INTERNATIONAL STANDARD

ISO 13909-8

First edition
2001-12-15

Hard coal and coke — Mechanical sampling —

Part 8: **Methods of testing for bias**

Houille et coke — Échantillonnage mécanique —

Partie 8: Méthodes de détection du biais

Reference number
ISO 13909-8:2001(E)

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.ch
Web www.iso.ch

Printed in Switzerland

Contents

Page

Annex

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 3.

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 part of ISO 13909 may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights.

International Standard ISO 13909-8 was prepared by Technical Committee ISO/TC 27, *Solid mineral fuels*, Subcommittee SC 4, *Sampling*.

ISO 13909 cancels and replaces ISO 9411-1:1994, *Solid mineral fuels — Mechanical sampling from moving streams — Part 1: Coal* and ISO 9411-2:1993, *Solid mineral fuels — Mechanical sampling from moving streams — Part 2: Coke*, of which it constitutes a technical revision. It also supersedes the methods of mechanical sampling of coal and coke given in ISO 1988:1975, *Hard coal — Sampling* and ISO 2309:1980, *Coke — Sampling*.

ISO 13909 consists of the following parts, under the general title *Hard coal and coke — Mechanical sampling*:

— *Part 1: General introduction*

— *Part 2: Coal — Sampling from moving streams*

— *Part 3: Coal — Sampling from stationary lots*

— *Part 4: Coal — Preparation of test samples*

— *Part 5: Coke — Sampling from moving streams*

— *Part 6: Coke — Preparation of test samples*

— *Part 7: Methods for determining the precision of sampling, sample preparation and testing*

— *Part 8: Methods of testing for bias*

Annex A of this part of ISO 13909 is for information only.

Introduction

It is not possible to lay down a standard method for field work by which a sampling procedure can be tested for bias because details of the procedure will inevitably be affected by local conditions. However, certain principles can be specified which should be adhered to whenever possible and these are discussed in this part of ISO 13909.

Testing for bias can be a tedious and expensive process, especially if testing of the primary increment sampler is included. All bias tests therefore include a thorough pre-test inspection, with appropriate action taken regarding any system deficiencies likely to cause bias.

Hard coal and coke — Mechanical sampling —

Part 8:
Methods of testing for bias

1 Scope

This part of ISO 13909 sets out principles and procedures for testing the bias of test samples of hard coals or cokes, taken in accordance with other parts of ISO 13909. The use of univariate statistical methods only is addressed.

The user is cautioned that the chance of falsely concluding that there is a bias, when no bias exists in any one of several variables measured on the same set of samples, is substantially greater than for a single variable. While several variables may be measured, the single variable on which the outcome of the test will be governed shall be designated in advance.

NOTE In the text the term 'fuel' is used where both coal and coke would be applicable in the context and either 'coal' or 'coke' where only one is applicable.

2 Normative references

The following normative documents contain provisions which, through reference in this text, constitute provisions of this part of ISO 13909. For dated references, subsequent amendments to, or revisions of, any of these publications do not apply. However, parties to agreements based on this part of ISO 13909 are encouraged to investigate the possibility of applying the most recent editions of the normative documents indicated below. For undated references, the latest edition of the normative document referred to applies. Members of ISO and IEC maintain registers of currently valid International Standards.

ISO 13909-1:2001, *Hard coal and coke — Mechanical sampling — Part 1: General introduction*.

ISO 13909-2:2001, *Hard coal and coke — Mechanical sampling — Part 2: Coal — Sampling from moving streams*.

ISO 13909-4:2001, *Hard coal and coke — Mechanical sampling — Part 4: Coal — Preparation of test samples*.

ISO 13909-5:2001, *Hard coal and coke — Mechanical sampling — Part 5: Coke — Sampling from moving streams*.

ISO 13909-6:2001, *Hard coal and coke — Mechanical sampling — Part 6: Coke — Preparation of test samples*.

ISO 13909-7:2001, *Hard coal and coke — Mechanical sampling — Part 7: Methods for determining the precision of sampling, sample preparation and testing*.

3 Terms and definitions

For the purposes of this part of ISO 13909, the terms and definitions given in ISO 13909-1 apply.

4 Principles

4.1 General

Testing for bias is not always done for a single purpose. The objectives of testing for bias may be for assessing conformity with contractual requirements, such as guarantees or purchase and acceptance specifications, or for diagnostic purposes or both, and may or may not involve multiple test parameters. This part of ISO 13909 uses univariate statistics for testing the performance of the system with respect to a single variable.

It is not possible for any scheme of sampling, or sample preparation and analysis, to be free of errors of measurement. For this reason, no statistical test can establish that there is no bias, but only that there is not likely to be a bias of more than a certain magnitude.

The testing of a sampling system for bias is based on taking a series of pairs of samples of essentially the same fuel; one member of each pair being sampled by the system or component under test, the other member being obtained by a reference method. For each pair, the difference between the analytical results is determined. The series of differences between the analytical results thus obtained are subjected to statistical analysis.

The procedure requires the sensitivity of the statistical test of significance to be such that the minimum bias that can be detected is less than or equal to the maximum tolerable bias, B. Therefore, B shall be established before the test begins.

NOTE In the absence of other information, a value of $B =$ 0,20 % to 0,30 % for ash or moisture may be appropriate, subject to commercial considerations.

The sensitivity of the statistical test used is dependent on the number of pairs compared and the variability of the differences between them.

The statistical analysis to which results will be subjected assumes three conditions:

— a normal distribution of the variable;

— independence of the errors of measurement;

— statistical homogeneity of the data.

The closeness with which these ideal conditions are achieved, in practice, governs the validity of the statistical analysis. The execution of the test, including sample reduction, division and laboratory analysis, shall be organized so as to ensure that deviations from these ideals do not invalidate the statistical analysis.

The statistical test used to make the final judgement is the t-test. A hypothesis is made that the observed mean of the differences between the two methods is drawn from a population whose mean is B. If the test shows that the observed difference is significantly less than B, then the sampler or component is declared free of bias.

In basing decisions on the outcomes of statistical tests, there is always the risk of making either one or the other of two types of error. If the hypothesis is rejected when it is true, e.g. a bias is not declared even though a bias really does exist, then an error of the first kind (Type I) has been made. On the other hand, if the hypothesis is accepted when it is false, e.g. a bias is declared even though a bias really does not exist, then an error of the second kind (Type II) has been made.

In any particular test, the probability of a Type I error can be arbitrarily set as a matter of discretion and the risk kept as small as desired. For a specific test, the probability of an error of Type II can only be quantified in relation to some other value than the original hypothesis. In this method the value of zero is used. The risk of a Type II error can be decreased at a fixed probability of an error of Type I only by increasing the number of observations. However, since the sample estimate of the population standard deviation must be used in the calculations, the risk of a Type II error is an estimated value. The final statistical test is not carried out until sufficient pairs of observations have been taken to limit both a Type I error in relation to B, and the estimated Type II error in relation to zero, to 5 %. Thus, if the observed value of the mean difference (the sample estimate of the population mean) is not significantly less than B, it shall also be significantly greater than zero.

The number of paired samples suitable for a test of the overall system relative to the maximum tolerable bias, B, may be insufficient for testing a given component. In such circumstances, if the performance of a given component is of critical importance, a separate test shall be considered. For components other than the primary increment sampler, such tests can usually be implemented with minimum disruption of normal operations and at a lesser cost than for a test for overall system bias (see 7.2 and 7.3).

