TECHNICAL SPECIFICATION

ISO/TS 20612

First edition
2007-12-01

Water quality — Interlaboratory comparisons for proficiency testing of analytical chemistry laboratories

Qualité de l'eau — Comparaisons interlaboratoires pour des essais de compétence de laboratoires de chimie analytique

Reference number
ISO/TS 20612:2007(E)

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

In other circumstances, particularly when there is an urgent market requirement for such documents, a technical committee may decide to publish other types of normative document:

In other circumstances, particularly when there is an urgent market requirement for such documents, a technical committee may decide to publish other types of document:

— an ISO Publicly Available Specification (ISO/PAS) represents an agreement between technical experts in an ISO working group and is accepted for publication if it is approved by more than 50 % of the members of the parent committee casting a vote;

— an ISO Technical Specification (ISO/TS) represents an agreement between the members of a technical committee and is accepted for publication if it is approved by 2/3 of the members of the committee casting a vote.

An ISO/PAS or ISO/TS is reviewed after three years in order to decide whether it will be confirmed for a further three years, revised to become an International Standard, or withdrawn. If the ISO/PAS or ISO/TS is confirmed, it is reviewed again after a further three years, at which time it must either be transformed into an International Standard or be withdrawn.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights.

ISO/TS 20612 was prepared by Technical Committee ISO/TC 147, *Water quality*, Subcommittee SC 2, *Physical, chemical and biochemical methods*.

Introduction

Participation in interlaboratory tests in various test fields offers a testing laboratory an opportunity to obtain an objective picture of its proficiency. Such tests serve as a confidence-building measure both for the laboratory itself and for prospective clients.

This Technical Specification is based on the following international recognized documents:

— ISO/IEC Guides 43-1 and 43-2;

— ISO 13528;

— *The International Harmonized Protocol for the Proficiency Testing of Analytical Chemistry Laboratories* – (IUPAC, ISO, AOAC);

— ILAC Guide 13;

— ISO/IEC 17025;

— ISO 5725-1 and ISO 5725-2.

As these documents only define a framework for design, execution and evaluation of proficiency testing by interlaboratory comparisons, this Technical Specification describes in detail an evaluation procedure which is especially suitable for the sector of water, waste water and sludge analysis, where results of interlaboratory comparisons play an important role in the admission of laboratories to certain analytical tasks. Therefore, the fairness of assessment of laboratories must be guaranteed. Assessment should not be dependent on the provider, the date, or the method of evaluation.

Water quality — Interlaboratory comparisons for proficiency testing of analytical chemistry laboratories

1 Scope

This Technical Specification specifies the criteria related to proficiency testing by interlaboratory comparisons in the field of water, waste water and sludge analysis. In particular, it specifies the requirements in respect to proficiency test providers and to the design, execution and evaluation of laboratory proficiency comparisons.

This document may be used if the determinands in the interlaboratory test may be regarded as capable of measurement with a certain degree of continuity. This is generally the case for chemical constituents and physicochemical determinands, but continuity does not always exist in the case of biological and/or microbiological determinands.

2 Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 3534-1, *Statistics — Vocabulary and symbols — Part 1: General statistical terms and terms used in probability*

ISO 5725-1, *Accuracy (trueness and precision) of measurement methods and results — Part 1: General principles and definitions*

ISO 5725-1:1994/Cor.1:1998, *Accuracy (trueness and precision) of measurement methods and results — Part 1: General principles and definitions — Technical Corrigendum 1*

ISO 5725-2, *Accuracy (trueness and precision) of measurement methods and results — Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method*

ISO 13528, *Statistical methods for use in proficiency testing by interlaboratory comparisons*

ISO/IEC Guide 43-1, *Proficiency testing by interlaboratory comparisons — Part 1: Development and operation of proficiency testing schemes*

ISO/IEC Guide 43-2, *Proficiency testing by interlaboratory comparisons — Part 2: Selection and use of proficiency testing schemes by laboratory accreditation bodies*

ISO/IEC 17025, *General requirements for the competence of testing and calibration laboratories*

ISO/IEC 17020, *General criteria for the operation of various types of bodies performing inspection*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 3534-1, ISO 5725-1, ISO 5725-2, ISO/IEC Guide 43-1, ISO/IEC Guide 43-2 and the following apply.

3.1
breakdown point
smallest percentage of outlier laboratories above which the estimation method may be entirely inapplicable

3.2
efficiency
ratio of the variance of the optimum estimation method for normal distribution to the variance of the estimation method under consideration, each assuming a normal distribution

NOTE This is expressed as a percentage.

3.3
sample
totality of a homogeneous analysis material with an identical composition or quality (similar to term batch)

3.4
subsample
defined portion of a sample obtained by suitable sample division and identical in terms of composition

4 Symbols

d_i	Absolute difference between log-linear variance function and the logarithm of the reproducibility standard deviation
$G_1(x_i)$	Generalized distribution function of interlaboratory differences with continuity correction (s_R)
$G_2(x_i)$	Generalized distribution function of intralaboratory differences with continuity correction (s_r)
g	Quality limit
$H_1(x_i)$	Generalized distribution function of interlaboratory differences (s_R)
$H_2(x_i)$	Generalized distribution function of intralaboratory differences (s_r)
i	Index denoting the serial number of one of p samples
J_i	Number of participants in the case of sample i
j	Index denoting the serial number of one of J participating laboratories
k_1,k_2	Correction factors for calculating z_U-score
μ	Overall mean
μ_i	Overall mean of i-th sample
x_a	Assigned value
n_j	Number of measurements made by laboratory j

q	Quantile parameter
p	Number of samples (levels)
PG_0, PG_1	Testing values for testing variance function
s_R	Reproducibility standard deviation
s_{R_i}	Reproducibility standard deviation of i-th sample
$\hat{s}_{R_i}$	Reproducibility standard deviation for sample i adjusted using a variance function
s_r	Repeatability standard deviation
$\hat{\sigma}$	Standard deviation for proficiency assessment
W	Weighting matrix
X	Design matrix
X^{T}	Transposed design matrix
x_i	Discontinuity point
$\bar{y}$	Arithmetic mean of test results of an unspecified laboratory
y_{ji}	Measurement result for the i-th measurement made by laboratory j
$\bar{y}_j$	Arithmetic mean of measurement results of laboratory j
z	z-Score, i.e. standardized deviation of laboratory result from assigned value
z_{U}	Corrected z-score
Ψ	Function for determining the Hampel estimator
$\chi^2_{p-2;0,95}$	95 % chi-squared quantile for $p - 2$ degrees of freedom
Φ	Distribution function of standard normal distribution
α	Significance level
ν	Relative reproducibility standard deviation
γ	Vector of logarithms of standard deviations
θ_0, θ_1	Parameters of log-linear variance function
$\tilde{\theta}_0$	Estimate of logarithm of relative reproducibility standard deviation if independent from concentration level

5 Requirements relating to proficiency test provider

Proficiency testing by interlaboratory comparisons must lie in the responsibility of specialists who are familiar not only with the requirements relating to the design, execution and evaluation of interlaboratory tests, but also with the analytical methods to be tested, and who have demonstrated their specialist knowledge. Against this background, it is recommended that the test provider regularly organizes interlaboratory tests in the relevant test field.

