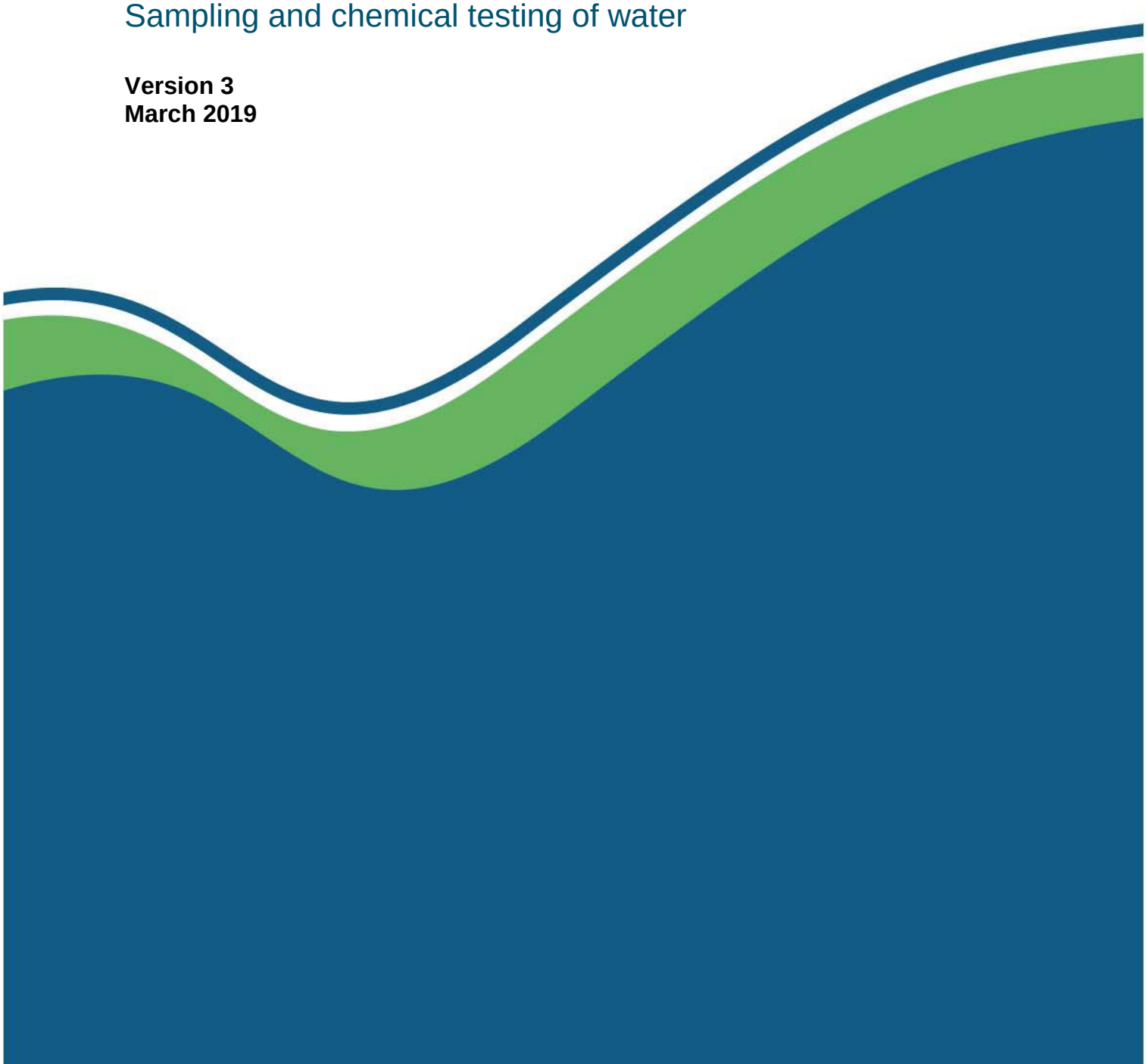
SEPA
Scottish Environment Protection Agency
Buidheann Dion Àrainneachd na h-Alba

Measurement Assurance and Certification Scotland

Performance Standard MACS-WAT-01
Sampling and chemical testing of water

Version 3
March 2019



Record of amendments

Version	Date	Amendment(s)
1	October 2016	First issue.
2	August 2017	4 - insertion of “body providing recognition” definition; alignment of terminology with MACS-WAT-02. 4 - alteration of “recovery” definition; for clarity and consistency of application. 5.2.1 - insertion of briefing note 2; requirement for annual review of sub-contracting arrangements. 6.6.5.4 - insertion of briefing note 2; clarification of validation requirements for filtered tests. 6.6.5.5 - insertion of briefing note 2; clarification of spiked/unspiked sample matrix requirements. 6.6.6 - inserted for clarity; in conjunction with insertion of Annex B. 6.6.6.1 - previously 6.6.5.8. 6.6.6.2 - previously 6.6.5.9. 6.6.6.3 - previously 6.6.5.10. 6.7.2.4 - insertion of briefing note 2; clarification of blank requirements for filtered tests. Annex A - updated. Annex B - inserted; to detail method validation assessment protocols. Annex C - previously Annex B. Annex D - previously Annex C; title updated to avoid possible conflict with new Annex B. Annex E - previously Annex D. Further minor alterations throughout to rectify typographical errors and improve consistency of application within and across MACS performance standards.
3	March 2019	4 - alteration of “10% standard” and “90% standard” definitions; for clarity. 7.6.2.6 b - alteration of requirements for setting of initial control chart limits. Annex A - updated; incorporating Addendum B (October 2017). Annex B - B.3.1.d iii - insertion of briefing note; clarification of requirements when comparing within-batch and between-batch variance. Annex B - B.4.2.b - amendment of formula for target bias (Z_b). Annex B - Table B1 - insertion of additional calculated values; for clarity. Document content reformatted and reordered to reflect ISO/IEC 17025:2017 update. Further minor alterations throughout to rectify typographical errors and improve consistency of application with ISO/IEC 17025:2017 and across MACS performance standards.

Contents

1 Introduction

As Scotland's principal environmental regulator, the Scottish Environment Protection Agency (SEPA) is responsible for protecting and improving Scotland's environment.

SEPA issues a range of authorisations designed to control operator activities which could lead to pollution or environmental damage. Compliance with these authorisations is important to ensure that the environment is protected. An operator's compliance is assessed by SEPA from information gathered from observations, sampling and analysis. These activities may be carried out by an operator under self-monitoring arrangements.

SEPA has established Measurement Assurance and Certification Scotland (MACS) to provide a range of performance standards which ensure data provided by self-monitoring operators is robust, and provides stakeholders with confidence that data is reliable.

Where an operator complies with the requirements of MACS, they will be deemed competent to supply self-monitoring data to SEPA.

SEPA requires all operators and associated organisations certified under MACS to be accredited by the United Kingdom Accreditation Service (UKAS) to ISO/IEC 17025.

Please direct questions regarding the MACS certification process to UKAS at:

United Kingdom Accreditation Service
2 Pine Trees
Chertsey Lane
Staines-upon-Thames
TW18 3HR

Tel: 01784 429 000

Email: info@ukas.com
Website: www.ukas.com

2 Scope

2.1 This performance standard lays out the detailed requirements that operators and laboratories must adhere to when producing data for submission to SEPA under MACS.

NOTE: SEPA requires that all data submitted by an operator is supplied in a consistent electronic format. Supplementary performance standard MACS-WAT-02 documents the detailed sample and data management requirements of MACS (ref. 3.2 a).

2.2 Accreditation to international standard ISO/IEC 17025 is a prerequisite for inclusion in MACS. All sampling and testing methods used by an operator or laboratory whilst producing data for submission to SEPA must be listed on the schedule of accreditation issued by UKAS.

NOTE 1: The requirements detailed in this MACS performance standard are in addition to those prescribed in ISO/IEC 17025, which must be complied with.

NOTE 2: Annex E of this document tabulates the cross references between this performance standard and ISO/IEC 17025.

NOTE 3: The numbering of this document does not directly align with that of ISO/IEC 17025.

2.3 This performance standard is applicable to the sampling and chemical testing of waters; specifically:

- untreated sewage influent and effluent;
- treated sewage effluent;
- water treatment works effluent;
- septic tank effluent;
- trade effluent;
- surface water outfall effluent.

NOTE: An operator's authorisation conditions may refer to sampling of 'effluent', 'discharge', or 'influent' rather than 'water'. For the purpose of this MACS performance standard the terms are to be considered equivalent.

3 References and bibliography

3.1 Normative references

a. BS EN ISO/IEC 17025:2017 – General requirements for the competence of testing and calibration laboratories, ISBN 978 0 580 88466 5.

3.2 Text references

a. MACS Performance Standard: Sample and data management, Scottish Environment Protection Agency, MACS-WAT-02, 2019.

b. BS EN ISO 5667-3:2018 – Water quality - Sampling. Part 3: Preservation and handling of water samples, ISBN 978 0 580 96052 9.

3.3 Bibliography

a. A Manual on Analytical Quality Control for the Water Industry (NS 30), R. V. Cheeseman and A. L. Wilson, revised M. J. Gardiner, Water Research Centre, 1989, ISBN 0 902156 85 3.

b. The Fitness for Purpose of Analytical Methods: A Laboratory Guide to Method Validation and Related Topics, EURACHEM, 2nd Ed., 2014.

c. Harmonized Guidelines for Single-Laboratory Validation of Methods of Analysis, M. Thompson, S. L. R. Ellison and R. Wood, IUPAC, 2002.

d. Harmonised Guidelines for the Use of Recovery Information in Analytical Measurement, M. Thompson, S. L. R. Ellison, A. Fajgelj, P. Willets and R. Wood, IUPAC, 1999.

e. Guide to Quality in Analytical Chemistry, CITAC/EURACHEM, 2002.

f. Statistics and Chemometrics for Analytical Chemistry, 4th Ed., J. N. Miller and J. C. Miller, Prentice Hall, 2000, ISBN 978 0 13 129192 8.

g. UKAS Policy on Participation in Proficiency Testing: TPS 47, Edition 2, 2013.

