INTERNATIONAL STANDARD

ISO
12099

First edition
2010-06-15

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

<div style="border: 1px solid black;">

PDF disclaimer

This PDF file may contain embedded typefaces. In accordance with Adobe's licensing policy, this file may be printed or viewed but shall not be edited unless the typefaces which are embedded are licensed to and installed on the computer performing the editing. In downloading this file, parties accept therein the responsibility of not infringing Adobe's licensing policy. The ISO Central Secretariat accepts no liability in this area.

Adobe is a trademark of Adobe Systems Incorporated.

Details of the software products used to create this PDF file can be found in the General Info relative to the file; the PDF-creation parameters were optimized for printing. Every care has been taken to ensure that the file is suitable for use by ISO member bodies. In the unlikely event that a problem relating to it is found, please inform the Central Secretariat at the address given below.

</div>

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.

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 12099 was prepared by the European Committee for Standardization (CEN) Technical committee TC 327, *Animal feeding stuffs — Methods of sampling and analysis*, in collaboration with ISO Technical Committee TC 34, *Food products*, Subcommittee SC 10, *Animal feeding stuffs*, in accordance with the Agreement on technical cooperation between ISO and CEN (Vienna agreement).

Introduction

This International Standard has been drafted using, as a basis, ISO 21543│IDF 201[15], 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 International Standard gives guidelines for the determination by near infrared spectroscopy of constituents such as moisture, fat, protein, starch, and crude fibre as well as 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   Terms and definitions

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

2.1
near infrared instrument
NIR instrument
apparatus which, when used under specified conditions, predicts **constituent contents** (2.3) and **technological parameters** (2.4) in a matrix through relationships to absorptions in the near infrared range

NOTE   In the context of this International Standard, the matrices are animal feeding stuffs, cereals and milled cereal products.

2.2
animal feeding stuff
any substance or product, including additives, whether processed, partially processed or unprocessed, intended to be used for oral feeding to animals

EXAMPLES   Raw materials, fodder, animal flour, mixed feed and other end products, and pet food.

2.3
constituent content
mass fraction of substances determined using the appropriate, standardized or validated chemical method

NOTE 1   The mass fraction is often expressed as a percentage.

NOTE 2   Examples of constituents determined include moisture, fat, protein, crude fibre, neutral detergent fibre, and acid detergent fibre. For appropriate methods, see, e.g., References [1] to [16].

2.4
technological parameter
property or functionality of a matrix that can be determined using the appropriate standardized or validated method(s)

EXAMPLE   Digestibility.

NOTE 1   In the context of this International Standard, the matrices are animal feeding stuffs, cereals and milled cereal products.

NOTE 2   It is possible to develop and validate NIR methods for other parameters and matrices than listed, as long as the procedure from this International Standard is observed. The measuring units of the parameters determined have to follow the units used in the reference methods.

3   Principle

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

4   Apparatus

4.1   Near-infrared instruments, based on diffuse reflectance or transmittance measurement covering the NIR wavelength region, 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 or wavenumber accuracy and wavelength or 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 pathlength (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. In reflectance measurements, a quartz window or other appropriate material to eliminate drying effects should preferably cover the interacting sample surface layer.

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

NOTE   Changes in grinding or milling conditions can influence NIR measurements.

5   Calibration and initial validation

5.1   General

The instrument has to be calibrated before use. 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, for example, Reference [17] and appropriate manufacturers' manuals. For the validation, it is important to have a sufficient number of representative samples, covering variations such as:

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

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

c)   processing techniques and conditions;

d)   storage conditions;

e)   sample and instrument temperature;

f)   instrument variations (differences between instruments).

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

5.2   Reference methods

Internationally accepted reference methods for determination of moisture, fat, protein, and other constituents and parameters should be used. See References [1] to [16] for examples.

The reference method used for calibration should be in statistical control, i.e. for any sample, the variability should consist of random variations of a reproducible system. It is essential to know the precision of the reference method.

5.3   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 the purpose of validation, samples are not to be regarded as outliers if:

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

b)   they are within the spectral variation of the calibration samples, e.g. as estimated by Mahalanobis distance;

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

d)   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. 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.

5.4   Validation of calibration models

5.4.1   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, at least 10 samples are needed; for the determination of standard error of prediction (SEP, see 6.5) at least 20 samples are needed. Validation shall be carried out for each sample type, constituent or parameter, and temperature. The calibration is valid only for the variations, i.e. sample types, range and temperature, used in the validation.

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 6.5) and the residual plot of data corrected for mean systematic error (bias) is examined for outliers, i.e. samples with a residual exceeding $\pm\, 3s_{SEP}$.

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

NOTE   What is acceptable depends on such criteria as the performance of the reference method, the range covered, and the purpose of the analysis and is up to the parties involved to decide.

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

5.4.2 Bias correction

The data are also examined for 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 6.3) in the calibration equation.

5.4.3 Slope adjustment

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

Adjusting the slope or 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.

5.4.4 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 performed. 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.

5.5 Changes in measuring and instrument conditions

Unless additional calibration is performed, a local validation of an 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, slope or intercept adjustments to calibration equations. In many cases, it is necessary to standardize the two instruments against each other before calibration equations can be transferred (Reference [17]). 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.

6 Statistics for performance measurement

6.1 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 is new batches; in agriculture, it is a new crop or a new experiment location.

This set of samples shall be carefully analysed following the reference methods. Care is essential in analysing validation samples 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.

6.2 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 as:

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

where

y_i is the ith reference value;

$\widehat{y}_i$ is the ith predicted value obtained when applying the multivariate NIR model.