If obtaining the number of pairs required is found to be impracticable, changes will need to be made to reduce the within-set variance. Investigate what improvements can be made in the closeness of members of pairs and/or what reduction can be achieved in the preparation and testing errors. If such improvements or reductions are not possible, give consideration to increasing the number of increments in the samples, taking into account the practical problems associated with taking increments and the relative costs and errors involved in sampling, sample preparation and testing.

If the required number of pairs of samples is still excessive, the maximum tolerable bias, B, may be reviewed.

When samples of more than one increment are compared, it is necessary that the reference samples and the samples from the system under test be constituted on the same basis, i.e. for a time-basis system, the individual increment masses shall be proportional to the flow rate and, for mass-basis systems, the individual increment mass shall be uniform (see ISO 13909-2 and ISO 13909-5).

4.2 Selection of sample pairs

4.2.1 Composition of sample pairs

The members of each pair of samples can each comprise one or more increments. Individual increments can be compared or samples compounded from increments taken by the two methods. The test shall be structured so that the expected mean of the differences of the result would be zero if no systematic error is present in the system or component under test.

4.2.2 Paired-increment samples

Paired-increment experimental design is the comparing of individual primary increments after being processed by the system, with the reference samples collected from the stopped belt.

For a given parameter, the variance of the differences between paired samples will normally be smaller than the variance of either of the two series of samples, taken by the system or component under test and the reference method respectively, except for fuels that are very homogeneous. For this reason, if the increments taken by the two methods are taken in close proximity to each other in the fuel stream (without overlapping), the variance of the differences between them will be minimized and the sensitivity of the test improved.

4.2.3 Paired-batch samples

It is often not practicable to obtain single increment samples from the system. Increments taken by the system can be compounded as samples, and compared with samples compounded from increments taken over the same period using the reference method. It is not necessary that the two samples have the same number of increments or that they are of similar mass. In the extreme, a single stopped-belt reference increment could be used as the reference sample.

4.3 Location of sampling points

For a test of the overall system, the reference sample shall be taken from the primary fuel stream using the stopped-belt reference method (see clause 7). The system sample shall be the final sample.

The primary sampler shall be tested by examining the differences between members of each pair consisting of the samples taken by the primary sampler and the reference method.

NOTE 1 For high volume, high capacity, conveyor systems, a test of a primary increment sampler as an individual component will require the collection and processing of increments of large mass in a short time interval. Before such a test is undertaken, therefore, it is necessary to consider carefully whether such a test can be justified.

With the exception of the primary increment sampler, when testing individual components and subsystems, the test compares the differences between the sample feed stream and the discharge stream of that component or subsystem.

NOTE 2 In some cases, the quality of the feed stream can only be obtained indirectly, for example, by calculation from the results of the divided sample and the corresponding reject-stream sample, weighted according to the division ratio.

For all crushing equipment, differences between samples taken from the fuel both before and after the crusher are used.

For subsystems and sample dividers, pairs obtained by one of the following methods shall be tested:

a) by taking samples from the feed stream and from the sample discharge stream;

b) by taking samples from the sample discharge stream and the reject stream;

c) by collecting the entire sample discharge stream and the entire reject stream.

When using either method a) or method b), great care shall be exercised to obtain unbiased samples; in the case of method a) being used, care shall be taken to minimize disturbance of the feed flow, as such disturbances may introduce bias or distort normal operating conditions.

5 Outline of procedure

The order of operations is as follows:

a) carry out a pre-test inspection (see clause 6);

b) for the overall system, determine where the stopped-belt reference sample will be collected (see 7.1); for diagnostic testing of the system components, see 7.2 or 7.3;

c) determine the variable for test (see clause 8);

d) choose the fuel to be used for the test (see clause 9);

e) decide on the maximum tolerable bias, B (see 4.1);

f) decide on the composition of the sample pairs, i.e. whether to compare sample pairs of one increment or more than one increment;

g) proceed with collection of samples and carry out the tests according to clauses 10 and 11.

6 Pre-test inspection

The primary sources of information regarding compliance with the sampling standard are the equipment specifications and drawings.

A thorough examination of the sampling system and a review of its component specification shall be made.

The party performing the test shall, however, verify performance by field measurements and observations. The operation of the sampling system shall be observed both with fuel flowing and with no fuel.

Pre-test inspections of all operations and equipment, both static and under load, should be carried out by persons experienced in the sampling of segregated, heterogeneous, lumpy bulk materials. It is recommended that operation under normal conditions be observed for an entire lot.

Do not execute a test for bias until all conditions known to cause bias are corrected, unless it is necessary to establish the performance of a system or component as it stands. In the latter case, the pre-test inspection provides essential documentation of what the conditions were at the time of the test.

7 Reference methods

7.1 Overall system

To test overall system bias, the use of a reference method which is known to be intrinsically unbiased is required. The preferred method is the stopped-belt method, i.e. the collection of increments from a complete cross-section of the fuel on the conveyor belt by stopping the belt at intervals. When properly collected, the stopped-belt increment can be considered as a reference increment.

NOTE 1 If it is not possible to collect stopped-belt increments, other reference methods may be used but, in these cases, an apparent absence of a lack of bias relative to the reference method may not be conclusive, and the use of such methods may compromise the validity and authority of the findings.

NOTE 2 With collection of stopped-belt increments, some disruption of normal operations can occur and therefore the plan of execution may need to be coordinated with the normal operations and organized to minimize such disruption. It should be recognized that the conveyor system involved may be used for only a few hours per day for normal operations and cannot be operated solely for bias tests, unless the fuel can be diverted to another discharge point. This can extend the time necessary for completion of the field work and require special arrangements for supplying fuel to the system for testing.

Stopped-belt increments shall be taken with a sampling frame (see Figure 1), or equivalent, from a complete cross-section of the solid mineral fuel on the belt at a fixed position, for a length along the belt which is at least three times the nominal top size of the fuel.

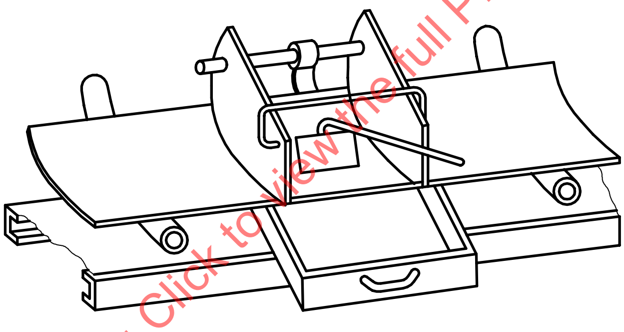


Figure 1 — Sampling frame

If single paired increment samples are tested, each stopped-belt increment shall be taken immediately after the system increment has been taken, preferably from a position located before the primary sampler. If the stopped-belt increment has to be taken from a position after the primary sampler, it is important to ensure that there is no change in conditions between the position at which the stopped-belt sample is taken and that at which the primary increment is taken.

The sampling frame (or equivalent) shall be placed on the stationary belt at the predetermined position so that the separator plates at each end are in contact with the belt across its full width. All particles lying inside the sampling frame end plates shall be swept into the sampling container.