4 Terms and definitions

For the purpose of this MACS performance standard, and unless the context requires otherwise, the following definitions shall apply:

10% standard – an ideal matrix, spiked with standard solution at 10% of the expected method range. It will not be prepared in real sample matrix. This test type will be analysed in the same way as samples.

90% standard – an ideal matrix, spiked with standard solution at 90% of the expected method range. It will not be prepared in real sample matrix. This test type will be analysed in the same way as samples.

analytical quality control (AQC) – the term used to describe the practical steps undertaken to ensure that analytical data is adequately free from error. The primary purpose of AQC is as an indicator of the performance of the analytical system, rather than as a guide to the error associated with an individual test result.

batch – those sample preparations which are performed as a discrete entity. Where appropriate, blank(s) and laboratory control samples will be prepared alongside routine samples.

bias – which may be a positive or negative value, is the difference (expressed as a percentage) between the mean number of determinations and the true or accepted concentration:

$$\%\boldsymbol{Bias} = \frac{(\bar{\boldsymbol{x}} - \boldsymbol{T})}{\boldsymbol{T}} \times \mathbf{100}$$

- $\bar{x}$ is the mean of the dataset.
- T is the true or accepted value (expected concentration) of the test sample.

blank – a blank is analysed with a run of samples to check for system contamination from the instrument. The blank does not go through any sample preparation steps. The blank may be a portion of deionised or interference free water or it may be neat solvent depending on the nature of the instrument.

In validation, this test type is the equivalent of a zero level standard. It will be made using the same ideal matrix which is used to make up routine method QC standards. It will not be prepared in real sample matrix. This test type will be analysed in the same way as samples.

This control measure will be used for direct methods. It will be used to blank correct results requiring to be reported as a final sample concentration.

NOTE : In validation, where it is not possible to source real matrix with sufficiently absent or low determinand levels, it may also be used to determine method detection limit for methods which are able to return numeric results less than zero.

body providing recognition – a body carrying out audits to ensure that the requirements of supporting quality standards, accreditation or certification are adhered to by an operator undertaking activities within the scope of MACS, e.g. UKAS.

certified reference material (CRM) – a sample of target matrix containing a specified concentration of the determinand(s) of interest; certified to a quoted uncertainty and traceable to a national/international standard. This test type will be analysed in the same way as samples.

concentration – as used in this performance standard is expressed as a mass determined per unit volume. (In certain circumstances the term concentration is not appropriate, for example in the determination of pH).

determinand – this is the measured analyte, compound or groups of compounds within a sample which require determination.

direct method – an analytical method where samples are analysed directly with no sample preparation (e.g. pH).

field data – information acquired on site at a monitoring location. May include observations, field based testing or measurements.

ideal matrix – deionised water (grade 1).

laboratory – a laboratory, or sub-contracting laboratory, that undertakes the chemical testing of samples. A laboratory may also undertake sampling activities.

laboratory manager – a person responsible for managing a laboratory.

limit of detection (LOD) – the lowest quantity or concentration of a determinand that can be reliably detected by a given analytical instrument.

method detection limit (MDL) – this is the minimum concentration that can be measured and reported for a determinand and, unlike LOD, covers the whole analytical process including any sample handling and preparation.

non-regulatory determinand - a determinand, where the concentration, level or presence of that determinand is controlled by a rule, limit or other condition set by legal statute of the Scottish Parliament, UK Parliament or directive of the European Commission.

organisation – in the context of this performance standard the term ‘organisation’ encompasses an operator or a body appointed (or sub-contracted) by the operator, including in both cases analytical laboratories undertaking related testing.

operator – a person or company who is responsible for the operation of an installation or plant monitored under MACS arrangements. In the context of this performance standard the term encompasses a body, company or person appointed or sub-contracted by an installation or plant’s responsible person or company, including in all cases analytical laboratories undertaking related testing.

NOTE: In relation to monitoring or assessment required by an authorisation under the Water Environment (Controlled Activities) (Scotland) Regulations 2011 (CAR), the operator is the ‘responsible person’ defined and identified as such in the CAR authorisation.

precision (relative standard deviation %) – this is the distribution of a number of repeated determinations. For the purpose of this performance standard, precision will be expressed as relative standard deviation % (%RSD):

$$\%\boldsymbol{RSD} = \frac{\boldsymbol{S_t}}{\boldsymbol{\bar{x}}} \times \mathbf{100}$$

- S_t is the total standard deviation of the dataset.
- $\bar{x}$ is the mean of the dataset.

pre-treatment method – an analytical method where samples undergo some form of sample preparation prior to instrumental analysis (e.g. organic solvent extraction methods)

process blank – is prepared with a batch of samples to check for process contamination. A process blank is defined as a quantity of clean matrix, i.e. one which does not contain the analyte(s) of interest, which is taken through the complete analytical process. In practice, the process blank is often a portion of deionised or interference free water.

In validation, this test type is the equivalent of a zero level laboratory control sample. It will be made using the same matrix which is used to make up routine method control samples. It will not be prepared in real sample matrix. This test type will be analysed in the same way as samples.

This control measure will be used for pre-treatment methods. It will be used to blank correct results requiring to be reported as a final sample concentration (with and without recovery correction).

NOTE 1: In certain circumstances it may not be possible to find a suitable clean matrix, i.e. where there is the potential for significant levels of background interferences. In these situations a process blank may not contain any matrix, but will comprise of only those reagents which are routinely taken through the entire analytical process.

NOTE 2: In validation, where it is not possible to source real matrix with sufficiently absent or low determinand levels, it may also be used to determine method detection limit for methods which are able to return numeric results less than zero.

recovery – the proportion of the amount of analyte, present in or added to the analytical portion of the test material, which is extracted and presented for measurement.

Correction for recovery is required for pre-treatment methods where there is known to be significant loss of analyte during sample preparation. Depending on the requirements of an individual method, the recovery factor used to perform this correction may be calculated on either a historic or per-batch basis.

regulated (or regulatory) determinand – a determinand, where the concentration, level or presence of that determinand is controlled by a rule, limit or other condition set by SEPA or other regulatory body.

run – a number of samples analysed as a discrete entity. A single run may contain multiple batches.

sample – a volume of water collected from a monitoring location and identified for the assessment or measurement of specific determinand(s).

standard deviation – the measure of how spread out a dataset is.

$$s = \sqrt{\frac{\sum(x_i - \overline{x})^2}{n}}$$

- x_i is each individual result.
- $\overline{x}$ is the mean of the dataset.
- n is the sample size.

supporting determinand – a determinand, the assessment of which may be required in order to mitigate the interferences, or other effects, that determinand may have upon the determination of another regulatory or non-regulatory determinand, so that integrity of the reported regulatory or non-regulatory determinand result is assured.

5 Structural requirements

5.1 It is the responsibility of the operator to ensure that sampling, calibration and testing activities are conducted in such a way as to meet the requirements of this performance standard, and satisfy both the needs of SEPA and the body providing recognition.

5.2 Provisions for the structural requirements of laboratories are laid out in ISO/IEC 17025. Organisations conducting sampling activities shall:

a. Have arrangements in place to ensure that management and personnel conducting these activities are free from any undue commercial, financial and other pressures and influences that may adversely affect the quality of their work.

b. Have policies and procedures to avoid involvement in any sampling or operational activities that would diminish confidence in its competence, impartiality, judgement or operational integrity.

NOTE: It is not acceptable for an organisation to manipulate the operation of their treatment plant, or effluent inputs/outputs to/from their treatment plant to take into account sampling dates. The sampling programme must be representative of the normal operation of that treatment plant.

c. Ensure that that its personnel are aware of the relevance and importance of their activities and how they contribute to achieving the objectives of this MACS performance standard.

d. Provide adequate supervision of personnel undertaking sampling, testing and calibration activities, including trainees, by persons competent in and authorised to undertake those activities.

5.3 For data to be submitted to SEPA under MACS, the organisation must ensure that the appropriate sampling and testing methods are selected and satisfy the requirements of this performance standard (see 7.5.3).

6 Resource requirements

6.1 Facilities and environmental conditions

6.1.1 In order to prevent adverse effects on analytical results, sample integrity must be maintained during collection, transport and subsequent storage in accordance with the general requirements of international standard ISO 5667-3 (ref. 3.2 b), or equivalent peer evaluated reference standard.

6.2 Externally provided services

6.2.1 An operator may sub-contract sampling and/or chemical testing to another organisation. It is the responsibility of the operator to ensure that the sub-contracted organisation is certified under MACS for the scope of work sub-contracted. Sub-contracting to an organisation not certified under MACS is only permitted where an operator has obtained the prior written approval of SEPA.

NOTE 1: The organisation may or may not be aware that the data it generates will be submitted to SEPA. However, the organisation's customer or procurer of the sampling and analytical service should

be aware that if it wishes to submit the data to SEPA, then the requirements of this performance standard need to be satisfied.

