INTERNATIONAL STANDARD

ISO 11095

First edition
1996-02-01

Linear calibration using reference materials

Étalonnage linéaire utilisant des matériaux de référence

Reference number
ISO 11095:1996(E)

Contents

International Organization for Standardization
Case Postale 56 • CH-1211 Genève 20 • Switzerland
Printed in Switzerland

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.

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.

International Standard ISO 11095 was prepared by Technical Committee ISO/TC 69, *Applications of statistical methods*, Subcommittee SC 6, *Measurement methods and results*.

Annexes A and B form an integral part of this International Standard. Annex C is for information only.

Introduction

Calibration is an essential part of most measurement procedures. It is a set of operations which establish, under specified conditions, the relationship between values indicated by a measurement system and the corresponding accepted values of some "standards". In this International Standard, the standards are reference materials.

A reference material (RM) is a substance or an artifact for which one or more properties are established sufficiently well to validate a measurement system. There exist several kinds of RMs:

a) an internal reference material is an RM developed by a user for his/her own internal use;

b) an external reference material is an RM provided by someone other than the user;

c) a certified reference material is an RM issued and certified by an organization recognized as competent to do so.

Linear calibration using reference materials

1 Scope

This International Standard:

a) outlines the general principles needed to calibrate a measurement system and to maintain that "calibrated" measurement system in a state of statistical control;

b) provides a basic method

— for estimating a linear calibration function under either one of two assumptions relating to the variability of the measurements,

— for checking the assumption of linearity of the calibration function and the assumptions on the variability of the measurements, and

— for estimating the value of a new unknown quantity by transforming the measured values obtained on that quantity with the calibration function;

c) provides a control method for extended use of a calibration function

— for detecting when the calibration function needs to be updated, and

— for estimating the uncertainty of the measured values after transformation with the calibration function;

d) provides two alternatives to the basic method under special conditions;

e) illustrates the basic method and the control method with an example.

This International Standard is applicable to measurement systems for which reference materials are available.

It is applicable to measurement systems with an assumed linear calibration function. It offers a method for examining the assumption of linearity. If it is known that the calibration function is nonlinear, then this International Standard is not applicable unless one uses the "bracketing technique" described in 8.3.

This International Standard does not make a distinction among the various types of RMs and considers that the accepted values of the RMs selected to calibrate the measurement system are without error.

2 Normative references

The following standards contain provisions which, through reference in this text, constitute provisions of this International Standard. At the time of publication, the editions indicated were valid. All standards are subject to revision, and parties to agreements based on this International Standard are encouraged to investigate the possibility of applying the most recent editions of the standards indicated below. Members of IEC and ISO maintain registers of currently valid International Standards.

ISO 3534-1:1993, *Statistics — Vocabulary and symbols — Part 1: Probability and general statistical terms.*

ISO 3534-2:1993, *Statistics — Vocabulary and symbols — Part 2: Statistical quality control.*

ISO Guide 30:1992, *Terms and definitions used in connection with reference materials.*

3 Definitions

For the purposes of this International Standard, the definitions given in ISO 3534-1 and ISO 3534-2 and the following definition apply.

3.1 reference material: A substance or an artifact for which one or more properties are established sufficiently well to be used to validate a measurement system.

4 General principles

Calibration is a procedure that determines the systematic difference that may exist between a measurement system and a "reference" system represented by the reference materials and their accepted values. In this International Standard, the term system (measurement system or reference system) is used to represent not only a measuring instrument but also the set of procedures, operators and environment conditions associated with that instrument.

The output of a calibration procedure is a calibration function that is used to make transformations of future measurement results. In this International Standard, the term "transformation" refers to

- either a correction of the future measurements if both the accepted values of the reference materials (RMs) and the observed values have the same units,
- or a translation from the units of the observed measurements to the units of the RMs.

The validity of the calibration function depends on two conditions:

a) that the measurements from which the calibration function was calculated are representative of the normal conditions under which the measurement system operates; and

b) that the measurement system is in a state of control.

The calibration experiment must be designed to ensure that point a) is met. The control method determines, as soon as possible, when the system has to be considered out of control.

