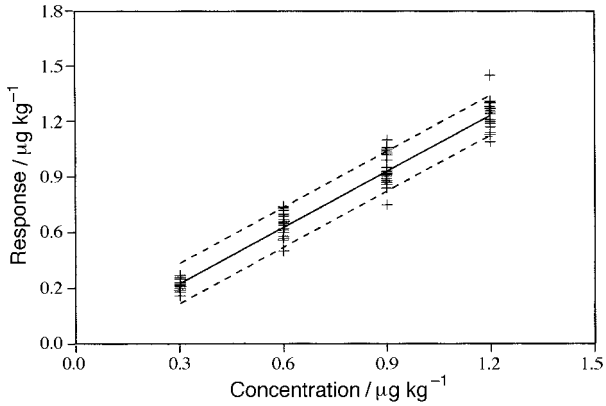

Fig. 3 98% prediction interval (without matrix- and time-induced deviations).

If the measured value exceeds the critical concentration, it will be concluded that the analyte is detected. The error probability α for making a wrong decision (false-positive rate)

Table 2 Variance components of the estimated calibration functions

Concentration x/ µg kg^{-1}	$s^2_{\hat{a}_j + \hat{b}_j x}$	Var(estimation error)	Var($a_j + b_j x$)
0.3	0.00119	0.00148	0
0.75	0.00236	0.00053	0.00188
1.2	0.00619	0.00148	0.00471

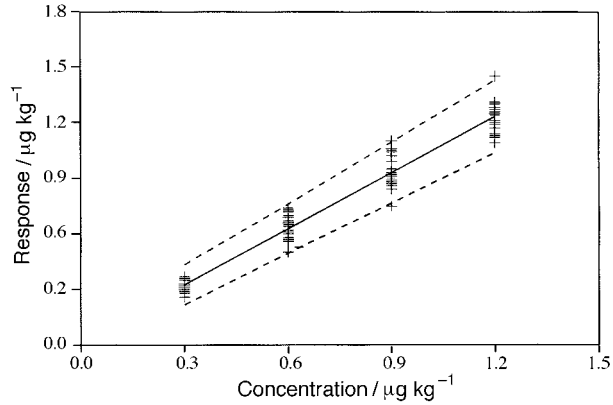

Fig. 4 98% prediction interval (with matrix- and time-induced deviations).

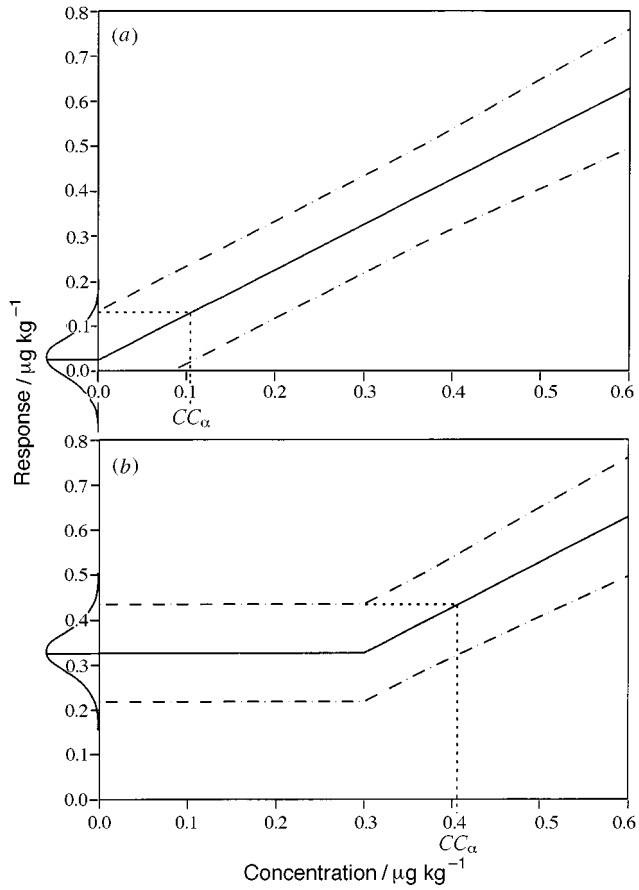

Fig. 5 (*a*) Graphical determination of CC_α using linear extrapolation. (*b*) Graphical determination of CC_α for confirmatory purposes.

depends on the prediction level. For a 98% prediction interval the error probability is 1%.