Particles obstructing the insertion of the end plate on the left-hand side shall be pushed into the increment while those obstructing the insertion of the end plate on the right-hand side shall be pushed out of the increment or vice-versa. Whichever practice is used initially, this practice shall be implemented throughout the test.

Take care to minimize the risk of bias being introduced in the course of preparation of the increments and samples and check all sample division equipment and procedures for bias with respect to relevant test parameters.

It is recommended that all increments/samples be weighed immediately after collection, as well as before and after each crushing and dividing operation.

Pay close attention to minimizing unintended mass losses.

Report all observed mass losses.

7.2 Primary increment samplers

Primary increment samplers shall be tested for bias using only the stopped-belt method described in 7.1 as the reference method.

7.3 Subsystems and components

Subsystems and components will ordinarily be tested under routine continuous-operating conditions by sampling the feed, discharge and reject streams using a previously validated sampler. At the final stage, the entire stream is collected (i.e. this stream is not sampled). The quality of the feed to the sample divider shall be calculated from the results of the divided sample and the corresponding reject-stream sample, weighted according to the division ratio.

If it is not practicable to test under normal operating conditions, a separate test may be arranged in which either of the following methods of sampling may be used, as appropriate.

a) Sample the feed and product streams simultaneously to obtain paired members. When using this method, great care should be exercised to obtain unbiased samples.

b) Alternately collect the entire feed stream and the entire product stream associated with successive primary increments.

8 Choice of variables for the test

Tests for bias can be carried out for ash, moisture or any other variable required, but tests for ash and moisture generally suffice. Bias in ash on a dry basis is most commonly caused by errors in size distribution. Bias in moisture may be caused by a wide variety of factors, including but not limited to errors in size distribution, moisture losses associated with crushers, excessive ventilation within the sampling system, less than the closest possible coupling between system components, excessive retention time in the system, or any combination of these.

Direct tests on particle-size distribution are often necessary for coke but the following points should be noted:

a) breakage may occur in the process of increment collection or between the sampling points, thus making zero size bias practically impossible to achieve;

b) bias in other variables, e.g. ash, may have complex and dissimilar relationships to bias in particle-size distribution, thus making meaningful interpretation of those variables difficult if not impossible.

It is therefore recommended that, for parameters other than size, a direct comparison for that parameter should be made.

9 Choice of fuel for the test

9.1 Coal

If more than one coal is to be sampled by the system, the coal chosen for the test shall be one that is expected to show up any bias in the sampling system. For example, bias for ash on primary samplers and sample dividers is commonly caused by the exclusion of larger sized particles (see clause 8). If a coal is chosen where the ash of such particles is similar to that of the coal as a whole, then no bias will be detected even though those particles are being excluded. If subsequently the sampler is used to sample a coal where the large particles have an ash which differs from the mean, then the results will be biased. In this case, therefore, examine the coals to be sampled to find the relationship between ash and the size fraction and the maximum tolerable bias (see clause 4) compared with the maximum bias possible, e.g. if the top 10 % of the size distribution were excluded.

It is recommended that the complete bias test be carried out on a coal from a single source. When this is not possible, the data shall be examined statistically to ensure that their combination for the statistical bias tests is valid (see 11.5 and 11.6).

9.2 Coke

The same requirements as those specified in 9.1 for coal with respect to its ash shall apply when choosing a coke for test with respect to its moisture content.

10 Conduct of the test

10.1 Establishing detailed test procedures

10.1.1 General

The test may encompass the primary sampler alone, individual components or subsystems, or the entire system. The design of the test will differ depending on the objectives. Users of these methods are cautioned to clearly define the objectives of the test as the first step in planning.

A bias test for the whole system is carried out by comparing the reference samples taken from the main flow with the samples collected at the final stage of the on-line system. The paired batch experimental design (see 4.2.3) is the preferred practice because it minimizes the disruption of normal operations. However, the paired-increment experimental design (see 4.2.2) is an acceptable method, provided precautions are taken to avoid introduction of bias caused by the test itself.

Bias may be hardware induced, system logic induced, or a combination of both. It follows that even with the paired increment experimental design, routine operating conditions are best simulated by operating sampling systems under the control of system logic at routine operating condition settings. Stop the conveyor for collection of reference increments by tripping it out with the safety line. Under such test conditions, the sampling system must not also be stopped by system interlocks, and a bias-test mode of operation must be provided that prevents this, as well as to run system timers continuously.

Conveyor belt systems for handling fuel are often not designed for repeated starting and stopping under load. The paired increment experimental design does not necessarily preclude collection of stopped-belt increments, provided arrangements are made to stop the feed to the belt from which the increment is to be collected before the belt is stopped, so that the belt will be only partially loaded on restart. Increments can then be collected off the belt from the points in the stream where conditions that prevailed before the feed was stopped still exist.

10.1.2 Testing the whole system

Test the whole system by comparing stopped-belt reference samples taken from the primary flow with the product at the final stage of the preparation system. The condition of the pairs of samples is different. Prepare and analyse the reference samples using methods specified in ISO 13909-4, ISO 13909-6 and ISO 13909-7 for sample preparation and analysis with respect to precision and bias. Prepare and analyse the system samples by the methods used routinely during regular operations.

10.1.3 Testing components

In an integrated system, the components are arranged in series and the performance of each is affected by that of preceding components. It may be desirable that individual components or groups of components (sub-systems), e.g. crushers and dividers, be tested for bias for diagnostic purposes to understand the effect of each on the overall bias. The principles of the statistical treatment are the same as for the whole system.

10.2 Collection of test sample

It is often possible, on the basis of previous knowledge, to collect a sufficient number of samples which may be in excess of the minimum number required. Furthermore, the use of a single collection will avoid the risk of data heterogeneity due to changes in operating conditions or the source of fuel between two sample collection events and hence being unable to combine the data from them.

Special care is required in planning the duties of each member of the sampling team and exactly how such duties will be performed. At this stage of planning, detailed operating protocols should be established to ensure uniform and consistent collection, weighing and processing of samples. This includes the preparation of facilities and assembly of all equipment necessary for collection of samples, processing and packaging them and for transporting them from collection and weighing sites to processing facilities and to the laboratory for analysis.

Special safety precautions are necessary for personnel working around the fuel handling and sampling machinery. Attention is drawn to the need to comply with all relevant safety regulations.

Stream-diversion gates in bypass chutes, reversible feed conveyors, or blanks inserted into the system facilitate the collection of increments within the sampling system.

The efficiency of the matched-pairs experimental design depends on the closeness with which reference and system samples are physically paired (correlated) to minimize the impact of variations in product quality within pairs. The extent to which within-pair variances are smaller than between-pair variances is an indication that the objective of the matched-pairs experimental design is being fulfilled. Prepare member pairs of samples together (even though through different steps) and analyse them in the same batch to avoid the introduction of systematic error resulting from variations in treatment during preparation and analysis. Exercise due care to preserve the identity of all samples.

Check the nominal top size of the fuel and the product of any sample crusher that is an integral part of the system or subsystem under test. Include this information in the test report. Samples for this purpose shall be separate from the matched pairs taken for the bias test.

When conducting bias tests for moisture content, take extreme care to minimize moisture change from the samples collected (see 10.3).

