FINAL
DRAFT

INTERNATIONAL
STANDARD

**ISO/FDIS
4259-4**

ISO/TC **28**

Secretariat: NEN

Voting begins on:
2021-09-03

Voting terminates on:
2021-10-29

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

Part 4:
Use of statistical control charts to validate 'in-statistical-control' status for the execution of a standard test method in a single laboratory

Produits pétroliers et connexes — Fidélité des méthodes de mesure et de leurs résultats —
Partie 4: Utilisation de cartes de contrôle statistique pour valider l'état 'sous maîtrise statistique' pour l'exécution d'une méthode d'essai normalisée dans un seul laboratoire

ISO/CEN PARALLEL PROCESSING

Reference number
ISO/FDIS 4259-4:2021(E)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

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 of 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 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*, in collaboration with the European Committee for Standardization (CEN) Technical Committee CEN/TC 19, *Gaseous and liquid fuels, lubricants and related products of petroleum, synthetic and biological origin*, in accordance with the Agreement on technical cooperation between ISO and CEN (Vienna Agreement).

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

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

In the current global business environment, measurement data 'trustworthiness' is a key business driver and an implicit expectation from customers and regulatory entities. Data trustworthiness means the data quality meets expectations and is 'fit-for-use'. Trustworthy data can only be produced by measurement systems that are demonstrated to be stable and are under common cause variation only.

This document describes the applications of specific statistical control charts selected from those that are widely used by the manufacturing sector for the purpose of monitoring and demonstrating the in-statistical-control status of a laboratory in the execution of a standardized test method to produce trustworthy data.

In ISO 4259-2[9], the requirement for assessment of product quality conformance to specification, is to be interpreted that each laboratory's test result is obtained from a test method that is in-statistical-control in terms of precision and bias, to be substantiated by in-house statistical quality control (SQC) charts or other equivalent statistical techniques. While in-house techniques are used by many laboratories for test method quality assurance, standardization on how to establish in-statistical-control is necessary to ensure consistency in application of ISO 4259-2[9]. Addressing the aforementioned necessity is the motivation of this document, which is based on ASTM D6299[1].

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

Part 4: Use of statistical control charts to validate 'in-statistical-control' status for the execution of a standard test method in a single laboratory

1 Scope

This document specifies the process and methodology for the construction, operation, and maintenance of statistical control charts to assess if a laboratory's execution of a standard test method is in-statistical-control and how to establish and validate the 'in-statistical-control' status.

It specifies control charts that are most appropriate for ISO/TC 28 test methods where the dominant common cause variation is associated with the long term, multiple operator conditions. The control charts specified for determination of in-statistical-control are: individual (I), moving range of 2 (MR_2), and either the exponentially weighted moving average (EWMA) or zone-based run rules [similar to Western Electric (WE) run rules[3]] as sensitivity enhancement strategy to support the I-chart.

The procedures in this document have been primarily designed for numerical results obtained from testing of control samples prepared from a homogenous source of petroleum and related products in a manner that preserves the homogeneity of properties of interest between control samples. If the test method permits, a certified reference material (CRM) sample is used as a control sample provided the sample composition is representative of the material being tested and is not a pure compound; if this is done then the laboratory best establishes its own mean for the CRM sample.

This document is applicable to properties of interest that are (known to be) stable over time, and for data sets with sufficient resolution to support validation of the assumption that the data distribution can be approximately represented by the normal (Gaussian) model. Mitigating strategies are suggested for situations where the assumption cannot be validated.

2 Normative references

The following documents are referred to in the text in such a way that some of their content support 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 4259-1:2017, *Petroleum and related products — Precision of measurement methods and results — Part 1: Determination of precision data in relation to methods of test*

3 Terms, definitions, symbols and abbreviated terms

For the purposes of this document, the terms and definitions given 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 https://www.electropedia.org/

3.1 Specific terms and definitions

3.1.1
common cause
factors that contribute to *common cause variation* (3.1.2)

Note 1 to entry: See Figure 1 for illustration.