The procedure in this International Standard is only applicable to measurement systems which are linearly related to their reference systems. To check whether the assumption of linearity is valid, more than two RMs must be used during the calibration experiment. This is illustrated in the basic method. Using several RMs, the basic method provides a strategy and techniques to analyse the data collected during the calibration experiment. If linearity is not in question, then an alternative method, simpler than the basic method, can be used to estimate a linear calibration function based on one point. This "one-point calibration" method (following a zero-level transformation) does not allow for any test of assumptions, but it is a quick and easy method to "recalibrate" a system that has been studied more thoroughly during previous experiments. If linearity is in question, then a second alternative can be used, called "bracketing".

The basic method and the one-point method are based on the assumption that the effort invested in calibration will be valid over a period of stability of the process. To study the period over which the calibration is valid, a control method has to be in place. The control method is designed to detect whether changes have taken place in the system that justify an investigation and/or a recalibration. The control method also provides a simple way to determine the precision of the values that have been transformed with a given calibration function.

The bracketing method is labour intensive but may provide greater accuracy in the determination of the values of unknown quantities. This method consists of surrounding as tightly as possible (bracketing) each unknown quantity by two RMs and extracting a transformed value for the unknown quantity from measurements of both the unknown quantity and the values of the two RMs. Only short-term stability of the measurement process is assumed (stability during the measurement of the unknown quantity and of the two RMs). Linearity is assumed solely in the interval between the values of the two RMs.

5 Basic method

5.1 General

This clause describes how to estimate and use a linear calibration function when several (more than two) RMs are available. The availability of several RMs allows the linearity of the calibration function to be verified.

5.2 Assumptions

5.2.1 It is assumed that there is no error in the accepted values of the RMs (this assumption will not be checked in this International Standard). In practice, accepted values of RMs are quoted with their uncertainties. The assumption of no error in the accepted values of the RMs can be considered valid if the uncertainties are small compared to the magnitude of the errors in the measured values of these RMs (see ref. [1]).

NOTE 1 In situations where the RMs have been treated chemically or, in some instances physically, before instrument readings are taken, this International Standard may underestimate the uncertainty associated with the transformation of a new measurement result.

5.2.2 The calibration function is assumed to be linear (this assumption will be examined).

5.2.3 Repeated measurements of a given RM are assumed to be independent and normally distributed, with variance referred to as "residual variance" (the independence and normality assumptions will not be checked in this International Standard). The square root of the residual variance is referred to as the residual standard deviation.

5.2.4 The residual standard deviation is assumed to be either constant or proportional to the accepted value of the RM (this assumption will be examined).

5.3 Calibration experiment

5.3.1 Experimental conditions

Experimental conditions should be the same as the normal operating conditions of the measurement system; i.e. if, for example, more than one operator uses the measuring equipment then there should be more than one operator represented in the calibration experiment.

5.3.2 Choice of RMs

The range of values spanned by the selected RMs should include (as far as is possible) the range of values encountered during normal operating conditions of the measurement system.

The composition of the selected RMs should be as close as possible to the composition of the targeted material to be measured.

The values of the RMs should be distributed approximately equidistantly over the range of values encountered during normal operating conditions of the measurement system.

5.3.3 Number of RMs, N

The number of RMs used to assess the calibration function should be at least 3.

For an initial assessment of the calibration function, a number larger than 3 is recommended (at least 3 over any subinterval where there is a doubt about the linearity of the calibration function).

5.3.4 Number of replicates, K

Each RM should be measured at least twice (as many replicates as is possible in practice is recommended). The number of replicates should be the same for all RMs.

The time and conditions at which the replicates are taken should cover as wide a range as is necessary to ensure that all operating conditions are represented.

5.4 Strategy for analysing the data

5.4.1 Plot the data to check

a) the state of control of the measurement system during the calibration experiment,

b) the assumption of linearity, and

c) the variability of the measurements as a function of the accepted values of the RMs.

5.4.2 Estimate the linear calibration function under the assumption of constant residual standard deviation.

5.4.3 Plot the calibration function and the residuals. The residuals plot is a strong indicator of departure from either the assumption of linearity or from the assumption of constant residual standard deviation. If the assumption of constant residual standard deviation does hold, skip step 5.4.4 and continue with step 5.4.5. Otherwise, execute step 5.4.4.

5.4.4 Estimate the linear calibration function under the assumption of proportional residual standard deviation and plot the calibration function and the residuals.

5.4.5 Evaluate the lack of fit of the calibration function. If the variability due to lack of fit is large relative to the variability due to replication of measurements, investigate the procedures followed during the calibration experiment and re-examine the assumption of linearity of the calibration function. If the assumption of linearity does not hold, then an alternative is to use the bracketing technique described in 8.3.