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

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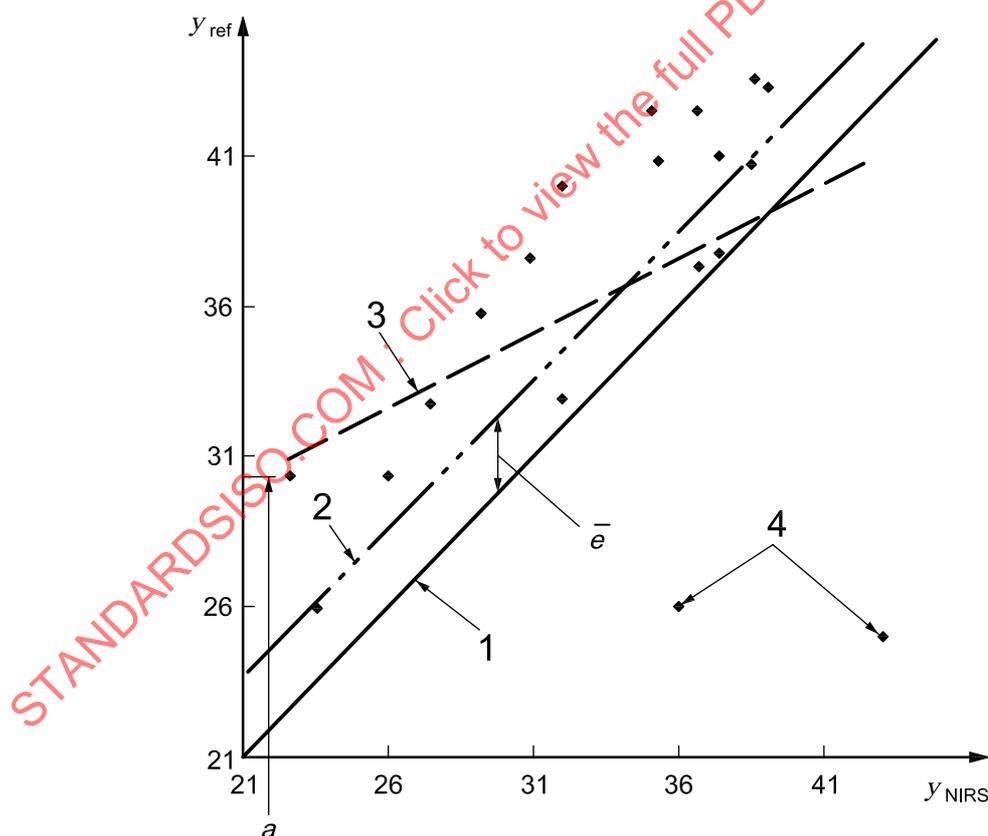
Key

1   45° line (ideal line with bias, $\overline{e}$ = 0 and slope, b = 1)

2   45° line displaced by bias, $\overline{e}$

3   linear regression line with y_{ref}-intercept, a

4   outliers

a intercept

$\overline{e}$ bias

y_{NIRS} near infrared spectroscopy predicted values

y_{ref} reference value

NOTE   The outliers 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($a + by_{NIRS}$)

6.3 The bias

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

With the number of independent samples, n, the bias (or offset) is the mean difference, $\bar{e}$, and can be defined as:

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

where e_i is the residual as defined in Equation (1), or

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

where

$\quad y_i \quad$ is the ith reference value;

$\quad \hat{y}_i \quad$ is the ith predicted value obtained when applying the multivariate NIR model;

and

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

$\quad \bar{y} \quad$ is the mean of the reference 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 equation performance on the small set of samples chosen from the new population.

$$T_b = \pm \frac{t_{(1-\alpha/2)} \, s_{SEP}}{\sqrt{n}} \tag{4}$$

where

$\quad \alpha \quad$ is the probability of making a type I error;

$\quad t \quad$ is the appropriate Student t-value for a two-tailed test with degrees of freedom associated with SEP and the selected probability of a type I error (see Table 1);

$\quad n \quad$ is the number of independent samples;

$\quad s_{SEP} \quad$ is the standard error of prediction (see 6.5).

EXAMPLE   With $n = 20$, and $s_{SEP} = 1$, the BCLs are

$$T_b = \pm \frac{2{,}09 \times 1}{\sqrt{20}} = \pm 0{,}48 \tag{5}$$

This means that the bias tested with 20 samples must 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, $\alpha = 0{,}05 = 5\ \%$

n	t	n	t	n	t	n	t
5	2,57	11	2,20	17	2,11	50	2,01
6	2,45	12	2,18	18	2,10	75	1,99
7	2,36	13	2,16	19	2,09	100	1,98
8	2,31	14	2,14	20	2,09	200	1,97
9	2,26	15	2,13	30	2,04	500	1,96
10	2,23	16	2,12	40	2,02	1 000	1,96
NOTE The Excel[1]) function TINV can be used.							

6.4   Root mean square error of prediction (RMSEP)

The RMSEP, s_{RMSEP}, (C.3.6) is expressed mathematically as

$$s_{\text{RMSEP}} = \sqrt{\frac{\sum\limits_{i=1}^{n} e_i^2}{n}} \tag{6}$$

where

  e_i is the residual of the ith sample;

  n is the number of independent samples.

This value can be compared with SEC (C.3.3) and SECV (C.3.4).

RMSEP includes the random error (SEP) and the systematic error (bias). It also includes the error of the reference methods (as do SEC and SECV).

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

where

  n is the number of independent samples;

  s_{SEP} is the standard error of prediction (see 6.5);

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

There is no direct test for RMSEP. This is the reason for separating the systematic error, bias or $\overline{e}$, and the random error, SEP or s_{SEP}.

---

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 comparable results.

6.5 Standard error of prediction (SEP)

The SEP, s_{SEP}, or the 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 Equation (8):

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

where

n is the number of independent samples;

e_i is the residual of the ith sample;

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

The SEP should be related to the SEC (C.3.3) or SECV (C.3.4) 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 2 variances) (see Reference [19] and Table 2).

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

where

s_{SEC} is the standard error of calibration (C.3.3);

α is the probability of making a type I error;

$v = n - 1$ is the numerator degrees of freedom associated with SEP of the test set, in which n is the number of samples in the validation process;

$M = n_c - p - 1$ is the denominator degrees of freedom associated with SEC (standard error of calibration);

in which

n_c is the number of calibration samples,

p is the number of terms or PLS factors of the model.