NOTE 2: An annual review of the suitability of existing sub-contracting arrangements must be performed by an operator whilst establishing their survey schedule for the following sampling year [see MACS-WAT-02, section 6.2 (ref. 3.2 a) for further information on survey schedule establishment]. In practice, this will require the operator to re-submit all sub-contracting request(s) to SEPA for assessment and approval by 1 December in the year preceding the start of a specified sampling year.

6.3 Business continuity arrangements

6.3.1 SEPA requires details of an operator's business continuity arrangements to demonstrate that plans are in place in the event of any laboratory (including sub-contractors) being unable to analyse samples within the timelines defined in the supplementary MACS performance standard (ref. 3.2 a).

7 Process requirements

7.1 Sampling

7.1.1 All sampling activity is required to be accredited to ISO/IEC 17025.

7.1.2 Operators must ensure appropriate quality assurance and management systems are in place for all sampling activities. Sampling activities may operate independently of the laboratory and procedures shall include, but are not limited to:

- a schedule for sample collection [see MACS-WAT-02 (ref. 3.2 a)];
- planned observations at regulated sites;
- sample collection methods;
- training and audit;
- appropriate sampling containers and preservation techniques;
- sample transport, receipt, handling, storage, disposal and chain of custody;
- operation, maintenance and calibration of sampling equipment;
- operation, maintenance and calibration of on-site test equipment.

7.1.3 All personnel engaged in sampling activity will be audited by their own organisation at least once annually.

7.1.4 SEPA reserves the right to send a SEPA officer to act as an observer at UKAS sampling audits.

7.2 Test methods

7.2.1 Only results generated using methods accredited to ISO/IEC 17025 will be considered suitable for submission to SEPA.

7.2.2 SEPA reserves the right to send a SEPA officer to act as an observer at a laboratory's UKAS surveillance audit.

7.3 Test method selection

7.3.1 SEPA will not prescribe specific analytical methods, but the operator must ensure that any method employed is fit for purpose and appropriate for the analyte, sample matrix and concentration range to be determined.

7.3.2 All methods are required to meet the target method detection limit (MDL) for a determinand.

7.3.3 The target MDL for a MACS determinand will be set by SEPA on an operator specific basis, at a level which permits SEPA to:

- assess an operator's compliance with the determinand's authorisation conditions;
- meet the reporting requirements of a non-regulatory determinand.

7.3.4 When setting a target MDL, SEPA will be mindful of the following:

- Authorisation conditions for the determinand at monitoring locations the operator has responsibility for.
- SEPA's internal target MDL for the determinand.
- SEPA's knowledge of analytical capability with respect to the determination of the determinand.
- The analytical capability of a third party laboratory, where testing is sub-contracted by an operator.
- The target MDL (or equivalent) set by an environmental body other than SEPA, for the reporting of determinand data by SEPA, e.g. OSPAR Commission.

7.3.5 The target MDL for a MACS determinand may be amended by SEPA where there is a change to:

- an operator's authorisation conditions;
- the reporting requirements set by an environmental body other than SEPA, for the reporting of determinand data by SEPA e.g. OSPAR Commission;
- other legal requirements.

7.3.6 Once set, target MDLs will be formally recorded in each operator's individual 'Operator specific criteria' document.

7.3.7 In order to allow SEPA to assess an operator's compliance, analytical methods must be able to provide absolute test result values for all determinands. The only exception being the use of a '<' qualifier when submitting a test result determined at less than the stated MDL.

7.3.8 A clear and concise summary of a method used to generate results submitted to SEPA shall be available and provided to SEPA upon request. This need not be fully comprehensive, but must contain sufficient detail to allow for direct comparison to similar methods.

7.3.9 For all methods, a fully documented analytical procedure shall be available and provided to SEPA upon request.

7.4 Handling of test items

7.4.1 In exceptional circumstance, SEPA will accept analytical results associated with sample handling or analysis that has not been undertaken according to documented procedures. In each case, the operator must record a non-conformance and obtain from SEPA a concession to report results.

The concession request shall include full assessment and justification that the non-conformance has had no impact on the quality of the data submitted to SEPA.

If it is not possible to justify a non-conformance, then the results will not be accepted by SEPA.

NOTE: Where analytical results are associated with a non-conformance, this must be clearly identified when data is returned to SEPA [see supplementary MACS performance standard (ref. 3.2 a)].

7.5 Method validation

7.5.1 The process of method validation is intended to deliver documented, objective evidence that the methods employed by a laboratory are suitable for the production of data for submission to SEPA under operator monitoring arrangements. It provides confidence that the established performance characteristics of a method are based on robust experimental determinations and are statistically sound.

It is implicit in the validation process that all studies to determine method performance characteristics are carried out using analytical equipment that is within specification, working correctly and adequately calibrated. Likewise, any analyst carrying out the studies must be competent in the field of work under study and have sufficient knowledge related to the work to be able to make appropriate decisions from the observations as the study progresses.

7.5.2 All analytical methods must be fully validated using appropriate matrices prior to use for generation of data for submission to SEPA. The specific matrix types applicable to this performance standard are listed in clause 2.4.

NOTE: Laboratories are not required to validate all of the matrix types listed in clause 2.4, only those matrices which are relevant to the analyses to be certified under MACS.

7.5.3 Target performance characteristics relevant to this performance standard are detailed in Annex A. In order for submitted data to be accepted by SEPA, laboratories must demonstrate that analytical performance measured during method validation meets these targets.

7.5.4 Validation records shall be made available and provided to SEPA upon request.

7.5.5 Validation procedure

7.5.5.1 As far as practicable, any validation exercise shall encompass the whole analytical procedure. This shall include, for example, any bottles normally used for sampling, any preservation reagent and all general equipment used in the process.

7.5.5.2 Validation shall be undertaken in a period of time of not less than six days and not more than three months.

7.5.5.3 No changes shall be made to the documented analytical procedure once a validation exercise has commenced. If circumstances indicate that significant changes are required then the validation exercise will be repeated.

NOTE: Assessment of the significance of a change is a matter of judgement for each individual laboratory. Where the laboratory deems that a change is not significant enough to warrant repeat validation the decision must be fully justified and documented.

7.5.5.4 Performance characteristics of a specified method, determinand and matrix shall be determined with a minimum of ten degrees of freedom. In practice, this can be achieved by analysis of 11 batches, each containing duplicates of the appropriate test sample types.

NOTE 1: This is often termed an '11×2' validation, as 11 batches containing two replicates of each test sample type are analysed.

NOTE 2: It is not necessary to perform full '11×2' validation for filtered determinands (e.g. dissolved metals), providing that equivalent performance can be demonstrated when compared against the unfiltered test (e.g. total metals).

7.5.5.5 Inclusion of the following test sample types is a mandatory requirement for a direct method:

- Blank.
- 10% standard.
- 90% standard.
- Method detection limit (MDL).
- Certified reference material (CRM); or
- Spiked sample matrix (spiked between 50-90% of method range), blank corrected by unspiked sample matrix.

Inclusion of the following test sample types is a mandatory requirement for a pre-treatment method:

- Process blank.
- 10% standard, taken through entire analytical process.
- 90% standard, taken through entire analytical process.
- Method detection limit (MDL).
- Certified reference material (CRM); or
- Spiked sample matrix (spiked between 50-90% of method range), blank corrected by unspiked sample matrix.

NOTE 1: It must be ensured that 10% and 90% standard concentrations are set using the appropriate range values, i.e. instrument working range for a direct method or method working range for a pre-treatment method. For example, a method with a range of 8 to 100mg/L would have a 10% standard set not at 10mg/L but at 17mg/L.

NOTE 2: When analysing spiked and unspiked sample matrix pairs, the sample matrix chosen should ideally contain negligible amounts of the determinand(s) of interest.

7.5.5.6 If all required duplicate test samples cannot be accommodated in a single batch then multiple batches shall be prepared, ensuring that all replicates of an individual test sample type are contained within the same batch. All test samples within a batch shall be analysed in random order within an analytical run.

7.5.5.7 Preferably, each individual validation batch will be analysed on separate days. However, where this proves impracticable, a maximum of two batches may be analysed on the same day. In these circumstances, the instrument must be allowed to return to 'ground state' between analytical runs to avoid obtaining falsely low estimates of precision.

NOTE: In practice, a return to 'ground state' will involve a break between analytical runs. Any routine daily or pre-use checks must be carried out, and instrument calibration performed if a part of the normal analytical procedure.

7.5.6 **Assessment of validation data**

7.5.6.1 The following performance characteristics must be assessed as part of any validation exercise:

- Precision (%RSD).
- Bias.

Further detail on assessment of validation data can be found in Annex B.

7.5.6.2 All performance characteristics must be calculated using final sample concentrations, i.e. corrected for volume, blank levels and, where applicable, recovery.

7.5.6.3 Each validation test sample type will return individual estimates of precision and bias. For each performance characteristic, overall method performance for a determinand will be quoted as the largest estimate taken from all relevant test sample types.

7.5.7 **MDL assessment**

7.5.7.1 MDL must be determined using within-batch performance data. This will be carried out with a minimum of ten degrees of freedom, using real matrix where possible. In practice, this can be achieved by analysis of 11 batches, each containing duplicates of the MDL test sample type.

Further detail on MDL assessment can be found in Annex C.

7.5.8 **Ongoing validation**

7.5.8.1 As a minimum, a reassessment of MDL is required every six years.

7.5.8.2 Overall method performance will be continually assessed by appropriate use of analytical quality control, and will be subject to annual review (see 7.6).

7.5.9 **Revalidation**

7.5.9.1 Any modification to a previously validated and accredited analytical method may affect the resulting performance. Where significant modifications are made, analytical methods will be subject to revalidation before any data generated is considered suitable for submission to SEPA.

