

INTERNATIONAL STANDARD

**ISO
21543**

**IDF
201**

First edition
2006-09-01

Milk products — Guidelines for the application of near infrared spectrometry

*Produits laitiers — Lignes directrices pour l'application de la
spectrométrie dans le proche infrarouge*

STANDARDSISO.COM : Click to view the full PDF of ISO 21543:2006



Reference numbers
ISO 21543:2006(E)
IDF 201:2006(E)

© ISO and IDF 2006

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. Neither the ISO Central Secretariat nor the IDF accepts any 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 and IDF national committees. In the unlikely event that a problem relating to it is found, please inform the ISO Central Secretariat at the address given below.

STANDARDSISO.COM : Click to view the full PDF of ISO 21543:2006

© ISO and IDF 2006

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from either ISO or IDF at the respective address below.

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

International Dairy Federation
Diamant Building • Boulevard Auguste Reyers 80 • B-1030 Brussels
Tel. + 32 2 733 98 88
Fax + 32 2 733 04 13
E-mail info@fil-idf.org
Web www.fil-idf.org

Published in Switzerland

Contents

Page

Foreword.....	iv
1 Scope	1
2 Terms and definitions.....	1
3 Principle	1
4 Reagents	1
5 Apparatus	2
6 Calibration and initial validation	2
6.1 Selection of calibration samples	2
6.2 Reference analyses and NIR measurements	3
6.3 Calibration	3
6.4 Outliers in calibration	4
6.5 Validation of calibration models	5
6.6 Changes in measuring and instrument conditions	6
6.7 Outlier detection	6
7 Statistics for performance measurement.....	6
7.1 Standard error of prediction (SEP) and bias.....	6
7.2 Root mean square error of prediction (RMSEP).....	7
7.3 Root mean square error of cross validation (RMSECV)	7
8 Sampling.....	7
9 Procedure	8
9.1 Preparation of test sample.....	8
9.2 Measurement.....	8
9.3 Evaluation of results.....	8
10 Checking instrument stability	9
10.1 Control sample.....	9
10.2 Instrument diagnostics	9
11 Running performance check of calibration.....	9
12 Precision and accuracy.....	10
12.1 Repeatability.....	10
12.2 Intralaboratory reproducibility	10
12.3 Accuracy.....	11
13 Test report	11
Annex A (informative) Examples of SEP and RMSEP values	12
Annex B (informative) Examples of figures	14
Bibliography	22

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 21543|IDF 201 was prepared by Technical Committee ISO/TC 34, *Food products*, Subcommittee SC 5, *Milk and milk products*, and the International Dairy Federation (IDF). It is being published jointly by ISO and IDF.

STANDARDSISO.COM : Click to view the full PDF of ISO 21543:2006

Foreword

IDF (the International Dairy Federation) is a worldwide federation of the dairy sector with a National Committee in every member country. Every National Committee has the right to be represented on the IDF Standing Committees carrying out the technical work. IDF collaborates with ISO in the development of standard methods of analysis and sampling for milk and milk products.

Draft International Standards adopted by the Action Teams and Standing Committees are circulated to the National Committees for voting. Publication as an International Standard requires approval by at least 50 % of the IDF National Committees casting a vote.

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

ISO 21543|IDF 201 was prepared by the International Dairy Federation (IDF) and Technical Committee ISO/TC 34, *Food products*, Subcommittee SC 5, *Milk and milk products*. It is being published jointly by IDF and ISO.

All work was carried out by Joint ISO-IDF Action Team on *Automated methods*, of the Standing Committee on *Quality assurance, statistics of analytical data and sampling*, under the aegis of its project leader, Mr L.K. Sørensen (DK).

Milk products — Guidelines for the application of near infrared spectrometry

1 Scope

This International Standard provides guidance on use of near infrared spectrometry in the determination of

- the total solids, fat and protein contents in cheese,
- the moisture, fat, protein and lactose contents in dried milk, dried whey and dried butter milk, and
- the moisture, fat, non-fat solids and salt contents in butter.

2 Terms and definitions

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

2.1

near infrared instrument

NIR instrument

proprietary apparatus which, when used under the conditions defined in this International Standard, estimates the mass fractions of the substances specified in Clause 1

2.2

total solids, moisture, non-fat solids, fat, protein, lactose and salt contents

mass fraction of substances determined using the method specified in this International Standard

NOTE These contents are expressed as mass fractions in percent.

3 Principle

The sample is pretreated to obtain a homogeneous test sample representing the chemical composition of the sample material. It is loaded into the sample holder of the NIR spectrometer. The absorbance at wavelengths in the near infrared region is measured and the spectral data are transformed to constituent concentrations by calibration models developed on representative samples from the population to be tested.

4 Reagents

Use only reagents of recognized analytical grade, unless otherwise specified, and distilled or demineralized water or water of equivalent purity.

4.1 Ethanol, or other appropriate solvent or detergent mixture, for cleaning re-usable sample cups.

5 Apparatus

5.1 Near-infrared (NIR) instrument, based on diffuse reflectance or transmittance measurement in the whole near infrared wavelength region of 700 nm to 2 500 nm or segments of this or at selected wavelengths.

The optical operation 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 instrument noise, wavelength accuracy and wavelength precision (for scanning spectrophotometers). The wavelength accuracy should be better than 0,5 nm and the repeatability standard deviation better than 0,02 nm.

The instrument should be equipped with a sample holder, which allows measurement of a sufficiently large sample volume or surface to eliminate any significant influence of inhomogeneity derived from the 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 recommendations 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.

The sample cup (cuvette) may be re-usable or made of disposable material.

5.2 Grinding or grating device, appropriate for preparing the sample (e.g. a food processor for semi-hard cheese).

Changes in grinding or grating conditions may influence the NIR measurements.

6 Calibration and initial validation

6.1 Selection of calibration samples

The instrument should be calibrated before being used. Because of the complex nature of near infrared spectral data, which are mainly overtones and combination bands of fundamental vibrations in the mid-infrared region, the instrument should be calibrated using a series of natural samples (often at least 120 samples).

The accuracy and robustness of calibration models are dependent on the strategies used for sample selection and calibration. Developed calibration models are only valid for samples covered by the domain of the calibration samples. The first step in calibration development is therefore to define the application (e.g. sample types and concentration ranges). When calibration samples are selected, care should be taken to ensure that all major factors affecting the accuracy of calibration are covered within the limits of the defined application area. These factors include the following:

- a) combinations and composition ranges of major and minor sample components: analytes (e.g. total solids, fat and protein) and non-analytes;
- b) seasonal, geographic and genetic effects on milk composition;
- c) processing techniques and conditions;
- d) ripening stages of cheeses;
- e) storage and storage conditions.