10.3 Special precautions for moisture-test samples

Bias testing for total moisture content requires extreme care to ensure that there is no change in the moisture content of samples collected. Consequently, the following precautions shall be taken:

a) avoid conditions that may cause changes in moisture content, such as rain, snow, heat, cold and wind;

b) avoid contamination from free water carried by conveyor belts;

c) ensure that sufficient labour is available to collect system and reference samples without delay;

d) seal samples immediately after collection;

e) carry out preparation and testing without delay to avoid moisture changes, if necessary by weighing samples at the collection point;

f) take account of any moisture condensation on sample containers.

NOTE Coal that has been treated to eliminate freezing or dust generation is not suitable for testing for moisture bias.

10.4 Documentation

Keep a record of the mass of each sample, reference and non-reference. Also keep a record of flow rates on the main belt if the system is equipped with a flow-rate indicator, or retain charts from flow-rate recorders, if available.

It is recommended practice that the flow rate/bias relationship be checked. It is also recommended that all samples be weighed before and after each stage of preparation to monitor preparation losses, and that a record be kept of these masses.

A detailed log should be kept during the test showing the clock time at the beginning and end of collection of each sample, the occurrence of any deviations from operating protocols or unusual events and delays, and the reasons for them.

11 Statistical analysis and interpretation

11.1 Outline of procedure

An outline of the procedure is as follows.

a) Fix the initial number of pairs of samples required, n_{P} [see 11.4 a)].

b) Collect n_{P} sets of paired samples.

c) Check for outliers and modify n_{P} if necessary (see 11.3).

d) Check for independence of differences (see 11.6). If this fails, repeat the bias test.

e) Calculate the standard deviation of the differences, s_{d} (see 11.2).

f) Calculate the factor g for the data set [see equation (6)] using the value of s_{d} obtained in step e) and establish a new value for n (see 11.4).

g) If $n_{\text{p}} \leqslant n$, proceed to step l).

h) Calculate the detection level, B', using the equation $B' = g \cdot s_{\text{d}}$ [see 11.4 d)].

i) Decide whether B' can be substituted for B, considering the estimated number of additional sets needed to attain the original B. If yes, proceed to l). If no, proceed to j).

j) More pairs of samples are required; take $n - n_{\text{P}}$ new pairs ("new data").

k) Carry out homogeneity checks between old and new data (see 11.5). If the tests are acceptable, combine the old and new data sets and return to step d).

l) Check that there is no obvious bias (see 11.7.1). If there is, the test is complete.

m) Test the hypothesis that the true bias, $|\mu| = B$ (see 11.7.2).

n) If the hypothesis $|\mu| = B$ is not rejected, there is bias of practical consequence to the extent considered in selecting the maximum tolerable bias, B.

o) If the hypothesis $|\mu| = B$ is rejected, test the null hypothesis that the true bias, $\mu = 0$ (see 11.7.3).

p) If the null hypothesis $\mu = 0$ is rejected, there is a statistically significant bias but the system may be accepted as unbiased, recognizing the statistical evidence of bias.

q) If the null hypothesis $\mu = 0$ is accepted, there is no statistical evidence of bias and the system under test may be accepted as unbiased.

11.2 Basic statistics

11.2.1 General

After results have been obtained from the initial sets of pairs, the following statistical procedure shall be carried out.

Let the values from the system or component being tested (the "system method") be known as A_i, and the values from the reference method be known as R_i, where $i = 1, 2, ..., n$, and i is the serial number and n is the total number of pairs.

For each pair of results, calculate the differences between the results for the two methods ($d_i = A_i - R_i$). Calculate the mean of the reference method values, $\overline{R}$, the mean of the differences, $\overline{d}$, and the standard deviation of the differences, s_{d}, (substituting d_i for x_i) using the general equations (1), (2) and (3).

11.2.2 Mean

The mean, $\overline{x}$ is calculated from equation (1):

$$\overline{x} = \frac{\sum_{i=1}^{i=n} x_i}{n} \tag{1}$$

where n is the number of observations.

11.2.3 Variance

The variance V is calculated from equation (2):

$$V = \frac{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}{(n-1)} \tag{2}$$

where $\sum x^2$ is an abbreviation for $\sum_{i=1}^{i=n} x_i^2$, i.e. the sum of the squared observations.

11.2.4 Standard deviation

The standard deviation, s, which is the square root of the variance, is calculated from equation (3):

$$s = \sqrt{V} \tag{3}$$

The standard deviation of the differences, s_{d}, is calculated from equation (4):

$$s_{\mathrm{d}} = \sqrt{\frac{\sum d^2 - \frac{\left(\sum d\right)^2}{n}}{(n-1)}} \tag{4}$$

11.3 Outliers

11.3.1 General

An outlier may be due to

a) an extreme manifestation of the random variability inherent in the data,

b) an error in calculating or recording, or

c) the result of gross deviation from the prescribed experimental procedure.

If, as a result of visual assessment of the data, possibly assisted by graphical analysis or identification by a statistical procedure (see 11.3.2), an observation is considered suspect, a further examination of the documentation of the test shall be carried out.

If there is an error in the calculation of a result, it shall be corrected. If there have been deviations from prescribed experimental procedures or hard evidence that field conditions were abnormal for an observation, then it shall be discarded whether or not it agrees with the rest of the data.

If there is no rational explanation why it should be so very different from the rest of the data then it shall be retained.

When a gross deviation from prescribed experimental procedure is known to have occurred for a given observation, such observation shall be discarded, whether or not it appears to be in agreement with other observations, and without recourse to statistical tests for outliers.

Statistical criteria for identifying an outlier are not a sufficient basis for discarding observations. When an outlier has been identified by a statistical test, an investigation shall be made to establish, if possible, the cause. Only if there is direct physical evidence that an outlier was caused by a gross deviation from the prescribed experimental procedure, shall an observation be discarded. Whenever observations are discarded, the direct physical evidence justifying the action shall be reported together with the value(s) so discarded.

In many cases, evidence of a deviation from prescribed procedures will consist only of the discordant value itself. When the experimenter cannot identify such observations with a gross deviation from prescribed procedures, the discordant value(s) shall be reported together with the extent to which these value(s) have been used in analysis of the data.

11.3.2 Statistical procedure for identifying outliers

Use the following procedure to identify possible outliers. This procedure is based on Cochran's maximum variance criterion. A statistic, Cochran's criterion, C, is calculated using equation (5):

$$C = \frac{d_{\text{max}}^2}{\sum_{i=1}^{i=n_{\text{p}}} d_i^2} \tag{5}$$

where

d_{max} is the highest absolute value in the set of differences;

n_{p} is the number of pairs in the set.

Critical values for Cochran's criterion at the 1 % level are given for $n = 20$ to $n = 40$ in Table 1. If the value of C as calculated is greater than the table value, d_{max} is a possible outlier.

11.4 Estimation of number of pairs required

Minimize the risk of a Type II error, using the following procedure.

NOTE The following procedure is applicable to testing for bias for a single variable. If more than one variable is tested, the risk of a Type I /Type II error may be greater.

a) Estimate the number of paired observations needed to set the risk of Type II error at 5 % by calculating the factor g for the data set from equation (6) and reading the minimum number of paired observations required, n_{p}, from Table 2.

$$g = \frac{B}{s_{\text{d}}} \tag{6}$$

where

B is the value of the preselected maximum tolerable bias as defined in 4.1.

s_{d} is the standard deviation of the differences.