The proficiency test provider must maintain an adequately documented quality management system based on the criteria specified in ISO/IEC 17020 or ISO/IEC 17025, covering all necessary framework conditions, responsibilities and standard operation procedures.

In addition all measurements within the framework of the provided proficiency test should fulfil the technical requirements as specified in ISO/IEC 17025.

An advisory group that includes specialists for all the fields involved should be appointed to enable the relevant interlaboratory test system to be brought into line with the state of the art and proper account to be taken of the specialist requirements relating to the interlaboratory tests. Keeping a written record of the group's decisions is recommended.

6 Participants

Only laboratories that have the requisite staffing and equipment for the tests to be performed shall take part in an interlaboratory test. Each participating laboratory should appoint a member of staff to be responsible for maintaining contact with the proficiency test manager and ensuring that the analyses are correctly carried out in accordance with the proficiency test manager's instructions.

7 Proficiency test design

7.1 Proficiency test plan

All details of the proficiency test design should be laid down in a plan prior to the start of the interlaboratory test. This includes especially details about:

— involved staff;

— sample matrix;

— determinands to be analysed;

— concentration level of the determinand;

— number of samples;

— sample containers;

— sample preservation;

— distribution of samples;

— communication with participants of the proficiency test (PT);

— homogenization method;

— homogeneity and stability check;

— method for stipulating the assigned value;

— schedule;

— evaluation and assessment procedure.

All relevant practices listed in ISO Guide 43-1 shall be fulfilled.

7.2 Sample selection

In selecting the sample material, account shall be taken of the objectives of the interlaboratory test, the target concentration levels, the required homogeneity and stability of the samples, and the transport and storage facilities. In general, real or spiked real samples shall be given preference over synthetic ones. Sample matrix and concentration levels should reflect routine conditions.

7.3 Selection of determinands

The determinands selected in a particular case and their number shall be defined precisely in accordance with the target group of participants or with the reason for the interlaboratory test. Determinands shall be defined accurately, i.e. whether a certain form (e.g. soluble) or the total concentration shall be determined.

7.4 Spiking

For the preparation of samples, the proficiency test provider may spike samples with low concentrations. This can be a useful way of establishing required combinations of concentrations of individual analytes in samples. However, it does not make sense or may not be possible in all cases, especially if the type of analyte binding in the original sample is significantly different from that in the spiked solutions and the degree of difficulty in performing the analytical methods is altered.

7.5 Number of participants

If statistical methods are used to calculate an assigned value from participants data, the number of participants has an influence on the reliability of the statistically calculated data. In this, it is therefore desirable to ensure that the number of laboratories participating in the interlaboratory tests is sufficiently large and never less than twelve if the assigned value is derived from the participants data.

7.6 Number of samples

Testing several samples for the same analyte yields a more reliable picture of the proficiency of a laboratory and it is therefore desirable for the participants to analyse several samples involving different concentrations of the individual analytes.

Steps shall be taken to ensure that no single participant receives only samples having a high (or low) concentration.

7.7 Multiple determinations and sample size

To ensure that the interlaboratory tests are performed under conditions that resemble routine operation as closely as possible, the participants shall make the same number of multiple determinations as in their routine work. Attention shall be drawn to any specification of the number of parallel determinations required by regulations or by the proficiency test provider.

To reduce the possibility that multiple determinations are not in line with routine or go beyond the number specified in the interlaboratory test, the proficiency test provider should, if practicable, limit the sample size to that required for the specified test.

Dilution of concentrates by the participants prior to testing should be avoided if possible.

8 Execution of proficiency tests

8.1 General

A written record should be kept confirming correct implementation of all the requirements of the proficiency test plan as well as any necessary deviations from the specified procedure.

8.2 Sample preparation

All the steps to be taken in obtaining the sample material, ranging from the selection and cleaning of the transport vessels, sampling and transportation to the laboratory to dispensing, labeling and packaging the subsamples, should be documented in standard operation procedures.

If synthetic samples are prepared or real samples are spiked, the proficiency test provider should provide evidence of the suitability of the materials/substances used in regard to traceability of the chemical composition and the stoichiometry.

All the procedures for ensuring correct spiking, e.g. determination of the pipettes precision or of volume measurements based on mass, should be clearly documented. In addition, contamination and analyte losses should be determined and taken into account. Responsibility for these steps should be specified before the interlaboratory test is started.

The variation in the concentrations of the subsamples should not be excessively increased by the preparation procedure adopted since the reproducibility standard deviation of the test data would otherwise assume unrealistically high values. This should be borne in mind, in particular in relation to unstable and highly volatile analytes.

The containers for samples and subsamples should be such as to ensure that contamination resulting from the material and losses due to adsorption, outgassing and the like are minimized.

8.3 Stability and homogeneity testing

The proficiency test provider should provide evidence of the stability and homogeneity of subsamples and, in particular, of the substances to be quantified, for every phase of the interlaboratory test. For this purpose, additional backup samples to be analysed at suitable time intervals during the interlaboratory test by the test provider for the purpose of checking stability should be prepared when dispensing the subsamples.

8.4 Prevention of collusion between participants

Examples of possible steps to be taken by the proficiency test provider to prevent improper contacts are given below:

a) requiring the laboratories to submit copies of the raw data printouts from their analytical equipment along with the analytical results so that the proficiency test manager can use them to perform plausibility tests;

b) each participant receives a subset of the samples prepared (e.g. 3 out of 12);

c) contact accreditation body requiring spot checks to be performed on raw data and other printouts in the course of auditing in the participants laboratory.