The accuracy of calibration is influenced by the extent of variation in the sample material and the analyte concentration range. A moderate variation is usually easier to fit than a large variation. If the required accuracy cannot be obtained by a single calibration, then the application area should be split up into static or dynamic sub-areas, each with an associated calibration, in order to fulfil the requirements. Dynamic sub-areas are used in locally weighted regression algorithms where calibration samples close in spectral space to the actual prediction sample are selected from a larger population to create a local calibration equation.

It is generally preferable that the whole calibration range be covered in a uniform way, with samples from low to high concentrations of analytes. The sample spread should also be as uniform as possible with respect to the other variables, including those mentioned above. Furthermore, the samples should be collected and measured over a certain period of time to ensure inclusion of time-dependent effects. This design will improve the ruggedness and give a more even performance of the calibration over the entire analyte concentration range.

Multivariate methods [1], [2] may be used as a tool in the selection of samples to ensure a homogeneous calibration set covering all variation in spectroscopic data induced by chemical, biological and physical factors without duplication of samples with similar information. In practice, a larger sample population is measured by NIR spectroscopy for collection of NIR data only. Then samples differing in spectral information are selected for reference analyses. Identification of differing samples may be obtained from inspection of score plots from principal component analysis (PCA) using, for example, the first three components. This may be less practical in the case of many samples. However, it is recommended always to perform a PCA and inspect score plots to obtain a visual overview of the sample set. More formal cluster analyses may be obtained using techniques based on distance measurements [2]. Further samples may be added over a period of time to this pool of selected samples using PCA space or distance measurement to identify differing samples.

6.2 Reference analyses and NIR measurements

Internationally accepted reference methods for the determination of analytes should be used. The reference method used for calibration should be in statistical control; i.e. the variability should consist of a constant system of random variations. To support assessment of outliers, it may be useful to perform replicate analyses in independent series (different analysts, different equipment, etc.).

All major variations in NIR measuring conditions that may appear in practice should be built into the calibration model. An important factor is sample temperature.

The sampling procedure used and the sample size measured by NIR spectroscopy may be critical for the accuracy obtained [3]. The test sample volume or surface interacting in measurements should be large enough to avoid sample inhomogeneity having a significant influence. Reflectance measurement at higher wavelengths normally requires a larger sample surface than transmittance measurement at shorter wavelengths because the light penetration is much less. The optimal sample size should be determined from experiments where the prepared sample material (see 9.1) is measured repeatedly after repacking of the sample cup.

Special care should be taken to avoid surface drying effects, particularly in reflectance measurements.

The NIR measurements and reference analyses should preferably be performed on the same test sample in order to eliminate effects related to sampling uncertainty. The NIR measurements and the initiation of reference analyses should also be performed with a minimum time lag (preferably less than one day). It is good practice to randomize the order in which the samples are presented for both the reference analysis and NIR measurement.

6.3 Calibration

Because NIR instruments apply different calibration systems, no specific procedure can be given for calibration. However, the person performing the calibration should be familiar with the statistical principles behind the calibration algorithm used.

The calibration may be performed using different techniques [e.g. multiple linear regression (MLR), multivariate algorithms such as partial least-squares regression (PLS), locally weighted regression (LWR) or artificial neural networks (ANN)]. The latter techniques are recommended if linearity problems between the spectral response and the constituent occur. Typically at least 120 calibration samples are needed to obtain rugged calibrations with MLR and PLS. When ANN are used for calibration, a substantially higher number of samples is required to avoid over fitting of data because ANN are very flexible functions with many parameters to be determined. Three different data sets are normally required for determining the architecture, fitting the parameters and validating the network. The concept of LWR also requires a considerably larger database from which local calibration samples can be selected.

Spectra should normally be preprocessed prior to calibration to remove or reduce the weighting of effects which are not related to the chemical absorption of light. Often used treatments are multiplicative scatter correction (MSC) [4], standard normal variate (SNV) [5], de-trending [5] and first or second derivatives [2]. The optimal transformation and other pretreatments of spectra (e.g. smoothing) should be determined from trials. Several techniques often give equivalent results. The optimal techniques should be assessed from cross validation where models are subsequently developed on parts of the data and tested on other parts [6]. Additional information may be obtained from testing on an independent test set.

An important issue is selection of the optimal number of variables (in MLR) or factors (in multivariate calibrations). If too few variables or factors are used, an under-fitted solution is obtained, which means that the model is not large enough to capture the important variability in the data. If too many variables or factors are used, an over-fitted solution may be obtained where much of the redundancy in the NIR data is modelled. Both cases can result in poor predictions on future samples. The optimal number can be determined by plotting RMSECV (see Clause 7) obtained from cross validation or RMSEP (see Clause 7) obtained from an independent test set versus the number of variables or factors (Figure B.1). Typically RMSECV (RMSEP) is large for small numbers of factors and decreases as the number increases, before it increases again when the number becomes too large. Generally, the best solution is the one giving the lowest RMSECV (RMSEP) with the fewest variables or factors.

The reference results should be plotted against predicted values obtained by cross validation. The plot should be examined for outliers. The plot should also be investigated for regions with different levels of prediction accuracy, random or systematic, which may indicate the need for more calibration samples or a segmentation of the calibration region.

6.4 Outliers in calibration

6.4.1 General

Outliers may be related to NIR data (x -outliers) or errors in reference data or samples with a different relationship between reference data and NIR data (y -outliers).

6.4.2 x -outliers

A homogeneous calibration set of spectrally similar samples is required for a robust predictive model. This can also form the basis of an outlier warning system. Any x -outliers should thus be removed before calibration. The projections of the five first PCA axes can be useful to reveal x -outliers either globally outside the population or falling in a gap in the PCA space. A more formal identification of outliers may be performed using, for example, the principle of Mahalanobis distance applied on PCA reduced data [7] or the so-called leverage [8].

Figure B.2 shows a case from practice without outliers. In Figure B.3, an x -outlier is present.

6.4.3 y -outliers

When a y -outlier is observed in the calibration set, the reference data should be checked for errors in sample identification, reference analyses, computations, data transfer, etc. However, it may be difficult to relate outliers to errors in reference analyses because the calibration step usually has to be performed at a later

stage than reference analyses, which may make it impossible to repeat analyses because of sample instability. There is no correct way to treat y -outliers, but outliers should generally be removed if the difference between NIR and reference results in cross validation exceeds three times the RMSECV (see Clause 7).

It is important to note that the removal of outliers can influence the future prediction of similar samples. Outliers should be removed as a batch before a new calibration model is created. The outlier removal step should only be performed one or two times in order not to reduce the robustness of the calibration and overestimate the accuracy. Care should be taken to preserve the optimum distribution of the calibration set when outliers are removed.

