INTERNATIONAL STANDARD

ISO 4259-2

First edition
2017-11

Petroleum and related products — Precision of measurement methods and results —

Part 2: Interpretation and application of precision data in relation to methods of test

Produits pétroliers — Fidélité des méthodes de mesure et des résultats —

Partie 2: Application des valeurs de fidélité relatives aux méthodes d'essai

Reference number
ISO 4259-2:2017(E)

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

Contents

Page

Foreword

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

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

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

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

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

This document was prepared by Technical Committee ISO/TC 28, *Petroleum and related products, fuels and lubricants from natural or synthetic sources.*

This first edition of ISO 4259-2, together with ISO 4259-1, cancels and replaces ISO 4259, which has been technically revised. This document provides the content of Clauses 7 to 10 of ISO 4259 and connected Annexes H and I. The remaining Clauses and Annexes A to G of ISO 4259:2006 are replaced by ISO 4259-1.

A list of all parts in the ISO 4259 series can be found on the ISO website.

Introduction

For purposes of setting product specifications, and to check product compliance against these specifications, standard test methods are usually referenced for specific properties of commercial petroleum and related products. Two or more measurements of the same property of a specific sample by a specific test method, or by different test methods that purport to measure the same property, will not usually give exactly the same result. It is, therefore, necessary to take proper account of this fact when setting product specifications, assessing if the differences between test results are within statistical expectation, and making specification compliance decisions based on limited test results. By using statistically-based estimates of the precision for a test method, the following can be achieved:

— an objective measure of the reliability of specification limits,

— a specification compliance decision, and

— the degree of agreement expected between two or more results obtained in specified circumstances.

This document describes the applications of the precision of test method as derived from ISO 4259-1. It is intended to be a companion document to ISO 4259-1. Additional normative and informative discussions on how to use this precision to assess the "in statistical control" status and precision capability of a specific laboratory in the execution of a test method are provided. Also, the general approach to the agreement between two different test methods that purport to measure the same property are given.

The two parts of ISO 4259 encompass both the determination of precision estimates and the application of precision data. It attempts to be aligned with ASTM D6300[1] regarding the determination of the precision estimates and with ASTM D3244[2] for the utilization of test data.

A glossary of the variables used in this document and ISO 4259-1 is included in ISO 4259-1:2017, Annex I.

Petroleum and related products — Precision of measurement methods and results —

Part 2: Interpretation and application of precision data in relation to methods of test

1 Scope

This document specifies the methodology for the application of precision estimates of a test method derived from ISO 4259-1. In particular, it defines the procedures for setting the property specification limits based upon test method precision where the property is determined using a specific test method, and in determining the specification conformance status when there are conflicting results between supplier and receiver. Other applications of this test method precision are briefly described in principle without the associated procedures.

The procedures in this document have been designed specifically for petroleum and petroleum-related products, which are normally homogeneous. However, the procedures described in this document can also be applied to other types of homogeneous products. Careful investigations are necessary before applying this document to products for which the assumption of homogeneity can be questioned.

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 4259-1, *Petroleum and related products — Precision of measurement methods and results — Part 1: Determination of precision data in relation to methods of test*

3 Terms and definitions

For the purposes of this document, the terms and definitions in ISO 4259-1 and the following 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
proficiency testing program
PTP
program designed for the periodic evaluation of participating laboratories' testing capability of a Standard Test Method through the statistical analysis of their test results obtained on aliquots prepared from a single batch of homogeneous material

Note 1 to entry: The frequency of such testing varies in accordance with the program objective. Each execution of testing involves testing of a single batch of material. Materials typically vary from test to test.

Note 2 to entry: This is also commonly referred to as Inter Laboratory Cross Check Program (ILCP).

3.2
recipient
individual or organization who receives or accepts a product delivered by the supplier

3.3
supplier
individual or organization responsible for the quality of a product just before it is taken over by the recipient

4 Application and significance of repeatability, *r*, and reproducibility, *R*

4.1 General

The value of these quantities is estimated from analysis of variance (two-factor with replication) performed on the results obtained in a statistically designed inter-laboratory programme in which different laboratories each test a range of samples. Repeatability and reproducibility values estimated in accordance with ISO 4259-1 or other statistical techniques shall be included in each published test method.

NOTE See Annex A for an account of the statistical reasoning underlying the formulae in this clause.

In the following clauses, it is assumed that the result(s) are obtained from a test method that is in statistical control. For determination of "in statistical control", see Annex C.

4.2 Repeatability, *r*

4.2.1 General

Most laboratories do not carry out more than one test on each sample for routine quality control purposes except in some circumstances, such as in cases of dispute or if the test operator wishes to confirm that his technique is satisfactory. In such circumstances, when multiple results are obtained, it is useful to check the consistency of repeated results against the repeatability of the method. The appropriate procedure is outlined in 4.2.2. It is also useful to know what degree of confidence may be placed on the average results, and the method of determining this is given in 4.2.3.

4.2.2 Acceptability of results

When only two results are obtained under repeatability conditions and their difference is less than or equal to *r*, the test operator may consider his work as being under control and may take the average of the two results as the estimated value of the property being tested.

