INTERNATIONAL STANDARD

ISO 12099

Second edition
2017-08

Animal feeding stuffs, cereals and milled cereal products — Guidelines for the application of near infrared spectrometry

Aliments des animaux, céréales et produits de mouture des céréales — Lignes directrices pour l'application de la spectrométrie dans le proche infrarouge

Reference number
ISO 12099:2017(E)

ISO copyright office
Ch. de Blandonnet 8 • CP 401
CH-1214 Vernier, Geneva, Switzerland
Tel. +41 22 749 01 11
Fax +41 22 749 09 47
copyright@iso.org
www.iso.org

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.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

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. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation on the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see the following URL: www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/TC 34, *Food products*, Subcommittee SC 10, *Animal feeding stuffs*.

This second edition cancels and replaces the first edition (ISO 12099:2010), which has been technically revised.

Introduction

This document has been drafted using, as a basis, ISO 21543 | IDF 201, which was prepared by Technical Committee ISO/TC 34, *Food products*, Subcommittee SC 5, *Milk and milk products*, and the International Dairy Federation (IDF).

Animal feeding stuffs, cereals and milled cereal products — Guidelines for the application of near infrared spectrometry

1 Scope

This document gives guidelines for the determination by near infrared spectroscopy of constituents such as moisture, fat, protein, starch and crude fibre and parameters such as digestibility in animal feeding stuffs, cereals and milled cereal products.

The determinations are based on spectrometric measurement in the near infrared spectral region.

2 Normative references

There are no normative references in this document.

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— IEC Electropedia: available at http://www.electropedia.org/

— ISO Online browsing platform: available at http://www.iso.org/obp

3.1
near infrared instrument
NIR instrument
apparatus which, when used under the conditions defined in this document, predicts *constituent contents* (3.3) and *technological parameters* (3.4) in *animal feeding stuffs* (3.2), cereals and milled cereal products through relationships to absorptions in the near infrared range

3.2
animal feeding stuffs
substance or product, including additives, whether processed, partially processed or unprocessed, intended to be used for oral feeding to animals

EXAMPLE Raw materials, fodder, meat and bone meal, mixed feed and other end products, pet food, etc.

3.3
constituent content
mass fraction of substances determined using the appropriate, standardized or validated chemical method

Note 1 to entry: The mass fraction is often expressed as a percentage.

Note 2 to entry: For examples of appropriate methods, see References [1] to [12].

EXAMPLE Moisture, fat, protein, crude fibre, neutral detergent fibre and acid detergent fibre.

3.4
technological parameter
property or functionality of *animal feeding stuffs* (3.2), cereals and milled cereal products that can be determined using the appropriate, standardized or validated method(s)

Note 1 to entry: It is possible to develop and validate NIR methods for other parameters and sample types than listed above, as long as the procedure from this document is observed. The measuring units of the parameters determined follow the units used in the reference methods.

EXAMPLE Digestibility.

4 Principle

Spectral data in the near infrared region are collected and transformed to constituent or parameter concentrations by calibration models developed on representative samples of the products concerned.

5 Apparatus

5.1 Near infrared instruments.

Instruments based on diffuse reflectance or transmittance measurement covering the near infrared wavelength region of 770 nm to 2 500 nm (12 900 cm^{-1} to 4 000 cm^{-1}) or segments of this or at selected wavelengths or wavenumbers. The optical principle may be dispersive (e.g. grating monochromators), interferometric or non-thermal (e.g. light emitting diodes, laser diodes and lasers). The instrument should be provided with a diagnostic test system for testing photometric noise and reproducibility, wavelength/wavenumber accuracy and wavelength/wavenumber precision (for scanning spectrophotometers).

The instrument should measure a sufficiently large sample volume or surface to eliminate any significant influence of inhomogeneity derived from chemical composition or physical properties of the test sample. The sample path length (sample thickness) in transmittance measurements should be optimized according to the manufacturer's recommendation with respect to signal intensity for obtaining linearity and maximum signal/noise ratio.

5.2 Appropriate milling or grinding device, for preparing the sample (if needed).

NOTE Changes in grinding or milling conditions can influence NIR measurements due, for example, to heating which can drive off volatile components such as water.

6 Calibration and initial validation

6.1 General

The instrument shall be calibrated before use. Calibration involves the comparison with a reference and adjustment processes to the instrument. Because a number of different calibration systems can be applied with NIR instruments, no specific procedure can be given for calibration.

For an explanation of methods for calibration development, see Reference [16] and the respective manufacturer's manual. For the validation, it is important to have a sufficient number of representative samples, covering variations such as the following:

a) combinations and composition ranges of major and minor sample components;

b) seasonal, geographic and genetic effects on forages, feed raw material and cereals;

c) processing techniques and conditions;

d) storage conditions;

e) sample and instrument temperature;

f) instrument variations (i.e. differences between instruments).

NOTE For a solid validation, at least 20 samples are needed.

6.2 Reference methods

Internationally accepted reference methods for determination of moisture, fat, protein and other constituents and parameters should be used. See References [1] to [12] for examples.

The reference method used for calibration should be in statistical control. It is essential to know the precision of the reference method.

Where possible, references that provide measurement traceability to the SI (International system of units), such as certified reference materials, should be used.

6.3 Outliers

In many situations, statistical outliers are observed during calibration and validation. Outliers may be related to NIR data (spectral outliers, hereafter referred to as "x-outliers") or errors in reference data or samples with a different relationship between reference data and NIR data (hereafter referred to as "y-outliers"); see Figures B.1 to B.5 for examples.

For the purpose of validation, samples are not to be regarded as outliers if they fulfil the following conditions:

a) if they are within the working range of the constituents/parameters in the calibration(s);

b) if they are within the spectral variation of the calibration samples, as, for example, estimated by Mahalanobis distance;

c) if the spectral residual is below a limit defined by the calibration process;

d) if the prediction residual is below a limit defined by the calibration process.

If a sample appears as an outlier, then it should be checked initially to see if it is an x-outlier. If it exceeds the x-outlier limits defined for the calibration, it should be removed. If it is not an x-outlier, then both the reference value and the NIR predicted value should be checked, e.g. by repeated measurements. If these confirm the original values, then the sample should not be deleted and the validation statistics should include this sample. If the repeat values show that either the original reference values or the NIR predicted ones were in error, then the new values should be used.

6.4 Validation of calibration models

6.4.1 General

Before use, calibration equations shall be validated locally on an independent test set that is representative of the sample population to be analysed. For the determination of bias, slope and for the determination of standard error of prediction (SEP, see 7.5), at least 20 samples are needed. Validation shall be carried out for each sample type, constituent/parameter, temperature and other factors known to affect or expected to have an effect the measurement. The calibration is valid only for the variations, i.e. sample types, range and temperature, used in the validation.