3.1.2
common cause variation
variation amongst results collected under *site precision conditions* (3.1.4) from repeated execution of a test method on the same test material attributable to known, unknown, or unknowable factors that are intentionally not or cannot be rigidly controlled as part of the normal and correct execution of all aspects of the test method

3.1.3
in-statistical-control
situation wherein the test results produced by the user on control samples are reasonably consistent with expectation over time with common cause variation scattered around a stable expected centre

3.1.4
site precision conditions
conditions under which single test results are obtained in time intervals, separated by at least 8 h, by the testing population in a single laboratory executing the same test method using the same apparatus on test specimens taken at random from the same material over the normal daily operating envelope

3.1.5
quality control sample
QC sample
specimen taken from a stable and homogeneous material with composition and properties similar to sample normally tested by the laboratory, prepared in a manner that preserves the homogeneity of property of interest between test specimens, stored in a manner that preserves the properties of interest over time, and available in sufficient quantity for repeated long-term testing

3.2 Symbols and abbreviated terms

AD	Anderson Darling
ARV	assigned reference value
CRM	certified reference material
EWMA	exponentially weighted moving average
GESD	generalized extreme studentized deviation
MR_2	moving range of two
$\overline{MR}_{known}$	moving range average associated with s_{known}
PT	proficiency testing
QC	quality control
q-q	quantile-quantile, term used to describe the plot type comparing the z-score of a data point with is numerical value
s_{known} [a]	statistically pooled standard deviation obtained using final achieved standard deviations from a group of retired control charts where all the final achieved control chart averages are determined to be not statistically significantly different using the appropriate clause(s) in stage 1

[a] The range spanned by the final achieved control chart averages are referred to as the working range associated with s_{known}. See concept illustrated in Table 1.

Table 1 — Statistically pooled standard deviation concept

Material	Property (unit)	s_{known}	df	Working range	$\overline{MR}_{known}$
summer gasoline	vapour pressure (kpa)	0,55	60	49,85 to 50,68	0,62
winter gasoline	vapour pressure (kpa)	0,83	85	104,67 to 105,91	0,93

4 Statistical control in the execution of a standard test method by a laboratory

4.1 General

The execution of a standard test method by a laboratory is in this document considered as the execution of a series of inter-connected work processes. Each work process is subject to variation caused by known, unknown, or sometimes unknowable causes that are inherent to a process over long time horizon such that every outcome of the process is affected. These causes are referred to as common causes. The effect on the final process outcome due to common causes are referred to as common cause variation.

Common causes for variation can be grouped into 5 categories (environment, operator, equipment, procedure and reagent material) using a technique known as a fishbone diagram. Due to common cause variation, repeated execution of the same test method on the same material over a long-time horizon yields results that are not numerically identical. This effect is illustrated in Figure 1.