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

EXAMPLE   With $n = 20$, $\alpha = 0,05$, $M = 100$, and $s_{SEC} = 1$,

$$T_{UE} = 1,30 \tag{10}$$

This means that, with 20 samples, a SEP can be accepted that is up to 30 % larger than the SEC.

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

---

2)   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.

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 is required to compare two or more models on the same data set.

Table 2 — F-values and square root of the F-values as a function of the degrees of freedom of the numerator associated with SEP and of the denominator associated with SEC
[see definitions under Equation (9)]

	$F_{(\alpha:\ \nu,\ M)}$						$\sqrt{[F_{(\alpha:\ \nu,\ M)}]}$				
Degrees of freedom (SEP)	Degrees of freedom (SEC)					Degrees of freedom (SEP)	Degrees of freedom (SEC)				
	50	100	200	500	1 000		50	100	200	500	100
5	2,40	2,31	2,26	2,23	2,22	5	1,55	1,52	1,50	1,49	1,49
6	2,29	2,19	2,14	2,12	2,11	6	1,51	1,48	1,46	1,45	1,45
7	2,20	2,10	2,06	2,03	2,02	7	1,48	1,45	1,43	1,42	1,42
8	2,13	2,03	1,98	1,96	1,95	8	1,46	1,43	1,41	1,40	1,40
9	2,07	1,97	1,93	1,90	1,89	9	1,44	1,41	1,39	1,38	1,37
10	2,03	1,93	1,88	1,85	1,84	10	1,42	1,39	1,37	1,36	1,36
11	1,99	1,89	1,84	1,81	1,80	11	1,41	1,37	1,36	1,34	1,34
12	1,95	1,85	1,80	1,77	1,76	12	1,40	1,36	1,34	1,33	1,33
13	1,92	1,82	1,77	1,74	1,73	13	1,39	1,35	1,33	1,32	1,32
14	1,89	1,79	1,74	1,71	1,70	14	1,38	1,34	1,32	1,31	1,30
15	1,87	1,77	1,72	1,69	1,68	15	1,37	1,33	1,31	1,30	1,29
16	1,85	1,75	1,69	1,66	1,65	16	1,36	1,32	1,30	1,29	1,29
17	1,83	1,73	1,67	1,64	1,63	17	1,35	1,31	1,29	1,28	1,28
18	1,81	1,71	1,66	1,62	1,61	18	1,30	1,31	1,29	1,27	1,27
19	1,80	1,69	1,64	1,61	1,60	19	1,34	1,30	1,28	1,27	1,26
29	1,69	1,58	1,52	1,49	1,48	29	1,30	1,26	1,23	1,22	1,22
49	1,60	1,48	1,42	1,38	1,37	49	1,27	1,22	1,19	1,17	1,17
99	1,53	1,39	1,32	1,28	1,26	99	1,24	1,18	1,15	1,13	1,12
199	1,48	1,34	1,26	1,21	1,19	199	1,22	1,16	1,12	1,10	1,09
499	1,46	1,31	1,22	1,16	1,13	499	1,21	1,14	1,11	1,08	1,07
999	1,45	1,30	1,21	1,14	1,11	999	1,20	1,14	1,10	1,07	1,05

6.6 Slope

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

Notice that the slope must 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, the slope is calculated as:

$$b = \frac{s_{\hat{y}y}}{s_{\hat{y}}^2} \tag{11}$$

where

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

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

The intercept is calculated as:

$$a = \overline{y} - b\overline{\hat{y}} \tag{12}$$

where

$\overline{\hat{y}}$ is the mean of the predicted values;

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

b is the slope.

As for the bias, a t-test can be calculated to check the hypothesis that $b = 1$

$$t_{\text{obs}} = |b - 1| \sqrt{\frac{s_{\hat{y}}^2 (n-1)}{s_{\text{res}}^2}} \tag{13}$$

where

n is the number of independent samples;

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

s_{res} is the residual standard deviation as defined in Equation (14).

$$s_{\text{res}} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - a + b\hat{y}_i)^2}{n-2}} \tag{14}$$

in which

n is the number of independent samples,

a is the intercept Equation (12),

b is the slope Equation (11),

y_i is the ith reference value,

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

(RSD is like the SEP when the predicted values are corrected for slope and intercept. Do not confuse bias and intercept — see also Figure 1.) The bias equals the intercept only when the slope is exactly one.

The slope, b, is considered as different from 1 when

$$t_{\text{obs}} \geqslant t_{(1-\alpha/2)}$$

where

t_{obs} is the observed t-value, calculated according to Equation (13);

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

Too narrow a range or an uneven distribution leads to inappropriate correction of the slope even when the 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 [Equation (14)] 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 from 1.

7 Sampling

Sampling is not part of the method specified in this International Standard. Recommended sampling procedures are given in ISO 6497[5] and ISO 24333[16].

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

8 Procedure

8.1 Preparation of test sample

All laboratory samples should usually be kept under conditions that maintain the composition of the sample from the time of sampling to the time of commencing the procedure.

Samples for routine measurements should be prepared in the same way as 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.

8.2 Measurement

Follow the instructions of the instrument manufacturer or supplier.

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

8.3 Evaluation of result

To be valid, routine results shall be within the range of the calibration model used.

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

9 Checking instrument stability

9.1 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 in 8.1 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.

9.2 Instrument diagnostics

For scanning spectrophotometers, the wavelength or wavenumber (see 4.1) 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 (4.1).

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

9.3 Instruments in a network

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

10 Running performance check of calibration

10.1 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 about 5 to 10 samples every week is quite 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.