Figure B.2 shows a case from practice without outliers. In Figure B.4, a y -outlier is present.

6.4.4 Combined x - and y -outliers

Samples which are both x - and y -outliers (influential outliers) have a very strong effect on the regression equation and can be very harmful. Such outliers may give slope effects and increase the prediction error considerably.

Figure B.2 shows a case from practice without outliers. In Figures B.5 and B.6, a sample is present which is both an x - and a y -outlier.

6.5 Validation of calibration models

When calibration equations have been developed, they should be validated on an independent test set, preferably sampled after the calibration period. The test set should cover all variations in the sample population and should contain at least 25 samples. The use of cross validation in the calibration process, where subsequent parts of the calibration set are reserved for validation, can give a good estimate of the uncertainty of the method when the calibration samples are properly selected.

However, the potential risk is that cross validation may underestimate the ruggedness of the calibration and the predicted uncertainty because cross validation samples are taken from the pool of samples used for calibration.

The results obtained on the independent test set are plotted, reference against NIR and residuals against reference, to give a visual impression of the performance of the calibration. The SEP is calculated (see Clause 7) and the residual plot of data corrected for mean systematic error (bias) is examined for outliers; i.e. samples with a residual exceeding $3 \times \text{SEP}$. If an outlier occurs and this cannot be classified as an x -outlier and re-analysis of the sample by NIR and reference methods confirms the result, the outlier should not be removed.

In this case, the ruggedness of the calibration is not sufficient and the calibration set should be expanded. The next step is to fit NIR and reference data by linear regression (reference = $b \times \text{NIR} + a$) to support the visual impression. If the slope (b) is significantly different from 1, the calibration is skewed. Adjusting the slope of the calibration is generally not recommended. If a re-investigation of the calibration does not detect outliers, especially influential outliers, it is preferable to expand the calibration set to include more samples. However, if the slope is adjusted, the calibration should be tested on a new independent test set. The data are also examined for a bias between the methods. An intercept (a) significantly different from 0 indicates that the calibration is biased. A bias may be removed by adjusting the constant term in the calibration equation. However, if the accuracy of the bias-adjusted calibration is significantly poorer than expected from cross validation on the calibration set, i.e. SEP is significantly larger than RMSECV, the calibration set should be expanded to include more samples. In all cases when a new calibration is developed on an expanded calibration set, the validation process should be repeated on a new independent test set. If necessary, expansion of the calibration set should be repeated until acceptable results are obtained on an independent test set.

6.6 Changes in measuring and instrument conditions

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

For example, the 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 cheeses from one dairy may not give the same accuracy on cheeses produced in another dairy if the processing and ripening 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.

Furthermore, 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 and slope adjustments to calibration equations. In some cases, it may even be necessary to standardize the two instruments against each other by mathematical procedures before calibration equations can be transferred [2]. Standardization procedures may be used to transfer calibrations between instruments of different types provided that samples are measured in the same way (reflectance, transmittance) in similar cups and that most of the spectral region is common. Adding a few samples scanned with the second instrument in the database can contribute to the transfer.

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.

6.7 Outlier detection

Use of NIR methods is generally limited to samples in the population covered by the calibration set with respect to sample material characteristics and analyte concentration. An outlier detection system should accompany the NIR method to reduce the risk of unintentional use of NIR spectrometry on samples outside this population. The system should be able to detect x -outliers and samples falling outside the concentration range. The principle of Mahalanobis distance [7] applied to PCA-reduced data or "leverage" [8] may be used for the detection of x -outliers. If a sample is detected as an outlier, the sample should be re-analysed by reference methods to obtain the final result.

7 Statistics for performance measurement

7.1 Standard error of prediction (SEP) and bias

The standard error of prediction (SEP), which expresses the accuracy of routine results corrected for the mean difference between routine and reference methods (bias), can be calculated by using the following equation:

$$SEP = \sqrt{\frac{1}{N-1} \sum (x_i - y_i - B)^2}$$

where

$(x_i - y_i)$ is the difference between results obtained by routine method (x_i) and reference method (y_i) on sample i ;

B is the bias, explained by the equation

$$B = \frac{1}{N} \sum (x_i - y_i) = \bar{x} - \bar{y}$$

N is the total number of samples in the test.

SEP is equal to the standard deviation of predicted residuals. This term expresses the accuracy which can be expected with the calibration equation when the bias is zero. Other symbols for this statistical term may be seen (e.g. s_d in ISO 8196-2 | IDF 128-2).

To obtain realistic estimates of the accuracy, SEP should be calculated on samples outside the calibration set and the distribution of the analyte concentration range should be uniform.

7.2 Root mean square error of prediction (RMSEP)

Instead of reporting SEP and bias in separate terms, they may be included in a single term, the root mean square error of prediction (RMSEP), defined as

$$\text{RMSEP} = \sqrt{\frac{1}{N} \sum (x_i - y_i)^2}$$

The relationship between SEP and RMSEP is $\text{RMSEP}^2 \approx \text{SEP}^2 + B^2$ (from Reference [2]). When the difference between the NIR method and the reference method is not clearly systematic, the SEP may overestimate the possibility of improving the accuracy using bias adjustment. In this case, the RMSEP gives a more realistic estimate of the prediction capability of the calibration. When the bias is insignificant, the RMSEP tends towards SEP with increasing data number.

To obtain realistic estimates of the accuracy, RMSEP should be calculated on samples outside the calibration set and the distribution of the analyte concentration range should be uniform.

7.3 Root mean square error of cross validation (RMSECV)

The equation is the same as for RMSEP (see 7.2). The difference is that the RMSECV is calculated from cross validation on the calibration set and not on an independent test set.

The SEP, RMSEP and RMSECV also contain the measurement uncertainty of the reference results. To reduce this to an insignificant level (less than 5 % relative), the imprecision of final reference results should be less than one-third of the SEP.

8 Sampling

A representative sample should have been sent to the laboratory. It should not have been damaged or changed during transport or storage.

All laboratory samples should usually be kept at a temperature of (0 to 4) °C from the time of sampling to the time of commencing the procedure.

Sampling is not part of the method specified in this International Standard. A recommended sampling method is given in ISO 707 | IDF 50.

9 Procedure

9.1 Preparation of test sample

9.1.1 Cheese

Before the analysis, remove the rind or coating on the cheese in such a way as to obtain a sample representative of the cheese as it is usually consumed.

Prepare the sample using an appropriate device (5.2). Quickly mix the ground or grated sample. If the sample cannot be ground or grated, mix it thoroughly by intensive kneading. Care should always be taken to avoid moisture loss.