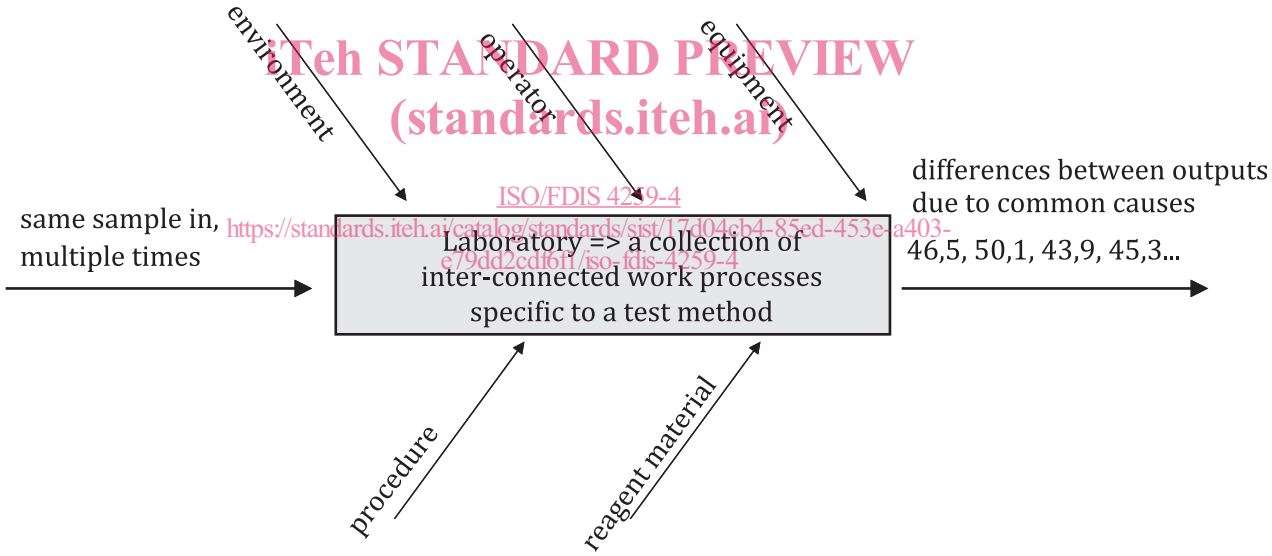

Figure 1 — Fishbone diagram representation of common cause variation in the execution of a test method

The complete process associated with the execution of the specific test method is said to be 'in-statistical-control' if the process outcomes (test results) from repeated analysis of control (QC) samples prepared from the same material are reasonably consistent with expectation over time, with random variation scattered around a stable centre due to common causes only.

To determine 'in-statistical-control', a multi-step and integrated work process involving use of statistical control charts and a quality control (QC) sample is required.

NOTE For simplicity, the word 'statistical' will be omitted and the term referred to as 'control charts' throughout the rest of this document.

4.2 Control chart description

4.2.1 General

Control charts appropriate for most petroleum industry test methods that yield numeric results are the individual (I) chart, and the moving range of 2 (MR_2) chart.

NOTE 1 The I-chart is also known as the X-chart, or the Shewhart chart [2].

NOTE 2 For simplicity, the MR_2 chart is referred to as the moving range (MR) chart throughout the rest of this document.

4.2.2 I- and MR-charts

The I-chart is a graphic display of individual control sample test results (X) collected under site precision conditions, plotted in chronological order, overlaid with a centre line, lower and upper decision limits that require action if exceeded. These limits are herein referred to as I-chart lower and upper control limits (LCL_X, UCL_X). The primary purpose of the I-chart is to monitor process centre stability over time.

NOTE Since the primary interest is to monitor the stability and common cause variation of the test process under site precision conditions over a long time horizon, replicate analysis obtained under repeatability conditions (see ISO 4259-1) does not contribute towards this objective, as the variation due to common causes of interest is not contained in replicate results collected under repeatability conditions.

The MR-chart is the successive difference (with no arithmetic sign) of two individual results in the I-chart, plotted in chronological order, also overlaid with a centre line and an upper decision limit for action, herein referred to as MR-chart upper control limit (UCL_MR). The primary purpose of the MR-chart is to monitor common cause variation stability between successive control sample results over time.

The decision limits for action for both charts as well as the conditions requiring action for the strategies in 4.2.3 are based on a very low theoretical probability (<0,3 %) of "action required" decision for a process that is in-statistical-control, using the Normal distribution (I-chart) and W distribution (MR-chart) as the reference statistical models. Hence, these limits represent the expectation limits for the process outcome if it is in-statistical-control.

4.2.3 I-chart sensitivity enhancement strategy

As a direct consequence of setting the action limits for the I-chart based on a low probability of exceedance for a process that is in-statistical-control, these limits are not sensitive to detection of small changes in the process centre. It is therefore necessary to support I-chart with additional sensitivity enhancement strategies to overcome this shortcoming.

This document requires use of one of the following strategies in conjunction with the I-chart:

a) Strategy 1: Zone-based run rules. Action is required if any of the following run rule conditions is present. For definitions of zones see 4.3.2:

   — two out of three consecutive individual results in Zone A on one side of the centre line;

   — four out of five consecutive individual results beyond Zone C on one side of the centre line;

   — nine consecutive individual results on the same side of the centre line (above or below);

or,

b) Strategy 2: Exponentially weighted moving average (EWMA). Action is required if either one of the following conditions occur:

   — any exceedance of the EWMA action limits;

— nine consecutive individual results on the same side of the centre line (above or below).