NOTE 1 Calibration models can only be used in the range they have been validated.

Results obtained on the independent test set are plotted, reference against NIR, and residuals against reference results, to give a visual impression of the performance of the calibration. The SEP is calculated (see 7.5) and the residual plot of data corrected for mean systematic error (bias) is examined for outliers, i.e. samples with a residual exceeding ±3 s_{SEP}.

If the validation process shows that the model cannot produce acceptable statistics, then it should not be used.

NOTE 2 What will be acceptable will depend, for example, on the performance of the reference method, the covered range, the purpose of the analysis, etc., and is up to the parties involved to decide.

Where available and suitable, reference materials or certified reference materials can be used as part of validation of calibration models.

The next step is to fit NIR, y_{NIRS}, and reference data, y_{ref}, by linear regression ($y_{ref} = a + b \times y_{NIRS}$) to produce statistics that describe the validation results.

6.4.2 Bias correction

The data are also examined for a bias between the methods. If the difference between means of the NIR predicted and reference values is significantly different from zero, then this indicates that the calibration is biased. A bias may be removed by adjusting the constant term (see 7.3) in the calibration equation.

6.4.3 Slope adjustment

If the slope, *b*, is significantly different from 1, the calibration is skewed.

Adjusting the slope/intercept of the calibration is generally not recommended unless the calibration is applied to new types of samples or instruments. If a reinvestigation of the calibration does not detect outliers, especially outliers with high leverage, it is preferable to expand the calibration set to include more samples. However, if the slope is adjusted, the calibration should then be tested on a new independent test set.

6.4.4 Expansion of calibration set

If the accuracy of the calibration does not meet expectations, the calibration set should be expanded to include more samples or a new calibration should be made. In all cases when a new calibration is developed on an expanded calibration set, the validation process should be repeated on a new validation set. If necessary, expansion of the calibration set should be repeated until acceptable results are obtained on a validation set.

6.5 Changes in measuring and instrument conditions

Unless additional calibration is performed, a local validation of a NIR method stating the accuracy of the method can generally not be considered valid if the test conditions are changed.

For example, calibrations developed for a certain population of samples may not be valid for samples outside this population, although the analyte concentration range is unchanged. A calibration developed on grass silages from one area may not give the same accuracy on silages from another area if the genetic, growing and processing parameters are different.

Changes in the sample presentation technique or the measuring conditions, e.g. temperature, not included in the calibration set may also influence the analytical results.

Calibrations developed on a certain instrument cannot always be transferred directly to an identical instrument operating under the same principle. It may be necessary to perform bias or slope / intercept adjustments to calibration equations. In many cases, it will be necessary to standardize the two instruments against each other before calibration equations can be transferred[16]. Standardization procedures can be used to transfer calibrations between instruments of different types provided that samples are measured in the same way (reflectance, transmittance) and that the spectral region is common.

If the conditions are changed, a supplementary validation should be performed.

The calibrations should be checked whenever any major part of the instrument (optical system, detector) has been changed or repaired.

7 Statistics for performance measurement

7.1 General

The performances of a prediction model shall be determined by a set of validation samples. This set consists of samples which are independent of the calibration set. In a plant, it will be new batches; in agriculture, it will be a new crop or a new experiment location.

This set of samples shall be carefully analysed following the reference methods. The care to analyse validation samples shall be emphasized and the precision of these results is more important for the validation set than for the samples used at the calibration phase.

The number of validation samples shall be at least 20 to compute the statistics with some confidence.

The NIR protocol used for the determination of the performances of the prediction model shall be the same as that used in routine (one measurement or two measurements).

7.2 Plot the results

It is important to visualize the results in plots, i.e. reference vs. predicted values or residuals vs. predicted values.

The residuals are defined by Formula (1):

$$e_i = y_i - \hat{y}_i \tag{1}$$

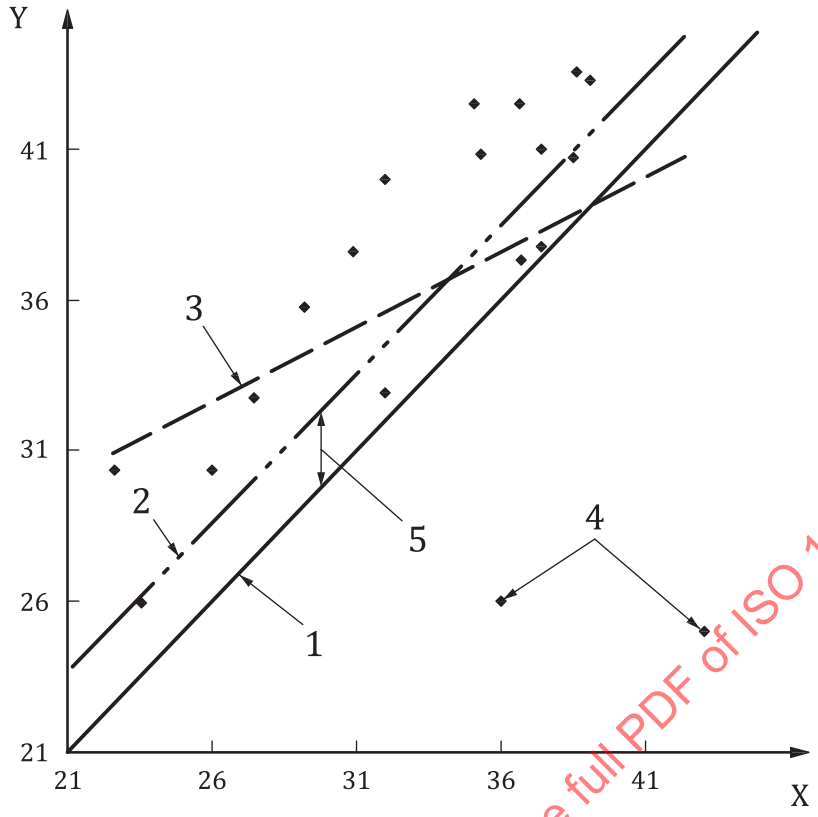
where

y_i is the i^{th} reference value (y_{ref});

$\hat{y}_i$ is the i^{th} predicted value (y_{NIRS}) obtained when applying the multivariate NIR model.

The way the differences are calculated will give a negative bias when the predictions are too high and a positive one when the predictions are too low compared to the reference values.

A plot of the data immediately gives an overview of the correlation, the bias, the slope and the presence of obvious outliers (see Figure 1).