Keep the prepared sample in an airtight container until the measurement, which should be carried out the same day. If delay is unavoidable, take every precaution to ensure proper storage of the sample. When refrigerated, ensure that any moisture condensation on the inside surface of the container is thoroughly and uniformly re-incorporated into the test sample.

9.1.2 Dried milk, dried whey and dried buttermilk

Thoroughly mix the sample by repeatedly rotating and inverting the sample container (if necessary, after having transferred all of the laboratory sample to an airtight container of sufficient capacity to allow this operation to be carried out).

9.1.3 Butter

Mixing is not necessary unless the sample gives evidence for it. In that case, the mixing temperature should not exceed 25 °C.

Keep the prepared sample in an airtight container until the measurement, which should be carried out the same day. If delay is unavoidable, take every precaution to ensure proper storage of the sample. When refrigerated, ensure that any moisture condensation on the inside surface of the container is thoroughly and uniformly re-incorporated into the test sample.

Mass reduction of the prepared sample to obtain the analytical sample should be performed by principles that keep the sampling error to a minimum. Use of incremental sampling techniques (e.g. riffle splitters for powdered products) may be useful for heterogeneous materials.

9.2 Measurement

The prepared sample should reach a temperature within the range included in the calibration, for example (20 to 30) °C for cheese and milk powder, and (8 to 12) °C for butter. A sub-sample is transferred to the sample cup without compressing the sample, and the sample surface is levelled off with a minimum of disturbance. The sample is measured following the instructions given by the manufacturer for operation of the instrument.

The number of scans, or sub-sample predictions averaged, or the dwell time on each wavelength should be large enough to reduce the noise of the measurement to an insignificant level. The calibration model valid for the measured sample type is then applied. If re-usable sample cups with windows are used, the windows should be cleaned between measurements [particles may be removed by a brush or a vacuum cleaner and sticky compounds may, for example, be removed using a cloth moistened with a suitable solvent (4.1)].

9.3 Evaluation of results

Results obtained on samples detected as spectral or concentration outliers may not be regarded as reliable.

10 Checking instrument stability

10.1 Control sample

At least one control sample should be measured before and after an uninterrupted series of sample analyses 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, as far as possible, resemble the samples to be analysed, and the parameter measured should be identical to, or at least biochemically close to, the sample analyte.

Cheese, milk powder and butter may be used as control material over short periods. A sample is prepared as in 9.1 and stored refrigerated as a series of sub-samples in closed containers. These samples should be stable for at least one week. Butter and especially milk powder is normally more stable than cheese. The stability should be tested in the actual cases. Shifts between 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 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). The photometric instrument noise should be tested daily to obtain warnings of lamp failure, mechanical instrument problems, light leaks in the sampling area, excessive temperature and humidity variations, etc. These tests may be part of a self-executing diagnostic system built into the instrument.

11 Running performance check of calibration

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. It may typically be (1 to 5) % of measured samples which have to be checked by reference methods.

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, for example, segmentation of concentration range and random selection of test samples within each segment.

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; $\pm 2 \times \text{SEP}$ (95 % confidence) and $\pm 3 \times \text{SEP}$ (99,9 % confidence) 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 laboratory are performing as they should, then only 1 in 20 points should plot outside the warning limits and 1 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 Shewhart control charts may be used in the assessment [34]. 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:

- one point outside either action limit;
- two out of three points in a row outside a warning limit;
- 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) and additional rules may be applied to strengthen decisions.

In 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 reduced to a negligible part if the imprecision of reference results is reduced to less than one-third of the SEP^[9].

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 too optimistic an SEP. An attempt to force the results within the limits by frequent adjustments of the calibration will 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 upgraded. 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.7). 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.8).

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. Such changes could in practice be caused by, for example, changes in the cheese processing parameters.

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 Intralaboratory reproducibility

The intralaboratory reproducibility, i.e. the difference between two individual single test results obtained on identical test material in the same laboratory 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 intralaboratory 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. SEP and RMSEP values reported in the literature are listed in Table A.1. The reported SEP and RMSEP values also include uncertainty of reference results which may vary from case to case.

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 International Standard, or regarded as optional,
- d) and any circumstances which may have influenced the results;
- e) the test result(s) obtained;
- f) the current SEP and bias (if statistically significant), estimated from running a performance test on at least 25 test samples (see Clause 11).

Annex A (informative)

Examples of SEP and RMSEP values

The following SEP and RMSEP values have been reported in the literature. The reported SEP and RMSEP values also include uncertainty of reference results which may vary from case to case.

Table A.1 — SEP and RMSEP values

Sample material	Analyte	Conc. range %	RMSEP %	SEP %	NIR technique	Ref.
Processed cheese, Gouda, Edam	Moisture	40 to 51		0,24	Reflection	10
	Fat	21 to 31		0,27	Ground samples	
Processed cheese	Moisture	48 to 51		0,21	Reflection	10
	Fat	21 to 23		0,23		
	Protein	20 to 24		0,35		
Cheddar	Moisture	35 to 40	0,34		Reflection	11
	Fat	31 to 35	0,33		Ground samples	
Tetilla, Arzúa, Edam	Total solids	45 to 62	0,61		Reflection	12
	Fat	18 to 32	0,47		Unground samples	
	Protein	16 to 30	0,50			
Danbo	Moisture	40 to 52	0,30		Transmission	13
	Fat	22 to 28	0,28		Unground samples	
	Protein	22 to 27	0,26			
Danbo	Total solids	46 to 62	0,58		Reflection	14
	Fat	14 to 36	0,52		Unground samples	
	Protein	22 to 31	0,38			
Edam	Total solids	50 to 61	0,20		Transmission Ground samples	15
Gouda	Total solids	40 to 43	0,12		Transmission	15
Brie	Total solids	41 to 55	0,33		Transmission	15
Colby	Total solids	38 to 41	0,23 to 0,27		Transmission	15
Cheddar	Total solids	37 to 40	0,31 to 0,35		Transmission	15
Danbo	Total solids	50 to 63	0,20		Transmission	16
	Fat	23 to 29	0,19		Unground samples	
Danbo	Total solids	47 to 52	0,16		Transmission	17
	Total solids	47 to 63	0,29		Unground samples	
	Fat	16 to 28	0,17			
Dried skim milk	Moisture	3,3 to 4,7		0,08	Reflection	10
	Fat	0,5 to 1,3		0,09		
	Protein	34 to 37		0,20		
	Lactose	48 to 50		0,44		

Table A.1 (continued)