Table 1 — Critical values for Cochran's maximum variance test

n[a]	99 % **Confidence level**
20	0,480
21	0,465
22	0,450
23	0,437
24	0,425
25	0,413
26	0,402
27	0,391
28	0,382
29	0,372
30	0,363
31	0,355
32	0,347
33	0,339
34	0,332
35	0,325
36	0,318
37	0,312
38	0,306
39	0,300
40	0,294
NOTE This table is an extract of the table in Table of ISO 5725-2:1994 [1].	
[a] n is the number of differences in the series.	

Since the test sample value of s_d cannot be known until the test is completed, a substitute value for s_d will have to be used. If no information on s_d is available, take a minimum number of 20 pairs. To avoid the necessity for remobilization to collect additional paired observations and the consequent problems of combining data, discount the value of s_d used in equation (6) by using a larger value than that expected from the test. Collect as many pairs in the first mobilization as is practicable.

NOTE The values in Table 2 come from the equation:

$$n_\text{P,R} = \frac{s_\text{d}^2 (t_\alpha + t_\beta)^2}{B^2}$$

From this it follows that:

$$g = \frac{(t_\alpha + t_\beta)}{\sqrt{n_\text{P,R}}}$$

where

t_α is the two-tailed value of Student's t at $(n_\text{P,R} - 1)$ degrees of freedom at the 95 % confidence level;

t_β is the one-tailed value of Student's t at $(n_\text{P,R} - 1)$ degrees of freedom at the 95 % confidence level;

$n_\text{P,R}$ is the number of pairs required.

Use of this relationship for calculation of the minimum number of observations required to limit the estimated risk of Type II error to 5 % yields values of $n_\text{P,R}$ identical to the non-central t distribution down to about five observations, well below the lowest number of observations needed for a viable test for bias.

Table 2 — Values of factor g for estimating the minimum number of sets of observations required

$n_{P,R}$	0	1	2	3	4	5	6	7	8	9
10	$>$ 1,295	1,218	1,154	1,099	1,051	1,009	0,971	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 body of the table gives the g values for the 95 % confidence level for $n_{P,R}$ observations.

b) After the test is completed, repeat the calculation of the factor g using the value of s_d obtained from the test, and establish a new value for $n_{P,R}$.

c) If $n_p \geqslant n_{P,R}$, the condition has been met in which the probability of a Type II error is 5 % or less and a sufficient number of pairs has been obtained. Continue the statistical analysis as described in 11.7.

d) If $n_p < n_{P,R}$, additional paired samples may be required. In order to make an informed decision concerning whether to collect additional sets in an attempt to attain the original, preselected maximum tolerable bias, B, calculate the detection level of the test, B', using the actual standard deviation of the differences obtained from the test data, using the equation

$$B' = g \cdot s_d$$

where

g is the factor from Table 2;

s_d is the standard deviation of the differences.

e) If B' can be substituted for B, considering the estimated number of additional sets needed to attain the original B, use B' and no more data sets are required. Otherwise, obtain additional paired data, calculate the mean, the standard deviation and the variance of the differences and check the additional data for outliers. Analyse for homogeneity as described in 11.5 before beginning anew with 11.3 and, if satisfactory, combine the data and repeat from step c).

To avoid having to take additional paired samples, collect a minimum number of samples initially to exceed the expected minimum number that will be required. Otherwise, it is recommended that at least 10 additional pairs be obtained, even if the required $(n_{P,R} - n_p)$ is less than 10. This will reduce the risk of having to make further collections.

11.5 Tests for homogeneity

11.5.1 General

Test additional data, henceforward called new data, for statistical homogeneity with the original data.

Table 3 — Values of F at 95 % confidence level

V_1/V_2	9	10	11	12	13	14	15	16	17	18	19	20	21	22
9	3,179	3,137	3,102	3,073	3,048	3,025	3,006	2,989	2,974	2,960	2,948	2,936	2,926	2,917
10	3,020	2,978	2,943	2,913	2,887	2,865	2,845	2,828	2,812	2,798	2,785	2,774	2,764	2,754
11	2,896	2,854	2,818	2,787	2,761	2,739	2,719	2,701	2,685	2,671	2,658	2,646	2,636	2,626
12	2,796	2,753	2,717	2,687	2,660	2,637	2,617	2,599	2,583	2,568	2,555	2,544	2,533	2,523
13	2,714	2,671	2,635	2,604	2,577	2,554	2,533	2,515	2,499	2,484	2,471	2,459	2,448	2,438
14	2,646	2,602	2,565	2,534	2,507	2,484	2,463	2,445	2,428	2,413	2,400	2,388	2,377	2,367
15	2,588	2,544	2,507	2,475	2,448	2,424	2,403	2,385	2,368	2,353	2,340	2,327	2,316	2,306
16	2,538	2,494	2,456	2,425	2,397	2,373	2,352	2,333	2,317	2,302	2,288	2,275	2,264	2,254
17	2,494	2,450	2,413	2,381	2,353	2,329	2,308	2,289	2,272	2,257	2,243	2,230	2,219	2,208
18	2,456	2,412	2,374	2,342	2,314	2,290	2,269	2,250	2,233	2,217	2,203	2,191	2,179	2,168
19	2,423	2,378	2,340	2,308	2,280	2,256	2,234	2,215	2,198	2,182	2,168	2,155	2,144	2,133
20	2,393	2,348	2,310	2,278	2,250	2,225	2,203	2,184	2,167	2,151	2,137	2,124	2,112	2,102
21	2,366	2,321	2,283	2,250	2,222	2,197	2,176	2,156	2,139	2,123	2,109	2,096	2,084	2,073
22	2,342	2,297	2,259	2,226	2,198	2,173	2,151	2,131	2,114	2,096	2,084	2,071	2,059	2,048
23	2,320	2,275	2,236	2,204	2,175	2,150	2,128	2,109	2,091	2,075	2,061	2,048	2,036	2,025
24	2,300	2,255	2,216	2,183	2,155	2,130	2,108	2,088	2,070	2,054	2,040	2,027	2,015	2,003
25	2,282	2,236	2,198	2,165	2,136	2,111	2,089	2,069	2,051	2,035	2,021	2,007	1,995	1,984
26	2,265	2,220	2,181	2,148	2,119	2,094	2,072	2,052	2,034	2,018	2,003	1,990	1,987	1,966
27	2,250	2,204	2,166	2,132	2,103	2,078	2,056	2,036	2,018	2,002	1,987	1,974	1,961	1,950
28	2,236	2,190	2,151	2,118	2,089	2,064	2,041	2,021	2,003	1,987	1,972	1,959	1,946	1,935
29	2,223	2,177	2,138	2,104	2,075	2,050	2,027	2,007	1,989	1,973	1,958	1,945	1,932	1,921
30	2,211	2,165	2,126	2,092	2,063	2,037	2,015	1,995	1,976	1,960	1,945	1,932	1,919	1,908
35	2,161	2,114	2,075	2,041	2,012	1,986	1,963	1,942	1,924	1,907	1,892	1,878	1,866	1,854
40	2,124	2,077	2,038	2,003	1,974	1,948	1,924	1,904	1,885	1,868	1,853	1,839	1,826	1,814
45	2,096	2,049	2,009	1,974	1,945	1,918	1,895	1,874	1,855	1,838	1,823	1,808	1,795	1,783
50	2,073	2,026	1,986	1,952	1,921	1,895	1,871	1,850	1,831	1,814	1,798	1,784	1,771	1,768
55	2,055	2,008	1,968	1,933	1,903	1,876	1,852	1,831	1,812	1,795	1,779	1,764	1,751	1,739
60	2,040	1,993	1,952	1,917	1,887	1,860	1,836	1,815	1,796	1,778	1,763	1,748	1,735	1,722