8.5 Analytical methods

Depending on the objective or context of the interlaboratory test, the proficiency test provider may restrict or specify the analytical methods to be used. If he does not, the person in charge in the participating laboratory shall use the method normally used by the laboratory for analysing this type of sample.

8.6 Specification of the assigned value

There are various ways of specifying the assigned value:

a) by preparing the samples from substances having a precisely known composition (synthetic samples) and determining the true values from the initial sample mass;

b) by preparing the samples from certified reference materials;

c) by using the results of reference laboratories;

d) by using the robust mean of the participating laboratories.

The proficiency test manager shall be responsible for choosing an optimum method of specifying the assigned value for a determinand in each individual case, variations and combinations of the abovementioned points being conceivable or useful. Proceed as described in ISO 13528.

8.7 Sample distribution

The proficiency test provider should organize sample distribution so as to avoid the stability of the subsamples being adversely affected while they await delivery. This may also mean that the subsamples have to be collected by the participating laboratories.

Preferably the subsamples should be shipped. Steps should be taken to ensure that the subsamples are received within a defined time window by all the participants, depending on the stability of the samples. The dispatch deadline should be such that the subsamples are delivered to all the participants under the specified conditions and any necessary steps relating to storage and pretreatment can be carried out without delay.

A suitable system should be instituted for checking that deadlines are met.

8.8 Communication with participants

The proficiency test provider should prepare a long-term plan for executing regular interlaboratory tests and should inform the interested laboratories in due time of when tests are to take place, the number of samples to be tested and the determinands, including any special features of the sample matrix. The assessment criteria should also be published before the test is started.

The proficiency test provider should provide the laboratories with the requirements relating to the test objective (analytes selected), sample pretreatment and, if required, the use of specified analytical methods (preferably standard methods) or, if applicable, to the possibility of choosing equivalent methods no later than the date on which the subsamples are delivered.

The results should be reported on standard forms and/or data media that have been supplied. The number of multiple determinations, the decimal places to be reported, and the units should be specified. The use of these forms should be obligatory, all the data required shall be entered on them and they shall be authorized by the person responsible.

The deadline for the submission of the test results shall be specified beforehand and shall be as short as possible since the measurements are to be performed under largely routine conditions. The laboratories should also be informed of the time that the proficiency test provider expects will be needed for the evaluation of the results.

After the test has been evaluated, the proficiency test provider shall inform all the participants of the results of the assessment of their laboratories and also of the statistical overall evaluation of the interlaboratory test (in an anonymous form).

The proficiency test provider should arrange meetings for the exchange of information and specialist criticism. This may promote, for example, the updating of the interlaboratory test system to the state of the art.

9 Proficiency test evaluation

9.1 General statistical evaluation procedure[1)]

9.1.1 General requirements

The statistical methods described below fulfil the following requirements.

a) The methods enable comparisons to be made over a range of concentrations.

b) The methods are robust in the sense that any outliers have only a limited effect on the overall result. Steps were taken to ensure that the results are still meaningful even if the proportion of outliers is 1/3, i.e. the breakdown point is not below 33 %.

c) The methods are fair in regard to the sign of the laboratory error. Adjustments of the analytical results towards higher or lower values does not result in an increase in the probability of a positive assessment.

d) The methods comply with international requirements, in particular with the joint ISO, IUPAC and AOAC protocol [1] and with ISO 13528.

9.1.2 Steps in the evaluation of an interlaboratory test

Evaluation of an interlaboratory test will, as a rule, involve the following four steps.

a) Definition of the standard deviation for proficiency assessment, $\hat{\sigma}$

The standard deviation for proficiency assessment, $\hat{\sigma}$, serves to calculate the quality limits for the analytical results. It may be specified as a quality requirement, but it is, as a rule, determined from the analytical results of the test participants using statistical methods, if it can be assumed that the majority of the participants competently uses suitable analytical methods. In 9.2.2 the Q-method is described, a robust statistical method for calculating the reproducibility standard deviation, s_R.

If steps are to be taken to ensure that the true standard deviation calculated in this way is not too wide or too narrow with regard to the analytical quality requirements, lower and upper limits can be defined for it. If the calculated true standard deviation is above or below one of these limits, the latter shall be defined as $\hat{\sigma}$.

b) Specification of the assigned value, x_a

As already described in 8.6, the specification of the assigned value depends on the sample preparation method. In 9.2.3 the Hampel estimator as a robust statistical method for use when the assigned value is to be determined from the participants' results is described.

1) Information on suitable software for the procedure and the statistical evaluations described in this International Standard is obtainable from the Normenausschuss Wasserwesen, DIN, 10772 Berlin, Germany.

c) Calculation and optional use of a variance function

In interlaboratory tests for assessing laboratories proficiency, samples of the same type having different concentrations are often distributed. A single evaluation of these various samples frequently reveals fluctuations in the variance of the different concentration levels that might result in laboratories being unfairly assessed. To be able to correct for such fluctuations, the variances of the various samples can be defined with the aid of a variance function determined by a regression calculation based on the individual variances. In 9.3, a suitable method requiring a minimum of four different concentration levels is described. In addition, the various samples have to be similar enough from an analytical point of view, in particular with regard to matrix, for it to be possible to assume that the concentration of the analyte is the main variable responsible for the difference in variance.

The proficiency test provider shall check whether the application of such a variance function is meaningful in a particular case. In addition, the method described also involves an additional statistical test that yields information about whether the calculated variance function is sufficiently precise.

d) Calculation of normalized deviations in the results (z- or z_U-scores)

To assess the quality of laboratory results, it is helpful to normalize the extent to which the results deviate from the assigned value, x_a, using the standard deviation for proficiency assessment, $\hat{\sigma}$. The z-scores described in 9.4, which yield tolerance limits in the assessment that are symmetrical with respect to the assigned value, x_a, can be used for this purpose. With a comparatively large standard deviation for proficiency assessment, $\hat{\sigma}$, and measurements close to the limit of determination, this may result in the lower tolerance limit being below the limit of determination. If this is the case, all those laboratories whose analytical findings are below the limit of determination automatically fulfil the quality criterion.

A marked preference for results that are unduly low can also be observed in the case of a smaller standard deviation.

The z_U-scores also described can be used to eliminate these disadvantages.

NOTE Annex C contains an example of the application of the evaluation methods described below.