Sample material	Analyte	Conc. range %	RMSEP %	SEP %	NIR technique	Ref.
Dried buttermilk	Moisture	2,7 to 6,3		0,10	Reflection	10
	Fat	5,3 to 11		0,13		
	Protein	29 to 35		0,21		
	Lactose	37 to 47		0,37		
Dried skim milk	Moisture	3,2 to 6,1		0,08 ^a	Reflection	18
	Fat	0,7 to 2,5		0,07 ^a		
	Protein	35 to 38		0,18 ^a		
Dried whole milk	Moisture	2,7 to 4,5		0,09 ^a	Reflection	18
	Fat	24 to 27		0,19 ^a		
	Protein	25 to 28		0,19 ^a		
Dried skim milk	Moisture	2,9 to 9,7		0,27	Reflection	19
	Fat	0,5 to 2,1		0,10		
	Protein	34 to 40		0,44		
	Lactose	53 to 58		0,59		
Dried whey	Moisture	2,7 to 5,7		0,37	Reflection	20
	Fat	0,2 to 7,2		0,52		
	Protein	9,5 to 42		1,3		
	Lactose	7,9 to 71		2,8		
Butter	Moisture	14 to 16	0,26		Reflection	21
	Non-fat solids	1,3 to 2,6	0,071			
	Fat	81 to 84	0,38			

^a Standard error from calibration without cross validation.

Annex B (informative)

Examples of figures

Full-scan calibrations (900 nm to 1 100 nm transmittance) were developed for the determination of fat (%) in semi-hard cheese containing ca. 30 % fat in total solids. The calibrations were developed using 110 samples and 6 cross-validation segments. Spectra were MSC treated before calibration. The calibrations were tested on an independent test set to obtain the plot shown in Figure B.1.

The optimal number of PLS factors is 10 ± 1 .

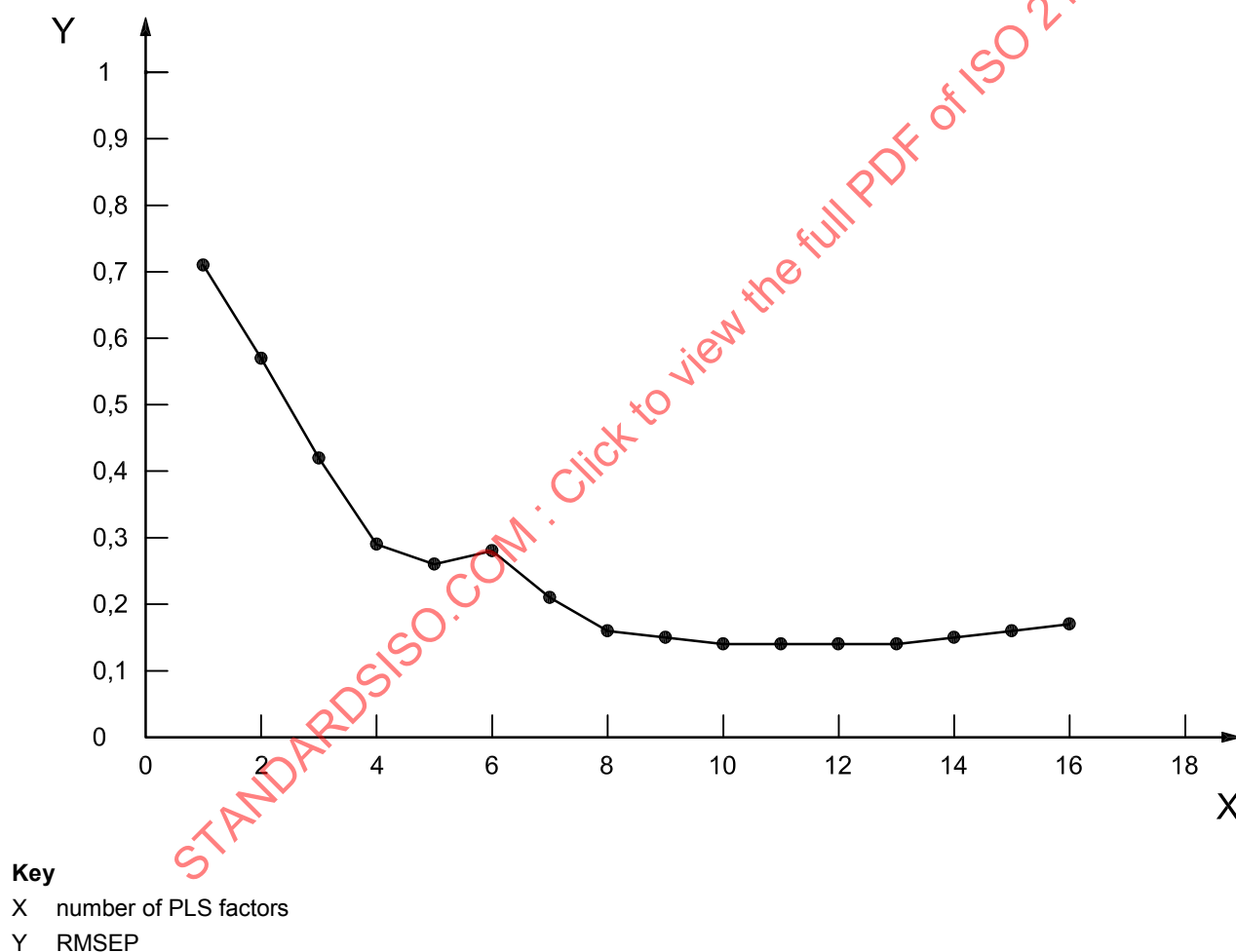


Figure B.1 — Example from practice showing a plot of RMSEP as function of the number of PLS factors

A full-scan calibration equation (900 nm to 1 100 nm transmittance) was developed for determination of fat (%) in semi-hard cheese containing ca. 30 % fat in total solids. Figure B.2 shows results from cross validation (6 segments, 110 samples) obtained after MSC treatment of spectra.

Results obtained on an independent test set (320 samples) using the developed calibration equation were

SEP 0,14; RMSEP 0,14; slope 1,00.