Table 3 — Values of F at 95 % confidence level (*continued*)

V_1/V_2	23	24	25	26	27	28	29	30	35	40	45	50	55	60
9	2,908	2,900	2,893	2,886	2,880	2,874	2,869	2,864	2,842	2,826	2,813	2,803	2,784	2,787
10	2,745	2,737	2,730	2,723	2,716	2,710	2,705	2,700	2,678	2,661	2,648	2,637	2,628	2,621
11	2,617	2,609	2,601	2,594	2,588	2,582	2,576	2,570	2,548	2,531	2,517	2,507	2,498	2,490
12	2,514	2,505	2,498	2,490	2,484	2,478	2,472	2,466	2,443	2,426	2,412	2,401	2,392	2,384
13	2,429	2,420	2,412	2,405	2,398	2,392	2,386	2,380	2,357	2,339	2,325	2,314	2,304	2,297
14	2,357	2,349	2,341	2,333	2,326	2,320	2,314	2,306	2,284	2,266	2,252	2,240	2,231	2,223
15	2,296	2,288	2,280	2,272	2,265	2,259	2,253	2,247	2,223	2,204	2,190	2,178	2,168	2,160
16	2,244	2,235	2,227	2,220	2,212	2,206	2,200	2,194	2,169	2,151	2,136	2,124	2,114	2,106
17	2,199	2,190	2,181	2,174	2,167	2,160	2,154	2,148	2,123	2,104	2,089	2,077	2,067	2,058
18	2,159	2,150	2,141	2,133	1,126	2,119	2,113	2,107	2,082	2,063	2,048	2,035	2,025	2,017
19	2,123	2,114	2,160	2,098	2,090	2,048	2,077	2,071	2,046	2,026	2,011	1,999	1,988	1,980
20	2,092	2,082	2,074	2,066	2,059	2,052	2,045	2,039	2,013	1,994	1,978	1,966	1,955	1,946
21	2,063	2,054	2,045	2,037	2,030	2,023	2,016	2,010	1,984	1,964	1,949	1,936	1,925	1,916
22	2,038	2,028	2,020	2,011	2,004	1,997	1,990	1,984	1,958	1,938	1,922	1,909	1,898	1,889
23	2,014	2,005	1,996	1,988	1,980	1,973	1,967	1,960	1,934	1,914	1,898	1,885	1,874	1,865
24	1,993	1,984	1,975	1,967	1,959	1,952	1,945	1,939	1,912	1,892	1,876	1,862	1,852	1,842
25	1,974	1,964	1,955	1,947	1,939	1,932	1,925	1,919	1,892	1,872	1,855	1,842	1,831	1,822
26	1,956	1,946	1,937	1,929	1,921	1,914	1,907	1,901	1,874	1,853	1,837	1,823	1,812	1,803
27	1,940	1,930	1,921	1,913	1,905	1,897	1,891	1,884	1,857	1,836	1,819	1,806	1,795	1,785
28	1,924	1,915	1,906	1,897	1,889	1,882	1,875	1,869	1,841	1,820	1,803	1,790	1,778	1,769
29	1,910	1,901	1,891	1,883	1,875	1,868	1,681	1,854	1,827	1,805	1,789	1,775	1,763	1,754
30	1,897	1,887	1,878	1,870	1,682	1,854	1,847	1,841	1,813	1,792	1,775	1,761	1,749	1,740
35	1,843	1,833	1,824	1,815	1,807	1,799	1,792	1,768	1,757	1,735	1,717	1,703	1,691	1,681
40	1,803	1,793	1,783	1,775	1,766	1,759	1,751	1,744	1,715	1,683	1,675	1,680	1,648	1,637
45	1,772	1,762	1,752	1,743	1,735	1,727	1,720	1,713	1,683	1,660	1,541	1,626	1,614	1,603
50	1,748	1,737	1,727	1,718	1,710	1,702	1,694	1,687	1,657	1,634	1,615	1,599	1,587	1,576
55	1,727	1,717	1,707	1,698	1,689	1,681	1,674	1,666	1,636	1,612	1,593	1,577	1,564	1,553
60	1,711	1,700	1,690	1,681	1,672	1,664	1,656	1,649	1,618	1,594	1,575	1,559	1,546	1,534

11.5.2 Identical variances test

Calculate the ratio F_c of the variances of the differences of the new and original data using equation (7):

$$F_c = \frac{V_1}{V_2} \tag{7}$$

where

V_1 is the variance of the set with the greater variance;

V_2 is the variance of the set with the smaller variance.

The value of F_c is compared with the value of F read from Table 3. The number of degrees of freedom is $n_1 - 1$ on the horizontal axis and $n_2 - 1$ on the vertical axis, where n_1 is the number of observations corresponding to V_1 and n_2 is the number of observations corresponding to V_2.

If $F_c < F$ (the value from Table 3), the two groups of data may be assumed to be drawn from populations with a common variance. If $F_c \geqslant F$, the test fails (see 11.5.4).

11.5.3 Identical means test

Calculate the pooled standard deviation, $\overline{s_x}$, using equation (8):

$$\overline{s_x} = \sqrt{\frac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2}{(n_1 + n_2 - 2)}} \tag{8}$$

where

n_1 is the number of observations in the original data;

n_2 is the number of observations in the new data;

s_1 is the standard deviation of the original data;

s_2 is the standard deviation of the new data.

Calculate the test statistic, t_m, using equation (9):

$$t_m = \frac{|\overline{x_1} - \overline{x_2}|}{\overline{s_x}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{9}$$

where

$\overline{x_1}$ is the mean of the original data;

$\overline{x_2}$ is the mean of the new data;

$\overline{s_x}$ is the pooled standard deviation [see equation (8)].

Read from Table 4 the value for t_α corresponding to two-tailed t with $(n_1 + n_2 - 2)$ degrees of freedom.

If $t_m < t_\alpha$, the new and original sets of paired data may be assumed to be drawn from populations with a common mean. If $t_m \geqslant t_\alpha$, the test fails.

11.5.4 Combination of data

If both of the tests for identical variance and identical means have demonstrated statistical homogeneity (the absence of statistically significant differences), the new data and original data can be combined. The statistical analysis may begin anew from 11.2.

If, on the other hand, either of the two tests fails, a serious problem is indicated since the new data are inconsistent with original data. Discard both sets of data. Determine the cause(s) of the lack of statistical homogeneity. Once such causes are believed to have been remedied, initiate an entirely new test programme.

