INTERNATIONAL STANDARD

ISO 11726

Second edition
2017-11

Solid mineral fuels — Guidelines for the validation of alternative methods of analysis

Combustibles minéraux solides — Lignes directrices pour la validation de variantes analytiques

Reference number
ISO 11726:2017(E)

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

Contents

Page

Foreword iv
1 Scope 1
2 Normative references 1
3 Terms and definitions 1
4 General 2
5 Preliminary work on the alternative method 3
  5.1 General 3
  5.2 Selectivity and specificity 3
  5.3 Range and linearity 3
  5.4 Sensitivity 3
  5.5 Limit of detection 3
  5.6 Limit of quantitation 4
  5.7 Ruggedness 4
  5.8 Accuracy 4
  5.9 Precision 4
6 Defining the alternative method to be validated 4
7 Procedure 5
  7.1 Measurement of precision and trueness 5
    7.1.1 Precision 5
    7.1.2 Trueness 5
  7.2 Comparison with reference materials 6
    7.2.1 Sources of reference materials 6
    7.2.2 Estimation of trueness 6
    7.2.3 Estimation of precision 9
  7.3 Comparison with International Standard method within a laboratory — Single fuel 10
    7.3.1 General 10
    7.3.2 Estimation of trueness 10
    7.3.3 Estimation of precision 12
  7.4 Comparison using the International Standard method within a laboratory — Range of fuels 13
  7.5 Comparison using the International Standard Method, between laboratories 13
8 Validation report 13
Annex A (informative) **Explanation of rationale of method A** 14
Annex B (informative) **Example of calculations** 16
Bibliography 23

Foreword

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

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

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

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

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

This document was prepared by Technical Committee ISO/TC 27, *Solid mineral fuels*, Subcommittee SC 5, *Methods of analysis*.

Solid mineral fuels — Guidelines for the validation of alternative methods of analysis

1 Scope

This document describes procedures for validating alternative methods of analysis for coal and coke either directly by comparison with the relevant International Standard method or indirectly by comparison with reference materials that have been exhaustively analysed using the relevant International Standard method.

The statistical analysis methods used are parametric, i.e. their use is possible only when the characteristic is expressed as a simple number on an approximately linear scale. The results from some methods, for example the Gray-King coke type, are not so expressed and the methods given here need to be used only if the data are converted to a parametric scale.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 5725-6:1994, *Accuracy (trueness and precision) of measurement methods and results — Part 6: Use in practice of accuracy values*

3 Terms and definitions

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

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

- ISO Online browsing platform: available at https://www.iso.org/obp
- IEC Electropedia: available at http://www.electropedia.org/

3.1
accuracy
closeness of agreement between a test result and the acceptable reference value

Note 1 to entry: The term accuracy, when applied to a set of results, describes a combination of random components and a common systematic error or bias component.

3.2
bias
difference between the expectation of the test results and an accepted reference value

Note 1 to entry: Bias is a systematic error as contrasted to random error. There may be one or more systematic error components contributing to the bias. A larger systematic difference from the accepted reference value is reflected by a larger bias value.

3.3
precision
closeness of agreement between independent test results obtained under prescribed conditions

Note 1 to entry: Precision depends only on distribution of random errors and does not relate to the accepted reference value.

Note 2 to entry: The measure of precision is usually expressed in terms of imprecision and computed as a standard deviation of the test results. Higher imprecision is reflected by a larger standard deviation.

Note 3 to entry: “Independent test results” means results obtained in a manner not influenced by any previous result on the same or similar material.

3.4

trueness

closeness of agreement between the average value obtained from a large series of test results and an accepted reference value

Note 1 to entry: The measure of trueness is usually expressed in terms of bias.

4 General

An International Standard method is a measurement method that has been subjected to a standardization process to satisfy various requirements. Among these requirements, taken from ISO 5725-6:1994, Clause 8, are the following:

- The method shall be applicable to a wide range of levels of characteristics to cover most materials that are internationally traded.

EXAMPLE A method for the determination of sulfur content in coal is applicable to as many internationally traded coals as possible.

- Equipment, reagents and personnel shall be available on an international basis.
- The costs of performing the tests shall be acceptable.
- The precision and trueness of the measurement method shall be acceptable for the users of the results.

Many analytical methods for coal and coke are based on traditional combustion or wet-chemical analysis and the results are highly dependent on the test conditions. They are frequently time-consuming, labour- and skill-intensive and unsuited to automation. However, they meet the requirements of International Standard measurement methods, both in being internationally available and in providing acceptable levels of trueness and precision in international coal trade.