9.2 Robust evaluation methods

9.2.1 General

In interlaboratory tests for determining proficiency, robust evaluation methods that have a high breakdown point and are sufficiently efficient both for normal distributions and for distributions having positive skew shall be used to determine the mean, μ, and the reproducibility standard deviation, s_R.

The evaluation methods described below (Q-method and Hampel estimator) have very favourable properties with regard to these criteria and, therefore, shall be applied in the sector of water, waste water and sludge analysis.

The method of determining the Hampel estimator and the reproducibility standard deviation can be applied both to multiple determinations and to those not involving replication. The determination of the repeatability standard deviation implies multiple determinations, but is necessary only for information and is not absolutely necessary for testing laboratories. For calculation details, see Annex B.

9.2.2 Determination of reproducibility standard deviation, s_R, using the Q-method

The Q-method is a robust method of determining standard deviations [2], [3], [4], whereas the Hampel estimator is used to determine the mean [4]. It can also be used independently of the Hampel estimator, for example if a specified reference value is to be used as the assigned value.

Annex A uses an example to explain the estimation principle of the Q-method.

In what follows, y_{ji} denotes the result of the i-th measurement made by laboratory j, where $j = 1, ..., J$ and $i = 1, ..., n_j$. Whether multiple determinations are made, i.e. whether n_j is 2 or over, or whether all the laboratories have each made only one measurement, i.e. n_j is 1 for all the values of j from 1 to J is unimportant. First of all, the function

$$H_1(x) = \frac{1}{\binom{J}{2}} \sum_{1 \leqslant j_1 < j_2 \leqslant J} \frac{1}{n_{j_1} \cdot n_{j_2}} \sum_{i_1=1}^{n_{j1}} \sum_{i_2=1}^{n_{j2}} 1\left\{ \left| y_{j_1 i_1} - y_{j_2 i_2} \right| \leqslant x \right\} \tag{1}$$

is calculated. The discontinuity points of this function are denoted by $x_1, ..., x_k$, where $x_1 < x_2 < ... < x_k$. The function is defined as:

$$G_1(x_i) = \begin{cases} 0{,}5 \cdot \left[H_1(x_i) + H_1(x_{i-1}) \right] & \text{if} \quad i \geqslant 2 \\ 0{,}5 \cdot H_1(x_1) & \text{if} \quad i = 1 \text{ and } x_1 > 0 \\ 0 & \text{if} \quad i = 1 \text{ and } x_1 = 0 \end{cases} \tag{2}$$

for all discontinuity points of x_i. Between the discontinuity points, this function is defined by linear interpolation.

Distortions due to rounding are suppressed in the following way:

$$q = 0{,}25 + 0{,}75 H_1(0) \tag{3}$$

and the reproducibility standard deviation, s_R, is calculated using Equation (4):

$$s_R = \frac{G_1^{-1}(q)}{\sqrt{2}\, \Phi^{-1}(0{,}5 + 0{,}5q)} \tag{4}$$

Depending on the number of multiple determinations and on the ratio of the repeatability and reproducibility standard deviations, the asymptotic efficiency of the estimation method is normally greater than 82 % for a normal distribution, but if the number of laboratories is small, this value may not quite be reached and is normally 65 % to 70 %. The breakdown point almost reaches the theoretical maximum of 50 %.

9.2.3 Determination of mean, μ

The mean shall be calculated using the Hampel estimator [4]. In the following, y_j denotes the arithmetic mean of the measurement results of laboratory j:

$$y_j = \frac{1}{n_i} \sum_{i=1}^{n_j} y_{ji} \tag{5}$$

If there are no multiple determinations, y_j denotes the measurement result itself, where $j = 1, ..., J$. The robust mean shall then be calculated using the conditional equation by the Hampel method:

$$\sum_{j=1}^{J} \psi\left(\frac{y_j - \mu}{s_R} \right) = 0 \tag{6}$$

where

$$\psi(x) = \begin{cases} 0 & x \leqslant -4{,}5 \\ -4{,}5 - x & -4{,}5 < x \leqslant -3 \\ -1{,}5 & -3 < x \leqslant -1{,}5 \\ x & -1{,}5 < x \leqslant 1{,}5 \\ 1{,}5 & 1{,}5 < x \leqslant 3 \\ 4{,}5 - x & 3 < x \leqslant 4{,}5 \\ 0 & x > 4{,}5 \end{cases} \tag{7}$$

The exact solution is then calculated in a finite number of steps, which means not iteratively, using the property that Ψ in the argument of μ is partially linear, bearing in mind that the interpolation nodes on the left-hand side of Equation (6) (interpreted here as a function of μ) are as follows:

$y_j + k \cdot s_R$ where $k = -4{,}5; -3; -1{,}5; 0; 1{,}5; 3$ and $4{,}5$.

The solution nearest the median is used, but if this does not yield a clear result, the median itself is used as position parameter.

NOTE If this estimation method is used, laboratory results differing from the mean by more than 4,5 times the reproducibility standard deviation no longer have any effect on the calculation result, i.e. they are treated as outliers.

9.3 Variance function

9.3.1 General

The method described below in four steps is used to determine a function describing the reproducibility standard deviation as a function of the concentration and to check whether this function is sufficiently precise. The method can be used if interlaboratory test results are available for at least four samples with different concentrations.

Let the assigned values or means, μ_1, ..., μ_p, and the associated reproducibility standard deviations, s_{R_1}, ..., s_{R_p}, of samples 1, ..., p, calculated by the Q-method, be given. Furthermore, J_1, ..., J_p denote the relevant numbers of participants.

9.3.2 Step 1: Identification of gross outliers

The slope parameter of the provisional log-linear variance function shall be calculated using Equation (8):

$$\theta_1 = \underset{i=1,\ldots,p}{\text{med}} \left\{ \underset{j \neq i}{\text{med}} \left\{ \frac{\ln s_{R_j} - \ln s_{R_i}}{\ln \mu_j - \ln \mu_i} \right\} \right\} \tag{8}$$

where “med” is a “median function”. The intercept on the axis shall be found using Equation (9):

$$\theta_0 = \underset{i=1}{\text{med}} \left\{ \ln s_{R_i} \right\} - \theta_1 \underset{i=1}{\text{med}} \left\{ \ln \mu_i \right\} \tag{9}$$

Gross outliers can be identified in the case of 4 to 15 samples (i.e. concentration levels) by using the approximate criterion that, if the absolute difference

$$d_i = \left| \ln s_{R_i} - \left(\theta_0 + \theta_1 \ln \mu_i \right) \right| \tag{10}$$

exceeds the value $5/\sqrt{J_i - 1}$ for a sample $i = 1, \ldots, p$, it can be assumed that it exhibits systematic deviations. Such samples shall be ignored in determining the final variance function. It is the responsibility of the proficiency test provider to determine whether the variance function described here or the original reproducibility standard deviation is to be used to calculate the z-scores for the samples.