10.2 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; $\pm 2s_{SEP}$ (95 % probability) and $\pm 3s_{SEP}$ (99,8 % probability) may be used as warning and action limits where the 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 one point in 20 points should plot outside the warning limits and two 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 (see ISO 8258[7]). 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, see ISO 9622[8]) and additional rules may be applied to strengthen decisions.

In the assessment of results, it should be remembered that SEP and measured differences between NIR and reference results also include the imprecision of reference results. This contribution can be neglected if the imprecision of reference results is reduced to less than one-third of the SEP (see Reference [19]).

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 reference analyses, for example.

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 SEP value that is too optimistic. An attempt to force the results within the limits by frequent adjustments of the calibration does not improve the situation in practice. The 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 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, e.g. using z-scores, may be used.

11 Precision and accuracy

11.1 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.

11.2 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.

11.3 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 SEP and RMSEP values also include uncertainty of reference results which may vary from case to case.

12 Test report

The test report shall contain at least the following information:

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 details not specified in this International Standard (ISO 12099:2010), or regarded as optional, together with details of any incidents which may have influenced the test results;

d)   the test result(s) obtained;

e)   the current SEP and bias (if statistically significant), estimated from running a performance test on at least 20 test samples (see Clause 10).

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 stuffs, cereals and milled cereal products by NIR spectrometry.

These standards should follow the ISO format and give specific information regarding:

a) type of samples and constituents or parameters determined followed by "near infrared spectrometry" and the calibration model(s) used in the title and the scope;

b) calibration model, preferably in the form of a table, including number of samples, range, s_{SEP} validation set and RSQ for each parameter (examples are given in Tables A.1 and A.2);

c) the reference methods used for the validation under "normative references";

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

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

f) precision data as determined by an interlaboratory test according to ISO 5725-2[22].
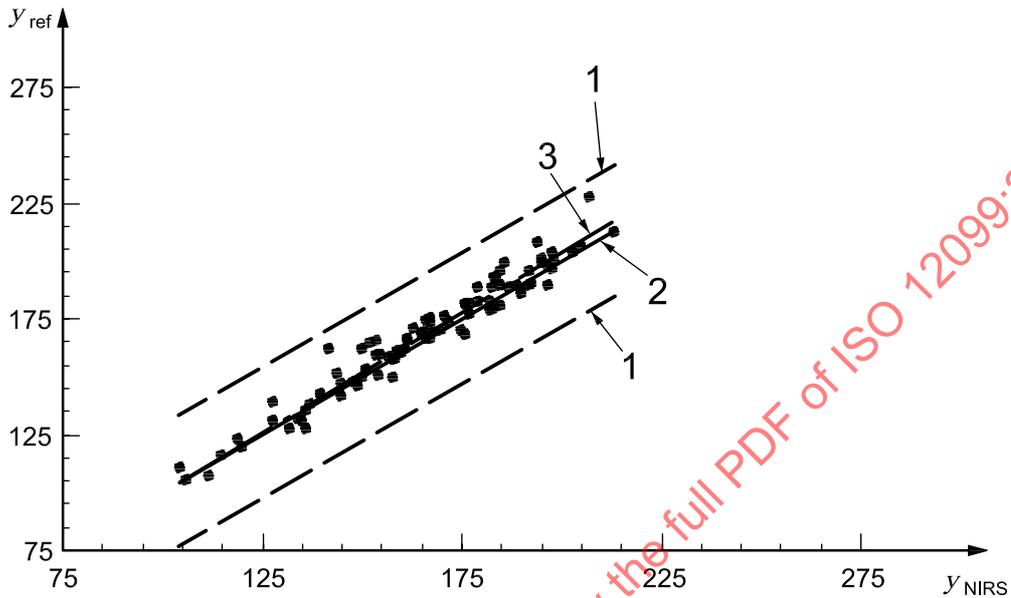
Table A.1 — Calibration set

Component	Moisture basis	Number of samples, N	Minimum content, % mass fraction	Maximum content, % mass fraction
Fat	as is	7 401	0,3	18,5
Moisture	as is	17 799	0,8	18,0
Protein	as is	17 165	6,0	74,1
Fibre	as is	2 892	0,2	26,8
Starch	as is	1 140	3,0	62,1

Table A.2 — Validation set

Component	Model	Number of samples, N	Accuracy, s_{SEP}	Minimum content, % mass fraction	Maximum content, % mass fraction	RSQ (C.3.9)
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 figures

Key

1   $\pm 3s$ limits, where s is standard deviation y_{ref} reference values

2   45° line (ideal line with slope, $b = 1$ and bias, $\bar{e} = 0$) y_{NIRS} near infrared predicted values

3   regression line

Determination of crude protein in forages: Results obtained on an independent test set (95 samples) using the developed calibration equation: standard error of prediction, $s_{SEP} = 4{,}02$; root mean square error of prediction, $s_{RMSEP} = 6{,}05$; slope, $b = 1{,}04$.