Other, non-standard methods of analysis are in use when

a) most of the material tested comes from the same source and the variation of its characteristics is relatively small. In such cases a simpler, less expensive method may be adequate;

b) an instrumental or automated version of the standard method provides much cheaper analysis of large numbers of samples. Such equipment may be much more expensive than the standard equipment or of a highly proprietary nature;

c) an instrumental method based on an analytical principle different from that of the standard method is available. Such methods have similar characteristics to 4 b) above.

In some cases, if it is possible to write a generic description of the method and the equipment is widely available, methods of types 4 b) and 4 c) above can be developed as International Standards. If an International Standard method already exists for analysis of a particular parameter, the new alternative method should be tested against the established method to ensure that it provides results that are comparable for trueness and accuracy. This is part of the process of issuing the alternative method as an International Standard.

Even if the equipment is widely available, it might not be possible to convert the method into an International Standard because of the proprietary nature of the equipment, speed of development and rapid obsolescence of such equipment.

The commercial pressure for cheaper, more rapid analysis has, however, meant that many analyses are carried out on equipment of this type. Some users develop their own in-house methods or use such methods for contractual purposes if agreed between both parties, provided that they can be ensured that the alternative method produces results that are comparable for accuracy and trueness with the International Standard method.

The purpose of this document is to give guidelines for such a validation, as applied to methods for testing of coal and coke. It is not intended to infer that the use of such alternative methods complies with the relevant International Standards nor is it for use in writing alternative International Standards. If the intention is to develop a new method into an International Standard, the procedures given in ISO 5725-6 should be used.

To summarize, alternative methods requiring validation range from simplified versions of the International Standard method to proprietary automated instrumental methods using principles entirely different from those of the International Standard method.

5 Preliminary work on the alternative method

5.1 General

Before any detailed comparison with the International Standard method is undertaken, it is necessary to investigate the performance characteristics of the alternative method. When buying specific commercial equipment, information on these aspects should be sought from the manufacturer. Many of the characteristics given below are applicable only to methods where the sample is in liquid form for the determination. For direct determination on solids (e.g. ash), little preliminary work is possible. Some of the main performance characteristics are given below, drawn from Reference[3].

5.2 Selectivity and specificity

Selectivity of a method refers to the extent to which it can determine particular analyte(s) in a complex mixture without interference from the other components in the mixture. A method that is perfectly selective for an analyte or group of analytes is said to be specific. The applicability of the method should be studied using various samples, ranging from pure standards to mixtures with complex matrices. Standard addition of pure analyte to coal/coke solutions should be used. In each case, the recovery of the analyte(s) of interest should be determined and the influences of suspected interferences stated. Any restrictions in the applicability of the technique should be documented in the method.

5.3 Range and linearity

The working range for a method is determined by examining samples with different analyte concentrations and determining the concentration range for which acceptable accuracy and precision can be achieved. While the working range of the analyte in solution may be determined using pure analyte or synthetic matrices containing analyte, the true range and linearity cannot be determined until a detailed comparison with the International Standard method is made on fuel samples.

5.4 Sensitivity

This is the difference in analyte concentration corresponding to the smallest difference in the response of the method that can be detected. It is represented by the slope of a calibration curve and can be determined by a least-squares procedure, or experimentally, using fuel samples containing various concentrations of the analyte.

5.5 Limit of detection

The limit of detection of an analyte is determined by repeat analysis of a blank test portion and is the analyte concentration whose response is the equivalent to the mean blank response plus three standard deviations. Its value is likely to be different for different types of sample.

5.6 Limit of quantitation

This is the lowest concentration of analyte that can be determined with an acceptable level of accuracy and precision, i.e. it is usually the lowest point on the calibration curve (excluding the blank). It should be established using an appropriate standard or sample; it should not be determined by extrapolation.

NOTE Within this document, “accuracy” is known as “trueness”.

5.7 Ruggedness

This is sometimes called robustness. Each time a method is used, small variations are inevitably introduced in the procedure, which may or may not have a significant influence on the performance of the method. The ruggedness of a method is tested by deliberately introducing small changes to the method, for example mass of sample and temperature of combustion, and examining the consequences. A large number of factors may need to be considered, but because most of these will have a negligible effect, it is normally possible to vary several at once.

5.8 Accuracy

The accuracy of a method is the closeness of the obtained analyte value to the true value. The overall accuracy can only be established by analysing suitable reference materials or comparison with the International Standard method (see Clause 7). For intermediate stages (i.e. solution finishes), an estimation of accuracy can be obtained by spiking test portions with chemical standards. The value of spiking is limited; it can only be used to determine the accuracy of those stages of the method following the spiking.