7.5.9.2 For methods used to generate data reported under MACS, both SEPA and UKAS must be:

- notified when revalidation has been performed;
- provided with full detail of any significant modification(s) made to a method.

7.5.9.3 The degree of revalidation necessary will be proportional to the significance of any modification. Assessment of the significance level of a modification is a matter of judgement for each individual laboratory.

7.5.9.4 Full method validation (see 7.5) is **always** required under the following circumstances:

- Introduction of a new determinand to an existing method.

If previous method performance is **not** to be retained and new performance characteristics are sought, then full method validation is also required under the following circumstances:

- Introduction of a new sample matrix to an existing method.
- Significant change to the range of a method.
- Direct replacement of a significant piece of test equipment.
- Relocation of existing test equipment.
- Transfer of method to second laboratory.

7.5.9.5 Where the circumstances above do not apply, or a laboratory judges that a modification is not significant enough to warrant full revalidation, then a partial revalidation must be performed. In these instances the approval of UKAS must be sought before proceeding.

NOTE: Where partial revalidation is approved it shall consist of a '6×2' exercise (i.e. six batches containing two replicates) comprising either a CRM; or Unspiked and Spiked sample matrix test sample types (see 7.5.5.5).

7.5.9.6 Where the laboratory judges that a modification is not significant enough to warrant any revalidation such decisions must be documented and fully justifiable.

NOTE: Care must be taken to ensure that the cumulative effects of several minor changes do not alter overall method performance [i.e. through close monitoring of internal AQC and PT performance (see 7.6)].

7.6 Ensuring the validity of results

7.6.1 Having met the required method performance criteria detailed in Annex A, on-going performance of a previously validated method must be continually monitored in order to:

- demonstrate that compliance with the performance criteria required by MACS is maintained in a statistically controlled manner;
- allow for early identification of any changes in method performance (especially deterioration in performance).

These objectives will be achieved by a laboratory through appropriate implementation of the following strategies:

- Internal analytical quality control.
- Participation in proficiency testing programmes.

7.6.2 Internal analytical quality control

7.6.2.1 The practice of analytical quality control (AQC) is dependent on the proper selection, application and monitoring of various quality control measures. Laboratories must ensure that:

- analytical equipment is calibrated and suitable for use;
- methods are free from the effects of interferences and contamination;
- methods are statistically under control and continue to meet performance targets.

These objectives will be achieved by appropriate implementation of the following checks and measures:

7.6.2.2 **System suitability check**

a. In order to ensure that a piece of equipment or instrument is performing acceptably a system suitability check (SSC) must be performed prior to the analysis of any sample(s).

b. Choice of SSC will be dependent on the analytical method in use, but must include assessment of appropriate physical measurement(s) or instrumental parameter(s) against predefined limits. These could include, for example:

- Instrument parameter - sensitivity, slope of calibration etc.
- Physical measurement (of an SSC standard) - absorbance, peak height, peak resolution etc.

c. The SSC assessment criteria selected shall be documented in the analytical procedure and initially based on performance measured during validation. Analysis of samples shall not commence until satisfactory SSC results have been obtained.

d. It must be ensured that SSC assessment criteria are set appropriately so that any deviation from acceptable performance is detected. Assessment criteria should be routinely reviewed; and revised when system performance permanently changes or revalidation is undertaken.

7.6.2.3 **Calibration of analytical equipment**

a. Where possible, instrument calibration must cover the range of the analysis being performed, and will ideally be linear over that range. A minimum of four calibration points (not including a blank) are required (more will be necessary if non-linear calibration is used).

NOTE: It is recognised that this may not be feasible for all determinands (e.g. pH) or when using certain types of analytical equipment (e.g. DO meter, FTIR spectrometer). Where this is the case, clause 7.6.2.3 a. will not apply, but it must be ensured that appropriate alternative calibration measures are put in place by the laboratory.

b. Depending on the method in use, solutions used for instrument calibration purposes may be taken through the entire analytical process or prepared for the determination stage only. Whichever approach is used, solutions shall be matched to the sample extract solution (e.g. prepared in the same solvent).

c. Instrument calibration shall be checked throughout a run by regular analysis of calibration check standards. Frequency of analysis will be dependent on the expected stability of the instrument in use, and will be defined in each individual analytical procedure. As a minimum, all samples must be bracketed by check standards.

d. Check standards must not be used to recalibrate or modify the instrument calibration in any way. If a check standard result fails to meet appropriate predefined control limits the root cause shall be investigated and recorded. Where necessary, the instrument shall be fully recalibrated. Affected samples must be reanalysed.

7.6.2.4 **Analysis of blanks**

a. In order to monitor interferences and contamination levels, and to allow for correction of sample results, at least one blank sample (for a direct method) or one process blank sample (for a pre-treatment method) shall be taken through the entire analytical process with each batch of samples.

NOTE 1: This may not be appropriate for determination of all determinands, e.g. pH.

NOTE 2: A filtered blank or process blank must be analysed alongside each batch of filtered samples (e.g. dissolved metals).

b. Laboratories shall have documented procedures demonstrating how blank samples are utilised. Blank results which indicate significant levels of contamination shall be investigated, and may require affected samples to be reanalysed.

7.6.2.5 **Laboratory control sample(s)**

a. For each analytical method used to generate data for submission to SEPA, method performance must be verified for each batch of samples by simultaneous analysis of the appropriate laboratory control sample(s).

b. Choice of laboratory control sample is a matter of judgement for each individual laboratory. Depending on the required application, the following types of laboratory control sample may be suitable:

- Certified reference material (CRM) - a sample of target matrix containing a specified concentration of the determinand(s) of interest; certified to a quoted uncertainty and traceable to a national/international standard.

- Reference material - a sample of target matrix containing a specified concentration of the determinand(s) of interest; characterised to a quoted uncertainty.

- Laboratory reference material (LRM) - a sample produced by the laboratory (which may be synthetic), containing a specified concentration of the determinand(s) of interest. Typically, LRMs will be prepared in advance of analysis with the intention of repeated use. The sample must be homogenised to ensure that only variations in analytical method performance are reflected by repeat analysis.

- Spiked sample - a sample, representative of the matrix being analysed, to which a known quantity of standard solution is added. Typically, spiked samples will be prepared immediately prior to analysis with the intention of single use only. Standards used for spiking shall be from a different source or lot to that used for calibration, unless other independent checks of calibration stocks are undertaken.

NOTE: In certain circumstances it may not be possible to find a clean matrix for spiked sample preparation, i.e. where there is potential for significant levels of background interference. In these situations, it will be necessary to analyse an unspiked sample (i.e. a portion of unspiked matrix) alongside the spiked sample and calculate method efficiency from the difference between these two control samples.

c. Control samples must be analysed within the analytical batch with which they have been prepared.

d. A minimum of 5% of samples within a batch must be control samples. Where the batch size is less than 20, one control sample per batch is still required.

e. If an analysis performed using an established method is considered to be infrequent, then a greater degree of quality control will be necessary to ensure control is maintained (e.g. duplicate sample analysis, multiple control samples within batches, use of standard addition techniques, use of surrogate compounds etc.)

NOTE: To monitor trends in analytical performance using a Shewhart chart a minimum of 30 points evenly spread over a 12 month period is recommended.

7.6.2.6 **Use of control charts**

a. To allow for easy identification of any deviation from a state of statistical control, and to immediately indicate where corrective action is required, results of control sample measurements must be plotted on statistically based control charts.

NOTE: For methods where multiple laboratory control sample types are employed then only one sample type is required to be plotted on a statistical chart.

b. Following completion of initial method validation control chart limits shall be set using the mean and standard deviation obtained during that validation, until such a time as 20 data points have been collected.

Once sufficient data are collated, control chart mean and standard deviation shall be calculated statistically from that data.

NOTE 1: If a laboratory chooses to also include any intended routine control samples as test sample types during an '11×2' validation then clause 7.6.2.6 b may be ignored; as it will be possible to statistically calculate initial control chart mean and standard deviation from validation data.

NOTE 2: Any clearly atypical, outlying data points with an assigned root cause (e.g. standard spiking error, incorrect sample dilution etc.) shall be omitted from any statistical calculation.

NOTE 3: It is expected that recovery corrected methods will not have a nominal target value. Control chart mean shall initially be set at the calculated mean result of the appropriate validation test sample type until sufficient data are collated to reset statistically.

c. In order to ascertain whether current mean and standard deviation values remain valid, control charts must be formally reviewed at least once annually (or sooner if checks indicate a change in current method performance).

To ensure chart parameters are set correctly any review should be performed using all relevant data. This shall include all data points prior to the most recent entry which can be considered as one continuous population.

NOTE: Any clearly atypical, outlying data points with an assigned root cause shall be omitted from any statistical calculation performed during chart review. Data points in excess of four standard deviations from the mean shall also be excluded from these calculations.

d. During review, the significance of any change in mean and standard deviation shall be tested using the statistical t-test and F-test at the 95% confidence interval, i.e. $\alpha = 0.05$ (see Annex D).

e. If, following review, it is determined that a statistically significant change to mean or standard deviation has occurred then the newly calculated values shall be used to establish new control limits on the control chart. Any decision made regarding an update to a control chart must be justified and recorded.

f. For any given determinand, the method performance targets detailed in Annex A must be met. Where a laboratory's internal AQC records indicate that these targets are being exceeded, then any data generated cannot be submitted to SEPA. If this occurs, appropriate method improvement must be carried out.

g. Laboratories shall have documented procedures that define the loss of statistical control and specify the actions to be taken when control limits are breached.

h. All AQC failures must be investigated immediately, with all findings and resulting actions recorded. Access to these records will be made available to SEPA upon request.