NOTE 2 There exist other techniques, beyond the scope of this International Standard, that allow the fitting of a quadratic or polynomial curve to the data (see refs. [2] and [3]).

5.4.6 Transform future measured values with the calibration function.

The next clause describes the six steps of this strategy. Clause 9 illustrates the basic method with an example.

6 The steps of the basic method

6.1 Plot of the data collected during the calibration experiment

Figure 1 shows a plot of the measured values versus the corresponding accepted values of the RMs. Figure 1 as well as figures 2 to 5 are obtained from simulated data. The purpose of these five plots is to illustrate the type of information one can extract from such plots. A complete example is treated in clause 8 with data, plots and analysis.

The major purpose of the plot shown as figure 1 is to detect visually any unusual behaviour of the measurement system during the calibration experiment, and to identify potential outliers. If possible, label the order of the data points and look for obvious time trends. If some of the data are considered suspicious, or if a time trend is obvious, then an investigation shall take place to discover causes of irregularities. As soon as the causes of irregularities are removed, the calibration experiment should be repeated and new data should be collected to establish a calibration function.

If the causes for one or a very few outliers are found, and if these causes do not affect the remaining measurements, then the outliers can be eliminated. The calibration experiment then becomes unbalanced; i.e. there is an unequal number of measurements K_n instead of K for each RM. Estimation of the calibration function can still proceed with the formulae given in 6.2, 6.4 and 6.5 replaced by the ones in annex B.

Figure 1 also allows an early diagnosis of the assumption of linearity of the calibration function, as well as a first look at the assumption of constant residual standard deviation. The linearity of the calibration function can be visually checked by visualizing a straight line through the data plotted in figure 1 (there seems to be some curvature in the data of figure 1). The assumption of constant residual standard deviation can be checked by looking at the spread of the points in figure 1 for a given RM. If it appears that this spread increases with the accepted values of the RMs, then the assumption of constant residual standard deviation is probably not correct (this does not seem to be the case for figure 1). A more sophisticated plot to check the assumptions of linearity and of constant residual standard deviation is presented in 6.3.

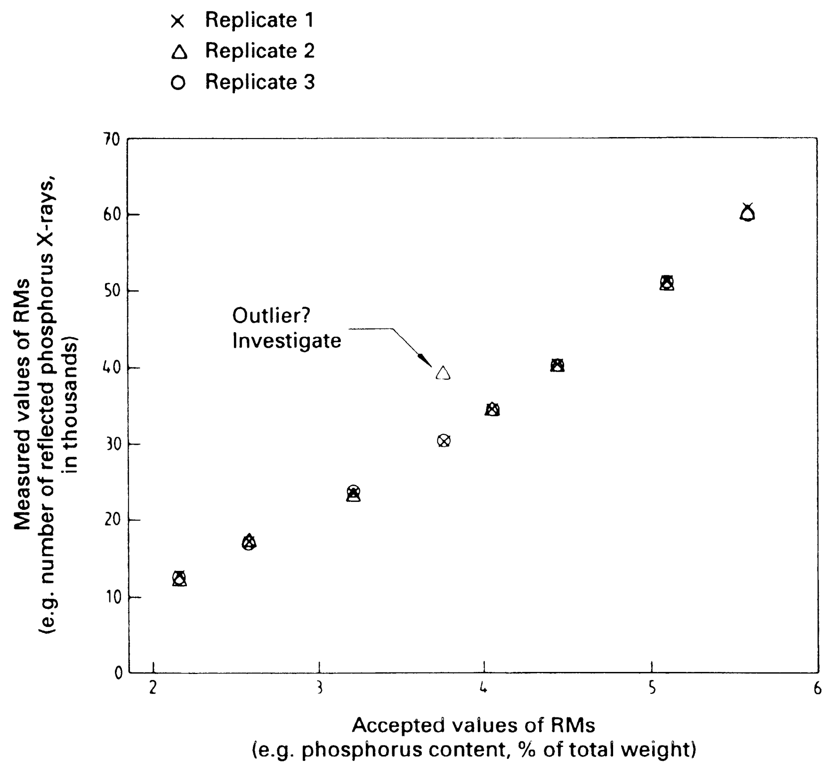

Figure 1 — Schematic diagram of data collected during the calibration experiment

6.2 Estimation of the linear calibration function under the assumption of constant residual standard deviation

6.2.1 Model

The assumptions of linearity of the calibration function and of constant residual standard deviation are captured by the model