5.9 Precision

The precision of a method is a statement of the closeness of agreement between mutually independent test results and is usually stated in terms of standard deviation. It is generally dependent on analyte concentration, and this dependence should be determined and documented. It may be stated in different ways depending on the conditions for which it is calculated. Repeatability is a type of precision relating to measurements made under repeatability conditions, i.e. same method, same material, same operator, same laboratory, different time but within a narrow time period. Preliminary estimations of precision of the alternative method may be made, for example, by comparing the results of duplicate samples for the ruggedness tests.

6 Defining the alternative method to be validated

Once the preliminary work on the alternative method (see Clause 5) has shown that it is likely to be suitable for the intended purpose, the test conditions for the method should be chosen and clearly and unambiguously defined in a manner similar to the way in which International Standard methods are defined. Critical test parameters vary with the type of test and cannot be exhaustively listed in this document. Examples of some parameters commonly found in coal and coke analysis are as follows:

a) mass of sample and solid reagents, plus critical range;

b) condition of sample, moisture content, particle size, particle size range;

c) accuracy of measuring equipment for temperature, mass, volume;

d) purity of reagents, accuracy of solution concentration;

e) furnace temperature, with critical dimensions of the hot zone where relevant;

f) length of time of combustion/heating;

g) atmosphere in the furnace/oven;

h) in spectroscopic determinations, cell path length, wavelength;

i) calibration procedures.

When automated instruments are operated under pre-set conditions, these conditions should be defined as closely as possible; all variable settings of the instrument should be defined. Particular attention should be paid to those instruments whose settings can be altered by modification of a computer program where any change might not be immediately apparent to the operator. Some method of checking that the program has not been altered from the standard conditions should be devised.

The test procedure should be written, again in a manner similar to that of an International Standard, so that subsequent operators are able to follow the method identical to that used during validation.

7 Procedure

7.1 Measurement of precision and trueness

7.1.1 Precision

Measure the precision of results in terms of the standard deviation of a set of analyses carried out under repeatability conditions. The precision of the alternative method is measured directly by making replicate analyses of samples.

If the alternative method is to be validated using reference samples, then calculate the standard deviation of the International Standard method from the repeatability limit given in the International Standard method.

If a direct comparison is to be made between the two methods, then determine the precision of the International Standard method directly by analysing samples in replicate. This will be a more accurate measure than that calculated from the International Standard because the precision of the two methods on the analysis of the same fuels is compared, rather than a precision determined at the time of development of the International Standard on fuels whose identity is unlikely to be known.

7.1.2 Trueness

Estimate the trueness either by comparing the results obtained by analysing a reference material using the alternative method with the reference value (see 7.2) or by comparing results on the same fuels using both the alternative and International Standard methods (see 7.3).

Measurement of trueness can only be an estimate, the errors of which are measured by considering the variability of the differences between the results. The greater the variability, the greater the estimation error, the more results that are compared, the lower the estimation error.

Two different statistical analysis methods (A and B) are given in this document.

Method A (see 7.2.2.1 and 7.3.2.1 for details) is recommended as the most rigorous. Decide before starting on the maximum tolerable bias, MTB, and design the test to be sufficiently sensitive to detect that bias, should it exist. Carry out a sufficient number of analyses to make the statistical test powerful enough to conclude either that

a) the bias is significantly greater than zero and not significantly less than MTB, or

b) the bias is significantly less than MTB and not significantly greater than zero.

In a simpler test (method B; see 7.2.2.2 and 7.3.2.2 for details), compare a fixed number of results and, unless the mean difference fails a null hypothesis test, it can be concluded that no bias exists between the two methods. If the estimation error is too great, however, it is also possible that a bias at an unacceptable level could exist. To obviate such an ambiguous conclusion, use method A.

The difficulty with method A is in deciding what value to assign for MTB. However, with method B, it is necessary to a make a judgement, after the result, as to what bias levels are tolerable. For either method, therefore, it is necessary to consider the practical implications of any possible bias and to make a rational judgement on what level is unacceptable.

7.2 Comparison with reference materials

7.2.1 Sources of reference materials

Coal and coke reference materials samples are widely available commercially. Before use, investigate the traceability, authority and methods of analysis. Obtain details of homogeneity trials, stability trials, the methods used for certification and the uncertainty and variations in the stated analyte values from the producer and use them to judge the pedigree. In order to compare the alternative method with the International Standard method, use only reference materials that have been analysed using the International Standard method. Where an International Standard method contains a major element of choice, quote the subclauses from the International Standard giving details of that variation.

Reference materials should be clearly labelled and stored under the specified conditions and should be safeguarded against contamination or loss of determinand.

Select reference materials to be of the same type of fuel and to have the same analyte concentration ranges as those which are to be analysed routinely by the alternative method. Some factors which should be considered are coal rank, coke type or manufactured fuel type.