Key

1	45° line (ideal line with bias = 0 and slope = 1)	X	Y_{NIRS}
2	(45° – bias) line	Y	Y_{ref}
3	linear regression line		
4	outliers		
5	bias		

NOTE The outliers (key 4) have a strong influence on the calculation of the slope and should be removed if the results are to be used for adjustments.

Figure 1 — Scatter plot for a validation set, $y_{ref} = f(a + b \times y_{NIRS})$

7.3 Bias

Most of the time, a bias or systematic error is observed with NIR models. Bias can occur due to several causes: new samples of a type not previously seen by the model, drift of the instrument, drift in wet chemistry, changes in the process, in the sample preparation, etc.

With n, the number of independent samples, the bias (or offset) is the mean difference and can be defined by Formula (2):

$$\bar{e} = \frac{1}{n}\sum_{i=1}^{n} e_i \tag{2}$$

where e_i is the residual as defined by Formula (1) resulting in Formula (3):

$$\bar{e} = \frac{1}{n}\left(\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \hat{y}_i\right) = \bar{y} - \bar{\hat{y}} \tag{3}$$

where

y_i is the i^{th} reference value;

$\hat{y}_i$ is the i^{th} predicted value;

$\bar{y}$ is the mean of the reference values;

$\bar{\hat{y}}$ is the mean of the predicted values.

The significance of the bias is checked by a t test. The calculation of the bias confidence limits (BCLs), T_b, determines the limits for accepting or rejecting formula performance on the small set of samples chosen from the new population; see Formula (4):

$$T_b = \pm\left[t_{(1-\alpha/2)} s_{SEP}\right] / \sqrt{n} \quad (4)$$

where

α is the probability of making a type I error;

t is the appropriate student's t value for a two-tailed test with degrees of freedom associated with s_{SEP} and the selected probability of a type I error;

n is the number of independent samples;

s_{SEP} is the standard error of prediction (defined in 7.5).

As an example, with n = 20, and s_{SEP} = 1, the BCLs are as in Formula (5):

$$T_b = \pm\left(2,09 \times 1\right) / \sqrt{20} = \pm 0,48 \quad (5)$$

This means that the bias tested with 20 samples shall be higher than 48 % of the standard error of prediction to be considered as different from zero.

Table 1 — Values of the t distribution with a probability α = 0,05 (5 %)

n	t values	n	t values
10	2,23	75	1,99
15	2,13	100	1,98
20	2,09	200	1,97
30	2,04	500	1,96
40	2,02	1000	1,96
50	2,01	—	—

NOTE The Excel function "TINV" can be used.[a]

[a] Excel is the trade name of a product supplied by Microsoft. This information is given for the convenience of users of this document and does not constitute an endorsement by ISO of the product named. Equivalent products may be used if they can be shown to lead to the same results.

7.4 Root mean square error of prediction (s_{RMSEP})

The s_{RMSEP} is defined by Formula (6):

$$s_{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n} e_i^{\,2}}{n}} \tag{6}$$

where

e_i is the residual of the i^{th} sample;

n is the number of independent samples.

This value can be compared with s_{SEC} and s_{SECV} (see Annex C).

S_{RMSEP} includes the random error (s_{SEP}) and the systematic error (bias). It includes also the error of the reference methods (as do s_{SEC} and s_{SECV}); see Formula (7):

$$s_{RMSEP} = \sqrt{\frac{(n-1)}{n} s_{SEP}^2 + \bar{e}^2} \tag{7}$$

where

n is the number of independent samples;

s_{SEP} is the standard error of prediction (defined in 7.5);

$\bar{e}$ is the bias or systematic error.

There is no direct test for S_{RMSEP}. This is the reason to separate the systemic error (bias or $\bar{e}$) and the random error s_{SEP}.

7.5 Standard error of prediction (s_{SEP})

The standard error of prediction (s_{SEP}), i.e. standard deviation of the residuals, which expresses the accuracy of routine NIR results corrected for the mean difference (bias) between routine NIR and reference method, can be calculated by using Formula (8):

$$s_{SEP} = \sqrt{\frac{\sum_{i=1}^{n} (e_i - \bar{e})^2}{n-1}} \tag{8}$$

where

n is the number of independent samples;

e_i is the residual of the i^{th} sample;

$\bar{e}$ is the bias or systematic error.

The s_{SEP} should be related to the s_{SEC} (respectively, s_{SECV}; see Annex C) to check the validity of the calibration model for the selected validation set.

The unexplained error confidence limits (UECLs), T_{UE}, are calculated from an *F*-test (ratio of two variances) (see Reference [18] and Table 2). See Formula (9):

$$T_{UE} = s_{SEC}\sqrt{F_{(\alpha,v,M)}} \tag{9}$$

where

s_{SEC} is the standard error of calibration (see Annex C);

α is the probability of making a Type I error;

v is $n - 1$ numerator degrees of freedom associated with s_{SEP} of the test set;

n is the number of samples in the validation process;

M is $n_c - p - 1$ denominator degrees of freedom associated with s_{SEC} (standard error of calibration)

[n_c is the number of calibration samples, p is the number of terms or PLS factors of the model or weights in the case of ANN (see Annex C). In ANN, weights are all unknown parameters in the model].

NOTE 1 s_{SEC} can be replaced by s_{SECV}, which is a better statistic than s_{SEC}; very often, s_{SEC} is too optimistic; ($s_{SEP} > s_{SECV} > s_{SEC}$).

EXAMPLE Where $n = 20$, $\alpha = 0{,}05$, $M = 100$, and $s_{SEP} = 1$, gives the following value: $T_{UE} = 1{,}30$.

With 20 samples, a s_{SEP} that is up to 30 % larger than the s_{SEC} can be accepted.

NOTE 2 The Excel[1)] function "FINV" can be used.

The *F*-test cannot be used to compare two calibrations on the same validation set. It needs (as here) two independent sets to work. Another test shall be used to compare two or more models on the same data set.

1) Excel is the trade name of a product supplied by Microsoft. This information is given for the convenience of users of this document and does not constitute an endorsement by ISO of the product named. Equivalent products may be used if they can be shown to lead to the same results.