$$y_{nk} = \beta_0 + \beta_1 x_n + \varepsilon_{nk}$$

where

- x_n is the accepted value of the n^{th} RM ($n = 1, \ldots, N$);
- y_{nk} is the k^{th} measurement of the n^{th} RM ($k = 1, \ldots, K$);
- $\beta_0 + \beta_1 x_n$ represents the expected value of the measurements of the n^{th} RM;
- ε_{nk} is the deviation between y_{nk} and the expected value of the measurement of the n^{th} RM (these deviations are assumed to be independent and normally distributed with mean 0 and with variance σ^2);
- β_0, β_1 and σ^2 are three parameters to be estimated from the data collected during calibration:
  - β_0 is the intercept of the calibration function,
  - β_1 is its slope,
  - σ^2 is a measure of the precision of the measurement system.

6.2.2 Estimates of the parameters

Estimates of the parameters β_0, β_1 and σ^2 can be obtained by using the formulae below or by running a linear regression software package with two columns of equal length as input, one for y and one for x.

NOTE 3 Estimates of parameters in this International Standard have a symbol ^ to differentiate them from the parameters themselves which are unknown.

$$\hat{\beta}_1 = \frac{\sum_{n=1}^{N} (x_n - \bar{x})(y_{n\cdot} - \bar{y})}{\sum_{n=1}^{N} (x_n - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\hat{\sigma}^2 = \frac{\text{SSE}}{(NK - 2)}$$

where

$$\bar{x} = \frac{1}{N} \sum_{n=1}^{N} x_n$$

$$y_{n\cdot} = \frac{1}{K} \sum_{k=1}^{K} y_{nk}$$

$$\bar{y} = \frac{1}{N} \sum_{n=1}^{N} y_{n\cdot}$$

$$NK = N \times K$$

$$\hat{y}_n = \hat{\beta}_0 + \hat{\beta}_1 x_n$$

$$e_{nk} = y_{nk} - \hat{y}_n$$

$$\text{SSE} = \sum_{n=1}^{N} \sum_{k=1}^{K} (e_{nk})^2$$

6.3 Plots of the calibration function and of the residuals

Figures 2 and 3 are recommended to test departures from the assumptions embedded in the model of 6.2.

6.3.1 Plot of the calibration function

In figure 2 the estimated calibration function is added to figure 1.

The plot shown as figure 2 primarily allows a check of the calculations given in 6.2.2. It also provides a visual check of the assumption of linearity of the calibration function.