   Examples of checks involved in an AQC investigation may include, but are not limited to:

   - Integrity of stock standard solutions and reagents.
   - Maintenance and calibration of analytical equipment.
   - Adherence to documented procedures.
   - Whether system suitability acceptance criteria were met.
   - Instrument performance during analysis.
   - Recent proficiency testing scheme results.

   AQC investigation records shall include details of:

   - AQC failure and associated control sample.
   - Control limits in operation at occurrence of failure.
   - Unique identifiers of affected analytical run and associated samples.
   - Investigation performed, conclusions made, corrective action taken and effectiveness of implemented correction.
   - Action taken with respect to affected sample results.

i. In exceptional circumstance, SEPA will accept analytical results associated with AQC failures. In each case, as part of the AQC investigation, the operator must obtain from SEPA a concession to report results.

   The concession request shall include full assessment and justification that the AQC failure has had no impact on the quality of the data submitted to SEPA.

   If it is not possible to justify the reporting of analytical results then the results will not be accepted by SEPA.

   NOTE: Where analytical results are associated with an AQC failure, this must be clearly identified when data is returned to SEPA [see supplementary MACS performance standard (ref. 3.2 a)].

7.6.3 **Proficiency testing scheme participation**

7.6.3.1 Any laboratory undertaking an analysis to generate data which will subsequently be submitted to SEPA under MACS must participate in an appropriate external proficiency testing (PT) programme.

7.6.3.2 PT sample(s) provided must reflect the typical sample matrices and determinand concentrations routinely analysed in the laboratory.

7.6.3.3 Where no appropriate external PT scheme is available, laboratories must demonstrate the on-going validity of their analysis methods by other means (e.g. interlaboratory comparisons other than proficiency testing, use of CRMs, replicate testing, intralaboratory comparisons).

7.6.3.4 Once a laboratory has subscribed to a scheme they must endeavour to meet the full requirements of that scheme and participate in the required number of distributions specified by the scheme provider, unless there is reasonable justification for altering the frequency.

NOTE: Partial participation in a scheme cannot be decided based on cost, but could be justified, for example, if a test is only run twice per year but five PT distributions are provided.

7.6.3.5 Upon receipt, a laboratory must treat all PT samples in the same manner as they would a routine sample.

NOTE: PT results are not required to be reported to SEPA. However, documented analytical procedures and laboratory management processes in place for sample registration, analysis, quality control and data recording must be followed.

7.6.3.6 The PT provider's guidelines must be followed with respect to PT sample storage and preparation prior to analysis.

7.6.3.7 All data submitted to a PT provider for evaluation purposes must be generated by the same method used to submit data to SEPA.

7.6.3.8 Appropriate, documented procedure(s) must be in place to allow for review, investigation and corrective action where results submitted for a PT sample are deemed unsatisfactory or questionable by the scheme organiser.

7.6.3.9 Periodically, but at least annually, the laboratory must review their on-going PT performance in order to examine trends in the data. Significant trends must be investigated.

If a review determines that PT performance has deteriorated to the extent that it is considered out of control then appropriate method improvement must be carried out. In this circumstance, data produced will not be eligible for submission to SEPA until such time as improvement is complete.

7.6.3.10 Full details of a laboratory's PT scheme programme will be made available to SEPA upon request. As a minimum, for each analytical method covered, this shall include:

- PT provider(s).
- PT sample product code(s).
- PT sample matrix.
- Number of sample distributions per annum.

- Determinands covered.

8 Management system requirements

8.1 Control of records

8.1.1 The organisation shall retain records for a period of time of not less than six years. When requested, an organisation shall submit copies of these records to SEPA within 28 calendar days from the date requested.

9 MACS document review and control

9.1 All MACS documentation will be subject to review and amendment. For the latest versions of all MACS performance standards, please refer to the SEPA website:

www.sepa.org.uk

Annex A

Performance characteristics

The minimum performance characteristics for all determinands included in this MACS performance standard are detailed in Tables A1, A2 and A3. This list is not exhaustive; targets and determinands will be amended as regulatory and environmental monitoring requirement changes dictate.

Table A1 – Inorganic determinands (wastewater matrix)

Determinand	Precision[1]	%Bias
Ammonia + TON Total (as N)[2]	5	10
Ammoniacal Nitrogen (as N)	5	10
Anionic detergents[3]	5	10
Biochemical Oxygen Demand - ATU suppressed (BOD)[4][5]	8	10
Chemical Oxygen Demand (COD)[5]	5	10
Chloride	5	10
Cyanide	5	10
Cyanide - free	5	10
Electrical Conductivity (25°C)	5	10
Fluoride	5	10
Nitrate (as N)	5	10
Nitrite (as N)	5	10
pH	0.1[6]	0.2[6]
Reactive Phosphorus (as P)	5	10
Suspended Solids (105°C)[7]	5	10
Total Nitrogen (as N)	5	10
Total Oxidised Nitrogen (as N)	5	10
Total Phosphorus (as P)	5	10

1. Expressed as %RSD.
2. Test determinand is a calculated result made up of a number of individual constituent determinands. The precision and %bias performance characteristics are for the individual determinands. Each individual parameter must meet the set targets.
3. Also known as MBAS (methylene blue active substances).
4. Standard 5 day analysis, Allylthiourea (ATU) suppressed.
5. Includes filtered BOD and/or filtered COD when stated as a monitoring requirement in the operator's Annual Monitoring Plan. Sample filtered through GF/C (1.2 µm) filter paper before analysis and filtrate analysed as per standard test.
6. Precision and bias for pH expressed in pH units not in percentage terms.
7. Sample filtered through GF/C (1.2 µm) filter paper. Filter dried for 1 hour at 105 °C.

Table A2 – Metal determinands (wastewater matrix)

Determinand	Precision[1]	%Bias
Aluminium	7.5	15
Aluminium - passing 0.45µm membrane[2]	7.5	15
Arsenic	7.5	15
Arsenic - passing 0.45µm membrane[2]	7.5	15
Cadmium	7.5	15
Cadmium - passing 0.45µm membrane[2]	7.5	15
Chromium	7.5	15
Chromium - passing 0.45µm membrane[2]	7.5	15
Copper	7.5	15
Copper - passing 0.45µm membrane[2]	7.5	15
Iron	7.5	15
Iron - passing 0.45µm membrane[2]	7.5	15
Lead	7.5	15
Lead - passing 0.45µm membrane[2]	7.5	15
Manganese	7.5	15
Mercury	7.5	15
Mercury - passing 0.45µm membrane[2]	7.5	15
Nickel	7.5	15
Nickel - passing 0.45µm membrane[2]	7.5	15
Zinc	7.5	15
Zinc - passing 0.45µm membrane[2]	7.5	15

1. Expressed as %RSD.
2. Sample filtered through 0.45 µm membrane filter (or equivalent) and filtrate analysed by standard method.

Table A3 – Organic determinands (wastewater matrix)

Determinand	Precision[1]	%Bias
Chlorfenvinphos	12.5	25
Chloroform	12.5	25
cis-Permethrin	12.5	25
Cyfluthrin	12.5	25
Diazinon	12.5	25
Dichloromethane	12.5	25
gamma - HCH[2]	12.5	25
Pentachlorophenol	12.5	25
Permethrin - all isomers total[3][4]	12.5	25
Propetamphos	12.5	25
Total Nonionic Detergents[3][5]	12.5	25
Total Petroleum Hydrocarbons	12.5	25
trans-Permethrin	12.5	25

1. Expressed as %RSD.
2. gamma-hexachlorocyclohexane (Lindane).
3. Test determinand is a calculated result made up of a number of individual constituent determinands. The precision and % bias performance characteristics are for the individual determinands. Each individual parameter must meet the set targets.
4. Required constituent determinands: cis-Permethrin and trans-Permethrin.
5. Required constituent determinands: 4-Nonylphenol monoethoxylate, 4-Nonylphenol diethoxylate, 4-Nonylphenol triethoxyate, p-tert-octylphenol monoethoxylate and p-tert-octylphenol diethoxylate.

Temperature

Temperature measurement is not listed in the tables above. Certification to this MACS performance standard can be granted for temperature provided the relevant requirements of ISO/IEC 17025 are met.

Annex B

Assessment of method validation data

B.1 MACS requires a common approach to the calculation and assessment of method validation data in order to ensure that method suitability can be evaluated in a consistent and comparable fashion across operators.

Upon the completion of a validation exercise, assessment must be made between method performance measured during validation and the agreed MACS target performance criteria. In order for a method to be considered acceptable for use under MACS, it must be demonstrated that each of these targets are achieved.

B.2 Where performance measured during method validation does not meet an agreed target, significance testing may be employed in order to ascertain whether the observed difference is statistically significant.

If it is determined that an observed difference **is** significant, then the performance of the analytical method employed is not considered satisfactory. Further method development or the use of an alternative analysis technique will be necessary to comply with the requirements of this performance standard.

B.3 Assessment of precision

Once a validation exercise has been completed, an estimation of the measured precision must be made using statistical Analysis of Variance (ANOVA) techniques.

MACS requires that two separate comparisons are made as part of the overall precision assessment:

- Comparison of within-batch and between-batch variance.
- Comparison of measured and target precision (%RSD).

The outcome of both comparisons must be acceptable in order for measured precision to be considered satisfactory.