Table 2 — F values and squared root of the F values in function of the degrees of freedom of the numerator associated with s_{SEP} and of the denominator associated with s_{SEC}

	$F(\alpha,v,M)$				
	df (s_{SEC})				
df (s_{SEP})	50	100	200	500	1 000
10	2,03	1,93	1,88	1,85	1,84
11	1,99	1,89	1,84	1,81	1,80
12	1,95	1,85	1,80	1,77	1,76
13	1,92	1,82	1,77	1,74	1,73
14	1,89	1,79	1,74	1,71	1,70
15	1,87	1,77	1,72	1,69	1,68
16	1,85	1,75	1,69	1,66	1,65
17	1,83	1,73	1,67	1,64	1,63
18	1,81	1,71	1,66	1,62	1,61
19	1,80	1,69	1,64	1,61	1,60
29	1,69	1,58	1,52	1,49	1,48
49	1,60	1,48	1,42	1,38	1,37
99	1,53	1,39	1,32	1,28	1,26
199	1,48	1,34	1,26	1,21	1,19
499	1,46	1,31	1,22	1,16	1,13
999	1,45	1,30	1,21	1,14	1,11

	$\sqrt{F(\alpha,v,M)}$				
	df (s_{SEC})				
df (s_{SEP})	50	100	200	500	1 000
10	1,42	1,39	1,37	1,36	1,36
11	1,41	1,37	1,36	1,34	1,34
12	1,40	1,36	1,34	1,33	1,33
13	1,39	1,35	1,33	1,32	1,32
14	1,38	1,34	1,32	1,31	1,30
15	1,37	1,33	1,31	1,30	1,29
16	1,36	1,32	1,30	1,29	1,29
17	1,35	1,31	1,29	1,28	1,28
18	1,30	1,31	1,29	1,27	1,27
19	1,34	1,30	1,28	1,27	1,26
29	1,30	1,26	1,23	1,22	1,22
49	1,27	1,22	1,19	1,17	1,17
99	1,24	1,18	1,15	1,13	1,12
199	1,22	1,16	1,12	1,10	1,09
499	1,21	1,14	1,11	1,08	1,07
999	1,20	1,14	1,10	1,07	1,05

NOTE 1 See explanation to Formula (9).

NOTE 2 df is the degree of freedom; $n - 1$ for s_{SEP}; and $n_c - p - 1$ for s_{SEC}.

7.6 Slope

The slope, b, of the simple regression $y = a + b \cdot \hat{y}$ is often reported in the NIR reports and publications.

Notice that the slope shall be calculated with the reference values as the dependent variable and the predicted NIR values as the independent variable, if the calculated slope is intended to be used for adjustment of NIR results (like in the case of the inverse multivariate regression used to build the prediction model).

From the least squares fitting, slope and intercept are calculated by Formula (10) and Formula (11), respectively:

$$b = \frac{s_{\hat{y}y}}{s^2_{\hat{y}}} \tag{10}$$

where

$s_{\hat{y}y}$ is the covariance between reference and predicted values;

$s^2_{\hat{y}}$ is the variance of the n predicted values.

$$a = \bar{y} - b\bar{\hat{y}} \tag{11}$$

where

$\bar{y}$ is the mean of the reference values;

b is the slope;

$\bar{\hat{y}}$ is the mean of the predicted values.

As for the bias, a t test can be calculated to check the hypothesis that $b = 1$ as in Formula (12):

$$t_{obs} = |b-1| \cdot \sqrt{\frac{s^2_{\hat{y}} \cdot (n-1)}{s^2_{res}}} \tag{12}$$

where

$s^2_{\hat{y}}$ is the variance of the n predicted values;

n is the number of independent samples;

s_{res} is the residual standard deviation.

The residual standard deviation, s_{res}, is defined in Formula (13):

$$s_{res} = \sqrt{\frac{\sum_{i=1}^{n} \left[y_i - (a + b\hat{y}_i) \right]^2}{n-2}} \tag{13}$$

where

n is the number of independent samples;

a is the intercept (see Formula (11));

b is the slope (see Formula (10));

y_i is the i^{th} reference value;

$\hat{y}_i$ is the i^{th} predicted value obtained when applying the multivariate NIR model.

NOTE s_{res} is like s_{SEP} when the predicted values are corrected for slope and intercept. Be aware to not confuse bias and intercept. See also Figure 1.

The bias equals the intercept only when the slope is exactly one.

The slope, b, will be considered as different from 1 when Formula (14) applies:

$$t_{obs} \geq t_{(1-\alpha/2)} \tag{14}$$

where

t_{obs} is the observed t value, calculated according to Formula (12);

$t_{(1-\alpha/2)}$ is the t value obtained from Table 1 for a probability of α = 0,05 (5 %).

A too-narrow range or uneven distribution will lead to useless correction of the slope even when the s_{SEP} is correct. The slope can only be adjusted when the validation set covers a large part of the calibration range.

EXAMPLE For n = 20 samples with a residual standard deviation [see Formula (13)] of 1, a standard deviation of the predicted values of $S_{\hat{y}} = 2$ and a calculated slope of b = 1,2, the observed t_{obs} value is 1,7 and then the slope is not significantly different from 1 as the t value (see Table 1) for n = 20 samples is 2,09. If the slope is 1,3, the t_{obs} value is 2,6 and then the slope is significantly different than 1.

8 Sampling

Sampling is not part of the method specified in this document.

NOTE Recommended sampling procedures are given in ISO 6497 and ISO 24333.

It is important that the laboratory receives a sample which is truly representative and has not been damaged or changed during transport or storage.

9 Procedure

9.1 Preparation of test sample

All laboratory samples should usually be kept under conditions that will not change the composition of the sample from the time of sampling to the time of commencing the procedure.

The preparation of samples for routine measurements needs to be made in the same way as the preparation of the validation samples. It is necessary to apply standard conditions.

Before the analysis, the sample should be taken in such a way as to obtain a sample representative of the material to be analysed.

For specific procedures, see specific NIR standards.

Guidelines for specific NIR standards are given in Annex A.

9.2 Measurement

Follow the instructions of the instrument manufacturer/supplier.

The prepared sample should reach a temperature within the range included in the validation.

9.3 Evaluation of result

For the routine results to be valid, they shall be within the range of the calibration model used.

Results obtained on samples detected as spectral outliers cannot be regarded as reliable.

If multiple measurements are made on the same sample, calculate the arithmetic mean if the repeatability conditions (see 12.1) are observed.

For the expression of results, refer to specific NIR standards.

10 Checking instrument stability

10.1 Control sample

At least one control sample should be measured at least once per day to check instrument hardware stability and to detect any malfunction. Knowledge of the true concentration of the analyte in the control sample is not necessary. The sample material should be stable and, as far as possible, resemble the samples to be analysed. The parameter(s) measured should be stable and, as far as possible, identical to or at least biochemically close to the sample analyte. A sample is prepared as described in 9.1 and stored in such a way as to maximize the storage life. These samples are normally stable for lengthy periods but the stability should be tested in the actual cases. Control samples should be overlapped to secure uninterrupted control.