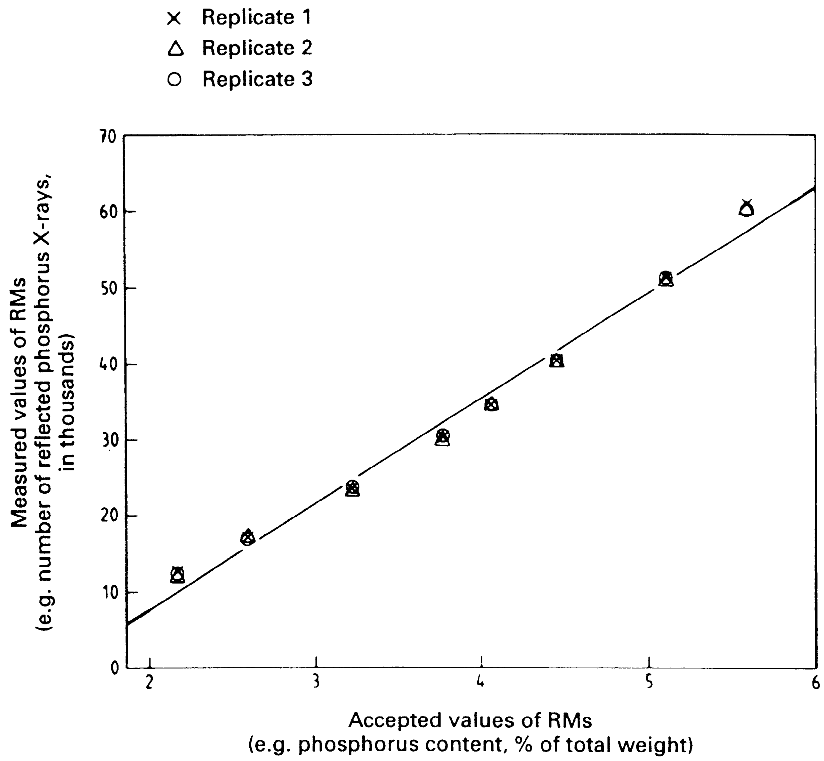

Figure 2 — Schematic diagram of a calibration curve

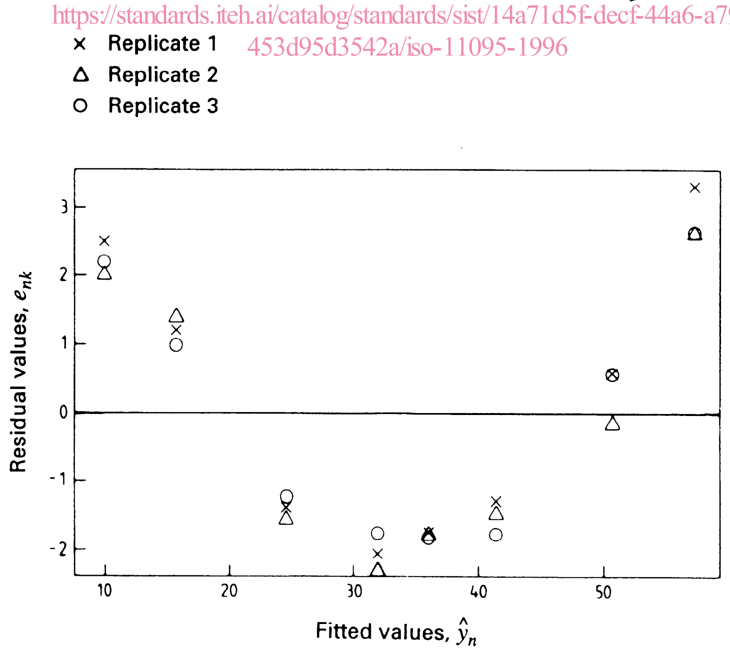

Figure 3 — Schematic diagram of a plot of residuals versus fitted values

6.3.2 Plot of the residuals versus the fitted values

The plot of the residuals e_{nk} versus the fitted values $\hat{y}_n$ (figure 3) is a powerful tool to detect departure from the two assumptions of linearity and of constant residual standard deviation. If these two assumptions hold, then figure 3 should display a plot of randomly distributed points centred around zero. Departure from the assumption of linearity is indicated by a systematic pattern between the residuals and the fitted values (as is the case in figure 3). Departure from the assumption of constant residual standard deviation is indicated by a dispersion in the data that increases or decreases with the fitted values. In figure 3, the dispersion of the residuals for any fitted value is almost constant throughout. Therefore, the assumption of constant residual standard deviation is tenable in this situation.

NOTE 4 Figure 8 illustrates the situation where the assumption of constant residual standard deviation is not tenable.

If the assumption of constant residual standard deviation does not hold, then the data collected during the calibration experiment must be re-analysed. A plot of the standard deviation of the replicated measurements of an RM versus the accepted value of that RM will indicate whether the assumption of proportional residual standard deviation is tenable. See figure 9 for such a plot.

a) If the assumption of proportional residual standard deviation seems to hold, then the data can be re-analysed according to step 6.4.

b) If the assumption of proportional residual standard deviation does not hold but there exists a model relating the residual standard deviation to the accepted values of the RMs (for example inverse proportionality), then an approach similar to the one presented in step 6.4 can be used.

If the assumption of linearity does not hold, then an alternative is to use the bracketing technique described in 8.3.

NOTE 5 There exist other techniques, beyond the scope of this International Standard, that allow the fitting of a quadratic or polynomial curve to these data (see refs. [2] and [3]).

Finally, testing the assumptions of independence and of normality of the ε_{nk} values is beyond the scope of this International Standard. These two assumptions are crucial to the validity of step 6.5 and can also be checked by studying the residuals. For example, a normal probability plot of the residuals allows a check of the normality assumption and a plot of the residuals against time allows a check of the assumption of independence of the measurements. Further information can be found in ref. [3].