For the method examined, the critical concentration is CC_α = 0.42 μg kg^{-1}, with an underlying error probability of α = 0.01. According to the model presented for confirmatory purposes (type II calibration curves, see Appendix), it is clear that the CC_α cannot be smaller than the lowest concentration in the calibration experiment. This has to be considered when defining the calibration points during the preliminary examinations.

Each analytical method involves not only a false-positive rate but also a false-negative rate which depends on the true concentration x of the analyte. The false-negative rate $\beta = 1 - p(x)$ can be calculated from the power function $p(x)$, which describes the probability of detecting the analyte when its true concentration is x. It characterizes the performance of the analytical method for detecting the analyte. The derivation of the power function will not be discussed in detail here. It is obvious that it depends on the decision criterion for the detection, and hence on the critical concentration CC_α and on the false positive rate α.

For confirmatory purposes, the power function can be calculated:

$$p(x) = 1 - F_{J-1;\delta_x^{(c)}}\left(t_{J-1,1-\alpha} \frac{\sqrt{\mathrm{Var}(a_j + b_j x_l) + s_0^2 + \mathrm{Var}(\hat{a} + \hat{b}x_1)}}{\sqrt{\mathrm{Var}(a_j + b_j x) + s_0^2 + \mathrm{Var}(\hat{a} + \hat{b}x_1)}} \right)$$

where x_1 denotes the lowest concentration level in the calibration experiment and $F_{J-1;\,\delta_x^{(c)}}$ denotes the distribution function of the t-distribution with $J - 1$ degrees of freedom and the non-centrality parameter

$$\delta_x^{(c)} = \frac{b(x - x_1)}{\sqrt{\mathrm{Var}(a_j + b_j x_1) + s_0^2 + \mathrm{Var}(\hat{a} + \hat{b}x_1)}}$$

Figure 6 shows the power function calculated by means of the measuring results of the calibration experiment. As illustrated, the power function provides the critical concentration CC_β at which the false-negative rate equals β, where β = 0.01 or β = 0.05 is given. The calculated values for the examined CAP confirmatory method are CC_β = 0.50 and 0.55 μg kg^{-1} for β = 0.05 and 0.01, respectively. The CC_β can be used as a validation criterion: as long as CC_β is below a given limit, it is guaranteed that for true concentrations above the limit the false-negative rate does not exceed β.

Discussion

In trace analysis, the analytical results are usually affected by the type of matrix or by time-related conditions of the analytical system. A validation based on the presented variance component model refers not only to samples that correspond to the particular matrix used in the validation procedure but also, within certain time limits, to all future samples belonging to the defined population.

As demonstrated in the Appendix, the model can be adapted not only to the different performance levels of methods, screening or confirmatory, but also to zero-tolerance and threshold decisions based on quantitative measurement data.

The proposed concept of a matrix-considering in-house validation procedure should be considered with regard to the current discussion about the uncertainty of measurement. At present, there are different approaches under discussion about how to quantify uncertainty.[18,19] The uncertainty of a measurement can be defined 'as the interval on the measurement scale within which the true value lies with a specified probability, when all sources of error have been taken into account'.[19] By taking into consideration matrix- and time-induced deviations, as done in the model presented, essential components of the uncertainty resulting from the systematic error are already covered by applying the in-house validation concept. By a recently published approach to the quantification of uncertainty,[19] the analytical error can be broken down into four components: the method bias, the laboratory bias, the run bias and the random measurement error.[19] According to the concept presented, the run bias refers to the time-induced deviations and the method bias is additionally split into two components, the matrix-induced deviations and remaining method-induced biases. The decomposition of the uncertainty of a measuring result can be illustrated by the uncertainty tree in Fig. 7. It should be noted that neither collaborative studies nor in-house validation procedures are able to cover all components of uncertainty. In-house validation procedures do not cover the laboratory bias, whereas collaborative studies do not cover matrix-induced deviations. (There are some hints that the laboratory bias in comparison with the time effects can sometimes be neglected.[20] When considering the laboratory bias as less important than other error components, it could be sufficient to determine it within proficiency tests, which are necessary anyway for the continuous assessment of the technical competence of the participating test laboratories.)