The recorded day-to-day variation should be plotted in control charts and investigated for significant patterns or trends.

10.2 Instrument diagnostics

For scanning spectrophotometers the wavelength/wavenumber (see 5.1) accuracy and precision should be checked at least once a week, or more frequently if recommended by the instrument manufacturer, and the results should be compared to specifications and requirements (see 5.1).

A similar check of the instrument noise shall be carried out weekly or at intervals recommended by the manufacturer.

10.3 Instruments in a network

If several instruments are used in a network, special attention shall be given to standardization of the instruments according to the manufacturer's recommendations.

11 Running performance check of calibration

11.1 General

The suitability of the calibration for the measurement of individual samples should be checked. The outlier measures used in the calibration development and validation can be applied, e.g. Mahalanobis distance and spectral residuals. In most instruments, this is done automatically.

If the sample does not pass the test, i.e. the sample does not fit into the population of the samples used for calibration and/or validation, it cannot be determined by the prediction model, unless the model is changed. Thus, the outlier measures can be used to decide which samples should be selected for reference analysis and included in a calibration model update.

If the calibration model is found to be suitable for the measured sample, the spectrum is evaluated according to the validated calibration model.

NIR methods should be validated continuously against reference methods to secure steady optimal performance of calibrations and observance of accuracy. The frequency of checking the NIR method should be sufficient to ensure that the method is operating under steady control with respect to systematic and random deviations from the reference method. The frequency depends *inter alia* on the number of samples analysed per day and the rate of changes in sample population.

The running validation should be performed on samples selected randomly from the pool of analysed samples. It may be necessary to resort to some sampling strategy to ensure a balanced sample distribution over the entire calibration range, e.g. segmentation of concentration range and random selection of test samples within each segment or to ensure that samples with a commercially important range are covered.

The number of samples for the running validation should be sufficient for the statistics used to check the performance. For a solid validation, at least 20 samples are needed (to expect a normal distribution of variance). One can fill in the results of the independent validation set for starting the running validation. To continue with about 5 to 10 samples every week is sufficient to monitor the performance properly. Using fewer samples it is hard to take the right decision in case one of the results is outside the control limits.

11.2 Control charts using the difference between reference and NIR results

Results should be assessed by control charts, plotting running sample numbers on the abscissa and the difference between results obtained by reference and NIR methods on the ordinate; ±2 s_{SEP} (95 % probability) and ±3 s_{SEP} (99,8 % probability) may be used as warning and action limits where the s_{SEP} has been obtained on a test set collected independently of calibration samples.

If the calibration and the reference laboratories are performing as they should, then only 1 point in 20 points should plot outside the warning limits and 2 points in 1 000 points outside the action limits.

Control charts should be checked for systematic bias drifts from zero, systematic patterns and excessive variation of results. General rules applied for Shewart control charts may be used in the assessment. However, too many rules applied simultaneously may result in too many false alarms.

The following rules used in combination have proved to be useful in detection of problems:

a) one point outside either action limit;

b) two out of three points in a row outside a warning limit;

c) nine points in a row on the same side of the zero line.

Additional control charts plotting other features of the running control (e.g. mean difference between NIR and reference results) and additional rules may be applied to strengthen decisions.

In the assessment of results, it should be remembered that s_{SEP} and measured differences between NIR and reference results also include the imprecision of reference results. This contribution can be reduced to a negligible part if the imprecision of reference results is reduced to less than one third of the s_{SEP}[17].

To reduce the risk of false alarms, the control samples should be analysed independently (in different series) by both NIR spectrometry and reference methods to avoid the influence of day-to-day systematic differences in, for example, reference analyses.

If the warning limits are often exceeded and the control chart only shows random fluctuations (as opposed to trends or systematic bias), the control limits may have been based on a too optimistic s_{SEP} value. An attempt to force the results within the limits by frequent adjustments of the calibration will not improve the situation in practice. The s_{SEP} should instead be re-evaluated using the latest results.

If the calibration equations after a period of stability begin to move out of control, the calibration should be updated. Before this is done, an evaluation should be made of whether the changes could be due to changes in reference analyses, unintended changes in measuring conditions (e.g. caused by a new operator), instrument drift or malfunction, etc. In some cases, a simple adjustment of the constant term in the calibration equation may be sufficient (an example is shown in Figure B.6). In other cases, it may be necessary to run a complete re-calibration procedure, where the complete or a part of the basic calibration set is expanded to include samples from the running validation, and perhaps additional samples selected for this purpose (an example is shown in Figure B.7).

Considering that the reference analyses are in statistical control and the measuring conditions and instrument performance are unchanged, significant biases or increased s_{SEP} values can be due to changes in the chemical, biological or physical properties of the samples compared to the underlying calibration set.

Other control charts, for example, using z-scores, may be used.

12 Precision and accuracy

12.1 Repeatability

The repeatability, i.e. the difference between two individual single test results, obtained with the same method on identical test material in the same laboratory by the same operator using the same equipment within a short interval of time, which should not be exceeded in more than 5 % of cases, depends on the sample material, the analyte, sample and analyte variation ranges, method of sample presentation, instrument type and the calibration strategy used. The repeatability should be determined in each case.

12.2 Reproducibility

The reproducibility, i.e. the difference between two individual single test results, obtained on identical test material by different laboratories and by different operators at different times, which should not be exceeded in more than 5 % of cases, depends on the sample material, the analyte, sample and analyte variation ranges, method of sample presentation, instrument type and the calibration strategy used. The reproducibility should be determined in each case.

12.3 Accuracy

The accuracy, which includes uncertainty from systematic deviation from the true value on the individual sample (trueness) and uncertainty from random variation (precision), depends *inter alia* on the sample material, the analyte, sample and analyte variation ranges, method of sample presentation, instrument type and the calibration strategy used. The accuracy should be determined in each case. The reported s_{SEP} and s_{RMSEP} values also include uncertainty of reference results which may vary from case to case.

12.4 Uncertainty

Uncertainty, *Ue*, is a parameter characterizing the dispersion of values that can reasonably be attributed to the result. For NIR predicted results, the uncertainty is usually expressed by Formula (15):

$$Ue = \pm 2 s_{RMSEP} \tag{15}$$

If a multiplier of 2 is used, this can be understood to indicate limits corresponding to 95 % probability of the true value lying within the range ± *Ue*.

The s_{RMSEP} shall be determined locally.

13 Test report

The test report shall specify:

a) all information necessary for complete identification of the sample;

b) the test method used, with reference to the relevant International Standard;

c) all operating conditions not specified in this document, or regarded as optional;

d) any circumstances which may have influenced the results;

e) the test result(s) obtained;

f) the current s_{SEP} and bias (if statistically significant), estimated from running a performance test on at least 20 test samples (see Clause 11).