Figure B.1 — Example: No outliers
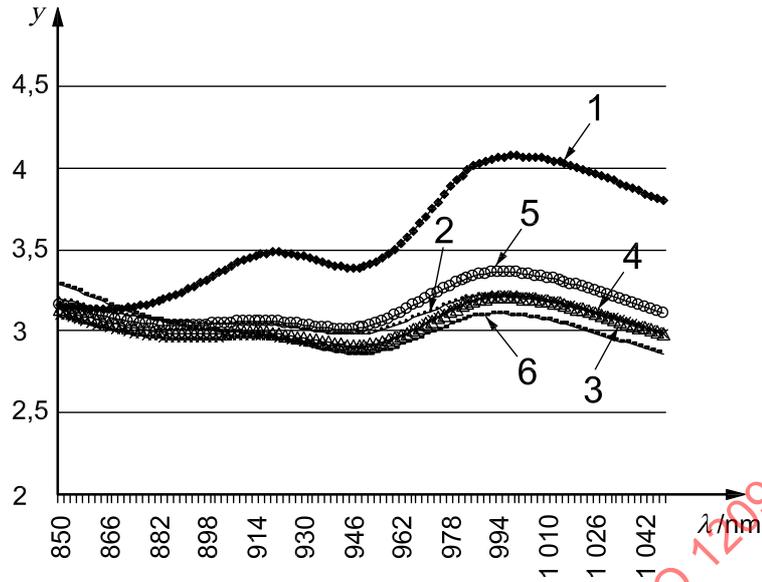
Key

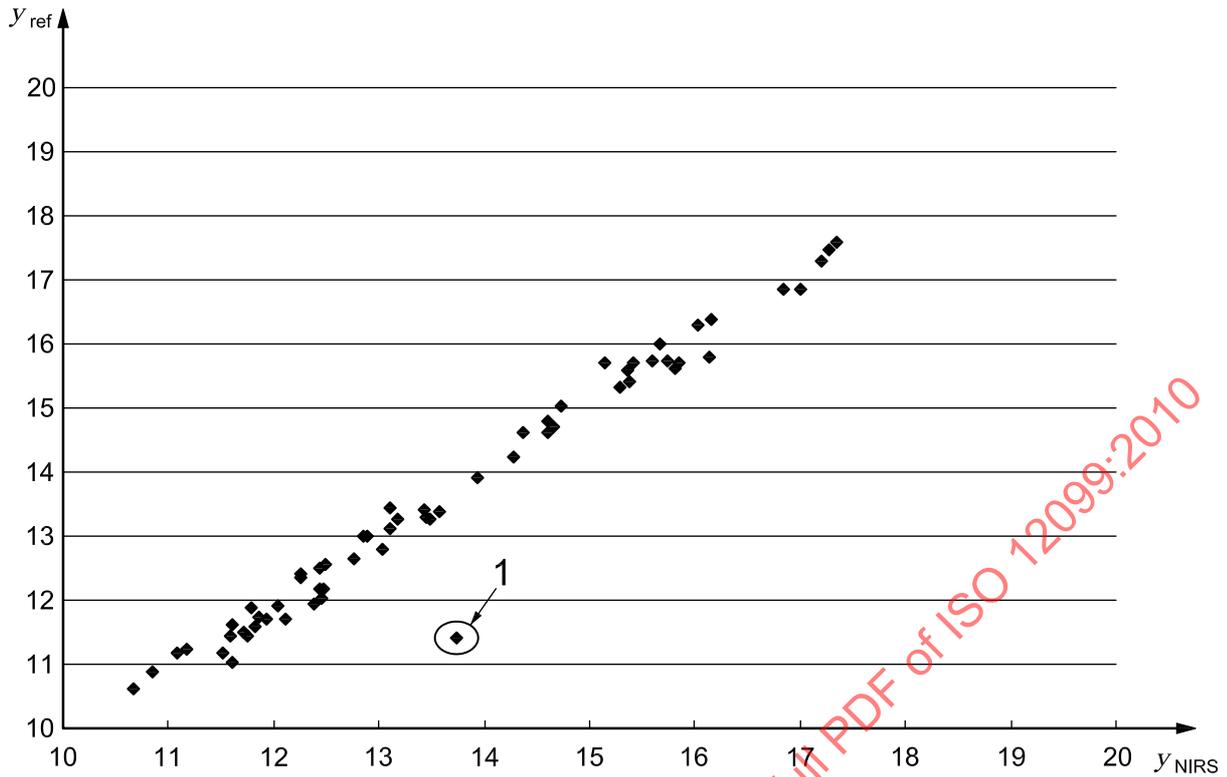|			
1	series 1, indicating a spectral outlier	5	series 5
2	series 2	6	series 6
3	series 3	y	absorbance
4	series 4	λ	wavelength

Figure B.2 — Absorbance spectra with an x-outlier
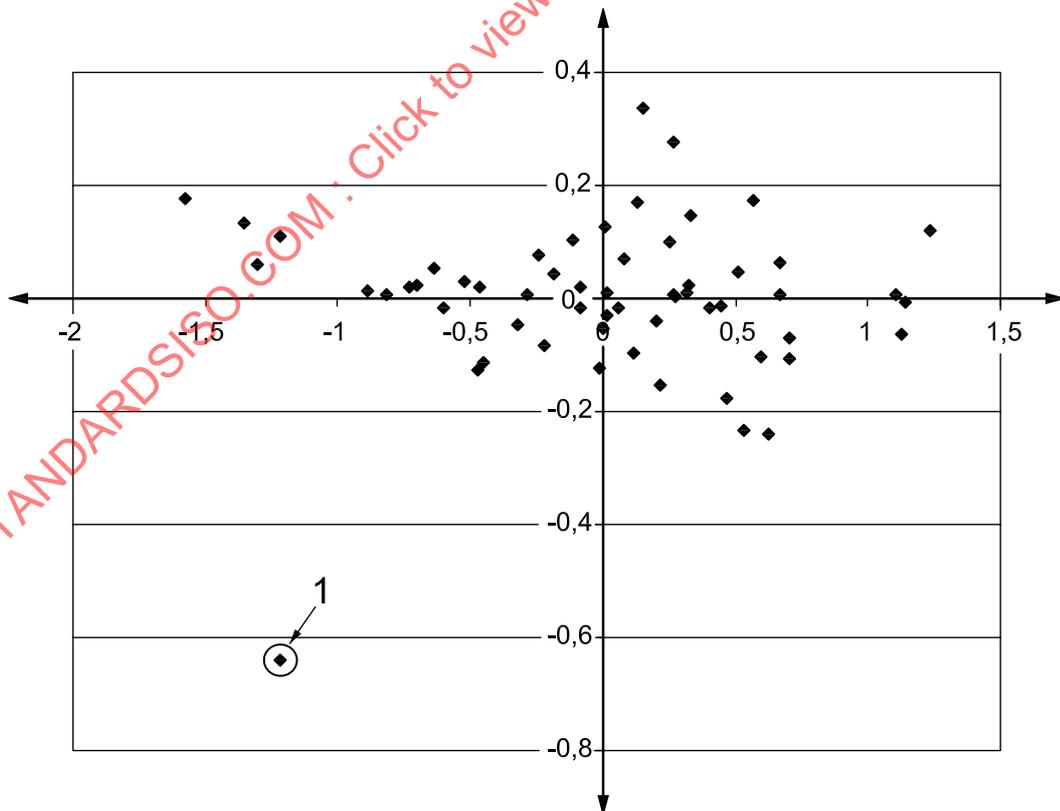
Key

1   outlier

Figure B.3 — Principal component analysis score plot with an x-outlier

Key

1   outlier

y_{ref} reference values

y_{NIRS} near infrared predicted values

The plot of reference vs predicted values (or vice-versa) shows one sample 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.

