INTERNATIONAL STANDARD

ISO 3725

First edition
2023-07

Ships and marine technology — Aquatic nuisance species — Methods for evaluating the performance of compliance monitoring devices for ballast water discharges

Reference number
ISO 3725:2023(E)

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

Published in Switzerland

Contents

Page

Foreword

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

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

ISO draws attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO takes no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO had not received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents. ISO shall not be held responsible for identifying any or all such patent rights.

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

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

This document was prepared by Technical Committee ISO/TC 8, *Ships and marine technology*.

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

Introduction

A compliance monitoring device (CMD) is an instrument intended to analyse samples of ballast water, to estimate whether the concentration of living or viable organisms present in the sample exceeds, or is at risk of exceeding, the regulated limit [i.e. the discharge standard, (DS)]. Typically, CMDs are designed for use in shipboard and field locations to provide results rapidly and with less effort relative to complex analyses. CMDs are instruments that are relatively new to their application in ballast water testing. They can rely upon standard optical, chemical, or physical measurements, but these technologies are deployed in unique configurations. They can be packaged in a rugged, transportable housing, or installed as shipboard equipment. A CMD may operate along a spectrum of water types with diverse assemblages of organisms. As intended, CMDs provide critical information to vessel inspectors, ballast water management system (BWMS) commissioning test teams, Port State Control Officers, ship owners, among others.

This document was developed in response to the need for a standardized approach to evaluate the performance of CMDs. This evaluation includes:

— laboratory-based tests using prepared sample water amended with cultured organisms as well as dissolved and particulate materials;

— laboratory-based tests using samples of natural assemblages of organisms, experimentally manipulated to achieve target concentrations of living or viable organisms (but without manipulation of dissolved and particulate materials);

NOTE 1 It is recognized that the end user can require laboratory testing with ambient organisms instead of, or in addition to, cultured organisms. Additionally, the end user can require that both types of laboratory-based tests are conducted using water that is treated by a BWMS or has undergone a simulated ballast water treatment, instead of, or in addition to, un-treated water.

— field-based tests using samples of water treated with a BWMS collected aboard a ship.

This standardized approach defines a general test procedure and minimum set of trials to evaluate the performance of a CMD. The key evaluation metrics are accuracy (hereafter, “trueness” - the agreement to a reference method), precision, and reliability. While a CMD may report numerical values or estimates of organism concentrations, trueness and precision are determined based upon the agreement between the CMD and reference method on the sample disposition (i.e. whether the sample meets or exceeds the DS).

NOTE 2 This approach is not appropriate to evaluate methods or devices intended to be used as an alternate to the reference method, i.e. with precise, numerical measurements across a wide range of organism concentrations.

The test methods are adaptable, such that additional factors which are deemed important — e.g. interferences, organism types, or water characteristics — may be addressed experimentally and included in the set of performance metrics. This flexibility allows end-users to supplement these minimal test requirements to examine additional characteristics, such as CMD performance under different types of BWMS treatments.

Ships and marine technology — Aquatic nuisance species — Methods for evaluating the performance of compliance monitoring devices for ballast water discharges

1 Scope

This document specifies methods to evaluate the performance of a specific class of analytical instruments, known as compliance monitoring devices (CMDs). These instruments are designed and intended to examine ballast water to determine whether a sample meets or exceeds limits for the concentration of living or viable organisms. These limits include those specified by the International Maritime Organization (IMO) Regulation D-2 in the International Convention for the Control and Management of Ships' Ballast Water and Sediments[4] or other discharge standards (DS) adopted by national or regional authorities.

The test methods measure the agreement between the CMD and a reference method to calculate trueness and precision. Both trueness and precision consider only simple, categorical outcomes (e.g. "meets" or "exceeds" the DS). The performance metric reliability is quantified by the frequency of instances when the CMD is not available or is not operating as expected.

The set of tests and trials is based upon the CMD manufacturer claims, such as the DS group(s) targeted by the CMD, and known limitations, including those based upon the salinity of the sample water.

NOTE Additional tests and trials, if required by the end-user, can follow this general test method. Guidance on determining experimental power is found in 7.5. This document provides guidance for customizing the tests to evaluate the claims of the manufacturer or to address optional factors of interest to the end-users.

This document does not set or recommend success criteria of any performance metric, as these are appropriately defined by the end-users.