Annex A
(informative)

Guidelines for specific NIR standards

Specific NIR standards may be developed for specific calibrations for the determination of specific constituents and parameters in animal feeding stuff, cereals and milled cereal products by NIR spectrometry.

A specific standard should

— respect the recommendation of this document for the process of validation,

— indicate particular specifications to be met as range of concentration, accuracy and precision,

— fix the sources of variation that shall be taken into account in the development of the model, and

— specify the procedure, the calculation and the expression of results.

A specific standard is not made for a particular apparatus and calibration.

These standards should follow the ISO format and provide information regarding

— the type of samples and constituents/parameters determined followed by "near infrared spectrometry" and the calibration model(s) used in the title and the scope,

— the reference methods used for the validation under "Normative references",

— the spectroscopic principle (e.g. NIR, NIT) and calibration principle (e.g. PLS, ANN),

— the procedure(s) including preparation of the test sample(s), measurement and quality control, and

— specifications to be followed in terms of accuracy, precision and range; for example, as in Table A.1 for a validation set.

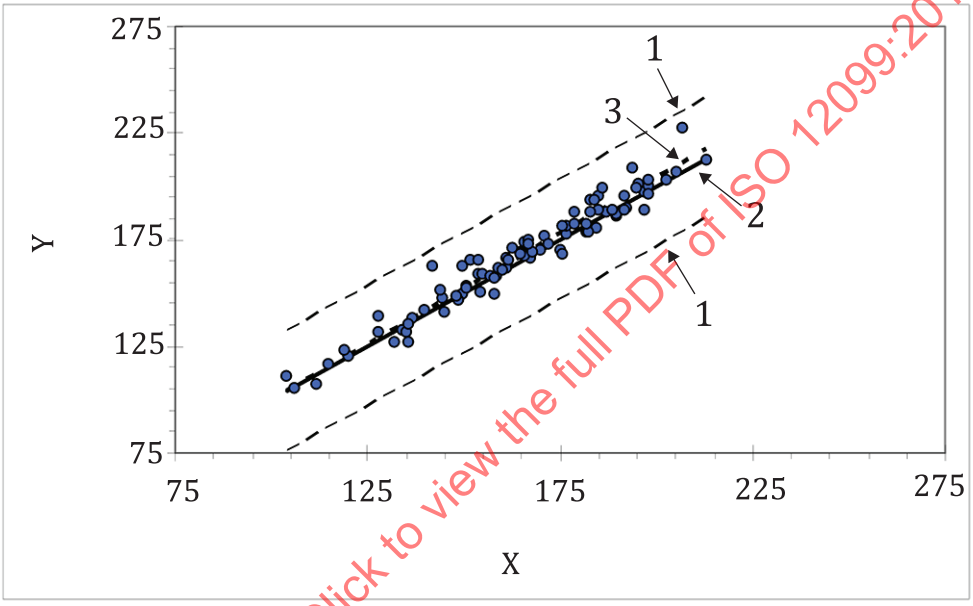
Table A.1 — Example for specifications given for a validation set

Component	Model	N	Precision (s_{SEP})	Min %	Max %	RSQ
Fat	ANN	183	0,50	2,8	12,9	0,94
Moisture	ANN	183	0,47	9,2	12,3	0,83
Protein	ANN	179	0,72	11,0	29,1	0,96
Fibre	ANN	123	1,11	0,5	18,0	0,90
Starch	PLS	113	1,80	7,8	50,2	0,92

Annex B
(informative)

Examples of outliers and control charts

Figure B.1 shows the determination of crude protein in forages, which is an example with no outliers. The results were obtained on an independent test set of 95 samples using the developed calibration equation: s_{SEP} = 4,02; s_{RMSEP} = 6,05; Slope, b = 1,04.

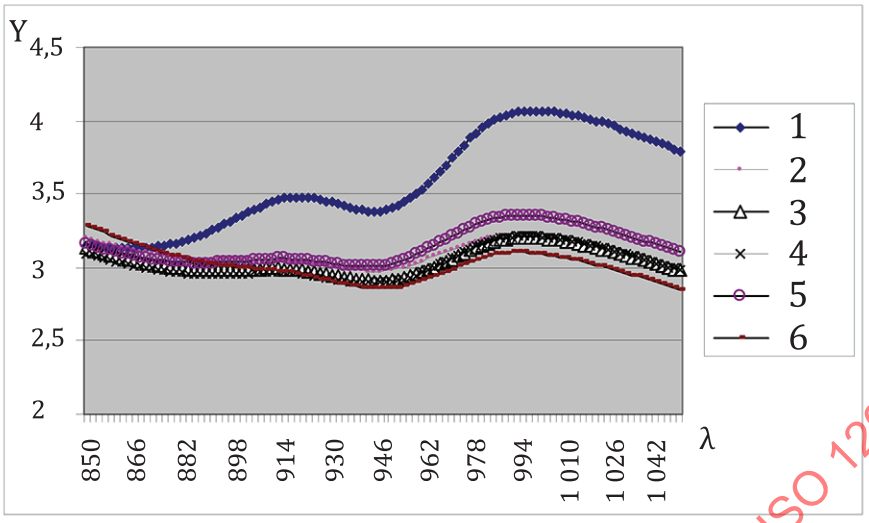


Key

1	±3 s_{SEP} limits,	X	NIR
2	45 degree line (ideal line with slope = 1 and bias = 0)	Y	reference
3	regression line		

Figure B.1 — No outliers

Figure B.2 shows an absorbance spectra with X-outlier. The series 1 (upper) spectrum indicates a spectral outlier.