If the two results differ by more than *r*, both shall be considered as suspect and at least three more results obtained. Including the first two, the difference between the most divergent result and the average of the remainder shall then be calculated and this difference compared with a new value, r_1, instead of *r*, given in Formula (1):

$$r_1 = r\sqrt{\frac{k}{2(k-1)}} \tag{1}$$

where k is the total number of results obtained.

If the difference is less than or equal to r_1, all the results shall be accepted. If the difference exceeds r_1, the most divergent result shall be rejected and the procedure specified in this section repeated until an acceptable set of results is obtained.

The average of the acceptable results shall be taken as the estimated value of the property. However, if two or more results from a total of not more than 20 have been rejected, the operating procedure and the apparatus shall be checked and a new series of tests made, if possible.

4.2.3 Confidence limits calculations using results collected under repeatability conditions

When a single test operator, who is working within the precision limits of the method, obtains a series of k results under repeatability conditions, giving an average, $\overline{X}$, and the results meet the repeatability requirement in 4.2.2, it can be assumed with 95 % confidence that the true value, μ, of the characteristic lies within the following limits:

$$\left(\overline{X}-\frac{R_1}{\sqrt{2}}\right)\leq\mu\leq\left(\overline{X}+\frac{R_1}{\sqrt{2}}\right) \tag{2}$$

where

$$R_1=\sqrt{R^2-r^2\left(1-\frac{1}{k}\right)} \tag{3}$$

When $k = 1$, use the single test result as the value for the $\overline{X}$ term as follows:

$$\left(X-\frac{R}{\sqrt{2}}\right)\leq\mu\leq\left(X+\frac{R}{\sqrt{2}}\right) \tag{4}$$

where R is the published test method reproducibility as discussed in 4.3.

Similarly, for the single limit situation, when only one limit is fixed (upper or lower), it can be assumed with 95 % confidence that the true value, μ, of the characteristic is limited as follows:

$$\mu\leq\left(\overline{X}+0{,}59R_1\right)\quad\text{(upper limit)} \tag{5}$$

or

$$\mu\geq\left(\overline{X}-0{,}59R_1\right)\quad\text{(lower limit)} \tag{6}$$

The factor 0,59 is the ratio $0{,}84/\sqrt{2}$, where 0,84 is derived in Annex A.

When r is much smaller than R, little improvement in the precision of the average is obtained by carrying out multiple testing under repeatability conditions.

4.3 Reproducibility, R

4.3.1 Acceptability of results

The procedure specified in 4.3 is intended for judging the acceptability, with respect to the reproducibility of the test method, of results obtained by different laboratories in normal, day-to-day operations and transactions. In cases of dispute between a supplier and a recipient, the procedure specified in Clauses 5 to 7 shall be adopted.

When single results are obtained in two laboratories and their difference is less than or equal to R, the two results shall be considered as acceptable, and used to calculate the average $\overline{X}$. The average $\overline{X}$, rather than either single result separately, shall be used as the estimated value of the tested property.

The true value μ of the characteristic is contained within the following limits with a 95 % confidence:

$$\left(\bar{X}-\frac{R}{2}\right)\leq\mu\leq\left(\bar{X}+\frac{R}{2}\right) \tag{7}$$

Similarly for the single limit situation, when only one limit is fixed (upper or lower), the true value μ of the characteristic is contained with the following limits with a 95 % confidence:

$$\mu\leq\left(\bar{X}+0{,}42R\right) \quad \text{(upper limit)} \tag{8}$$

or

$$\mu\geq\left(\bar{X}-0{,}42R\right) \quad \text{(lower limit)} \tag{9}$$

The factor 0,42 is the ratio of $0{,}59/\sqrt{2}$ as it is an average of two results.

If the two results differ by more than R, both shall be considered as suspect. Each laboratory shall then obtain at least three other acceptable results (see 4.2.2).

In this case, the difference between the averages of all acceptable results of each laboratory shall be judged for conformity using a new value, R_2, instead of R, as given by Formula (10):

$$R_2=\sqrt{R^2-r^2\left(1-\frac{1}{2k_1}-\frac{1}{2k_2}\right)} \tag{10}$$

where

R is the reproducibility of the method;

r is the repeatability of the method;

k_1 is the number of results of the first laboratory;

k_2 is the number of results of the second laboratory.

If the difference between the averages is less than or equal to R_2, then these averages are acceptable and their overall average shall be considered as the estimated value of the tested property. If the difference between the averages is greater than R_2, and there is a dispute on the specification conformance of the tested property, then the procedure specified in Clause 7 shall be adopted.

If circumstances arise in which there are more than two laboratories, each supplying one or more acceptable results, the difference between the most divergent laboratory average and the average of the remaining N laboratory averages shall be compared to R_3:

where

$$R_3=\sqrt{\frac{R_1^2}{2}+\frac{R_4^2}{2N}} \tag{11}$$

$$R_4=\sqrt{R^2-\frac{r^2}{N}\left(N-\frac{1}{k_1}-\frac{1}{k_2}-...-\frac{1}{k_N}\right)} \tag{12}$$

R_1 is given in Formula (3), and corresponds to the most divergent laboratory average.