Where the range of fuels to be tested is very wide or reference materials of the appropriate type cannot be obtained, test the method against the International Standard method (see 7.3 or 7.4).

Two methods are given for the estimation of trueness. The recommended method, method A, requires the specification, before any analysis is done, of the greatest bias (the maximum tolerable bias) that the user is prepared to risk; testing continues until an unambiguous conclusion is reached. In a simpler method, method B, a fixed number of analyses is performed and a conclusion drawn from the result. In the latter method, there is a possibility that an alternative method is considered to be unbiased even though there is a risk that it is biased to an unacceptable extent.

7.2.2 Estimation of trueness

7.2.2.1 Method A

Decide on a value for the maximum tolerable bias, B (see 7.1.2).

Calculate the standard deviation of the International Standard, s_{ISO}, method using Formula (1):

$$s_{\text{ISO}} = \frac{r}{2\sqrt{2}} \tag{1}$$

where

s_{ISO} is the standard deviation of the International Standard method under repeatability conditions;

r is the repeatability limit for the International Standard method.

Calculate the value of g using Formula (2):

$$g = \frac{B}{s_{\text{ISO}}} \tag{2}$$

Using Table 1, calculate the number of replicate tests, n, necessary to identify the MTB.

Table 1 — Values for factor g for calculating the number of analyses required

	0	1	2	3	4	5	6	7	8	9
0	—	—	—	4,170	2,728	2,195	1,872	1,659	1,506	1,389
10	1,295	1,218	1,154	1,099	1,051	1,009	0,972	0,938	0,907	0,880
20	0,855	0,832	0,810	0,790	0,772	0,755	0,739	0,724	0,710	0,696
30	0,684	0,672	0,660	0,649	0,639	0,629	0,620	0,611	0,602	0,594
40	0,586	0,579	0,571	0,564	0,558	0,551	0,545	0,539	0,533	0,527
50	0,521	0,516	0,511	0,506	0,501	0,496	0,491	0,487	0,483	0,478
60	0,474	0,470	0,466	0,463	0,459	0,455	0,451	0,448	0,445	0,441
70	0,438	0,435	0,432	0,429	0,426	0,423	0,420	0,417	0,414	0,411
80	0,409	0,406	0,404	0,401	0,399	0,396	0,394	0,392	0,389	0,387
90	0,385	0,383	0,380	0,378	0,376	0,374	0,372	0,370	0,368	0,366

NOTE The number of sets required corresponding to a given g factor is the sum of the column and row headings.

Analyse the reference material using the alternative method n times.

Calculate s_{ALT} using Formula (3):

$$s_{\mathrm{ALT}} = \sqrt{\frac{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}{(n-1)}} \qquad (3)$$

where

$\sum x$ is the sum of all the results;

$\sum x^2$ is the sum of the squares of the results;

n is the number of the results.

Recalculate g and hence n. If the new value for n is greater than the number of analyses already done, then carry out additional analyses to reach n. Continue with this process until enough analyses have been carried out.

Calculate the value of the mean difference of the analytical values from the reference value, $\bar{d}$, using Formula (4):

$$\bar{d} = \frac{\sum x_i}{n} - R \qquad (4)$$

where

x_i is the analytical value of the ith determination;

R is the reference value.

Calculate the statistic, t_{c}, from Formula (5):

$$t_{\mathrm{c}} = \frac{\bar{d}\sqrt{n}}{s_{\mathrm{ALT}}} \qquad (5)$$

Compare with the value of t_{t} from Table 3 at (n − 1) degrees of freedom.

If $t_c > t_t$ then it can be concluded that the alternative method is biased in comparison with the International Standard method. Otherwise, it can be concluded that any bias is significantly less than the maximum tolerable bias.

Repeat this process for reference fuels taken from across the range of fuels to be tested.

7.2.2.2 Method B

Using the method as defined in Clause 6, analyse the reference material using the alternative method n times (at least 5) under repeatability conditions.

Calculate the mean, $\bar{x}$, of the results using Formula (6), and hence its difference from the reference value, $\bar{d}$, and the standard deviation of the results, s_{ALT}, using Formula (3).

$$\bar{x} = \frac{\sum x}{n} \tag{6}$$

Calculate the statistic, t_c, using Formula (5).

Compare with the value of t_t from Table 3 at (n − 1) degrees of freedom.

If $t_c > t_t$ then it can be concluded that the alternative method is biased in comparison with the International Standard method.

Formula (5) implies that the sensitivity of the test to detect any bias is dependent on the number of analyses, n, and the standard deviation s_{ALT}.