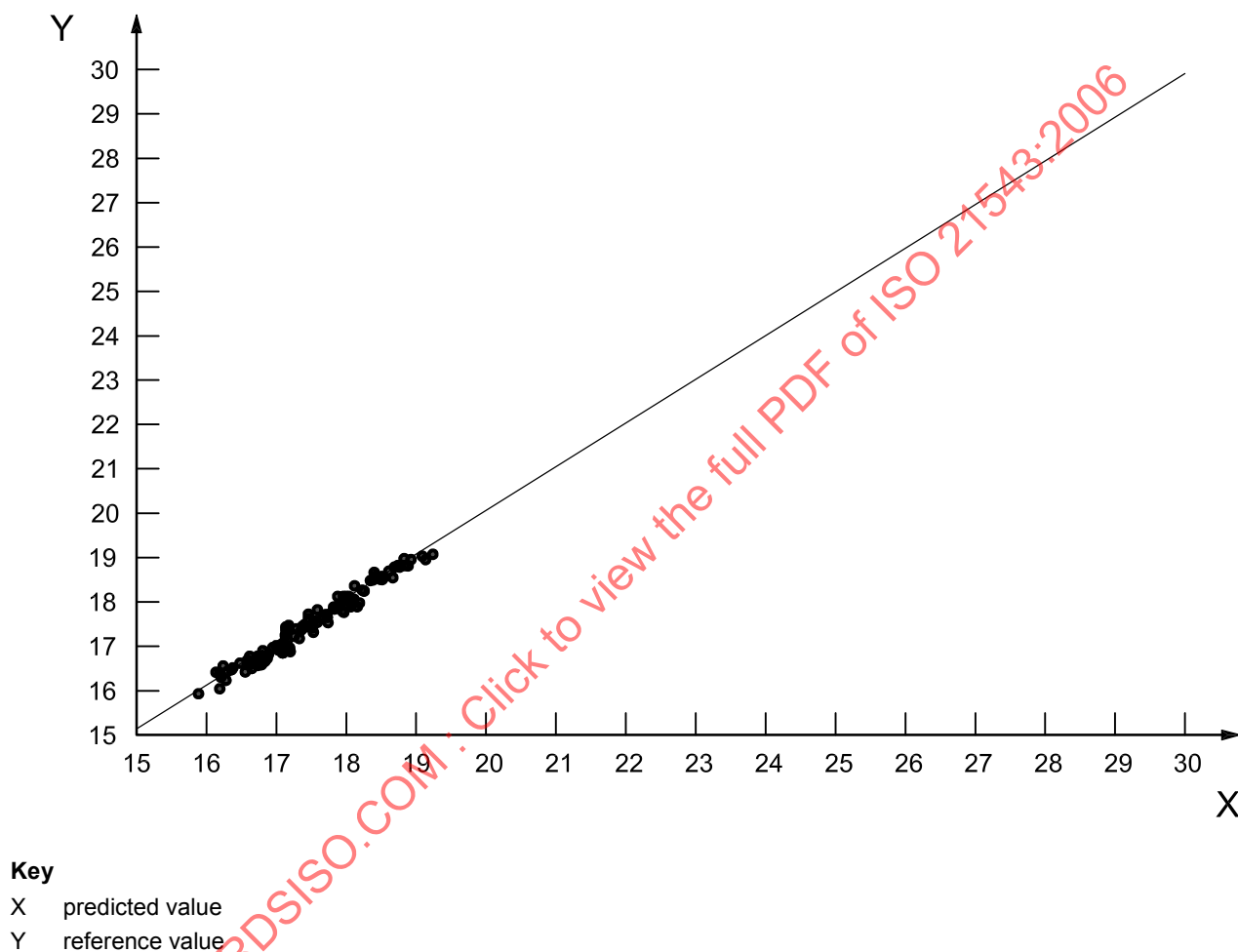


Figure B.2 — No outliers

Figure B.3 was obtained with calibration conditions as given in Figure B.2. The x -outlier was a cheese containing ca. 45 % fat in total solids.

Results obtained on an independent test set (320 samples) using the developed calibration equation were:

SEP 0,16; RMSEP 0,16; slope 1,03.

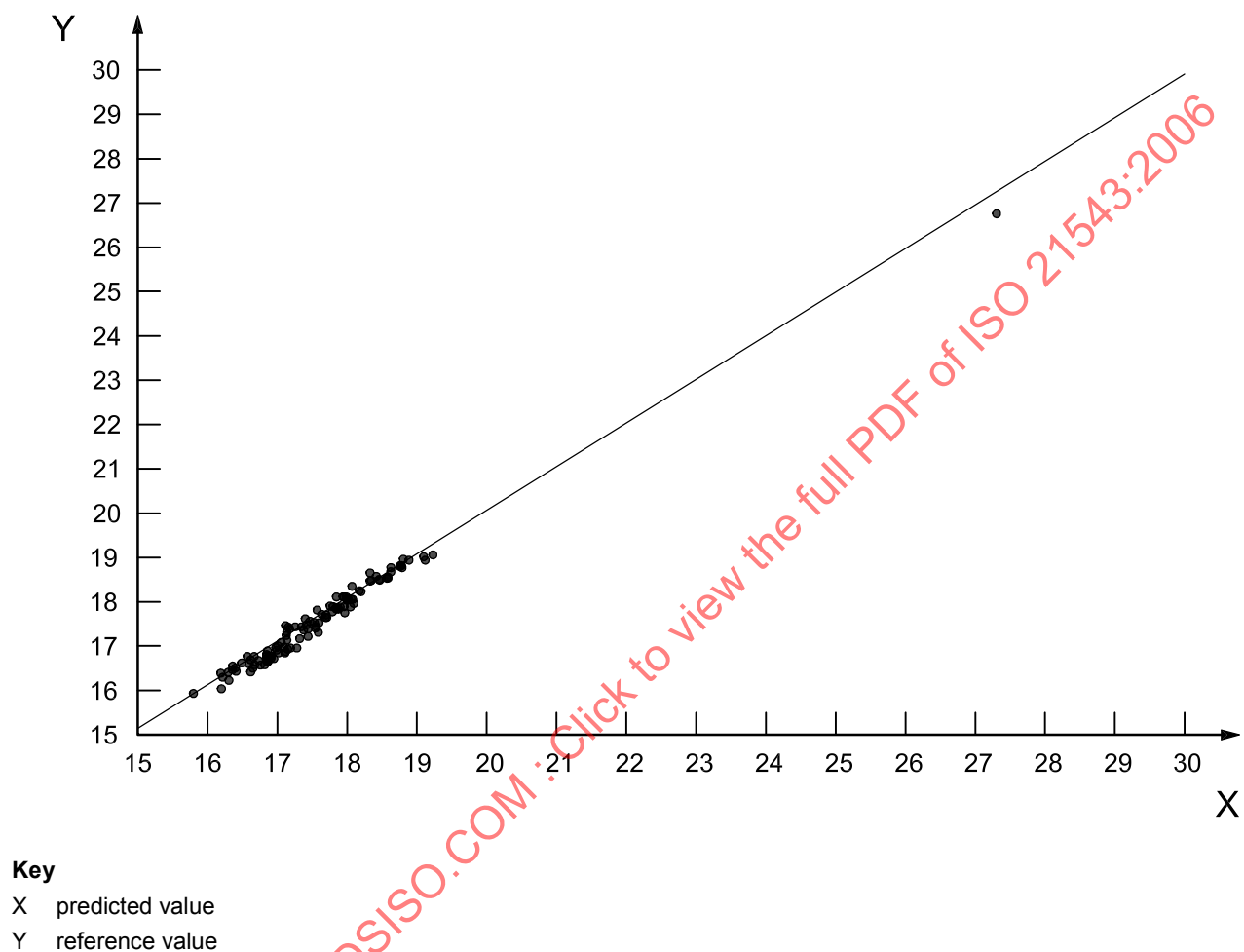


Figure B.3 — x -outlier

Figure B.4 was obtained with calibration conditions as given in Figure B.2. The y -outlier was a cheese with a 2,0 % unit error in the reference result.