B.3.1 Comparison of within-batch and between-batch variance

This comparison assesses whether a significant difference exists between observed within-batch and between-batch variances for each validation test sample type.

a. In practice, this first requires the calculation of the within-batch and between-batch mean squares, M_0 and M_1 respectively:

$$\boldsymbol{M_O} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{m}} \frac{\boldsymbol{s_i}^2}{\boldsymbol{m}}$$

where:

- s_i is the standard deviation of an individual batch.
- m is the total number of batches.

$$\boldsymbol{M_1 = n.{s_{bm}}^2}$$

where:

- s_{bm} is the standard deviation of the batch means.
- n is the number of replicates in each batch.

b. A two-tailed F-test at the 95% confidence interval (see NOTE 2, below) is then applied to determine whether there is a statistically significant difference between the calculated variances:

$$\boldsymbol{F_{(obs)} = \frac{\sigma_1}{\sigma_2}}$$

where:

- σ_1 and σ_2 are the within-batch and between-batch mean squares (M_0 and M_1 respectively), assigned by $\sigma_1 > \sigma_2$ (see NOTE 1, below).

NOTE 1: In a two-tailed F-test the highest variance should always be used as the numerator when calculating the observed F value ($F_{(obs)}$) in order to ensure a result greater than one.

NOTE 2: Use of a two-tailed F-test requires that the significance level is halved when determining the critical value of F ($F_{(crit)}$) i.e. for this performance standard α = 0.025.

c. In determining the critical value of F ($F_{(crit)}$), degrees of freedom for each variance are to be calculated as follows:

within-batch (M_0): $\boldsymbol{df = m(n-1)}$

between-batch (M_1): $\boldsymbol{df = m-1}$

where:

- m is the total number of batches.
- n is the number of replicates in each batch.

d. There are three possible outcomes:

i. No significant difference exists between M_0 and M_1 (i.e. $F_{(obs)} \leq F_{(crit)}$) - this is considered a pass.

ii. M_1 is significantly greater than M_0 (i.e. $F_{(obs)} > F_{(crit)}$; and between-batch variance > within-batch variance) - this is a common situation in many methods and may also be considered a pass, providing the target %RSD is also met (see B.3.2).

iii. M_0 is significantly greater than M_1 (i.e. $F_{(obs)} > F_{(crit)}$; and within-batch variance > between-batch variance) - this is considered a fail, and is indicative of a potential problem with the method. The laboratory should investigate, assess,

and perform additional method development and/or repeat the validation exercise as required.

NOTE: It is recognised that in exceptional circumstance M_0 may be significantly greater than M_1 but method performance cannot be further improved by additional development (e.g. when total standard deviation (s_t) is very low). In such instances, the laboratory may comply with the requirements of this performance standard provided that both the target %RSD is met, and they are able to justify acceptance of the validation data to the body providing recognition.

B.3.2 Comparison of measured and target precision (%RSD)

This comparison assesses whether the measured precision, expressed as percent relative standard deviation % (%RSD), meets the required target precision (%RSD) detailed in Annex A.

B.3.2.1 Calculation of measured precision (%RSD)

a. By manipulating the mean square values obtained from ANOVA (see B.3.1) using the calculation detailed below, an estimate of total standard deviation (s_t) will be made for each validation test sample type:

$$\boldsymbol{s_t = \sqrt{\frac{(M_1 + (n-1)M_0)}{n}}}$$

where:

- M_0 is the within-batch mean square.
- M_1 is the between-batch mean square.
- n is the number of replicates in each batch.

b. The measured %RSD of each test sample type may then be calculated as follows:

$$\boldsymbol{\%RSD = \frac{s_t}{\bar{x}} \times 100}$$

where:

- s_t is the total standard deviation.
- $\bar{x}$ is the mean of results.

c. The measured %RSD for each validation test sample type must then be assessed against the appropriate target %RSD detailed in Annex A.

If the measured value is less than or equal to the target value, the required precision has been achieved, performance is considered satisfactory and no further action is required.

If the measured value is greater than the target value, it is still possible to comply with the requirements of this performance standard if statistical significance testing indicates that the exceedance is not significant (see B.3.2.2).

B.3.2.2 **Significance testing of precision (%RSD)**

a. A one-tailed F-test at the 95% confidence interval (α = 0.05) is applied to determine whether the difference between the measured precision (%RSD) and the target precision (%RSD) is statistically significant:

$$\boldsymbol{F_{(obs)} = \frac{{s_t}^2}{{Z_p}^2}}$$

where:

- s_t is the measured total standard deviation.
- Z_p is the target standard deviation.

b. The target standard deviation (Z_p) can be calculated from both the MACS target %RSD and the operator's target MDL.

The value used when determining the observed F value ($F_{(obs)}$) will be whichever of the two calculated Z_p values below is the greater:

$$\boldsymbol{Z_p = \overline{x} \times \frac{target\ \%RSD}{100}}$$

or

$$\boldsymbol{Z_p = \frac{target\ MDL}{4}}$$

where:

- $\bar{x}$ is the mean of results.

c. In determining the critical value of F ($F_{(crit)}$), an estimated number of degrees of freedom for s_t are to be calculated as follows, with the final value rounded to the nearest whole number:

$$\boldsymbol{df = \frac{m(m-1)(M_1 + (n-1)M_0)^2}{m{M_1}^2 + (m-1)(n-1){M_0}^2}}$$

where:

- M_0 is the within-batch mean square.
- M_1 is the between-batch mean square.
- m is the total number of batches.
- n is the number of replicates in each batch.

Degrees of freedom for Z_p are infinite, although for calculation purposes a value of ≥ 10^{10} is considered sufficient for the requirements of this performance standard.

d. There are two possible outcomes:

i. The measured precision **is not** significantly greater than the target precision (i.e. $F_{(obs)} \leq F_{(crit)}$) - this is considered a pass, the required precision has been achieved and performance is considered satisfactory.

ii. The measured precision **is** significantly greater than the target precision (i.e. $F_{(obs)} > F_{(crit)}$) - this is considered a fail, the required precision has not been achieved and performance is not considered satisfactory.

B.4 Assessment of bias (systematic error)

An assessment of bias, or systematic error, need only be made if the assessment of precision (see B.3) has proved acceptable.

B.4.1 Comparison of measured and target bias

This comparison assesses whether the measured bias, expressed as a percentage (%Bias), meets the required target %Bias detailed in Annex A.

Assessment of measured %Bias for a method is based on the difference of the actual mean of results from a 'true' or expected concentration. This assessment may be made using data generated from the analysis of reference materials or from the results of spiked/unspiked sample matrix pairs.

B.4.1.1 Calculation of measured bias

a. The theory behind the calculation of measured %Bias is identical regardless of whether the analysis of reference materials or the results of spiked/unspiked pairs are used:

$$\%\boldsymbol{Bias} = \frac{(\bar{\boldsymbol{x}} - \boldsymbol{E})}{\boldsymbol{E}} \times \mathbf{100}$$

where:

- $\bar{x}$ is the mean of results.
- E is the expected, or 'true' concentration.

b. It is important to note that the expected concentration (E) used in the calculation above is defined differently depending on which experimental approach is used.

- Where analysis of reference materials has been used to generate the result mean, the expected concentration is the accepted reference value of the material(s) used.

- Where spiked/unspiked pairs have been used to generate the result mean, the expected difference in concentration is to be calculated as follows:

$$\boldsymbol{E} = \frac{\boldsymbol{v}(\boldsymbol{C} - \boldsymbol{U})}{\boldsymbol{V} + \boldsymbol{v}}$$

where:

- C is the concentration of the spiking solution.
- U is the mean of the unspiked sample results.
- v is the volume of spiking solution added.
- V is the volume of sample which has been spiked.

c. The measured %Bias for each validation test sample type must then be assessed against the appropriate target %Bias detailed in Annex A.

If the measured value is less than or equal to the target value, the required bias has been achieved, performance is considered satisfactory and no further action is required.

If the measured value is greater than the target value, it is still possible to comply with the requirements of this performance standard if statistical significance testing indicates that the exceedance is not significant (see B.4.2).

B.4.2 **Significance testing of bias**

a. A one-tailed t-test at the 95% confidence interval (α = 0.05) is applied to determine whether the difference between the measured bias (expressed as a concentration) and the target bias is statistically significant:

$$\boldsymbol{t_{(obs)} = \frac{|(|measured\ bias| - Z_b)|}{SE}}$$

where:

- Z_b is the target bias (expressed as a concentration).
- SE is the standard error of the batch means, calculated as:

$$\boldsymbol{SE = \frac{s_{bm}}{\sqrt{m}}}$$

where:

- s_{bm} is the standard deviation of the batch means.
- m is the total number of batches.

NOTE: The symbol $|measured\ bias|$ signifies the value of measured bias regardless of sign. Likewise, the symbol $|(|measured\ bias| - Z_b)|$ signifies the value of $(|measured\ bias| - Z_b)$ regardless of sign.

b. The target bias (Z_b) can be calculated from both the MACS target %RSD and the operator's target MDL.

The value used when determining the observed t value ($t_{(obs)}$) will be whichever of the two calculated Z_b values below is the greater:

$$Z_b = E \times \frac{target\ \%Bias}{100}$$

or

$$Z_b = \frac{target\ MDL}{2}$$

where:

- E is the expected, or 'true' concentration.

c. In determining the critical value of t ($t_{(crit)}$), degrees of freedom are to be calculated as follows:

$$df = m - 1$$

where:

- m is the total number of batches.

d. There are two possible outcomes:

i. The measured bias **is not** significantly different from the target bias (i.e. $t_{(obs)} \leq t_{(crit)}$) - this is considered a pass, the required bias has been achieved and performance is considered satisfactory.

ii. The measured bias **is** significantly different from the target bias (i.e. $t_{(obs)} > t_{(crit)}$) - this is considered a fail, the required bias has not been achieved and performance is not considered satisfactory.