If the estimated bias is $\bar{d}$, then the true bias will lie within the 95 % confidence limits given by Formula (7):

$$\bar{d} \pm \frac{t_t \cdot s_{\mathrm{ALT}}}{\sqrt{n}} \tag{7}$$

If $t_c \le t_t$, these limits will encompass the value zero indicating that the test has not been able to find significant bias at the 95 % confidence level used for the test. However, if the value at either confidence limit is a level of bias that is unacceptable, such a conclusion is ambiguous.

Table 2 — Variance ratio (*F* distribution) at 95 % probability

F_t	Degrees of freedom
6,39	4
5,05	5
4,28	6
3,79	7
3,44	8
3,18	9
2,98	10
2,82	11
2,69	12
2,58	13
2,48	14
2,40	15
2,33	16
2,27	17
2,22	18

Table 2 *(continued)*

2,17	19
2,12	20
2,08	21
2,05	22
2,01	23
1,98	24
1,96	25
1,93	26
1,91	27
1,88	28
1,86	29
1,84	30
1,76	35
1,69	40
1,64	45
1,60	50
1,56	55
1,53	60

7.2.3 Estimation of precision

Derive the standard deviation, s_{ISO}, of the International Standard method from the quoted repeatability limit of that method, using Formula (1).

Compare s_{ISO} with s_{ALT}.

There are two possibilities:

a) if $s_{\text{ISO}} \geq s_{\text{ALT}}$, then it can be concluded that the precision of the alternative method is as good, or better than that of the International Standard method;

or

b) if $s_{\text{ISO}} < s_{\text{ALT}}$, then calculate the value of F_c from Formula (8).

$$F_c = \frac{(s_{\text{ALT}})^2}{(s_{\text{ISO}})^2} \qquad (8)$$

Compare F_c with the value of F in Table 2 at $(n - 1)$ degrees of freedom, F_t. If $F_c > F_t$ then the precision for the alternative method is worse than that of the International Standard method.

NOTE A worse precision does not necessarily invalidate the alternative method. Precision can be improved by increasing the number of determinations on a sample and taking the mean. Thus, if the precision of a single result is P, then the precision of the mean of n determinations will be $P/\sqrt{n}$. For methods which are automated or more rapid, this might still represent improved efficiency over doing a single analysis using the International Standard method.

7.3 Comparison with International Standard method within a laboratory — Single fuel

7.3.1 General

When a laboratory shall analyse a fuel from a single source on a regular basis, then use the procedure described in 7.3.2 and 7.3.3.

The principle of the procedure is that duplicate analyses are carried out by both methods A and B on a series of routine samples.

If the test sample specifications for both methods are the same, then take all four test portions required from the same test sample.

If the test sample specifications are different, for example a different particle size specification, then prepare a test sample for the method requiring the larger particle size and divide it into two. Carry out further preparation on one-half in order to obtain the second test sample. Take two test portions from each test sample.

7.3.2 Estimation of trueness

7.3.2.1 Method A

Decide on a value for the maximum tolerable bias, B (see 7.1.2).

Calculate the value of g using Formula (2) using the standard deviation of the International Standard method calculated using Formula (1) for s_{ISO}.

Using Table 1, calculate the number, n, of replicate tests necessary to identify B.

Using both the alternative method as defined in Clause 6 and the International Standard method, analyse in duplicate a series of n routine samples, under repeatability conditions.

Calculate the standard deviation of the differences between the methods, s_{d}, recalculate g and then n. If the new value for n is greater than the number of analyses already done, then carry out the additional analyses to reach n. Continue with this process until enough analyses have been carried out.

Calculate the mean difference between the methods, $\bar{d}$, from the differences between the means of the duplicates for each sample, taking account of the sign of the differences and the standard deviation of the differences, s_{d}.

Calculate the statistic t_{c} from Formula (9):

$$t_{\mathrm{c}} = \frac{\bar{d}\sqrt{n}}{s_{\mathrm{d}}} \tag{9}$$

Compare t_{c} with the value of t_{t} from Table 3 at $(n - 1)$ degrees of freedom.

If $t_{\mathrm{c}} > t_{\mathrm{t}}$, then it can be concluded that the alternative method is biased in comparison with the International Standard method.