Conclusion and outlook

The application of the concept presented provides the analyst with comprehensive information about the performance of the method in question. Critical concentrations and error probabilities are calculated by means of a power function, which is based on a variance component model. Ongoing validation

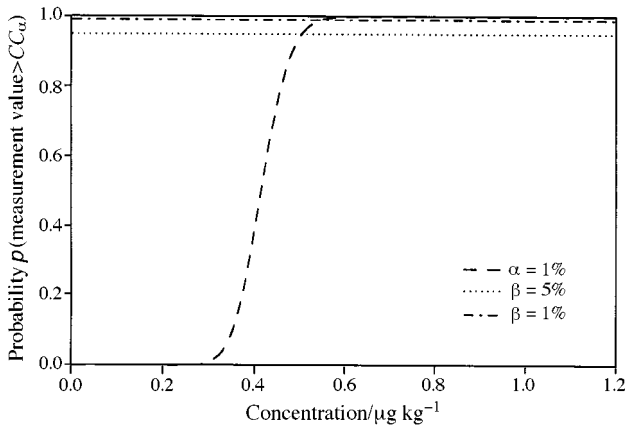

Fig. 6 Power function of the examined CAP confirmatory method.

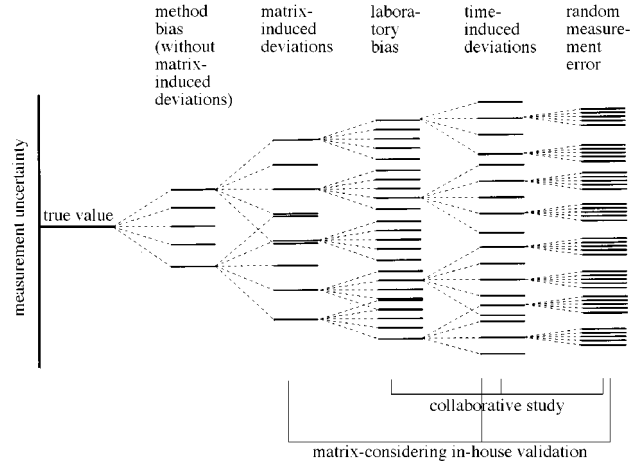

Fig. 7 Uncertainty tree.

experiments using different matrices, substance groups and analytical techniques are expected to confirm the practicability of the presented concept. Future work will be dedicated to developing the necessary statistical procedures for the integration of this validation concept into the overall framework of quality management.

Financial support by the European Commission is gratefully acknowledged.

Appendix: derivation of formulae

Calibration function

The statistical model considered in this paper is

$$Y = Y_{ij} = \mu(x_i) + \tilde{a}_j + \tilde{b}_j x_i + \varepsilon_{ij} \quad (1)$$

where Y_{ij} denote the measurement value at contaminant level i, $i = 1, \ldots, I$ and matrix j, $j = 1, \ldots, J$. $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_I$ denote the contaminant values in the calibration experiment and $\tilde{a}_j + \tilde{b}_j x$ represents the matrix specific correction term of the calibration function $\mu(x)$. The latter is assumed to be linear within the calibration interval, *i.e.*, $\mu(x) = a + bx$ for $x_1 \leqslant x \leqslant x_I$. Out of the interval two different extrapolations are applied. For screening methods the curve will be linearly extrapolated (type I calibration), as long as $a + bx \geqslant 0$, *i.e.*, $\mu(x) = \max\{a + bx, 0\}$. For confirmatory methods (type II calibration) a worst case scenario is considered:

$$\mu(x) = \begin{cases} a + bx_1 & \text{if } x < x_1 \\ a + bx & \text{if } x \geqslant x_1 \end{cases}$$

The calculation of the critical concentration depends on the assumptions concerning the calibration function and on the threshold value. However, in each case it is based on a significance test: The critical concentration represents the measured concentration from which on the threshold value $x_{\text{threshold}}$ is exceeded significantly. The threshold value is assumed to be greater or equal to zero. If it is zero (as in the example presented), the significance test consists in investigating the question whether the substance can be detected in the sample. However, one should take into account that the threshold value could also be a target value or a maximum admissible concentration.

The underlying test hypotheses can be formulated as follows:

$$H_0 : x \leqslant x_{\text{threshold}} \text{ against } H_1 : x > x_{\text{threshold}}$$

where x denotes the unknown concentration of a sample with measurement value Y. Additionally a modified threshold is defined as

$$x_0 = \begin{cases} x_{\text{threshold}} & \text{at type I calibration} \\ \max\{x_{\text{threshold}}, x_1\} & \text{at type II calibration} \end{cases}$$

which is equivalent to $x_{\text{threshold}}$ with respect to the calibration function, *i.e.*, $\mu(x_0) = \mu(x_{\text{threshold}})$.