B.5 Worked example

B.5.1 The example on the following pages is presented to demonstrate the application of the theory, statistical tests and assessments described above.

It considers a hypothetical method validation exercise for a determinand with the following minimum performance criteria:

- Precision (%RSD) target: 5%
- Bias target: ±10%
- Required MDL: 0.5 mg/L

NOTE 1: The corrected test sample results used in Table B1 have been generated manually for illustrative purposes only, and do not represent real analytical validation data.

NOTE 2: Test sample results are included for both a reference material (CRM) and a spiked/unspiked sample matrix pair (Spiked minus Unspiked sample matrix). In practice, concurrent analysis of both of these test types is not a mandatory requirement for a validation exercise (see 7.5.5.5).

Table B1 – '11×2' validation results (corrected)

Batch	Replicate	10% standard	90% standard	CRM	Unspiked Sample Matrix	Spiked Sample Matrix	Spiked minus Unspiked
1	1	10.090	92.910	43.231	4.133	79.899	75.766
	2	9.730	87.070	43.556	4.550	79.330	74.780
batch mean	$\bar{x}_i$	9.9100	89.9900	43.3935	4.3415	79.6145	75.2730
within batch st dev	s_i	0.25456	4.12950	0.22981	-	-	0.69721
within batch variance	s_i^2	0.0648	17.0528	0.0528	-	-	0.4861
2	1	10.220	90.100	43.086	4.688	80.227	75.539
	2	11.450	88.330	39.914	4.376	79.380	75.004
batch mean	$\bar{x}_i$	10.8350	89.2150	41.5000	4.5320	79.8035	75.2715
within batch st dev	s_i	0.86974	1.25158	2.24294	-	-	0.37830
within batch variance	s_i^2	0.7564	1.5664	5.0308	-	-	0.1431
3	1	9.730	92.270	46.674	4.560	79.637	75.077
	2	9.500	92.790	45.165	4.417	80.336	75.919
batch mean	$\bar{x}_i$	9.6150	92.5300	45.9195	4.4885	79.9865	75.4980
within batch st dev	s_i	0.16263	0.36770	1.06702	-	-	0.59538
within batch variance	s_i^2	0.0265	0.1352	1.1385	-	-	0.3545
4	1	9.370	90.740	45.585	4.770	77.871	73.101
	2	9.840	91.720	37.062	4.564	77.039	72.475
batch mean	$\bar{x}_i$	9.6050	91.2300	41.3235	4.6670	77.4550	72.7880
within batch st dev	s_i	0.33234	0.69296	6.02667	-	-	0.44265
within batch variance	s_i^2	0.1105	0.4802	36.3208	-	-	0.1959
5	1	9.890	92.150	44.693	5.189	79.114	73.925
	2	10.440	89.180	45.247	5.882	79.565	73.683
batch mean	$\bar{x}_i$	10.1650	90.6650	44.9700	5.5355	79.3395	73.8040
within batch st dev	s_i	0.38891	2.10011	0.39174	-	-	0.17112
within batch variance	s_i^2	0.1512	4.4105	0.1535	-	-	0.0293
6	1	9.840	91.830	50.017	5.055	79.389	74.334
	2	10.000	90.340	46.385	5.720	78.773	73.053
batch mean	$\bar{x}_i$	9.9200	91.0850	48.2010	5.3875	79.0810	73.6935
within batch st dev	s_i	0.11314	1.05359	2.56821	-	-	0.90580
within batch variance	s_i^2	0.0128	1.1100	6.5957	-	-	0.8205
7	1	10.200	89.840	46.369	4.239	78.304	74.065
	2	10.210	92.160	44.948	4.678	79.836	75.158
batch mean	$\bar{x}_i$	10.2050	91.0000	45.6585	4.4585	79.0700	74.6115
within batch st dev	s_i	0.00707	1.64049	1.00480	-	-	0.77287
within batch variance	s_i^2	0.0001	2.6912	1.0096	-	-	0.5973
8	1	10.590	92.550	42.043	5.271	79.437	74.166
	2	9.980	88.600	42.905	5.310	79.736	74.426
batch mean	$\bar{x}_i$	10.2850	90.5750	42.4740	5.2905	79.5865	74.2960
within batch st dev	s_i	0.43134	2.79307	0.60953	-	-	0.18385
within batch variance	s_i^2	0.1861	7.8013	0.3715	-	-	0.0338
9	1	9.320	91.240	50.800	4.501	78.513	74.012
	2	9.330	85.530	49.954	5.149	79.835	74.686
batch mean	$\bar{x}_i$	9.3250	88.3850	50.3770	4.8250	79.1740	74.3490
within batch st dev	s_i	0.00707	4.03758	0.59821	-	-	0.47659
within batch variance	s_i^2	0.0000	16.3021	0.3579	-	-	0.2271
10	1	10.690	88.750	47.608	4.802	78.552	73.750
	2	9.780	86.950	46.678	4.920	79.382	74.462
batch mean	$\bar{x}_i$	10.2350	87.8500	47.1430	4.8610	78.9670	74.1060
within batch st dev	s_i	0.64347	1.27279	0.65761	-	-	0.50346
within batch variance	s_i^2	0.4141	1.6200	0.4325	-	-	0.2535
11	1	9.850	86.950	45.255	5.172	78.952	73.780
	2	10.860	87.080	41.990	5.277	78.642	73.365
batch mean	$\bar{x}_i$	10.3550	87.0150	43.6225	5.2245	78.7970	73.5725
within batch st dev	s_i	0.71418	0.09192	2.30870	-	-	0.29345
within batch variance	s_i^2	0.5101	0.0084	5.3301	-	-	0.0861

conc. of spiking solution (wt/L)	C	-	-	-	-	85000	-
vol. of spiking solution added (L)	v	-	-	-	-	0.001	-
vol. of sample spiked (L)	V	-	-	-	-	1	-
expected concentration (wt/L)	E	10	90	50	-	-	84.910

mean	$\bar{x}$	10.041	89.958	44.962	4.874	79.170	74.297
%bias		0.414	-0.046	-10.076	-	-	-12.500
within batch mean square	M_0	0.2030	4.8344	5.1631	-	-	0.2934
between batch mean square	M_1	0.3569	5.5204	16.3282	-	-	1.3784
st dev of batch means	s_{bm}	0.4224	1.6614	2.8573	-	-	0.8302
standard error of batch means	SE	0.1274	0.5009	0.8615	-	-	0.2503
total st dev	s_t	0.5291	2.2754	3.2780	-	-	0.9143

B.5.2 Applying the protocols for assessment of precision and bias previously outlined in B.3 and B.4 to the corrected test sample results presented in Table B1 produces the statistical summary in Table B2, below.

Table B2 – Summary statistics

		10% standard	90% standard	CRM	Spiked minus Unspiked	
between-batch mean square	M_1	0.3569	5.5204	16.3282	1.3784	
within-batch mean square	M_0	0.2030	4.8344	5.1631	0.2934	
observed F value	$F_{(obs)}$	1.759	1.142	3.162	4.698	Precision - ANOVA
critical F value	$F_{(crit)}$	3.526	3.526	3.526	3.526	
significant?		N.S.	N.S.	N.S.	*	
assessment		PASS	PASS	PASS	CHECK %RSD	
mean	$\bar{x}$	10.0414	89.9582	44.9620	74.2966	
total st dev	s_t	0.529	2.275	3.278	0.914	
measured relative st dev	%RSD	5.27	2.53	7.29	1.23	
st dev from RSD		0.502	N/A	2.248	N/A	
st dev from MDL		0.125	N/A	0.125	N/A	
target st dev	Z_p	0.502	N/A	2.248	N/A	Precision - %RSD
observed F value	$F_{(obs)}$	1.110	N/A	2.126	N/A	
critical F value	$F_{(crit)}$	1.587	N/A	1.666	N/A	
estimated degrees of freedom	d.f.	19	N/A	16	N/A	
significant?		N.S.	N/A	*	N/A	
assessment		PASS	PASS	FAIL	PASS	
measured %bias		0.41	-0.05	-10.08	-12.50	
measured bias (conc.)		N/A	N/A	-5.038	-10.614	
bias (conc.) from RSD		N/A	N/A	5.000	8.491	
bias (conc.) from MDL		N/A	N/A	0.250	0.250	
target bias (conc.)	Z_b	N/A	N/A	5.000	8.491	%Bias
observed t value	$t_{(obs)}$	N/A	N/A	0.044	8.480	
critical t value	$t_{(crit)}$	N/A	N/A	1.812	1.812	
significant?		N/A	N/A	N.S.	*	
assessment		PASS	PASS	PASS	FAIL	

N/A significance testing not applicable

N.S. not significant

* significant at the 0.05 level

In this example, the summary statistics are interpreted as follows:

a. **'*Precision - ANOVA*' assessment**

Comparison of within-batch and between batch variance is acceptable for the 10% standard, 90% standard and CRM, as no significant difference is found between M_0 and M_1.

On this occasion, the assessment may also be considered a pass for the Spiked minus Unspiked results as, although M_1 is found to be significantly greater than M_0, the target precision (%RSD) is also met.

b. **'*Precision - %RSD*' assessment**

The required target precision is met for the 90% standard (2.53%) and the Spiked minus Unspiked results (1.23%), so significance testing is not necessary.