If this difference is equal to or less than R_3 in absolute value, all results shall be regarded as acceptable and their average taken as the estimated value of the property.

If the difference is greater than R_3, the most divergent laboratory average shall be rejected and the comparison using Formulae (11) and (12) repeated until an acceptable set of laboratory averages is obtained. The average of these laboratory averages shall be taken as the estimated value of the property. However, if two or more laboratory averages from a total of not more than 20 have been rejected, the operating procedure and the apparatus shall be checked and a new series of tests made, if possible.

4.3.2 Confidence limits calculations using results collected under reproducibility conditions

When N laboratories obtain one or more results under conditions of repeatability and reproducibility, giving an average of laboratory averages $\overline{X}$, the true value μ of the characteristic is contained within the following limits with 95 % confidence:

$$\overline{X} - \frac{R_4}{\sqrt{2N}} \le \mu \le \overline{X} + \frac{R_4}{\sqrt{2N}} \qquad (13)$$

Similarly for the single limit situation, when only one limit is fixed (upper or lower), the true value μ of the characteristic is contained with the following limits with 95 % confidence:

$$\mu \le \left(\overline{X} + 0{,}59 \frac{R_4}{\sqrt{N}} \right) \quad \text{(upper limit)} \qquad (14)$$

or

$$\mu \ge \left(\overline{X} - 0{,}59 \frac{R_4}{\sqrt{N}} \right) \quad \text{(lower limit)} \qquad (15)$$

4.4 Use of reproducibility to determine bias between two different test methods that purport to measure the same property

4.4.1 General

For the situation where two different test methods purport to measure the same property, the reproducibility estimates (R) from the respective test methods shall be used in conjunction with the averages obtained from multiple laboratories for the same material to determine if a bias correction can be applied to improve statistically the agreement between the two methods for that material. For example, results collected through Proficiency Testing Programs (PTP) for different test methods using the same sample can be analysed in this fashion.

NOTE Discussion on methodology for this type of assessment for the simultaneous analysis of multiple materials / property levels that span the intersecting scope of two different test methods is beyond the scope of this document. Interested readers are encouraged to consult ASTM D6708[3] for a detailed presentation on the subject. Annex D provides a brief overview on the general statistical approach for the aforementioned situation.

4.4.2 Process

Assume that Test Method A and Test Method B are test methods that purport to measure property C.

Calculate the following statistic:

$$Z = \frac{|\overline{Y}_A - \overline{Y}_B|}{\sqrt{\dfrac{R_A^2}{7{,}683 L_A} + \dfrac{R_B^2}{7{,}683 L_B}}} \qquad (16)$$

where

Y_A is the average from L_i results for property C for a material using Test Method A, where each result is a single result obtained under reproducibility conditions;

L_A is the total number of laboratories (results) for Test Method A and should be >20;

R_A is the reproducibility of Test Method A;

Y_B is the average from L_B results for property A using Test Method B on the same material tested by Test Method A, where each result is a single result obtained under reproducibility conditions;

L_B is the total number of laboratories (results) for Test Method B, and should be >20;

R_B is the reproducibility of Test Method B.

If Z > 2, it shall be concluded, with 95 % confidence, that a constant bias correction statistically improves the degree of agreement between Test Method 1 and Test Method 2 for property C for this material.

5 Specifications

5.1 Aim of specifications

The purpose of a specification is to specify an acceptable limit or limits to the true value, μ, of the property as determined by a specified test method. In practice, however, this true value can never be established exactly since the results obtained by applying the specified test method in a single or multiple laboratories can show acceptable scattering as defined by the repeatability and reproducibility. There is, therefore, always some uncertainty as to the true value of the tested property determined from a finite number of test results.

Petroleum product compliance with specifications is assessed in accordance with Clauses 6 and 7. By prior agreement a supplier and recipient may use the alternative procedures described in Annex B.

It is important that the test method specified for the determination of the property governed by the specification limit(s) is sufficiently precise to reliably determine whether or not the product meets the specifications.

5.2 Construction of specifications limits in relation to scope and precision of the specified test method

The specification limits shall not be outside the method scope limits as determined in ISO 4259-1.

The lower specification limit shall not be less than the lower scope limit of the test method, and the upper specification limit shall not be greater than the upper scope limit of the test method (see §6.5 in ISO 4259-1).

In addition, the distance between lower and upper specification limit shall also satisfy the following condition: upper specification limit minus lower specification limit shall not be less than the quantity 2R evaluated at lower method scope limit plus 2R evaluated at upper method scope limit. See Figure 1 for an illustration of this concept.

Usually, specifications deal with limits for the values of the properties. To avoid uncertainty, such limits are normally expressed as “not less than” or “not greater than”. Limits are of two types:

— a double limit, upper and lower, for example viscosity not less than 5 mm^2/s and not greater than 16 mm^2/s; boiling point 100 °C ± 0,5 °C;

— a single limit, upper or lower, for example water content not greater than 2 %; sulfur content not greater than 10 mg/kg ; solubility of bitumen not less than 99 %.