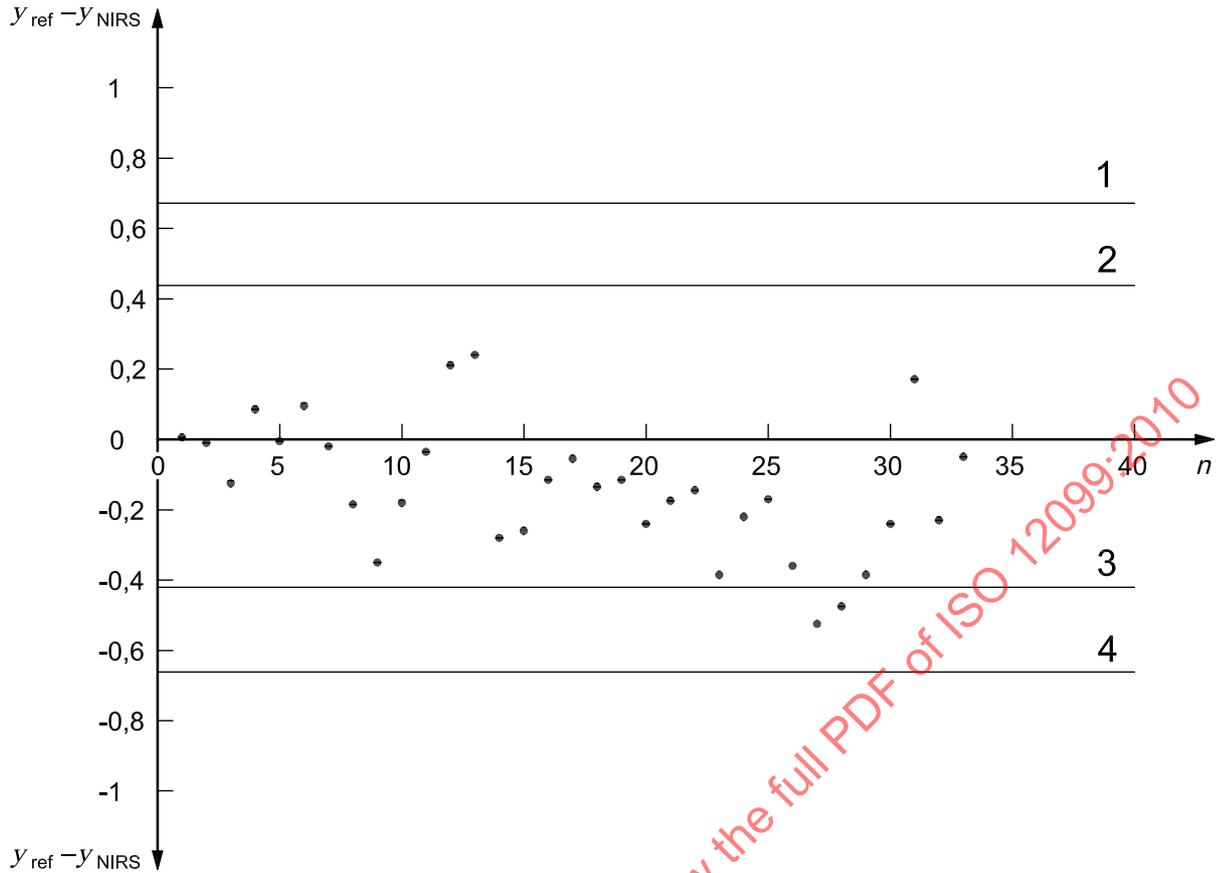**Figure B.4 — Scatter plot with a y-outlier**
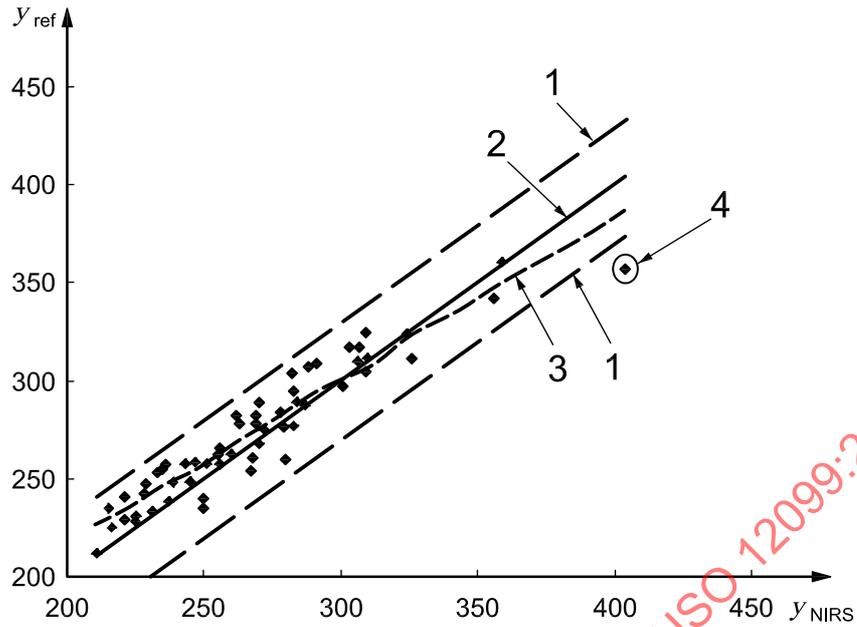
Key