The measured %RSD of the 10% standard (5.27%) does not meet the required target, but is found not to be significantly different (i.e. $F_{(obs)} \leq F_{(crit)}$) once an F-test is performed. It can therefore also be considered acceptable.

The measured %RSD of the CRM (7.29%) does not meet the required target precision and is found to be significantly different once an F-test is performed (i.e. $F_{(obs)} > F_{(crit)}$). As a result, performance is not considered acceptable for this test type.

c. **'*%Bias*' assessment**

The required target bias is met for the 10% standard (+0.41%) and the 90% standard (-0.05%). Significance testing is not necessary.

The measured %Bias of the CRM (-10.08%) does not meet the required target, but is found not to be significantly different (i.e. $t_{(obs)} \leq t_{(crit)}$) once a t-test is performed. It can therefore also be considered acceptable.

NOTE: Had the CRM results in this example been generated from a real validation, assessment of bias would not be required as the precision assessment has already been deemed unsatisfactory (see B.4). Bias assessment has been performed in this case for indicative purposes only.

The measured %Bias of the Spiked minus Unspiked results (-12.50%) does not meet the required target bias and is found to be significantly different once a t-test is performed (i.e. $t_{(obs)} > t_{(crit)}$). As a result, performance is not considered acceptable for this test type.

Annex C

Method detection limit

C.1 MACS requires the adoption of a common approach to method detection limit (MDL) assessment in order to ensure that all operator supplied data can be evaluated in a consistent and comparable fashion.

Assessment of MDL must be undertaken during method validation. MDL shall be determined with a minimum of ten degrees of freedom, using the within-batch performance data generated by analysis of duplicate MDL test sample types during an '11×2' validation exercise.

C.2 **MDL test sample type**

Wherever possible, the MDL validation test sample type will be prepared using real sample matrix. Choice of appropriate MDL test sample is dependent on the type of analytical method to be employed:

a. For methods which **are** capable of returning numeric values at levels below the instrument detection limit (i.e. negative values), no determinand(s) of interest should be present in the chosen test sample type.

   MDL is to be determined using measurements obtained from an unspiked, blank real sample matrix.

b. For methods which are **not** capable of returning numeric values at levels below the instrument detection limit, a measurable amount of the determinand(s) of interest should be present in the chosen test sample type.

   MDL is to be determined using measurements obtained from a blank real sample matrix, spiked with determinand(s) of interest at a level approximately two to five times the instrument detection limit.

   Alternatively, a real sample matrix may be used with sufficiently low levels of determinand(s) of interest naturally present. There would be no requirement to spike this sample matrix.

NOTE 1: In certain circumstances it may not be possible to find a suitable real matrix for MDL assessment, i.e. where there is potential for the presence of significant natural levels of the determinand(s) of interest. In these situations, use of ideal matrix is acceptable.

NOTE 2: During validation, the MDL test sample type must **not** be used as the blank or process blank.

C.3 MDL calculation

C.3.1 Theory

For the purposes of this MACS performance standard, MDL is defined by the equation:

$$\boldsymbol{MDL = 4.65 \times s_w}$$

where:

- s_w is the pooled within-batch standard deviation of the MDL test sample type.

$$\boldsymbol{s_w = \sqrt{M_0}}$$

where:

- M_0 is the within-batch mean square (also known as the pooled estimate of within-batch variance).

$$\boldsymbol{M_0 = \sum^{m} \frac{s_i^2}{m}}$$

where:

- s_i is the standard deviation of an individual batch.
- m is the total number of batches.

So, when 11 batches of duplicate MDL test type samples are analysed during an '11×2' validation exercise:

$$\boldsymbol{s_w = \sqrt{\sum s_i^2}}$$

where:

- s_i is the standard deviation of an individual batch.

NOTE 1: Before accepting the calculated method detection limit, it must be ensured that s_w is calculated from data consisting of final sample concentrations i.e. recovery corrected (where applicable). Data used in MDL calculations should **not** be blank corrected.

NOTE 2: Quoted MDL values must always be reported in the same units as the determinand represented. The calculated MDL value for a determinand may be rounded up for convenience and ease of use.

C.3.2 Worked example

Table C1 – '11×2' MDL test sample results (µg/L)

Batch No.	1	2	3	4	5	6	7	8	9	10	11
Replicate 1(1) Replicate 2(1)	22.92 23.38	21.81 22.27	23.99 23.68	23.23 22.77	23.39 22.78	22.57 22.90	22.25 21.22	22.70 22.15	22.83 22.48	22.74 23.78	23.39 21.87
Within-batch st. dev. (s_i)	0.32527	0.32527	0.21920	0.32527	0.43134	0.23335	0.72832	0.38891	0.24749	0.73539	1.07480
Within-batch variance (s_i^2)	0.1058	0.1058	0.04805	0.1058	0.18605	0.05445	0.53045	0.15125	0.06125	0.5408	1.1552

1. Final sample concentrations. Not blank corrected.

Applying the theory previously outlined in C.3.1 to the example test sample data from Table C1, above, produces the following results:

$$M_0 = \sum_{i=1}^{m} \frac{{s_i}^2}{m} = \frac{3.0449}{11} = 0.2768$$

$$s_w = \sqrt{M_0} = \sqrt{0.2768} = 0.5261$$

$$MDL = 4.65 \times s_w = 4.65 \times 0.5261 = 2.45 \text{ µg/L}$$

Annex D

Use of the statistical t-test and F-test during control chart review

D.1 When performing a control chart review, the statistical significance tests, t-test and F-test, are used to decide whether or not two sets of data are significantly different at a given significance level (typically 5%, or α = 0.05; this is also called the 95% confidence interval).

t and F values shall be calculated (using sample standard deviation and sample size) and compared with critical values of t and F. If the calculated values are less than the critical values then there is no significant difference detected. If the calculated values are greater than the critical values then there is a significant difference detected.

D.2 F-test assessment, calculation and comparison to critical F value

The F-test is used to compare the standard deviation of two datasets and test whether or not they are significantly different from one another.

$$F = \frac{s_1^2}{s_2^2}$$

- s_1 and s_2 are sample 1 and 2 standard deviation (1 and 2 are assigned by $s_1 > s_2$).

The observed value of F is compared with a critical value from statistical reference tables. The critical value of F determining whether difference is significant depends on the size of both samples and the significance level (for this performance standard, α = 0.05).

D.3 t-test assessment

The t-test is used to compare the mean of two datasets and test whether or not they are significantly different from one another. Before applying a t-test, the F-test must first be performed and its result used to inform selection of the appropriate t-test to use.

$$t = \frac{|(\overline{x} - \overline{y})|}{s_m}$$

- $\overline{x}$ and $\overline{y}$ are sample 1 and 2 means.
- s_m is the estimated standard deviation of the difference between the two means.

NOTE: The symbol $|(\overline{x} - \overline{y})|$ signifies the value of $(\overline{x} - \overline{y})$ regardless of sign.

The choice of which t-test calculation to use depends on whether there is a significant difference in the two sample standard deviations as observed by the F-test. In the two scenarios s_m is calculated in slightly different ways:

D.3.1 t-test when F-test does not show significant difference

If the F-test demonstrates no significant difference between the two sets of data then perform the following t-test calculation using a pooled estimate of the standard deviation. This method assumes that the samples are drawn from populations with equal standard deviations.

$$t = \frac{|(\overline{x} - \overline{y})|}{\sqrt{\quad}}$$

- $\overline{x}$ and $\overline{y}$ are sample 1 and 2 means.
- n_1 and n_2 are the sample sizes.
- s is the pooled estimate of the standard deviation calculated from the individual sample standard deviations:

$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{\quad}}$$

- n_1 and n_2 are the sample sizes.
- s_1 and s_2 are the standard deviations for the two populations.

The observed value of t is compared with a critical value from statistical reference tables. The critical value of t determining whether difference is significant depends on the size of both samples and the significance level (for this performance standard, α = 0.05).

D.3.2 t-test when F-test does show significant difference

If the F-test demonstrates a significant difference between the two sets of data then perform the following t-test calculation where standard deviations from the individual datasets are applied. This approximate method is used when it cannot be assumed that the two samples come from populations with equal standard deviations.

$$t = \frac{|(\overline{x} - \overline{y})|}{\sqrt{\quad}}$$

- $\overline{x}$ and $\overline{y}$ are sample 1 and 2 means.
- n_1 and n_2 are the sample sizes.
- s_1 and s_2 are the standard deviations for the two populations.

NOTE: The symbol $|(\overline{x} - \overline{y})|$ signifies the value of $(\overline{x} - \overline{y})$ regardless of sign.

The observed value of t is compared with a critical value from statistical reference tables. The critical value of t determining whether difference is significant depends on the size of both samples and the significance level (for this performance standard, α = 0.05).

Annex E

Nominal cross references with ISO/IEC 17025:2017

The table below cross references the clauses in this MACS performance standard with the clauses of ISO/IEC 17025.

This MACS performance standard	ISO/IEC 17025
2	1
3.1	2
4	3
5	5
5.1	5.1
5.2	5.4
5.3	5.6
6	6
6.1	6.3
6.2	6.6
7	7
7.1	7.3
7.1.2	7.3.1
7.2	7.2
7.3	7.2.1
7.4	7.4
7.5	7.2.2
7.6	7.7
7.6.2	7.7.1 (a)
7.6.3	7.7.2 (a)
8	8
8.1	8.4
8.1.1	8.4.2