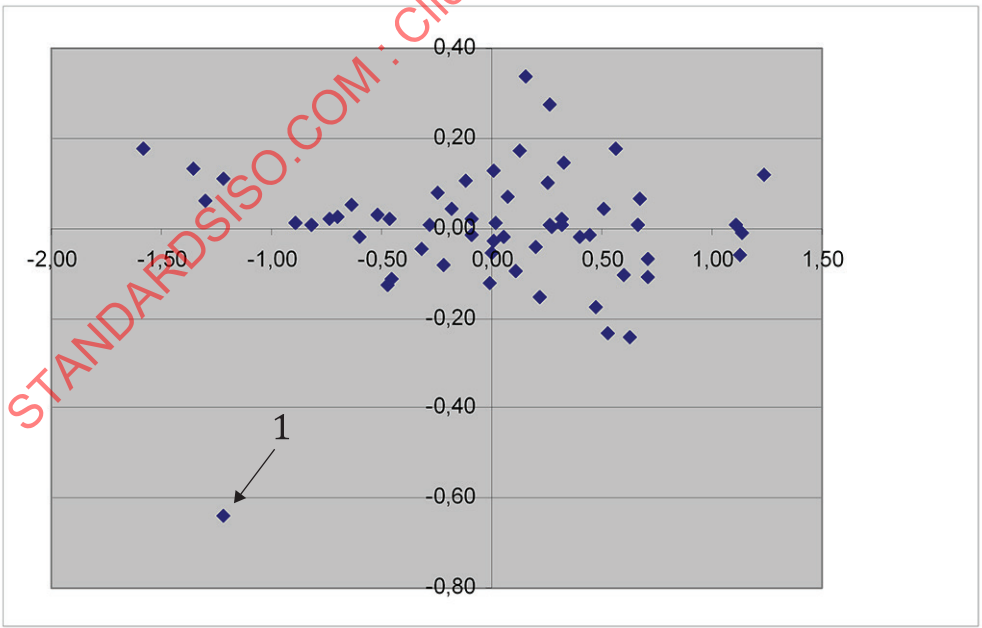


Key

1	series 1	5	series 5
2	series 2	6	series 6
3	series 3	λ	wavelength (nm)
4	series 4	Y	absorbance

Figure B.2 — Absorbance spectra with X-outlier

Figure B.3 shows a PCA score plot with X-outlier (1).

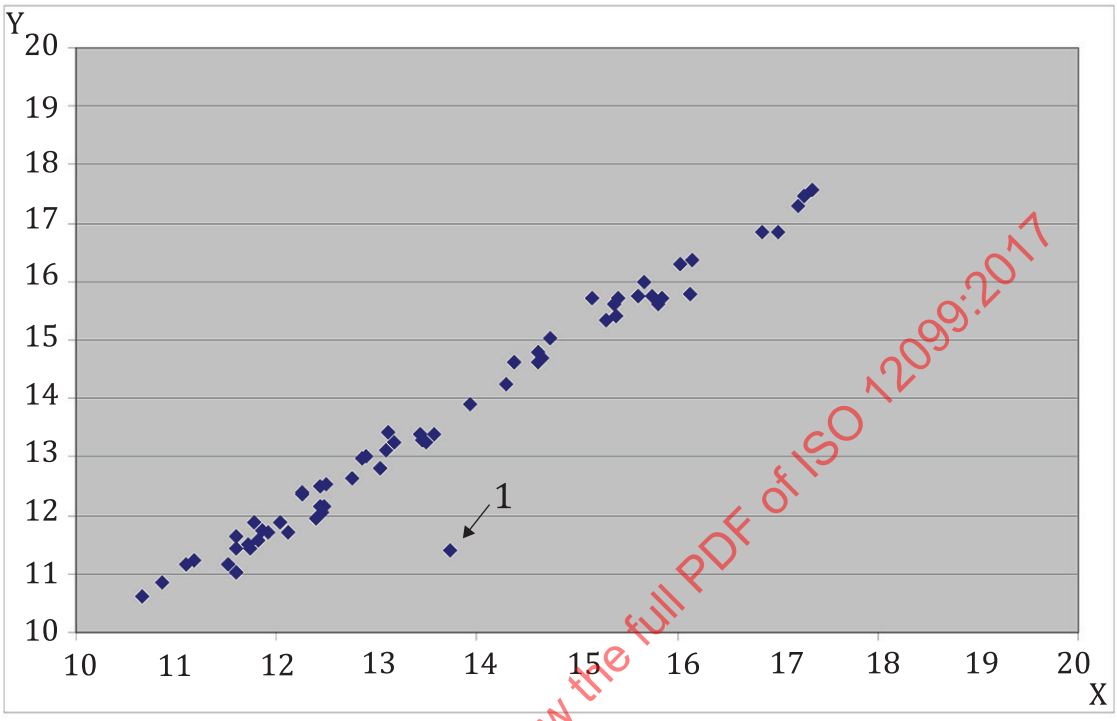


Key

1 outlier

Figure B.3 — PCA score plot with X-outlier (1)

Figure B.4 shows a scatter plot with Y-outlier (1). The plot of reference vs. predicted values (or vice versa) shows one sample (1) that strongly deviates from the other samples. If the reason for this deviation is not related to NIR data (X-outlier), this sample will be a Y-outlier due to erroneous reference data or a different relationship between reference data and spectral data.

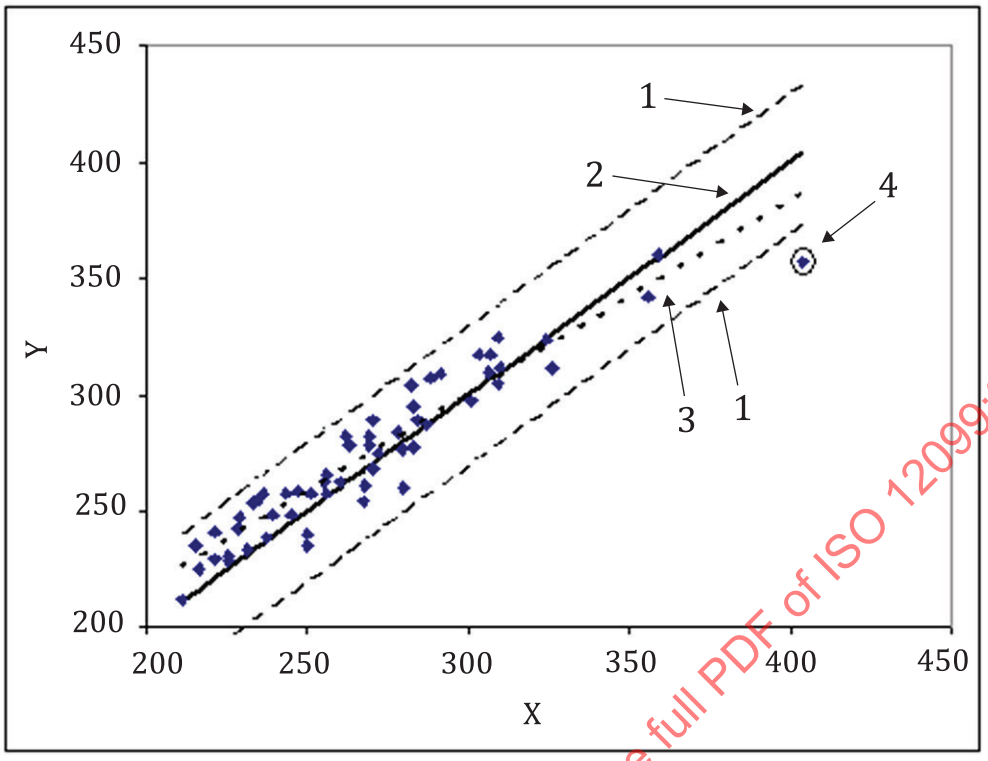


Key

1 outlier
X NIR
Y reference

Figure B.4 — Scatter plot with Y-outlier (1)

Figure B.5 shows the determination of ADF in forages with Y-outlier (1).