9.3.3 Step 2: Determination of variance function

After the gross outliers have been eliminated, the final variance function shall be determined in the following way. Using the matrices:

$$X = \begin{pmatrix} 1 & \ln \mu_1 \\ \vdots & \vdots \\ 1 & \ln \mu_p \end{pmatrix} \quad \text{and} \quad W = \begin{pmatrix} J_1 - 1 & & \\ & \ddots & \\ & & J_p - 1 \end{pmatrix} \tag{11}$$

and the vector

$$\gamma = \begin{pmatrix} \ln s_{R_1} \\ \vdots \\ \ln s_{R\,p} \end{pmatrix} \tag{12}$$

the parameters θ_0 and θ_1 of the log-linear variance function are determined as:

$$\begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix} = \left(X^{\mathsf{T}} W X \right)^{-1} X^{\mathsf{T}} W \gamma \tag{13}$$

The adjusted reproducibility standard deviation for sample i ($i = 1, ..., p$) is given by

$$\hat{s}_{R_i} = \exp\left(\theta_0 + \theta_1 \ln \mu_i\right) \tag{14}$$

9.3.4 Step 3: Testing the variance function for adequate precision

The testing values

$$PG_1 = \sum_{i=1}^{p} \left(J_i - 1\right)\left(\theta_0 + \theta_1 \ln \mu_i - \ln s_{R_i}\right)^2 \tag{15}$$

is used to test whether

$$PG_1 \leqslant \chi^2_{p-2;0,95} \tag{16}$$

is valid, where $\chi^2_{p-2;0,95}$ is the 95 % quantile of the chi squared distribution with $p - 2$ degrees of freedom (see Table 1). If this inequality is satisfied, the variance function found may be regarded as sufficiently precise. If not, a check shall be made as to whether the variance is also determined by other influencing factors or whether the functional relationship is more complex.

In the latter case, if at least ten concentrations have been investigated, extended procedures (e.g. a LOESS procedure in which the linear relationship between the logarithm of the concentration and the logarithm of the standard deviation is only assumed locally [5]) can be used. The final decision as to whether the variance function is used rests with the proficiency test manager, who has to weigh the chemical analytical findings against the results of the statistical calculations.

Table 1 — Chi squared distribution having $p - 2$ degrees of freedom

$p-2$ (number of levels − 2)	$\chi^2_{p-2;0,95}$	$p-2$ (number of levels − 2)	$\chi^2_{p-2;0,95}$
1	3,84	16	26,3
2	5,99	17	27,6
3	7,82	18	28,9
4	9,49	19	30,1
5	11,1	20	31,4
6	12,6	21	32,7
7	14,1	22	33,9
8	15,5	23	35,2
9	16,9	24	36,4
10	18,3	25	37,7
11	19,7	26	38,9
12	21,0	27	40,1
13	22,4	28	41,3
14	23,7	29	42,6
15	25,0	30	43,8

9.3.5 Step 4: testing the concentration dependence for significance

If the test criterion $PG_1 \leqslant \chi^2_{p-2;0,95}$ is satisfied and if it can be assumed that the relative standard deviation is independent of concentration, i.e. that $\theta_1 = 1$, the concentration dependence can be tested by determining the testing value, PG_0:

$$PG_0 = 1{,}64 \sum_{i=1}^{p} (J_i - 1)\left(\tilde{\theta}_0 - \frac{\ln s_{R_i}}{\ln \mu_i}\right)^2 \tag{17}$$

where

$$\tilde{\theta}_0 = \frac{1}{\sum_{i=1}^{p} (J_i - 1)} \sum_{i=1}^{p} (J_i - 1) \ln \frac{s_{R_i}}{\mu_i} \tag{18}$$

if

$$PG_0 - PG_1 < 3{,}84$$

the concentration level cannot be shown statistically to influence the relative reproducibility standard deviation. In that case, $\theta_0 = \tilde{\theta}_0$ and $\theta_1 = 1$, i.e. $\exp\left(\tilde{\theta}_0\right)$ is equal to the adjusted relative reproducibility standard deviation.

9.4 z-Scores

Laboratory proficiency assessment shall be based on normalized deviations of the individual analytical results (in the case of multiple determinations, mean values) from an assigned value (8.6), x_a, using a defined standard deviation for proficiency assessment, $\hat{\sigma}$ [1]. Various criteria or standardizations are suitable for measuring such deviations.

The use of so-called z-scores [1] is the recommended way of measuring the deviations of the individual laboratory analytical results from the assigned value, x_a. With the standard deviation for proficiency assessment, $\hat{\sigma}$, a z-score is calculated as follows:

$$z = \frac{y - x_a}{\hat{\sigma}} \tag{19}$$

Assuming that the analytical results have a normal distribution, the z-score is interpreted as meaning that the probability that the absolute value of z does not exceed 2 is 0,954 5, i.e. about 95 %. Consequently, it seems reasonable to define $g = 2$ as the quality limit. The quality criterion is therefore satisfied precisely if the absolute value of z does not exceed 2; otherwise, it may be assumed, with a significance level of $\alpha = 0{,}045\,5$ or about 5 %, that the laboratory has performed the analysis incorrectly. In individual cases, the limits 2,5 or 3 may be used instead of the quality limit $g = 2$. This will be decided by the proficiency test provider taking the requirements into account.

The z-score is sensitive to fluctuations in the reproducibility standard deviation and substantial estimate errors may be expected, particularly if the latter has been estimated using only a few laboratories.