Table 4 — Value of Student's t at 95 % confidence level for two- and one-tailed distributions

df	two-tail, t_α	one-tail, t_β
5	2,571	2,015
6	2,447	1,943
7	2,365	1,895
8	2,306	1,860
9	2,262	1,833
10	2,228	1,812
11	2,201	1,796
12	2,179	1,782
13	2,160	1,771
14	2,145	1,761
15	2,131	1,753
16	2,120	1,746
17	2,110	1,740
18	2,101	1,734
19	2,093	1,729
20	2,086	1,725
21	2,080	1,721
22	2,074	1,717
23	2,069	1,714
24	2,064	1,711
25	2,060	1,708
26	2,056	1,706
27	2,052	1,703
28	2,048	1,701
29	2,045	1,699
30	2,042	1,697

Table 4 — Value of Student's t at 95 % confidence level for two- and one-tailed distributions

df	two-tail, t_α	one-tail, t_β
31	2,040	1,695
32	2,037	1,694
33	2,035	1,692
34	2,033	1,691
35	2,031	1,690
36	2,029	1,688
37	2,027	1,687
38	2,025	1,686
39	2,023	1,685
40	2,021	1,684
41	2,020	1,683
42	2,019	1,682
43	2,017	1,681
44	2,016	1,680
45	2,015	1,679
46	2,013	1,679
47	2,012	1,678
48	2,011	1,678
49	2,010	1,677
50	2,009	1,676
55	2,005	1,673
60	2,000	1,671
70	1,995	1.,667
80	1,990	1,664
90	1,987	1,662
100	1,984	1,660

11.6 Test for independence of differences

To draw inferences correctly about sampling system bias using the procedures of this part of ISO 13909, the sample differences shall be independent. The following test for independence is a test for randomness based on the expected number of runs above and below the sample median (see A.2.5). A run is a sequence of values all above, or all below the median. The test is run after outliers have been removed.

Determine the number of runs, r, of the population of differences as follows. Subtract the sample median from each of the observed differences; record a plus sign for each value that is positive, and a minus sign for each value that is negative; ignore differences equal to the median; count the number of changes of sign. Specimen calculations are given in informative annex A.

Let n_1 denote the smallest number of like signs and n_2 denote the largest number of like signs. If there is an equal number of positive and negative signs then $n_1 = n_2$.

Obtain the lower, l, and upper, u, significance values from Table 5, corresponding to n_1 and n_2.

If $r < l$ or $r > u$, the observations have failed the test for independence and the following statement shall be included in the bias test report together with anything known about assignable causes for the indicated lack of independence:

"There is evidence that the series of differences between reference and system values are not independent."

11.7 Final assessment of bias

11.7.1 Obvious bias condition

For the matched-pairs experimental design of the test for bias, the expected value of the mean of the differences is zero. If $\bar{d} \leqslant -B$ or $\bar{d} \geqslant +B$, there is evidence of bias and no further statistical analysis is required.

Table 5 — Table of significance values for number of runs

n_1	n_2	Lower l	Upper u
3	5	3	—
3	6	3	—
3	7	3	8
4	4	3	7
4	5	3	8
4	6	4	8
4	7	4	8
4	8	4	8
5	5	4	8
5	6	4	9
5	7	4	9
5	8	4	10
5	9	5	10
6	6	4	10
6	7	5	10
6	8	5	11
6	9	5	11
6	10	6	11
7	7	5	11
7	8	5	12
7	9	6	12
7	10	6	12
7	11	6	13
7	12	7	13
8	8	6	12
8	9	6	13
8	10	7	13
8	11	7	14
8	12	7	14
9	9	7	13
9	10	7	14
9	11	7	14
9	12	8	15
9	13	8	15
9	14	8	16
10	10	7	15
10	11	8	15
10	12	8	16
10	13	9	16
10	14	9	16
10	15	9	17
11	11	8	16
11	12	9	16
11	13	9	17
11	14	9	17

Table 5 — Table of significance values for number of runs

n_1	n_2	Lower l	Upper u
11	15	10	18
11	16	10	18
11	17	10	18
12	12	9	17
12	13	10	17
12	14	10	18
12	15	10	18
12	16	11	19
12	17	10	19
12	18	11	20
13	13	10	18
13	14	10	19
13	15	11	19
13	16	11	20
13	17	11	20
13	18	12	20
13	19	12	21
14	14	11	19
14	15	11	20
14	16	12	20
14	17	12	21
14	18	12	21
14	19	13	22
14	20	13	22
15	15	12	20
15	16	12	21
15	17	12	21
15	18	13	22
15	19	13	22
15	20	13	23
16	16	12	22
16	17	13	22
16	18	13	23
16	19	14	23
16	20	14	24
17	17	13	23
17	18	14	23
17	19	14	24
17	20	14	24
18	18	14	24
18	19	15	24
18	20	15	25
19	19	15	25
19	20	15	26
20	20	16	26

11.7.2 Test for significant difference from B

If $-B < \bar{d} < +B$, calculate the statistic, t_{nz}, for the difference between $\bar{d}$ and B, from equation (10):

$$t_{nz} = \frac{B - |\bar{d}|}{\left(\frac{s_d}{\sqrt{n_p}}\right)} \tag{10}$$

where

B is the maximum tolerable bias;

$\bar{d}$ is the mean of the differences;

s_d is the standard deviation of the differences;

n_p is the number of differences.

Compare t_{nz} with the corresponding value for a one-tailed t (t_β) in Table 4, with $(n - 1)$ degrees of freedom.

If $t_{nz} < t_\beta$, there is evidence of bias which is both significantly greater than zero and not significantly less than B. The test result indicates the existence of relevant bias.

If $t_{nz} \geqslant t_\beta$, any bias is significantly less than B. The test result does not indicate the existence of relevant bias.

11.7.3 Test for significant difference from zero

If the test results are within the range $-B < \bar{d} < +B$, calculate the statistic, t_z, for the mean of the differences from equation (11):

$$t_z = \frac{|\bar{d}|}{\left(\frac{s_d}{\sqrt{n_p}}\right)} \tag{11}$$

where

$\bar{d}$ is the mean of the differences;

s_d is the standard deviation of the differences;

n_p is the number of differences.

Compare t_z with the corresponding value for a two-tailed t (t_α) in Table 4, with $(n - 1)$ degrees of freedom.

If $t_z < t_\alpha$, the observed mean of the differences is not significantly different from zero and the system or component under test can be accepted as free of bias.

If $t_{nz} \geqslant t_\alpha$, there is a bias which is less than B.

12 Test report

The format and detail included in the test report will depend on the purpose of the test. It is recommended that the report should include the following:

a) statement of the purpose of the test;

b) summary statement of the findings;

c) description and technical specification of the sampling system;

d) findings of the pretest inspection together with corrective action taken if the test is not to determine existing performance;

e) characterisation of the fuel by preparation, size and quality;

f) specification of test conditions, maximum tolerable bias, B, flow rate, etc.;

g) details of procedures adopted for the conduct of the test and tabulation of analysis values obtained;

h) explanation and discussion of any deviations from normal operating conditions, test specifications or of events occurring during conduct of the test that could affect results;

i) determination of number of observations required;

j) statistical analysis and interpretations of results.

An annex should include all field logs, analytical results and mass data.

Where this method is used for certification of performance of systems, components, or procedures with respect to absolute bias, it is likely that all of the above items will be required.

Annex A
(informative)

Specimen calculations

A.1 Data

The data used for the specimen calculations in A.2 and A.3 and given in Table A.1 are taken from an analysis for ash in a coal sample. The maximum tolerable bias is set at 0,2 % with regard to ash.