6.4 Estimation of the calibration function under the assumption of proportional residual standard deviation and plot of the calibration function and the residuals

6.4.1 Model

An alternate model to the one given in step 6.2.1 is one where the calibration function is linear but the residual standard deviation increases with the accepted values of the RMs. This is captured in the model

$$y_{nk} = \gamma_0 + \gamma_1 x_n + \eta_{nk}$$

where

x_n is the accepted value of the n^{th} RM ($n = 1, \ldots, N$);

y_{nk} is the k^{th} measurement of the n^{th} RM ($k = 1, \ldots, K$);

$\gamma_0 + \gamma_1 x_n$ represents the expected value of the measurement of the n^{th} RM;

η_{nk} is the deviation between y_{nk} and the expected measurement of the n^{th} RM (these deviations are assumed to be independent and normally distributed with mean 0 and with a variance proportional to x_n^2); i.e.

$$\text{var}(\eta_{nk}) = \text{var}(y_{nk}) = x_n^2\, \tau^2$$

γ_0, γ_1 and τ^2 are three parameters to be estimated from the data collected during calibration:

- γ_0 and γ_1 are, respectively, the intercept and the slope of the calibration function,
- τ^2 is a measure of the relative precision of the measurement system.

This model can be transformed into a model equivalent to the one given in 6.2.1; i.e. with errors having constant variance. The transformation consists of dividing by x_n both sides of the equation

$$y_{nk} = \gamma_0 + \gamma_1 x_n + \eta_{nk}$$

This gives

$$\frac{y_{nk}}{x_n} = \frac{\gamma_0}{x_n} + \gamma_1 + \frac{\eta_{nk}}{x_n}$$

or, equivalently,

$$z_{nk} = \gamma_1 + \gamma_0 w_n + \varepsilon_{nk}$$

where

$$z_{nk} = y_{nk}/x_n$$

$$w_n = 1/x_n$$

$$\varepsilon_{nk} = \eta_{nk}/x_n$$

The new model can be analysed as in 6.2 after making the correct substitutions of terms.

6.4.2 Estimates of the parameters

The estimates of the parameters γ_0, γ_1 and τ^2 can be obtained by using the formula below or by running a weighted linear regression software package with three columns of equal length as input, one for y, one for x, and one for the weights ($= 1/x^2$). The same outputs can also be obtained by using a linear regression software package without weights but with the two input columns being z and w.

$$\hat{\gamma}_0 = \frac{\sum_{n=1}^{N} (w_n - \bar{w})(z_{n\cdot} - \bar{z})}{\sum_{n=1}^{N} (w_n - \bar{w})^2}$$

$$\hat{\gamma}_1 = \bar{z} - \hat{\gamma}_0 \bar{w}$$

$$\hat{\tau}^2 = \frac{\text{WSSE}}{(NK - 2)}$$

where

$$NK = N \times K$$

$$z_{nk} = \frac{y_{nk}}{x_n}$$

$$w_n = \frac{1}{x_n}$$

$$\bar{w} = \frac{1}{N} \sum_{n=1}^{N} w_n$$

$$z_{n\cdot} = \frac{1}{K} \sum_{k=1}^{K} z_{nk}$$

$$\bar{z} = \frac{1}{N} \sum_{n=1}^{N} z_{n\cdot}$$

$$\hat{z}_n = \hat{\gamma}_1 + \hat{\gamma}_0 w_n$$

$$u_{nk} = z_{nk} - \hat{z}_n$$

$$\text{WSSE} = \sum_{n=1}^{N} \sum_{k=1}^{K} (u_{nk})^2$$

6.4.3 Plot of the calibration function and residuals

As in 6.3, two plots are recommended:

a) a plot of the estimated calibration function $\hat{y} = \hat{\gamma}_0 + \hat{\gamma}_1 x$ with the data of figure 1;

b) a plot of the weighted residuals u_{nk} versus the weighted fitted values $\hat{z}_n$.

The interpretation of these plots is the same as that for figures 2 and 3.

6.5 Evaluation of the lack of fit of the calibration function

6.5.1 General

A comparison between

— the variability due to lack of fit of the model selected either in 6.2 or in 6.4 and

— the variability of the pure error representing the inability of the system to repeat measurements exactly

is carried out after constructing an ANOVA table. Such a comparison is possible because the measurements of each RM have been replicated.

The selection of the significance level α depends on particular applications and is left to the user of this International Standard.

6.5.2 Model with constant residual standard deviation (defined in 6.2)

6.5.2.1 The ANOVA table shown as table 1 can be obtained by using the formulae below or as an output of most linear regression software packages.