The use of z_U-scores as the quality criterion is recommended for determinands that can never assume negative values, the z-scores being modified as follows:

$$z_U = \begin{cases} \frac{g}{k_1} \cdot z & \text{if} \quad z < 0 \\ \frac{g}{k_2} \cdot z & \text{if} \quad z \geqslant 0 \end{cases} \tag{20}$$

where k_1 and k_2 are unambiguously determined by Equations (21) and (22):

$$\left(k_2 + \frac{1}{\nu}\right) \exp\left\{-\frac{1}{2}k_2^2\right\} = \left(-k_1 + \frac{1}{\nu}\right) \exp\left\{-\frac{1}{2}k_1^2\right\} \tag{21}$$

$$\left[1 - \Phi\left(-\frac{1}{\nu}\right)\right]^{-1} \left[\Phi(k_2) - \Phi(-k_1)\right] = 1 - \alpha \tag{22}$$

The two equations cannot be solved for k_1 and k_2, with the result that it is only possible to calculate these values iteratively.

10 Presentation of results

The participants shall receive a complete set of measurement results, with graphs of the individual results and the quality limits. Anonymity shall be by using a randomly coded number for each laboratory, the identities of the participating laboratories being treated as strictly confidential by the proficiency test provider and rendered accessible to third parties only with the agreement of the laboratory concerned.

At the same time, statistical features, such as the reproducibility standard deviation, mean and assigned value, and, if applicable, the spiked amount of analyte, shall be specified. Measurements lying outside the quality limits (9.4) shall also be clearly identified.

In addition, every participant shall be informed of his laboratory number, the results for his subsamples and an overall assessment of the results of his laboratory and of the interlaboratory test in a suitable form.

New participants shall be provided with brief information relating to the evaluation methods used.

11 Archiving and managing the results

The proficiency test provider shall keep a data base in which it is possible to search for the individual results of the laboratories. Requests by authorized third parties, for example relating to the fulfilling of quality criteria for permits, notifications/accreditations, should be answered quickly. The documentation of all the steps in the methods shall be detailed enough for quality audits by external inspectors.

Annex A
(informative)

Example of Q-method estimation principle

For reasons of clarity, contrary to the stipulation described in 7.5 an example will be reproduced that involves a reduced number of laboratories, namely the measurement results of eight laboratories without multiple determinations:

6, 7, 8, 9, 11, 13, 14, 50.

First all the absolute differences for the selected example:

|6-7| |6-8| |6-9| |6-11| |6-13| |6-14| |6-50|

|7-8| |7-9| |7-11| |7-13| |7-14| |7-50|

|8-9| |8-11| |8-13| |8-14| |8-50|

|9-11| |9-13| |9-14| |9-50|

|11-13| |11-14| |11-50|

|13-14| |13-50|

|14-50|

are sorted by magnitude:

1, 1, 1, 1, 2, 2, 2, 2, 3, 3, 3, 4, 4, 5, 5, 5, 6, 6, 7, 7, 8, 36, 37, 39, 41, 42, 43, 44.

The generalized distribution function, $H_1(x_i)$, is found for this ordered random sample and the continuity-corrected distribution function, $G_1(x_i)$, is determined from it (see Table A.1).

Table A.1 — Empirical distribution function of the interlaboratory differences, $H_1(x_i)$, and the continuity-corrected distribution function, $G_1(x_i)$ — Absolute differences for the chosen example

x_i	$H_1(x_i)$	$G_1(x_i)$
1	4/28	4/56
2	8/28	12/56
3	11/28	19/56
4	13/28	24/56
5	16/28	29/56
6	18/28	34/56
7	20/28	38/56
8	21/28	41/56
36	22/28	43/56
37	23/28	45/56
39	24/28	47/56
41	25/28	49/56
42	26/28	51/56
43	27/28	53/56
44	1	55/56

The quantile parameter, q, has the value $0,25 + H_1(0) = 0,25$, for which

$$G_1^{-1}(q) = G_1^{-1}(0,25) = 2,285\ 7$$

$$\Phi^{-1}(0,5q + 0,5) = \Phi^{-1}(0,625) = 0,318\ 6$$

and therefore

$$s_R = 5,072\ 9.$$

This estimated value is not (directly) affected by the outlier value 50. A reasonable estimate result is ensured even if there are two outlier values in the set of data. The breakdown point of the method, i.e. the maximum proportion of outlier values that still ensures reasonable estimate results, approaches the 50 % value as the number of measurement results increases.

It can be shown that the asymptotic efficiency of the estimation method is 82 % for a normal distribution, whereas, for example, the MAD (median of the median deviations) [6] only reaches about 36,7 %. Even in the case of skew distributions, the estimation method is very efficient.

Annex B
(informative)

Determination of repeatability standard deviation, s_r

If multiple determinations are available, the Q-method may also be used to estimate the repeatability standard deviation, s_r. This is used primarily for information and is not needed for the proficiency test described. It relates not to differences between the laboratories, but to those within the laboratory. The associated empirical distribution function for the intralaboratory differences is defined as follows:

$$H_2(x) = \frac{1}{J}\sum_{j=1}^{J} \quad \frac{2}{n_j(n_j-1)}\sum_{1 \leqslant i_1 < i_2 \leqslant n_j} 1\left\{ \left| y_{ji_1} - y_{ji_2} \right| \leqslant x \right\} \tag{23}$$

The discontuinity points of this function are denoted by x_1, ..., x_k. The second function associated with it is defined as follows:

$$G_2(x_i) = \begin{cases} 0{,}5 \cdot \left[H_2(x_i) + H_2(x_{i-1}) \right] & \text{if} \quad i \geqslant 2 \\ 0{,}5 \cdot H_2(x_1) & \text{if} \quad i = 1 \text{ and } x_1 > 0 \\ 0 & \text{if} \quad i = 1 \text{ and } x_1 = 0 \end{cases} \tag{24}$$

for all the discontuinity points of x_i. Between the discontuinity points, this function is defined by linear interpolation.

Distortions due to rounding are suppressed in the following way:

$$q = 0{,}5 + 0{,}5\, H_2\,(0) \tag{25}$$

and the repeatability standard deviation, s_r, is calculated using Equation (26):

$$s_r = \frac{G_2^{-1}(q)}{\sqrt{2}\Phi^{-1}(0{,}5 + 0{,}5q)} \tag{26}$$

Annex C
(informative)

Example of evaluation method in Clause 9

NOTE The results in this example were each calculated using all the decimal places available. Different results are obtained if rounded intermediate results are used for the calculation.

Table C.1 shows the individual results of a duplicate determination (measurements 1 and 2) for 33 laboratories participating in an interlaboratory test relating to the determination of cadmium in water samples, while Table C.2 shows the laboratory means.