Table A.1 — Raw data, % ash, dry basis

i	System A_i	Reference R_i	Difference d_i
1	9,55	9,63	−0,08
2	8,99	8,99	0,00
3	8,74	8,62	0,12
4	9,08	9,12	−0,04
5	9,83	9,14	0,69
6	9,70	9,57	0,13
7	8,71	8,83	−0,12
8	8,50	8,29	0,21
9	8,83	8,60	0,23
10	8,29	8,15	0,14
11	8,51	8,76	−0,25
12	8,80	8,69	0,11
13	8,69	8,60	0,09
14	8,81	8,67	0,14
15	8,60	8,70	−0,10
16	9,23	8,97	0,26
17	8,56	8,52	0,04
18	8,35	8,23	0,12
19	9,01	9,09	−0,08
20	9,13	9,14	−0,01

A.2 Specimen calculation No. 1

A.2.1 Basic statistics

$n = 20$ pairs

Parameter		**System, A**	**Difference, d**
mean	$\overline{x}$	8,895 50	0,080 0
variance	s_{d}^2	—	0,037 9
standard deviation	s_{d}	—	0,194 8

A.2.2 Graphical analysis

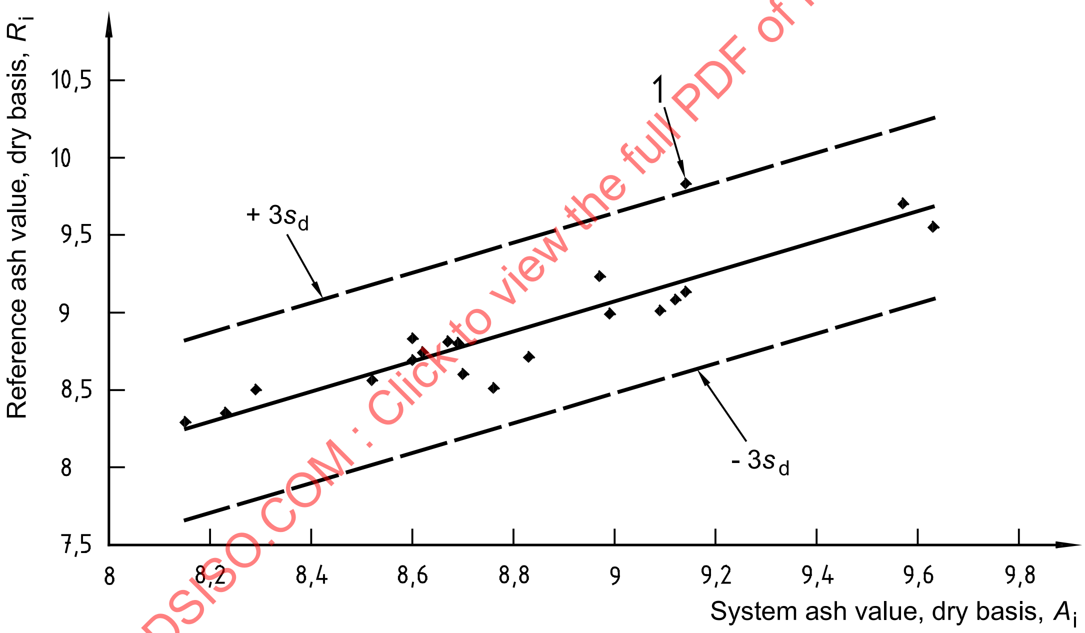
A.2.2.1 General

Graphical analysis is not required but can be helpful in identifying outliers, trends, lack of statistical control and bias. Examples are given in A.2.2.2 to A.2.2.5.

A.2.2.2 Reference against system values

Plot the data pairs with the vertical axis as the reference value R_i, and the horizontal axis as the system value, A_i, using the same scale for both axes. Add the base reference line $R_i = A_i$. Additional reference lines at $\pm\, 3s_\text{d}$ may be added. This is shown in Figure A.1.

The points should cluster around the base reference line. A very few points deviating by more than three standard deviations from the base reference line may indicate the presence of outliers. If the points trend away from the base reference line, other systematic problems may be present. A broad scattering of data indicates poor correlation.



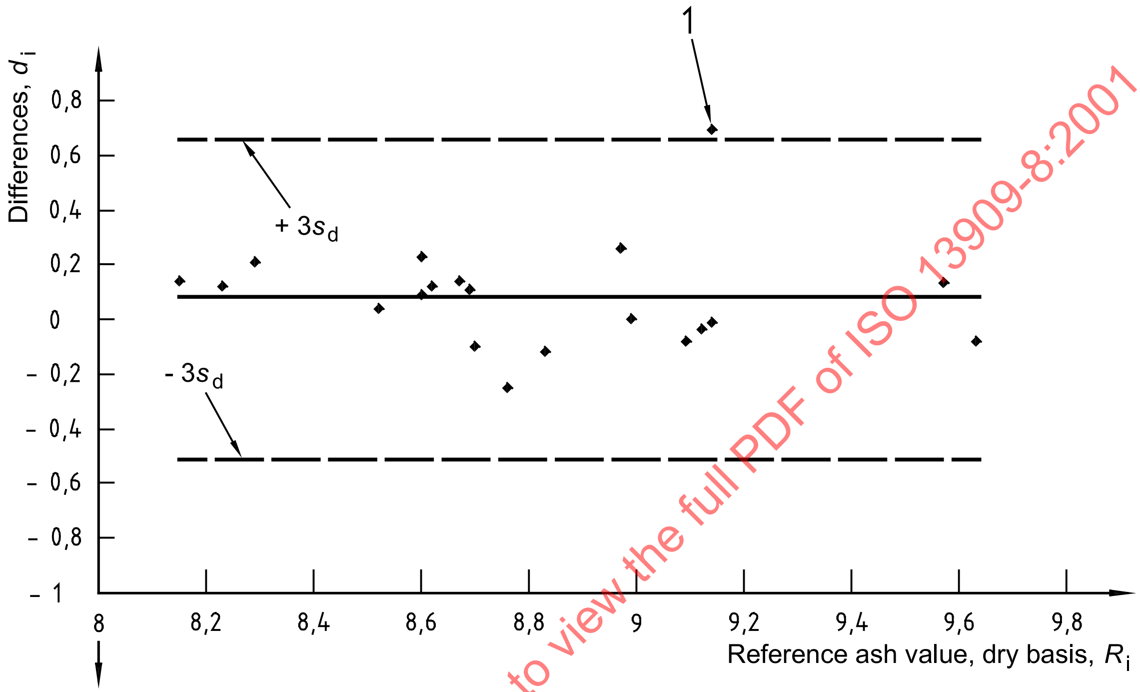
Key

1 Suspected outlier

Figure A.1 — Scatter plot: reference against system

A.2.2.3 Differences against reference values

Plot the difference values, d_i, on the vertical axis against the corresponding reference value, R_i, on the horizontal axis. Draw a horizontal line at $\bar{d}$ (the mean of the differences). Additional reference lines at $\pm 3s_d$ may be drawn (see Figure A.2). The points should cluster about the $\bar{d}$ reference line. Points deviating by more than three standard deviations from d may indicate outliers. Trends in the data indicate serious problems.



Key

1 Suspected outlier

Figure A.2 — Differences against reference values