The EWMA is a 'time-weighted moving average' calculated using all data points up to the most current one, the weighting of each datum reduced with age exponentially. The rate of this weight decay is controlled by λ. It is re-calculated with the arrival of each new datum and is judged against its own action limits (herein referred to as the EWMA-action limits). An EWMA with λ = 0,4 has similar detection power as Strategy 1[4].

Use of Strategy 2 is recommended due to the ease of implementation and lower expected false alarm rate than Strategy 1.

4.2.4 In-statistical-control conditions

The complete process associated with the execution of a test method is deemed to be in-statistical-control if all of the following conditions are met:

a) all individual control sample results are within the I-chart action limits,

b) less than five out of 12 successive MR results exceed MR-chart upper control limit, and

c) no action required for the sensitivity enhancement chosen (Strategy 1 or 2).

4.3 Control chart work process

4.3.1 General

To determine if the complete process associated with the execution of a test method is 'in-statistical-control', a two-stage multi-step work process involving use of control charts and QC material is specified.

Stage 1 comprises visual and statistical assessment of initial test results for a new batch of QC material plotted in a chronological order (known as a run chart). This is followed by construction of the I-chart and MR-chart using these results by overlaying the mean and the action limits onto the respective charts. The action limits represent the boundaries within which the current and future test results and MR for this QC material are expected to lie, on the assumption that the process is in-statistical-control and the QC material remains unchanged. The control charts (I and MR) constructed in Stage 1 are deployed for Stage 2 if all in-statistical-control conditions (see 4.2.4) are met.

Stage 2 comprises of two modes, operation and maintenance. Under 'operation', future test results (for the QC material tested in Stage 1) as they arrive in chronological order are compared against the established action limits and chosen enhancement strategy in Stage 1. Under 'maintenance', the statistics used in the computation of the control chart action limits from Stage 1 are re-assessed periodically using newly accrued in-statistical-control results and updated as appropriate.

For this practice, the root-mean-square technique is used to compute the sample standard deviation statistic, s.

4.3.2 Stage 1 of control chart work process

The primary objective of Stage 1 is to establish initial means and action limits for the I, MR control charts and implement the chosen enhancement strategy (see 4.3.2) for a specific batch of QC material using chronologically obtained data and the normal distribution as the reference model. Figure 2 is a flow chart of the 15-step process defined in this subclause. The main steps are:

1) Prepare multiple control (QC) samples from a stable and homogeneous material with composition and properties similar to samples normally tested by the laboratory.

2) Collect a minimum of 20 QC sample test results under site precision conditions.

3) Plot the individual results chronologically (this is called a run chart) and study the plot for any visually discernible transcription errors. Correct or discard obvious transcription errors.

4) Construct a quantile-quantile (q-q) plot (as shown in Annex A) and compute the Anderson-Darling (AD) statistic. Examine this plot in conjunction with the data for variation sufficiency. Proper application of the control charts in this document require at least 6 unique values to provide sufficient observable common cause variation. Insufficient variation in the data set will manifest into the following two outcomes:

   a) q-q plot shows several distinct horizontal data clusters where each horizontal cluster represent numerically identical data (see Clause 5),

   b) AD value exceeds the 0,01 sig. level of 1,0 by a large margin.

   If the total number of unique values is less than six, proceed directly to Clause 5 for guidance.

5) Perform a formal statistical assessment for outliers using generalized extreme studentized deviation (GESD) technique for outliers similar to ISO 4259-1 (or refer to ASTM D7915[5]). The recommended maximum number of outliers for 20 to 25 observations is 3 at 0,01 significance.

   Reject identified outliers, obtain replacement results and repeat from step 4).

   While not always possible, it is recommended that rejection of outliers be justified by corresponding root cause(s).

6) Confirm the goodness-of-fit of the normal distribution using the Anderson-Darling (AD) statistic computed from at least 20 non-outlier results.

   — If AD is less than 1,0, continue to step 7);

   — if AD is between 1,0 and 1,5, proceed to Clause 5 for guidance;

   — if AD is greater than 1,5, do not proceed with this document as this is strong statistical evidence that either the system is not in-statistical-control, or, the process results distribution is severely non-normal.