Table C.1 — Interlaboratory test for the determination of cadmium in water samples

Values in micrograms per litre (µg/l)

Laboratory	1	2	3	**4**	5	6	7	8	9	10	11
Measurement 1	41,41	39,22	47,29	**82,46**	45,24	49,96	38,20	45,41	39,82	48,17	39,67
Measurement 2	38,10	38,10	46,82	**90,11**	45,74	53,40	42,65	47,92	42,02	49,47	43,89
Laboratory	12	13	14	15	16	17	18	19	20	21	22
Measurement 1	47,55	35,75	46,13	52,18	45,15	41,57	43,39	49,38	45,67	41,08	49,28
Measurement 2	50,05	37,41	53,64	56,30	47,13	39,08	44,73	47,00	50,53	44,22	47,83
Laboratory	23	24	25	26	**27**	28	29	30	31	32	33
Measurement 1	49,48	48,37	33,96	49,40	**24,4**	41,55	37,43	40,63	49,92	47,88	43,73
Measurement 2	46,04	47,00	36,30	46,44	**24,79**	46,26	39,88	38,64	50,19	48,61	44,18

Table C.2 — Laboratory means of the determination of cadmium in water samples in an interlaboratory test

Values in micrograms per litre (µg/l)

Laboratory	1	2	3	**4**	5	6	7	8	9	10	11
Mean	39,755	38,660	47,055	**86,285**	45,490	51,680	40,425	46,665	40,920	48,820	41,780
Laboratory	12	13	14	15	16	17	18	19	20	21	22
Mean	48,800	36,580	49,885	54,240	46,140	40,325	44,060	48,190	48,100	42,650	48,555
Laboratory	23	24	25	26	**27**	28	29	30	31	32	33
Mean	47,760	47,685	35,130	47,920	**24,595**	43,905	38,655	39,635	50,055	48,245	43,955

The statistical calculations using the method described in 9.2 yielded the following (intermediate) results.

1) Reproducibility standard deviation: $s_R = 5{,}768$ µg/l

   with the intermediate results:

   $H_1(0) = 0{,}000\ 947$;

   $q = 0{,}250\ 71$;

   $\Phi^{-1}(0{,}5\ q + 0{,}5) = 0{,}319\ 576$;

   $G_1^{-1}(q) = 2{,}606\ 7$.

2) Overall mean (Hampel estimator): $\mu = 44{,}707\ 2$ µg/l. The solutions of Equation (6) were as follows: $-\infty$; −1,359; 44,707; 75,256; 86,285; 112,239; ∞. The median is 46,14 and is closest to the value 44,707 2. (In this connection, it must be emphasized that only laboratory 4 differed from the overall mean to such an extent that the Hampel estimator ignored the value completely.)

3) With the quality limit $g = 2$, this yields (for the correction factors of the z_U-scores) the values $k_1 = 1{,}887$ and $k_2 = 2{,}146$.

4) The z-scores found using the laboratory means (Table C.2) and the z_U-scores are shown in Table C.3.

Table C.3 — z- and z_U-scores for the data of an interlaboratory test for determining cadmium in water samples

Laboratory	1	2	3	**4**	5	6	7	8	9	10	11
z-Score	−0,859	−1,048	0,407	**7,209**	0,136	1,209	−0,742	0,339	−0,657	0,713	−0,508
z_U-Score	−0,910	−1,111	0,379	**6,717**	0,126	1,126	−0,787	0,316	−0,696	0,664	−0,538
Laboratory	12	13	14	15	16	17	18	19	20	21	22
z-Score	0,710	−1,409	0,898	1,653	0,248	−0,760	−0,112	0,604	0,588	−0,357	0,667
z_U-Score	0,661	−1,494	0,836	1,540	0,231	−0,805	−0,119	0,563	0,548	−0,378	0,622
Laboratory	23	24	25	26	**27**	28	29	30	31	32	33
z-Score	0,529	0,516	−1,660	0,557	**−3,487**	−0,139	−1,049	−0,879	0,927	0,613	−0,130
z_U-Score	0,493	0,481	−1,760	0,519	**−3,696**	−0,147	−1,112	−0,932	0,864	0,572	−0,138

In this example, the differences between z-scores and z_U-scores are comparatively small since, at 12,9 % of the overall mean, the relative reproducibility standard deviation is also relatively small. Only laboratories 4 and 27, (highlighted) also during the calculation of the Hampel estimator, exceed the quality limit $g = 2$.

Annex D
(informative)

Example of variance function calculation (9.3)

D.1 General

NOTE The results in this example were each calculated using all the decimal places available. Different results are obtained if rounded intermediate results are used for the calculation.

The metolachlor concentration of a total of nine different samples was determined by each of 34 to 38 randomly selected laboratories and the mean and standard deviation calculated. The results are shown in Table D.1.

Table D.1 — Results of an interlaboratory test for nine different metolachlor concentrations

Measurements in micrograms per litre (µg/l)

i	J_i	μ_i	s_{R_i}
1	35	0,1282	0,046 7
2	36	0,1693	0,043 4
3	38	0,2256	0,060 0
4	35	0,2818	0,069 5
5	37	0,3380	0,112 7
6	38	0,4672	0,090 8
7	36	0,5423	0,115 7
8	34	0,5953	0,132 9
9	38	0,6826	0,100 7

The statistical calculations made using the method described in 9.3 yielded the following (intermediate) results.

D.2 Step 1: Determination of gross outliers (9.3.2)

Parameters of the provisional log-linear variance function:

$\theta_0 = -1{,}635$ and $\theta_1 = 0{,}705$.

This yields the absolute differences, d_i, shown in Table D.2. Gross outliers cannot evidently be detected since the absolute differences are less than the associated critical values $5/\sqrt{J_i - 1}$.