Table 3 — Values of Student's *t* for 95 % confidence and two-sided distribution

Degrees of freedom	t
4	2,780
5	5,570
6	2,447
7	2,365
8	2,306

Table 3 *(continued)*

9	2,262
10	2,228
11	2,201
12	2,179
13	2,160
14	2,145
15	2,131
16	2,120
17	2,110
18	2,101
19	2,093
20	2,086
21	2,080
22	2,074
23	2,069
24	2,064
25	2,060
26	2,056
27	2,052
28	2,048
29	2,045
30	2,042
31	2,040
32	2,037
33	2,035
34	2,032
35	2,030
36	2,028
37	2,026
38	2,024
39	2,023
40	2,021
41	2,020
42	2,018
43	2,017
44	2,015
45	2,014
46	2,013
47	2,012
48	2,011
49	2,010
50	2,009
55	2,004
60	2,000
70	1,994

Table 3 *(continued)*

80	1,990
90	1,987
100	1,984

7.3.2.2 Method B

Using both the alternative method as defined in Clause 6 and the International Standard method, analyse in duplicate a series of n routine samples, where n is at least 10, under repeatability conditions.

Calculate the mean difference between the methods, $\bar{d}$, from the differences between the means of the duplicates for each sample, taking account of the sign of the differences. Calculate the standard deviation of the differences, s_d.

Calculate the statistic t_c from Formula (9).

Compare t_c with the value of t_t from Table 3 at (n − 1) degrees of freedom.

If $t_c > t_t$, then it can be concluded that the alternative method is biased in comparison with the International Standard method.

Formula (9) implies that the sensitivity of the test to detect any bias is dependent on the number of analyses, n, and the standard deviation, s_d.

If the estimated bias is d, then the true bias will lie, with 95 % confidence, within the limits calculated from Formula (10):

$$\bar{d} \pm t_t \cdot \frac{s_d}{\sqrt{n}} \tag{10}$$

If $t_c \leq t_t$, these limits will encompass the value zero, indicating that the test has not been able to find significant bias at the 95 % confidence level used for the test. However, if the value at either confidence limit is a level of bias that is unacceptable, such a conclusion is ambiguous.

7.3.3 Estimation of precision

Calculate the variance of the duplicate results, V_{ISO} and V_{ALT}, for both methods, using Formula (11).

$$v = \frac{\sum d^2}{2n} \tag{11}$$

where

d is the difference between duplicate pairs;

n is the number of pairs.

Compare the International Standard method variance, V_{ISO}, with the alternative method variance V_{ALT}.

If $V_{ISO} \geq V_{ALT}$, then it can be concluded that the precision of the alternative method is as good, or better than that of the International Standard method.

If $V_{ISO} < V_{ALT}$, calculate the value of F_c from Formula (8).

Compare $F_c = \frac{V_{ALT}}{V_{ISO}}$ with the value in Table 2 at (n − 1) degrees of freedom, F_t.

If $F_c > F_t$, then the precision for the alternative method is worse than that of the International Standard method.

NOTE A worse precision does not necessarily invalidate the alternative method. Precision can be improved by increasing the number of determinations on a sample and taking the mean. Thus, if the precision of a single result is P, then the precision of the mean of n determinations will be $P/\sqrt{n}$. For methods which are automated or more rapid, this might still represent improved efficiency over doing a single analysis using the International Standard method.

7.4 Comparison using the International Standard method within a laboratory — Range of fuels

Where a range of fuels is to be analysed, it should not be assumed that either the precision or the trueness will be consistent across that range. If the alternative method uses a measurement principle different from that of the International Standard method, it is possible that different matrices will have differing effects on the trueness, even to the extent of producing biases in opposite directions in different parts of the range. The relationship between the analyte concentration and precision and accuracy might also vary. A single experiment on a series of samples across the range is unsatisfactory because of the possibility that, across the range, the results might indicate that precision and trueness are no different from those of the International Standard method, whereas in fact, for individual fuels within the range, that is not the case. For example, positive and negative biases might cancel each other out.

Separate tests of the type given in 7.3.2 and 7.3.3 should therefore be carried out on representative fuels across the range. It is possible that an alternative method, while satisfactory for some fuels, is unsatisfactory for others.

7.5 Comparison using the International Standard Method, between laboratories

If reference materials are not available or appropriate and the laboratory wishing to validate an alternative method does not have the equipment for the International Standard method, another approach is to compare results using the alternative method with those on the same samples from another laboratory using the International Standard method.

The test samples should be prepared by one laboratory in the manner indicated in 7.3.1. In order to avoid differential loss of sample integrity, it should be arranged that analyses on the same samples are carried out within an agreed period of 24 h.

In all other respects, the procedures given in 7.3 should be followed.

8 Validation report

Before the alternative method is used as the equivalent of the International Standard method, a validation report should be produced. This should contain the following information:

a) definition of the alternative method which has been validated (see Clause 6);

b) method of validation (see 7.2, 7.3, 7.4 or 7.5);

c) details of the reference material(s) used (see 7.2) or International Standard method (see 7.3, 7.4 or 7.5), where there is a choice of procedures;

d) raw analytical data and the statistical analysis;

e) mean difference between the alternative method and the reference material or between methods, and the confidence interval of that difference;

f) precision of the alternative method and the repeatability critical difference;

g) clear statement of conclusion, including any restriction on the analyte range within which the method is valid.