According to the statistical model, the variance of the measurement value $Y = Y(x)$ at concentration x can be written

$$\text{Var}(Y) = \sigma_0^2 + \sigma_1^2(x),$$

where $\sigma_1^2(x)$ denotes the variance of the matrix-induced error and σ_0^2 denotes the measurement variance. In practice, these variance components are unknown and have to be estimated. For the moment we assume that they are exactly known.

Let $c = \mu(x_0) = a + bx_0$ denote the expected signal at the threshold level and let $\hat{c} = \hat{a} + \hat{b}x_0$ denote the estimate of c based on the data of the calibration experiment. The corresponding error variance of $\hat{c}$ is denoted by $\sigma_{\hat{c}}^2$. Then at the significance level α the null hypothesis H_0 can be rejected if

$$\frac{Y - \hat{c}}{\sqrt{\sigma_{\hat{c}}^2 + \sigma_1^2(x_0) + \sigma_0^2}} > z_{1-\alpha},$$

where $z_{1-\alpha}$ denotes the $(1-\alpha)$-quantile of the standard normal distribution. This formula does not take into account that $\hat{c}$ might be negative, which is not allowed since $\mu(x) \geqslant 0$. Replacing $\hat{c}$ by $\max\{\hat{c}, 0\}$ provides the rejection rule

$$\frac{Y - \max\{\hat{c}, 0\}}{\sqrt{\sigma_{\hat{c}}^2 + \sigma_1^2(x_0) + \sigma_0^2}} > z_{1-\alpha},$$

The critical concentration CC_α is the corresponding value on the x-axis, *i.e.*

$$CC_\alpha = \frac{z_{1-\alpha}}{\hat{b}} \sqrt{\sigma_{\hat{c}}^2 + \sigma_1^2(x_0) + \sigma_0^2} + \frac{\max\{\hat{c}, 0\} - \hat{a}}{\hat{b}}$$

Random sampling

In practice the variance components $\sigma_{\hat{c}}^2$, σ_0^2 and $\sigma_1^2(x_0)$ have to be estimated. We assume that $\boldsymbol{J}$ matrices are chosen at random. Each matrix sample is divided into I test portions, which will be spiked at concentration levels $x_1 \ldots, x_I$. Linear regression of submodels $Y_{ij} = a_j + b_j x_i + \varepsilon_{ij} = (a + a_j) + (b + \hat{b}_j)x + \varepsilon_{ij}$ with fixed j provides estimates $\hat{a}_j$, $\hat{b}_j$ and the corresponding residual variances $s_j^2 = \Sigma_i^I (Y_{ij} - \hat{a}_j - \hat{b}_j x_i)^2/(I - 2)$. These parameters provide parameters of the overall calibration curve, the constant $\hat{a} = \frac{1}{J}\Sigma_j^J \hat{a}_j$, the slope $\hat{b} = \frac{1}{J}\Sigma_j^J \hat{b}_j$ and the estimated error variance $\hat{\sigma}_0^2 = \Sigma_{j=1}^J s_j^2/J$.

In order to obtain an estimator of the variance of $\sigma_1^2(x_0)$ of $c_j = a_j + b_j x_0$, we consider the estimation error γ_j of the estimator $\hat{c}_j = \hat{a}_j + \hat{b}_j x_0 = a_j + b_j x_0 + \gamma_j$. Using the information matrix $\boldsymbol{M} = \sigma^2 \begin{pmatrix} I & \sum_i^I x_i \\ \sum_i^I x_i & \sum_i^I x_i^2 \end{pmatrix}$ the variance of the estimation error γ_j can be calculated according to

$$\sigma_\gamma^2 = \sigma_0^2 \cdot (1, x_0) \boldsymbol{M}^{-1} \begin{pmatrix} 1 \\ x_0 \end{pmatrix}$$

$$= \sigma_0^2 \left\{ \frac{1}{I} + (x_0 - \frac{1}{I}\sum_{i=1}^I x_i)^2 \Big/ \left[\sum_{i=1}^I x_i^2 - \frac{1}{I}\left(\sum_{i=1}^I x_i \right)^2 \right] \right\}$$

Replacing σ_0^2 by the estimator $\hat{\sigma}_0^2$ provides an unbiased estimator $\hat{\sigma}_\gamma^2$ of σ_γ^2. Since the variance of $\hat{c}_j$ can be estimated by the empirical variance $s_{\hat{c}}^2$ of $\hat{c}_1, \ldots, \hat{c}_J$ an unbiased estimator of $\sigma_1^2(x_0)$ is $\hat{\sigma}_1^2(x_0) = s_{\hat{c}}^2 - \hat{\sigma}_\gamma^2$. If the result is negative, let $\hat{\sigma}_1^2(x_0) = 0$.