Key

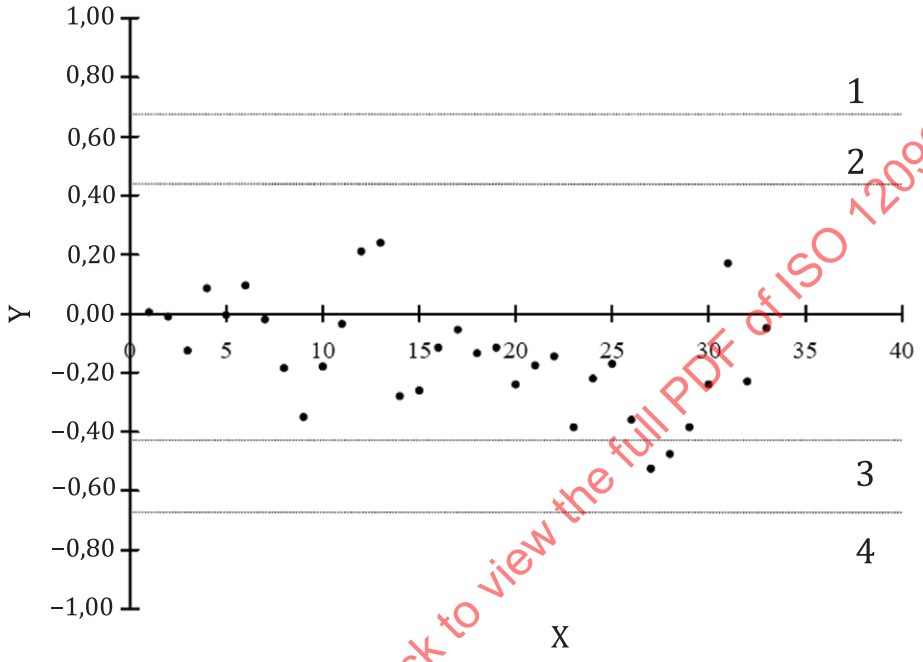
1	±3 s_{SEP} limits	X	NIR
2	45 degree line	Y	reference
3	regression line		
4	outlier		

Figure B.5 — Determination of ADF in forages with Y-outlier (1)

Figure B.6 shows a control chart for determination of percent fat in cereals. No points are outside the upper action limit (UAL) or the lower action limit (LAL). However, 9 points in row (e.g. 14 to 22) are on the same side of the zero line. That indicates a bias problem. Two points (27 and 28) out of 3 points are outside the lower warning limit (LWL) but none are outside the upper warning limit (UWL). This also indicates a bias problem. No increase in random variation is observed. The spread is still less than 3 s_{SEP}.

In conclusion, the calibration should be bias adjusted.

(Difference Reference – NIR is plotted. UAL/LAL = 3 s_{SEP} and UWL/LWL = 2 s_{SEP})



Key

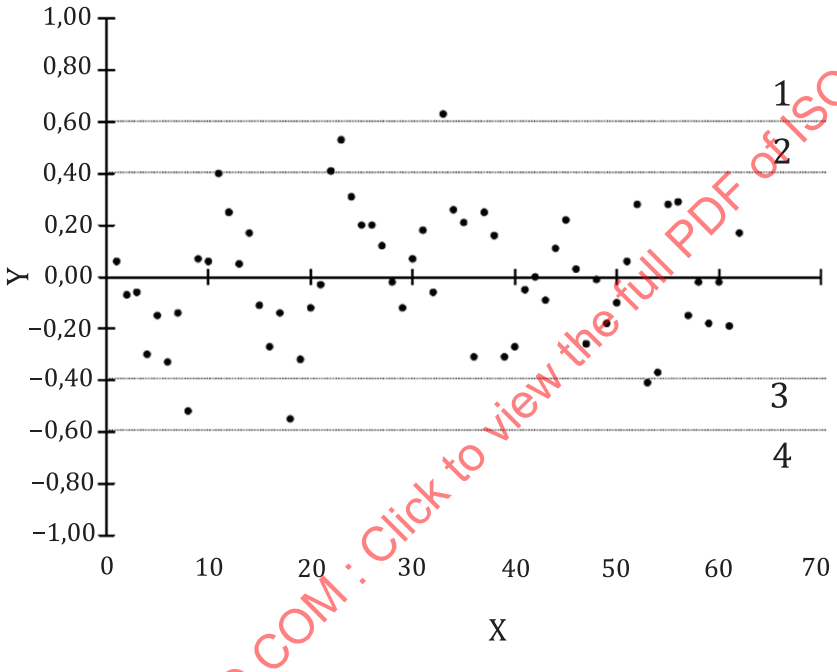
1	upper action limit (UAL)	X	run number
2	upper warning limit (UWL)	Y	reference - NIR
3	lower warning limit (LWL)		
4	lower action limit (LAL)		

Figure B.6 — Control chart for determination of percent fat in cereals

Figure B.7 shows a control chart for determination of XXXX in YYY in the range 44 % to 57 %, where recalibration was performed at point 35.

Viewing the first 34 points, one point is outside the UAL. This indicates a serious problem. Two points (22 and 23) out of 3 points are outside the UWL. Two separate points are also outside the LWL. The spread is uniform around the zero line (the 9-points rule is obeyed) but 5 point out of 34 points are outside the 95 % confidence limits (UWL, LWL) and 1 point out of 34 points is outside the 99,9 % confidence limits (UAL, LAL). This is much more than expected.

One reason for this picture could be that the s_{SEP} value behind calculation of the limits is too optimistic. This means the limits should be widened. Another reason could be that the actual samples are somewhat different from the calibration samples. To test this possibility, the calibration set was extended to include the control samples and a new calibration was developed. The performance of this calibration was clearly better, as shown by the control samples number 35 to 62.



Key

1	upper action limit (UAL)	X	run number
2	upper warning limit (UWL)	Y	reference - NIR
3	lower warning limit (LWL)		
4	lower action limit (LAL)		

Figure B.7 — Control chart for determination of XXXX in YYY (range 44 % to 57 %)