Annex A
(informative)

Explanation of rationale of method A

The outcome of a series of analyses in order to estimate bias, whether by repeat analysis of a reference material or by using two different methods on a series of samples, will be a set of differences. It is assumed that these differences are drawn from a normal distribution whose mean is the true bias and whose standard deviation is estimated by the standard deviation of the set, s_d. The mean of the set, $\bar{d}$, is an estimate of the true bias.

If the true bias is zero, then the expected values of $\bar{d}$ will be normally distributed about zero with 95 % confidence limits of $0 \pm ts_\mathrm{d}/\sqrt{n}$ where n is the number of analyses making up the set.

NOTE This is a two-sided distribution and the value for t is taken from a table of Student's t at 97,5 % confidence limits.

Under the normal protocol of a null hypothesis test, unless the value of $\bar{d}$ lies outside these limits, it is concluded that the hypothesis is not disproved, i.e. it cannot be concluded that bias is other than zero.

It is also possible to test whether or not the bias is less than some predetermined value of the maximum tolerable bias, B. If the true bias equals B, then the expected values for $\bar{d}$ will be normally distributed about B with a lower confidence limit of $B - ts_\mathrm{d}/\sqrt{n}$. This is a single-sided distribution and t is drawn from a table of Student's t at 95 % confidence limits. Unless the value of $\bar{d}$ is lower than this limit, the hypothesis "bias = B" is not disproved and hence it cannot be concluded that the true bias is less than B.

If n analyses are carried out, then three possible alternatives exist with regard to the location of the confidence limits. Let the upper confidence limit of the "bias = 0" distribution be z and the lower limit of the "bias = B" distribution be y. The locations are represented graphically on the right in Figure A.1.

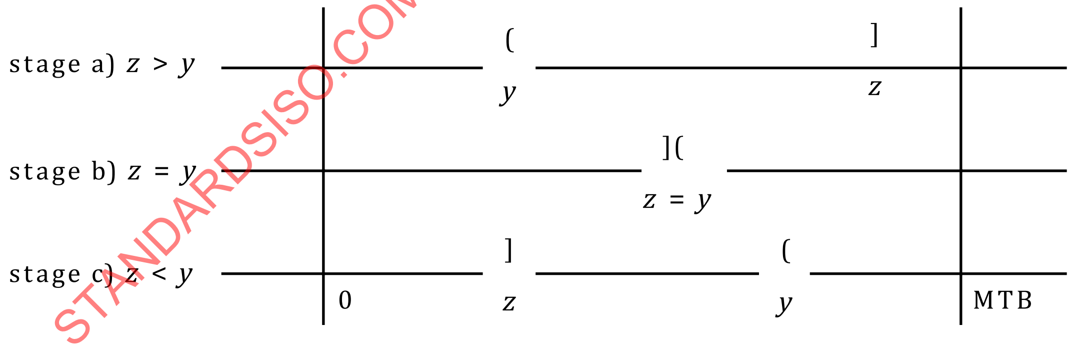

Figure A.1 — Changes in the relationship between the upper and the lower confidence limit with increasing *n*

Assuming that the value for s_d remains constant, then as n increases so the situation corresponds successively to stage a), then b) and c) in Figure A.1.

The ideal situation is stage b) because wherever the value of $\bar{d}$ lies, the conclusion will be unambiguous. There are two possibilities:

a) if $\bar{d} > (z = y)$, then the bias is greater than zero and not less than B; or

b) if $\bar{d} < (z = y)$, then the bias is less than B and not greater than zero.

When an insufficient number of analyses has been carried out and situation a) exists, it is possible that, if $\bar{d}$ falls in the overlapping region ($z > \bar{d} > y$), the ambiguous conclusion is reached that the bias is neither greater than zero, nor less than B. If B is the maximum level of bias which the method users can tolerate, then this result will be unsatisfactory. Method A, by using Table 1 for g, estimates how many analyses are needed to achieve stage b). Unfortunately, since the estimate of the value for s_d used will change as more analyses are done, it is unlikely that b) is achieved exactly, but eventually stage c) will be reached.

At stage c), when $\bar{d}$ falls in the “gap” between the limits, a different problem is posed. This will need to be resolved. Either the null hypothesis or the “bias = B” hypothesis should take precedence, or possibly a judgement made on which side of the equiprobability value $\bar{d}$ lies.