Table D.2 — Test for gross outliers

i	d_i	$5/\sqrt{J_i - 1}$	**Status**
1	0,018	0,857	OK
2	0,251	0,845	OK
3	0,129	0,822	OK
4	0,139	0,857	OK
5	0,216	0,833	OK
6	0,228	0,822	OK
7	0,091	0,845	OK
8	0,018	0,870	OK
9	0,392	0,822	OK

D.3 Step 2: Determination of the variance function (see 9.3.3)

Using the matrices

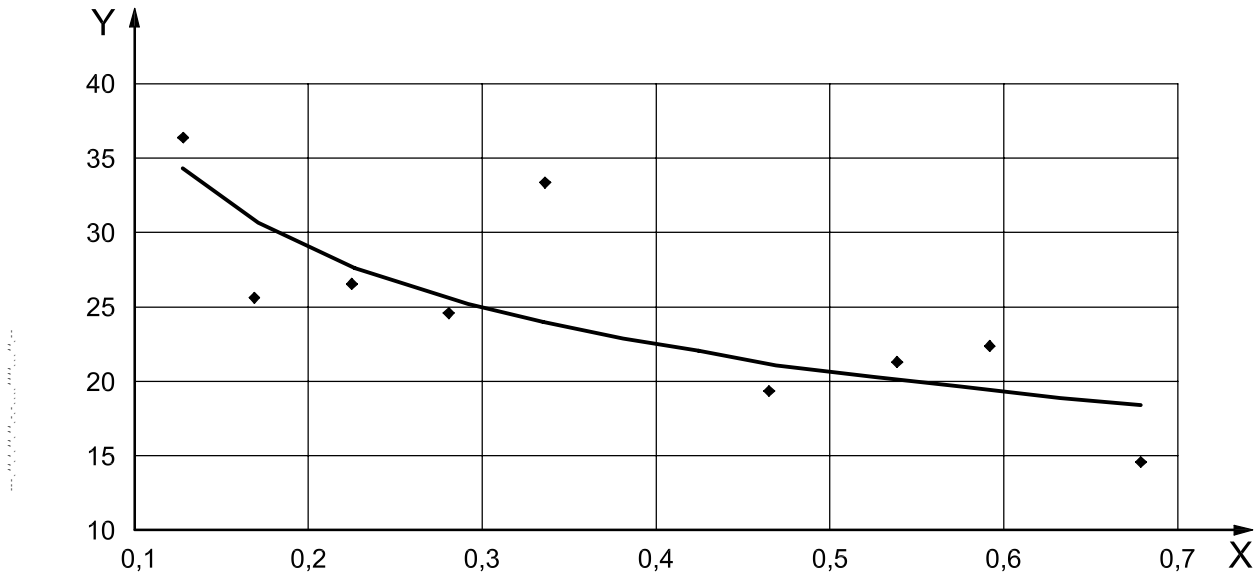
$$X = \begin{pmatrix} 1 & -2{,}054 \\ 1 & -1{,}776 \\ 1 & -1{,}489 \\ 1 & -1{,}267 \\ 1 & -1{,}085 \\ 1 & -0{,}761 \\ 1 & -0{,}612 \\ 1 & -0{,}519 \\ 1 & -0{,}382 \end{pmatrix} \text{ and } W = \begin{pmatrix} 34 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 35 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 37 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 34 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 36 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 37 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 35 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 33 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 37 \end{pmatrix}$$ and the vector

$$\gamma = \begin{pmatrix} -3{,}064 \\ -3{,}137 \\ -2{,}813 \\ -2{,}666 \\ -2{,}183 \\ -2{,}399 \\ -2{,}157 \\ -2{,}018 \\ -2{,}296 \end{pmatrix}$$

the parameters θ_0 and θ_1 of the log-linear variance function are found to be –1,831 and 0,631, respectively. This yields the adjusted reproducibility standard deviations, $\hat{s}_{R_i}$, and the associated relative reproducibility standard deviations shown in Table D.3. The evaluation is also shown graphically in Figure D.1 for the purpose of illustration.

Table D.3 — Adjusted reproducibility standard deviations

i	μ_i	s_{R_i}	$\hat{s}_{R_i}$	$\frac{s_{R_i}}{\mu_i}$ %	$\frac{\hat{s}_{R_i}}{\mu_i}$ %
1	0,128 2	0,046 7	0,043 8	36,43	34,19
2	0,169 3	0,043 4	0,052 2	25,63	30,86
3	0,225 6	0,060 0	0,062 6	26,60	27,76
4	0,281 8	0,069 5	0,072 1	24,66	25,57
5	0,338 0	0,112 7	0,080 8	33,34	23,92
6	0,467 2	0,090 8	0,099 2	19,43	21,22
7	0,542 3	0,115 7	0,108 9	21,34	20,09
8	0,595 3	0,132 9	0,115 5	22,32	19,41
9	0,682 6	0,100 7	0,126 0	14,75	18,45



Key

X concentration in micrograms per litre, µg/l

Y $\frac{s_{R_i}}{\mu_i}$

Figure D.1 — Variance function with adjusted reproducibility standard deviations for a metolachlor interlaboratory test

D.4 Step 3: Testing the variance function for adequate precision (9.3.4)

$PG_1 = 13{,}68$ was obtained for the test variable. Since it is below the critical value $\chi^2_{7;0,95}$, the variance function can be regarded as sufficiently precise.

D.5 Step 4: Testing the concentration dependence for significance (9.3.5)

$PG_0 = 35{,}17$ was obtained for the test variable. Consequently, $PG_0 - PG_1 = 21{,}48 > 3{,}84$, and the effect of concentration on the relative reproducibility standard deviation may be regarded as capable of statistical demonstration.

Bibliography

[1] THOMPSON, M.; Ellison, S.L.R. and WOOD, R. The International Harmonized Protocol for the Proficiency Testing of Analytical Chemistry Laboratories (IUPAC Technical Report) *Pure Appl. Chem*., **78** (1), 2006, pp. 145-196

[2] UHLIG, S. Robust estimation of variance components with high breakdown point in the 1-way random effect model. Published in Kitsos, C.P. and Edler, L.. *Industrial Statistics: Physica-Verlag*, 1997, pp. 65-73

[3] UHLIG, S. *Robuste Schätzung in Kovarianzmodellen mit Anwendung im Qualitätsmanagement und in Finanzmärkten (Robust estimation in covariance models with applications in quality management and financial markets)*, Postdoctoral thesis, Berlin: Free University of Berlin, 1997

[4] MÜLLER, C. H. and UHLIG, S. Estimation of variance components with high breakdown point and high efficiency. *Biometrika*, **88** (2), 2001, pp. 353-366

[5] HASTIE, T.J. and TIBSHIRANI, R.J. *Generalized additive models*. London: Chapman & Hall, 1991

[6] DAVIS, L. and GATHER, U. The Identification of Multiple Outliers. *J. Am. Stat. Ass*., **88**, 1993. pp. 782-792

ICS 13.060.45

Price based on 25 pages