Results obtained on an independent test set (320 samples) using the developed calibration equation were:

SEP 0,16; RMSEP 0,16; slope 1,03.

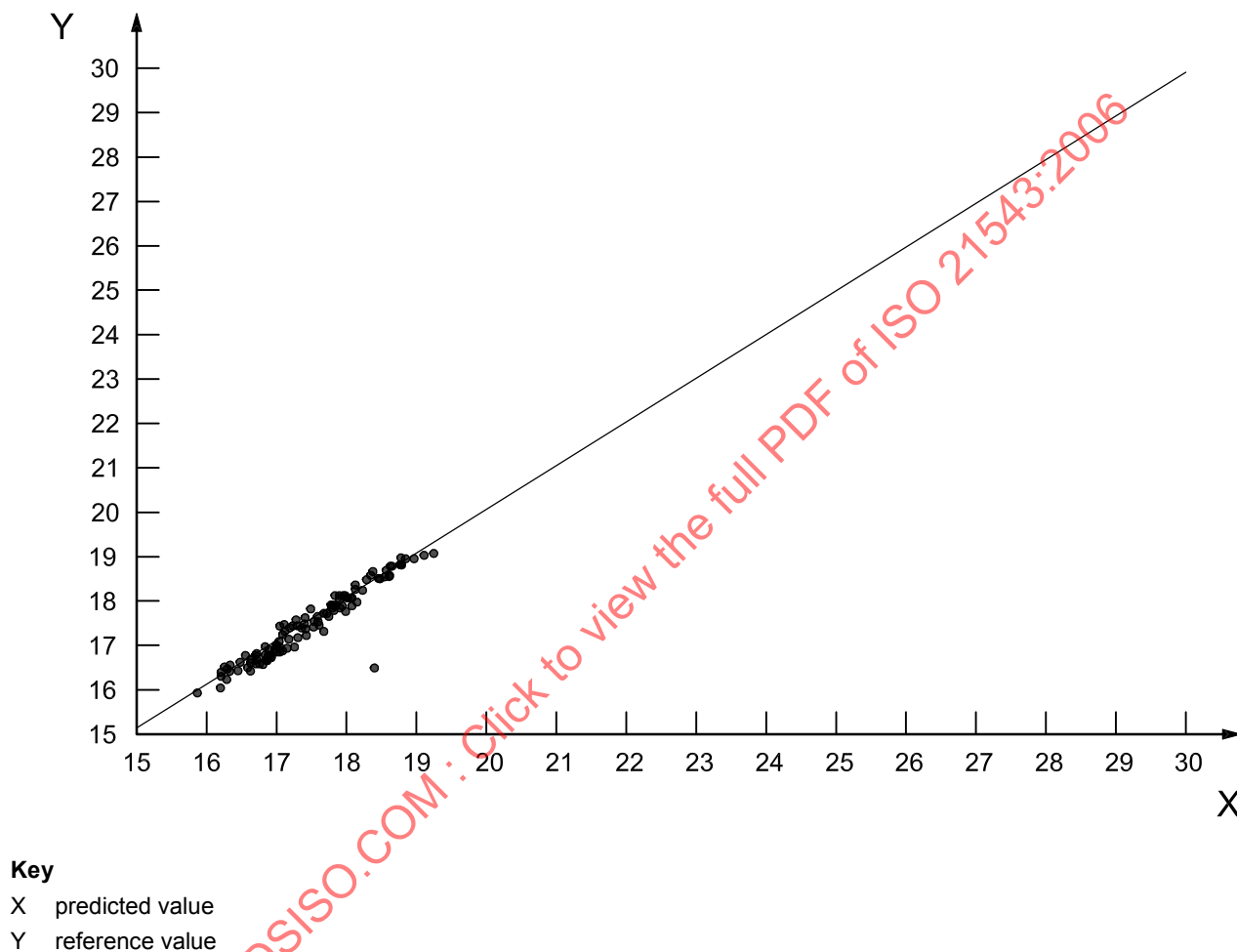


Figure B.4 — y -outlier

Figure B.5 was obtained with calibration conditions as given in Figure B.2. The sample was the outlier shown in Figure B.3 (cheese with ca. 45 % fat in total solids) assigned a 2,0 %-unit error in reference result. Significantly increased residuals and a slight tilting of the main group in order to fit this group and the outlier simultaneously are observed.

Results obtained on an independent test set (320 samples) using the developed calibration equation were:

SEP 0,28; RMSEP 0,29; slope 1,12.