Annex B

(informative)

Example of calculations

B.1 Comparison with a reference material

B.1.1 General

An alternative method is to be examined for determination of sulfur. A reference material has been obtained with a reference value of 1,04 %.

B.1.2 Method A

B.1.2.1 Number of analyses

A maximum tolerable bias of 0,02 % has been decided.

Calculate the standard deviation implied from the critical repeatability difference, i.e. 0,05 %, given in the International Standard method, using Formula (1).

$$s_{\mathrm{ISO}} = \frac{0{,}05}{2\sqrt{2}} = 0{,}0177$$

Calculate the value of g using Formula (2).

$$g = \frac{0{,}02}{0{,}0177} = 1{,}130$$

From Table 1, look up the value of n corresponding to g, i.e. 13.

Analyse the reference material 13 times using the alternative method. The results are given in Table B.1, together with the standard deviation of the results, s_{ALT}. Recalculate g from Formula (2) using the value of s_{ALT}.

$$g = \frac{0{,}02}{0{,}0221} = 0{,}905$$

From Table 1, the new value for n is 18.

Carry out 5 additional analyses. The results are given as 14 through 18 in Table B.1, together with the new value of s_{ALT}, i.e. 0,0204. Recalculate g from Formula (2) using the value of s_{ALT}.

$$g = \frac{0{,}02}{0{,}0204} = 0{,}980$$

From Table 1 the new value for n is now 15, which is less than the number of analyses already done. Therefore no additional analyses are required.

Table B.1 — Comparison with reference material — Examples data for Method A

Analysis number	Result	Result squared
1	1,06	1,124
2	1,05	1,103
3	1,04	1,082
4	1,02	1,040
5	1,06	1,124
6	1,02	1,040
7	1,09	1,188
8	1,03	1,061
9	1,03	1,061
10	1,06	1,124
11	1,08	1,166
12	1,04	1,082
13	1,03	1,061
sum 1	13,61	14,254 5
mean 1	1,047	—
standard deviation, (s_{ALT}) 1	0,022 1	—
14	1,06	1,123 6
15	1,02	1,040 4
16	1,04	1,081 6
17	1,03	1,060 9
18	1,05	1,102 5
sum 2	18,81	19,663 5
mean 2	1,045	—
standard deviation, (s_{ALT}) 2	0,020 4	—

B.1.2.2 Trueness

Calculate the mean value of the 18 measurements, i.e. 1,045, the difference of the mean from the reference value, i.e. 0,005 and the value for t_c from Formula (3):

$$t_c = 0{,}005 \times \frac{\sqrt{18}}{0{,}020\,4} = 1{,}040$$

Since the value of t_c is less than the value of t from Table 3 (2,110), it can be concluded that the method is unbiased.

B.1.2.3 Precision

Calculate the F ratio from Formula (8).

$$F_c = \frac{0{,}020\,4^2}{0{,}017\,7^2} = 1{,}328$$

Since the value of F_c is less than the value of F from Table 2, it can be concluded that the precision of the alternative method is not worse than the International Standard Method.

B.1.3 Method B

B.1.3.1 Number of analyses

The sample is analysed five times using the proposed alternative method. The results are given in Table B.2, with the mean using Formula (6), difference of the mean from the reference value and standard deviation using Formula (3).

$$\text{mean} = \frac{5{,}240}{5} = 1{,}048$$

$$s_{\text{ALT}} = \sqrt{\left(\frac{5{,}493 - \frac{5{,}240^2}{5}}{4}\right)} = 0{,}019\,2$$

Table B.2 — Comparison with reference material — Method A example data

Analysis number	x	x_2
1	1,06	1,124
2	1,05	1,103
3	1,04	1,082
4	1,02	1,040
5	1,07	1,145
sum	5,24	5,493

B.1.3.2 Trueness

Calculate the test statistic t_c from Formula (5).

$$t_c = \frac{0{,}008\sqrt{5}}{0{,}019\,2} = 0{,}932$$

Compare the value of t_c with the value of t found in Table 3 for four degrees of freedom ($n - 1$), i.e. 2,780. Since t_c is less than the table value, it is concluded that no bias has been established.

Calculate the confidence interval for the value of the mean from Formula (7).

$$0{,}008 \pm 2{,}780 \times \frac{0{,}019\,2}{\sqrt{5}} = 0{,}008 \pm 0{,}024$$

The true bias lies between +0,032 % and −0,016 %.

B.1.3.3 Precision

Calculate the standard deviation implied from the critical repeatability difference, 0,05 % given in the International Standard method from Formula (1).

$$s_{\text{ISO}} = \frac{0{,}05}{2\sqrt{2}} = 0{,}017\,7$$

Compare this with the standard deviation of the results, s_{ALT}.