7) Compute the average ($\bar{x}_{\text{stage1}}$) and standard deviation (s_{stage1}) statistics using the non-rejected results from step 5).

8) Assess if additional data from historic results for this test method can be used to improve the estimate of s_{known}, this should be done using an F-test at the 0,025 significance level with the numerically larger number in the numerator. The reproducibility statement function determines how this should be done:

   a) For methods with a constant published reproducibility: use the F-test assess if s_{stage1} is statistically indistinguishable versus s_{known}.

   b) For methods with a non-constant published reproducibility: calculate the reproducibility at the $\bar{x}_{\text{stage1}}$ level and the reproducibility at the mean of the previous chart using $\bar{x}_{\text{known}}$.

      NOTE Annex A shows a worked example including the F-test to pool the sigmas.

      If the ratio of Reproducibility_$\bar{x}_{\text{stage1}}$ / Reproducibility_$\bar{x}_{\text{known}}$ is between 0,85 and 1,15 then use the F-test to assess if s_{stage1} is statistically indistinguishable versus s_{known}.

   If the F-Test to pool s_{stage1} and s_{known} passes, pool both standard deviations (follow the procedure in Annex B). Assign the pooled standard deviation s_{pool} as the standard deviation s_{chart} to be used to construct the limits in step 9).

   If there is no prior information on s_{known} for this type of material, or, if the F-test fails then use s_{stage1} as s_{chart} for step 9).

9) Create the I-Chart for this batch of QC samples by assigning $\bar{x}_{\text{stage1}}$ from step 7) as $\bar{x}_{\text{chart}}$ for this step; then, overlaying the centre line represented by $\bar{x}_{\text{chart}}$, and the two control limits represented by $\bar{x}_{\text{chart}} \pm 3 \cdot s_{\text{chart}}$ onto the run chart in step 3). If the EWMA sensitivity enhancement strategy is used, construct and plot the EWMA line with its associated control limits placed at $\bar{x}_{\text{chart}} \pm 1{,}5 \cdot s_{\text{chart}}$ onto the run chart as well.

10) Label the zones in the I-chart as follows:

   — Zone C: $\bar{x}_{\text{chart}} \pm 1 \cdot s_{\text{chart}}$ (exclusive) - Zone B: $\bar{x}_{\text{chart}} + 1 \cdot s_{\text{chart}}$ (inclusive) to $2 \cdot s_{\text{chart}}$ (exclusive); $\bar{x}_{\text{chart}} - 1 \cdot s_{\text{chart}}$ (inclusive) to $- 2 \cdot s_{\text{chart}}$ (exclusive).

   — Zone A: $\bar{x}_{\text{chart}} + 2 \cdot s_{\text{chart}}$ (inclusive) to $3 \cdot s_{\text{chart}}$ (exclusive); $\bar{x}_{\text{chart}} - 2 \cdot s_{\text{chart}}$ (inclusive) to $- 3 \cdot s_{\text{chart}}$ (exclusive).

11) Compute the *MR* results using all non-rejected results from Step 5).

12) Compute the average using all *MR* results and designate this as $\overline{MR}_{\text{stage1}}$.

13) If s_{pool} is used from step 8), use the weighted average $\overline{MR}_{\text{wtd}}$ computed from $\overline{MR}_{\text{known}}$ corresponding to the s_{known} and $\overline{MR}_{\text{stage1}}$ as the $\overline{MR}_{\text{chart}}$ to be used to construct the MR-chart in step 14). Otherwise, use $\overline{MR}_{\text{stage1}}$ as $\overline{MR}_{\text{chart}}$ for this batch of QC material.

14) Create the MR-chart by overlaying the lines represented by $\overline{MR}_{\text{chart}}$ and upper control limit at 3,27 • $\overline{MR}_{\text{chart}}$ onto a run chart plot of all the MR values computed in step 11).

15) If all of the in-statistical-control conditions (see 4.2.4) are met, proceed to Stage 2. Otherwise, investigate and mitigate root causes for failure, then repeat from step 1).

NOTE See Annex A and B for a detailed illustration of Stage 1 steps 3) to 15) and related statistical techniques.