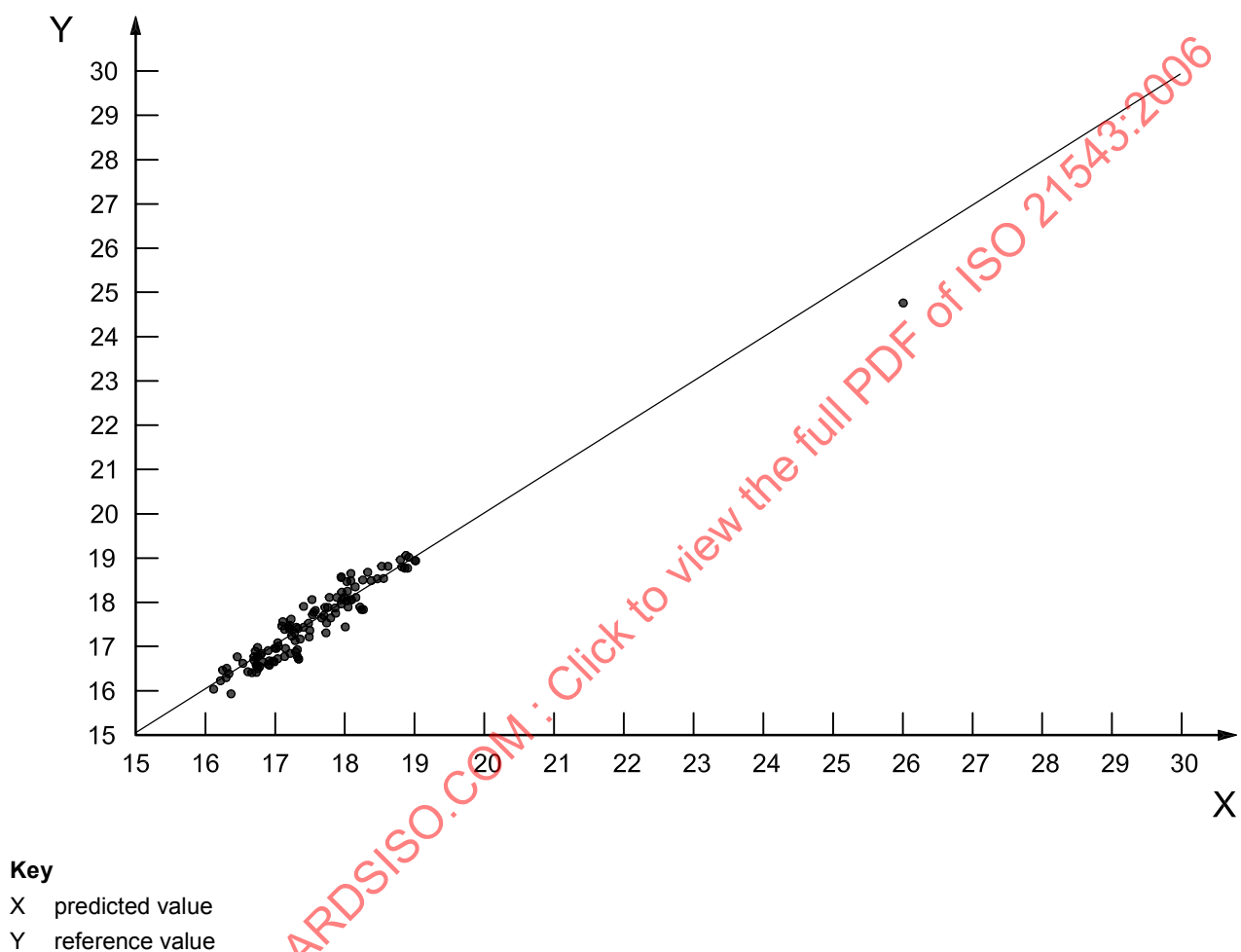


Figure B.5 — Combined x - and y -outlier

Figure B.6 was obtained with the same calibration conditions as given in Figure B.2. The sample was the outlier shown in Figure B.3 (cheese with ca. 45 % fat in total solids) assigned a wrong result (a result for a cheese containing 30 % fat in total solids). Dramatically increased residuals and a tilting of the main group in order to fit this group and the outlier simultaneously are observed.

Results obtained on an independent test set (320 samples) using the developed calibration equation were:

SEP 0,62; RMSEP 0,64; slope 1,61.

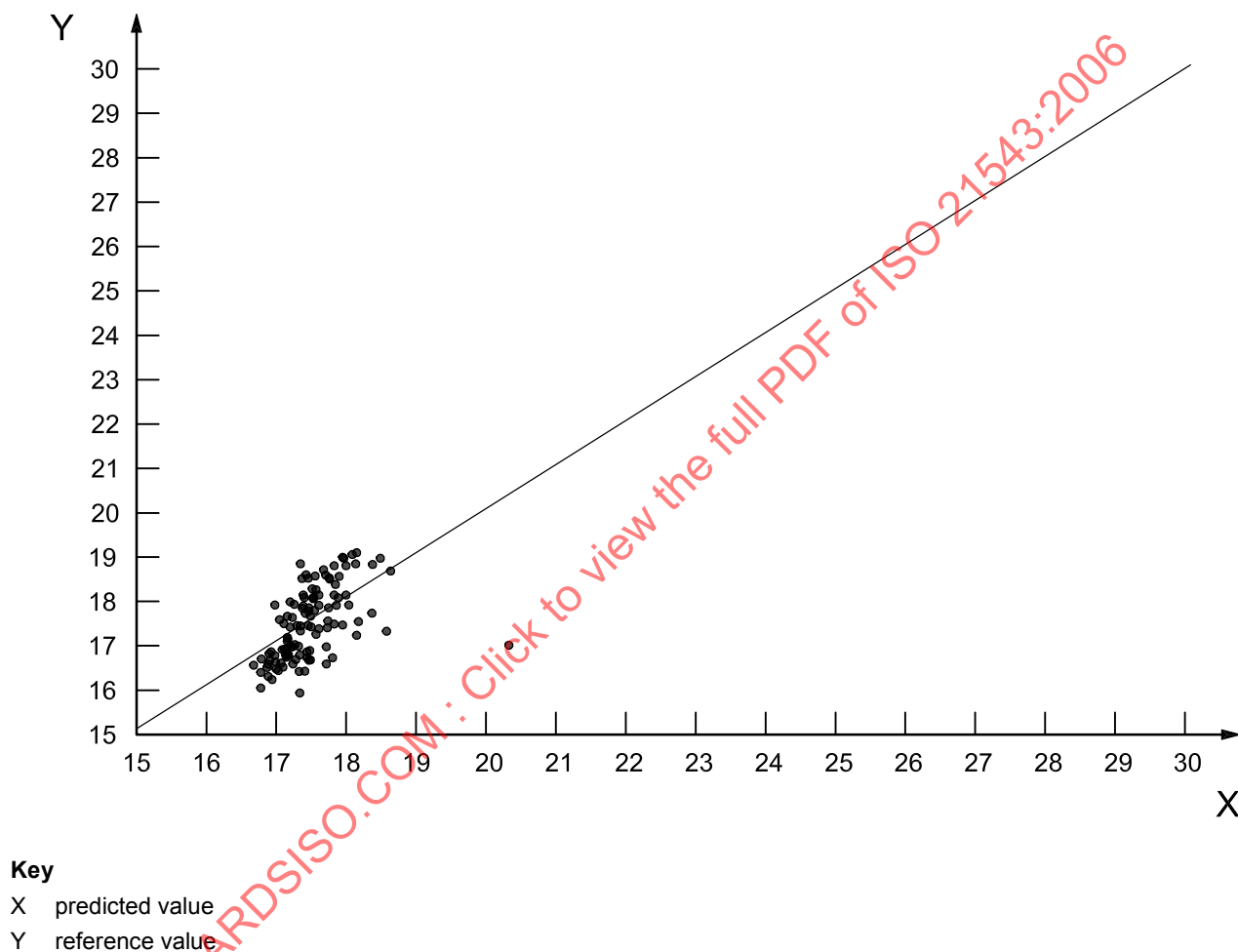


Figure B.6 — Another combined x - and y -outlier