1	$\pm\,3s$ limits	4	outlier	
2	45° line	y_{ref}	reference values	
3	regression line	y_{NIRS}	near infrared predicted values	

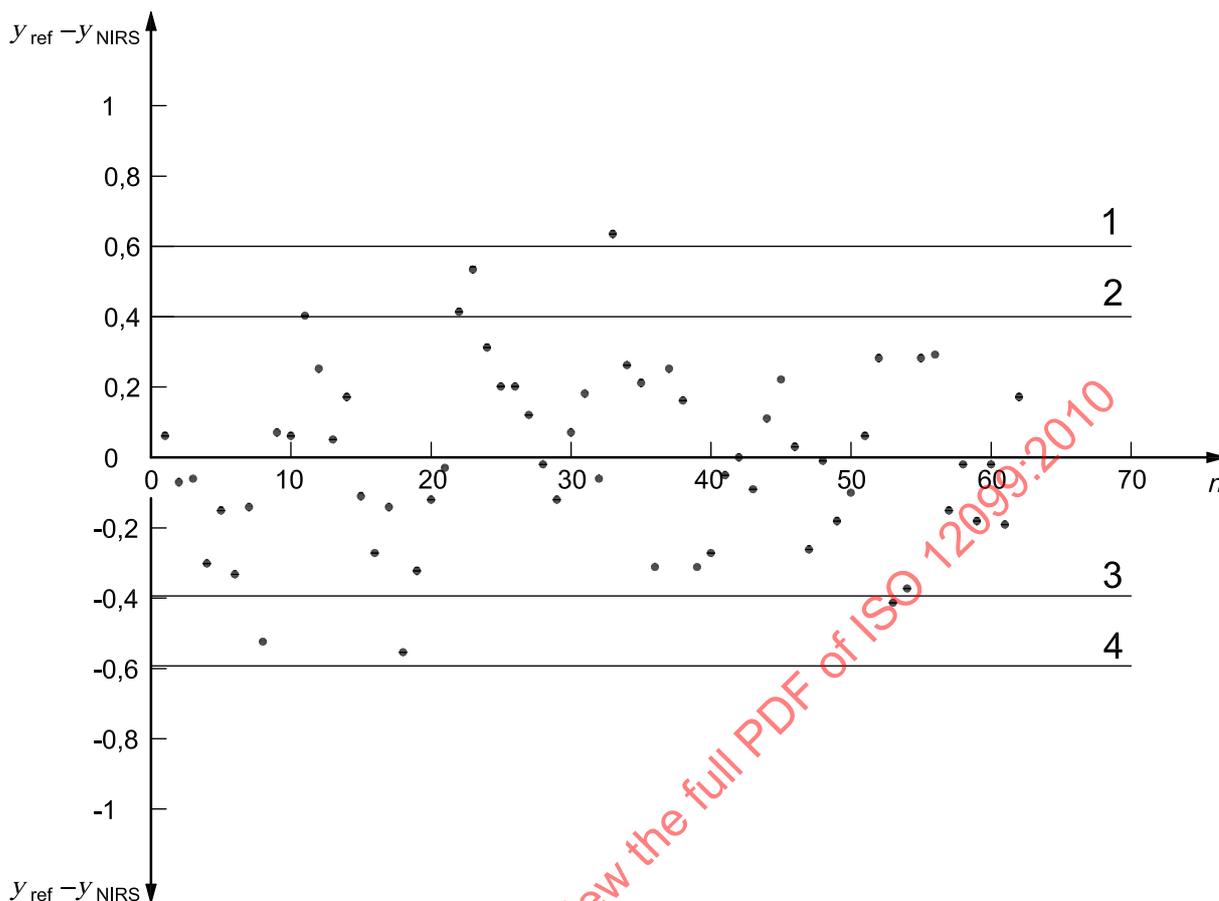**Figure B.5 — Example determination of ADF in forages with a y-outlier**

Key

1   upper action limit (UAL, $+3s_{SEP}$)

2   upper warning limit (UWL, $+2s_{SEP}$)

3   lower warning limit (LWL, $-2s_{SEP}$)

4   lower action limit (LAL, $-3s_{SEP}$)

n run number

y_{ref} reference values

y_{NIRS} near infrared predicted values

No points are outside the UAL or the LAL. However, nine points in a 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 three points are outside the LWL but none are outside the UWL. This also indicates a bias problem. No increase in random variation is observed. The spread is still less than $3s_{SEP}$.

In conclusion, the calibration should be bias adjusted.

**Figure B.6 — Example: Control chart for determination of fat content,
as a percentage mass fraction, in cereals**

Key

1   upper action limit (UAL, $+3s_{SEP}$)
2   upper warning limit (UWL, $+2s_{SEP}$)
3   lower warning limit (LWL, $-2s_{SEP}$)
4   lower action limit (LAL, $-3s_{SEP}$)
n run number
y_{ref} reference values
y_{NIRS} near infrared predicted values

Recalibration was performed at point 35.

Viewing the first 34 points, one point is outside the UAL. This indicates a serious problem. Two (22 and 23) out of three points are outside the UWL. Two separate points are also outside the LWL. The spread is uniform around the zero line (the nine points rule is obeyed) but five out of 34 points are outside the 95 % confidence limits (UWL, LWL) and one 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 SEP value behind the 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 numbers 35 to 62.

Figure B.7 — Control chart for determination of a parameter in a matrix (range 44 % to 57 %**).**

Annex C
(informative)

Supplementary terms and definitions

C.1 General

C.1.1
reference method
validated method of analysis internationally recognized by experts or by agreement between parties

NOTE 1   A reference method gives the "true value" or "assigned value" of the quantity of the measurand.