These estimators may be applied for the estimation of the sum of variances $\sigma_{\hat{c}}^2 + \sigma_1^2(x_0) + \sigma_0^2$ by $\hat{\sigma}_{\hat{c}}^2 + \hat{\sigma}_1^2(x_0) + \hat{\sigma}_0^2$. The latter is a linear combination of stochastically independent χ^2-distributed random variables. If these estimators replace the true variances in the formula for the CC_α, the quantile $z_{1-\alpha}$ has to be corrected, too. A conservative correction (which guarantees that the actual significance level never exceeds the given α) replaces $z_{1-\alpha}$ by the critical value of the t-distribution with $J - 1$ degrees of freedom, *i.e.*, the CC_α can be calculated

$$CC_\alpha = \frac{t_{J-1,1-\alpha}}{\hat{b}} \sqrt{\hat{\sigma}_{\hat{c}}^2 + \hat{\sigma}_1^2(x_0) + \hat{\sigma}_0^2} + \frac{\max\{\hat{c}, 0\} - \hat{a}}{\hat{b}}$$

Power function

The power function $p(x)$ represents the probability of the exceedance of the critical concentration CC_α in the case the sample measured has a true concentration $x > x_0$. The power function depends on α, σ_0, σ_1, the type of calibration curve, the sampling design, *etc.* In order to determine the power function we have to compute the probability $p(x)$ of

$$\frac{Y - \max\{\hat{c}, 0\}}{\sqrt{\hat{\sigma}_{\hat{c}}^2 + \hat{\sigma}_1^2(x_0) + \hat{\sigma}_0^2}} > t_{J-1,1-\alpha}$$

where Y represents the measurement value of a sample with concentration x. It is assumed that according to the statistical model Y is normally distributed with expectation $\max\{a + bx, 0\}$ and variance $\mathrm{Var}(Y) = \sigma_0^2 + \sigma_1^2(x)$, where $\sigma_1^2(x)$ denotes the variance component induced by random matrix effects at concentration x. Moreover it is assumed that the probability of negative estimates $\hat{c}$ is neglectable.

At first we consider the distribution of $\hat{\sigma}_{\hat{c}}^2 + \hat{\sigma}_1^2(x) + \hat{\sigma}_0^2$. Unfortunately, there is no closed form expression for its distribution. In fact, it is a linear combination of χ^2-distributed variables. Worst case considerations lead to the conservative approximation

$$\frac{\hat{\sigma}_{\hat{c}}^2 + \hat{\sigma}_1^2(x) + \hat{\sigma}_0^2}{\sigma_{\hat{c}}^2 + \sigma_1^2(x) + \sigma_0^2} \sim \chi_{J-1}^2/(J-1)$$

and bearing in mind that

$$\frac{Y - \hat{c}}{\sqrt{\sigma_{\hat{c}}^2 + \sigma_1^2(x) + \sigma_0^2}} \sim N\left(\frac{b(x - x_0)}{\sqrt{\sigma_{\hat{c}}^2 + \sigma_1^2(x) + \sigma_0^2}}\right), 1$$

we obtain

$$\frac{Y - \hat{c}}{\sqrt{\hat{\sigma}_{\hat{c}}^2 + \hat{\sigma}_1^2(x) + \hat{\sigma}_0^2}} \sim t_{J-1}(\delta_x)$$

$$\text{where } \delta_x = \frac{b(x - x_0)}{\sqrt{\sigma_{\hat{c}}^2 + \sigma_1^2(x) + \sigma_0^2}}$$

(non-centrality parameter).