NOTE 2   Adapted from ISO 8196-1│IDF 128-1:2009[23], 3.1.2.

C.1.2
indirect method
method that measures properties that are functionally related to the parameter(s) to be determined and whose obtained signal is related to the "true" value(s) as determined by the reference method(s)

C.1.3
near infrared spectroscopy
NIRS
measurement of the intensity of the absorption of near-infrared light by a sample within the range 770 nm to 2 500 nm (12 900 cm^{-1} to 4 000 cm^{-1})

NOTE   NIRS instruments use either part of, the whole, or ranges that include this region (e.g. 400 nm to 2 500 nm). Multivariate calibration techniques are then used to relate a combination of absorbance values either to composition or to some property of the samples.

C.1.4
near infrared reflectance
NIR
type of near infrared spectroscopy where the basic measurement is the absorption of near-infrared light diffusely reflected back from the surface of a sample collected by a detector in front of the sample

C.1.5
near infrared transmittance
NIT
type of near infrared spectroscopy where the basic measurement is the absorption of near-infrared light that has travelled through a sample and is then collected by a detector behind the sample

C.1.6
NIRS network
number of near infrared instruments, operated using the same calibration models, which are usually standardized so that the differences in predicted values for a set of standard samples are minimized

C.1.7
standardization of an instrument
process whereby a group of near infrared instruments are adjusted so that they predict similar values when operating the same calibration model on the same sample(s)

NOTE   A number of techniques can be used but these can be broadly defined as either pre-prediction methods where the spectra of samples are adjusted to minimize the differences between the response of a "master" instrument and each

instrument in the group and "post-prediction" methods where linear regression is used to adjust the predicted values produced by each instrument to make them as similar as possible to those from a "master" instrument.

C.1.8
***z*-score**
performance criterion calculated by dividing the difference between the near infrared predicted result and the true or assigned value by a target value for the standard deviation, usually the standard deviation for proficiency assessment

NOTE   This is a standardized measure of laboratory bias, calculated using the assigned value and the standard deviation for proficiency assessment.

C.2 Calibration techniques

C.2.1
principal component analysis
PCA
form of data compression, which for a set of samples works solely with the x (spectral) data and finds principal components (factors) according to a rule that says that each PC expresses the maximum variation in the data at any time and is uncorrelated with any other PC

NOTE   The first PC expresses as much as possible of the variability in the original data. Its effect is then subtracted from the x data and a new PC derived again expressing as much as possible of the variability in the remaining data. It is possible to derive as many PCs as there are either data points in the spectrum or samples in the data set, but the major effects in spectra can be shown to be concentrated in the first few PCs and therefore the number of data that need to be considered is dramatically reduced.

PCA produces two new sets of variables at each stage: PC scores represent the response of each sample on each PC; PC loadings represent the relative importance of each data point in the original spectra to the PC.

PCA has many uses, e.g. in spectral interpretation, but is most widely used in the identification of spectral outliers.

C.2.2
principal component regression
PCR
technique which uses the scores on each principal component as regressors in a multiple linear regression against y values representing the composition of samples

NOTE   As each PC is orthogonal to every other PC, the scores form an uncorrelated data set with better properties than the original spectra. While it is possible to select a combination of PCs for regression based on how well each PC correlates to the constituent of interest, most commercial software forces the regression to use all PCs up to the highest PC selected for the model ("the top down approach").

When used in NIRS, the regression coefficients in PC space are usually converted back to a prediction model using all the data points in wavelength space.

C.2.3
partial least squares regression
PLS
form of data compression which uses a rule to derive the factors consisting of allowing each factor in turn to maximize the covariance between the y data and all possible linear combinations of the x data

NOTE   PLS is a balance between variance and correlation with each factor being influenced by both effects. PLS factors are therefore more directly related to variability in y values than are principal components. PLS produces three new variables, loading weights (which are not orthogonal to each other), loadings, and scores which are both orthogonal.

PLS models are produced by regressing PLS scores against y values. As with PCR, when used in NIRS, the regression coefficients in PLS space are usually converted back to a prediction model using all the data points in wavelength space.

C.2.4
multiple linear regression
MLR
technique using a combination of several x variables to predict a single y variable

NOTE In NIRS, the x values are either absorbance values at selected wavelengths in the NIR or derived variables such as PCA or PLS scores.

C.2.5
artificial neural network
ANN
non-linear modeling technique loosely based on the architecture of biological neural systems

NOTE The network is initially "trained" by supplying a data set with several x (spectral or derived variables such as PCA scores) values and reference y values. During the training process, the architecture of the network may be modified and the neurons assigned weighting coefficients for both inputs and outputs to produce the best possible predictions of the parameter values.

Neural networks require a lot of data in training.

C.2.6
multivariate model
any model where a number of x values are used to predict one or more y variables

C.2.7
outlier
member of a set of values which is inconsistent with the other members of that set

[ISO 5725-1:1994[21], 3.21]

NOTE For NIRS data, outliers are points in any data set that can be shown statistically to have values that lie well outside an expected distribution. Outliers are normally classified as either x- (spectral) outliers or y- (reference data) outliers.

C.2.8
x-outlier
outlier related to the NIR spectrum

NOTE An x-outlier can arise from a spectrum with instrumental faults or from a sample type that is radically different from the other samples or in prediction, a sample type not included in the original calibration set.

C.2.9
y-outlier
outlier related to error in the reference data, e.g. an error in transcription or in the value obtained by the reference laboratory

C.2.10
leverage
measure of how far a sample lies from the centre of the population space defined by a model

NOTE Samples with high leverage have high influence on the model. Leverage is calculated by measuring the distance between a projected point and the centre of the model.

C.2.11
Mahalanobis distance
global h-value
distance in PC space between a data point and the centre of the PC space

NOTE 1 Mahalanobis distance is a non linear measurement. In PC space, a set of samples usually form a curve shaped distribution. The ellipsoid that best represents the probability distribution of the set can be estimated by building