Here $t_{J-1}(\delta_x)$ denotes the t-distribution with $J - 1$ degrees of freedom and non-centrality parameter δ_x. Because

$$p(x) = P\left(\frac{Y - \hat{c}}{\sqrt{\hat{\sigma}_{\hat{c}}^2 + \hat{\sigma}_1^2(x_0) + \hat{\sigma}_0^2}} > t_{J-1,1-\alpha}\right)$$

$$= 1 - F_{J-1,\delta_x}\left(t_{J-1,1-\alpha}\sqrt{\frac{\sigma_{\hat{c}}^2 + \sigma_1^2(x_0) + \sigma_0^2}{\sigma_{\hat{c}}^2 + \sigma_1^2(x) + \sigma_0^2}}\right)$$

the power function at the true concentration value x can be estimated

$$p(x) = 1 - F_{J-1,\delta_x}\left(t_{J-1,1-\alpha}\sqrt{\frac{\hat{\sigma}_{\hat{c}}^2 + \hat{\sigma}_1^2(x_0) + \hat{\sigma}_0^2}{\hat{\sigma}_{\hat{c}}^2 + \hat{\sigma}_1^2(x) + \hat{\sigma}_0^2}}\right)$$

where F_{J-1,δ_x} denotes the distribution function of the t-distribution with $J - 1$ degrees of freedom and non-centrality parameter

$$\delta_x = \frac{b(x - x_0)}{\sqrt{\hat{\sigma}_{\hat{c}}^2 + \hat{\sigma}_1^2(x) + \hat{\sigma}_0^2}}$$

References

1 *Codex Alimentarius, Volume Three, Residues of Veterinary Drugs in Foods*, Codex Alimentarius Commission, Rome, 2nd edn., 1996.
2 *Analytical Chemistry Laboratory, Guidebook—Residue Chemistry*, USDA Food Safety and Inspection Service, Washington, DC, 1991.
3 EEC Directive 93/256, *Off. J. Eur. Commun.*, 1993, No. L 118/64.
4 EEC Directive 85/591, *Off. J. Eur. Commun.*, 1985, No. L 372/50.
5 *Accreditation for Chemical Laboratories: Guidance on the Interpretation of the EN 45000 Series of Standards and ISO/IEC Guide 25*, EURACHEM Secretariat, Teddington, 1993.
6 Food and Drug Administration, *Guideline for Industry, Text on Validation of Analytical Procedures*, 60 FR 11260, Center for Drug Evaluation and Research, Rockville, MD, 1995.
7 Food and Drug Administration, *Draft Guideline on the Validation of Analytical Procedures: Methodology; Availability; Notice*, 61 FR 9315, Center for Drug Evaluation and Research, Rockville, MD, 1996.
8 IUPAC, ISO and AOAC, *J. Assoc. Off. Anal. Chem.*, 1989, **72**, 694.
9 *Capability of Detection—Part 1: Terms and Definitions*, ISO/DIS 11843-1, International Standards Organization, Geneva, 1995.
10 Deutsches Institut für Normung, *DIN 32645, Chemical Analysis; Decision Limit, Detection Limit and Determination Limit; Estimation in Case of Repeatability, Terms, Methods, Evaluation*, Beuth Verlag, Berlin, 1994.
11 Deutsches Institut für Normung, *DIN 55350, Teil 34, Concepts in the Field of Quality and Statistics; Limit of Detection, Limit of Determination and Capability of a Method for Determination*, Beuth Verlag, Berlin, 1991.
12 IUPAC, *Pure Appl. Chem.*, 1990, **62**(1), 149.
13 Guidelines for Collaborative Studies, *J. Assoc. Off. Anal. Chem.*, 1989, **72**(4), 694.
14 *Accuracy (Trueness and Precision) of Measurement Methods and Results; General Principles and Definitions*, ISO 5725-1, International Standards Organization, Geneva, 1994.
15 Bundesgesundheitsamt, *Planning and Statistical Evaluation of Ring Tests*, Collection of official methods under Article 35 of the German Federal Foods Act, Beuth Verlag, Berlin, 1983.
16 Wernimont, G. T., in *Use of Statistics to Develop and Evaluate Analytical Methods*, ed. Spendley, W., Association of Official Analytical Chemists, Arlington, VA, 4th edn., 1993.
17 Stephany, R. W., in *Residues of Veterinary Drugs in Food: Proceedings of the EuroResidue Conference in Noordwijkerhout, The Netherlands, May 21–23, 1990*, ed. Haagsma, A., Ruiter, A., and Czedik-Eysenberg, P. B., University of Utrecht, Utrecht, 1990, pp. 76–85.
18 *Quantifying Uncertainty in Analytical Measurement*, EURACHEM Secretariat, Teddington, 1995
19 Analytical Methods Committee, *Analyst*, 1995, **120**, 2303.
20 Zimmermann, R., and von Lengerken, J., *Arch. Anim. Nutr.*, 1987, **37**, 723.

Paper 7/07281C
Accepted October 10, 1997