2 Normative references

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

ISO 11711-1:2019, *Ships and marine technology — Aquatic nuisance species — Part 1: Ballast water discharge sample port*

ISO 11711-2:2022, *Ships and marine technology — Aquatic nuisance species — Part 2: Ballast water sample collection and handling*

ISO 21748, *Guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty evaluation*

ISO/IEC 17025, *General requirements for the competence of testing and calibration laboratories*

ASTM D1141-98, *Standard Practice for Preparation of Substitute Ocean Water*

3 Terms and definitions

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

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

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at https://www.electropedia.org/

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

Note 1 to entry: The more specific term, *trueness* (3.24), is used as a metric throughout this standard.

[SOURCE: ISO 5725-1:1994, 3.6, modified — Note 1 to entry has been replaced.]

3.2
agreement
concurrence between two independent measurements on the outcome of analysis

Note 1 to entry: Analysis outcomes are qualitative or categorical descriptions of whether a sample meets or exceeds the *discharge standard* (3.13).

3.3
ambient water
natural water
water collected directly from the natural environment that 1) contains natural communities of organisms, dissolved and particulate constituents, and 2) has intrinsic characteristics, such as temperature and salinity

3.4
ballast water
water with its abiotic and biotic constituents taken on board a ship to control trim, list, stability or stresses of the ship

3.5
ballast water management system
BWMS
equipment that processes *ballast water* (3.4) such that the water discharged (the treated water) is intended to meet the specified performance requirements for eliminating, inactivating, or reducing aquatic organisms

3.6
calibration
analysis, in water, of standards to develop a relationship between raw output of an analytical system and analyte concentration

3.7
calibration standard
sample containing the analyte of interest at a known concentration either purchased from an external source or prepared in-house from materials of known purity or concentration, or both, and used to calibrate the measurement system

3.8
challenge water
water prepared or manipulated (e.g. by adding organisms and abiotic constituents) to achieve minimum test criteria when testing the performance of equipment, in this case *compliance monitoring devices* (3.11)

Note 1 to entry: This protocol shares some characteristics with the minimum water quality requirements for challenge water for *type approval testing* (3.25) of the International Maritime Organization and USA, such as salinity and temperature ranges and abiotic constituents. However, requirements for concentrations and diversity of organisms are unique to this application.

3.9
colonial organisms
collection of multiple, clonal individuals that are physically connected

Note 1 to entry: Clusters of connected, but non-clonal individuals are typically referred to as aggregated organisms.

3.10
compliance monitoring device
CMD
instrument and its associated analytical methodology typically used as a rapid assessment of the concentration of living or viable organisms in treated ballast water for the purpose of determining compliance or non-compliance with a *discharge standard* (3.13)

3.11
detection limit
method detection limit
lowermost quantity or concentration measurable by the *compliance monitoring device* (3.10)

Note 1 to entry: In the context of compliance monitoring device (CMD) evaluation, the detection limit is specified by the manufacturer. The CMD evaluation may include test samples with concentrations reflecting the stated detection limit to verify the *manufacturer claim* (3.16).

Note 2 to entry: In the context of *reference method* (3.21), the method detection limit is according to the definition in ISO/IEC Guide 99:2007.

3.12
dissolved organic matter
DOM
mass of organic compounds present in water that are not separated by particle (≤0,7 µm) filtration

Note 1 to entry: Dissolved organic carbon is a related quantity that is commonly measured directly. Although the two quantities are related, they are distinct and should be clearly identified.

3.13
discharge standard
DS
regulated limits of organism concentrations allowable in discharged ballast water

Note 1 to entry: Regulation D-2 of the International Maritime Organization's Ballast Water Management Convention.

Note 2 to entry: The term is generic unless a particular DS is specified.

Note 3 to entry: A DS is also known as a performance standard.

3.14
independent testing organization
testing organization that is free of any conflict of interest with the manufacturer of the *compliance monitoring device* (3.10)

3.15
living organism
organism that demonstrates characteristics of life (movement, membrane integrity, etc.)

Note 1 to entry: It is possible that living organisms are not *viable* (3.27).

3.16
manufacturer claims
specific characteristics of the *compliance monitoring device* (3.10) that are asserted by the manufacturer or vendor of the device

Note 1 to entry: Claims typically include the organisms size class(es) or indicator microbe(s) targeted by the device, limitations based upon organism characteristics (such as autotrophy), water temperature and salinity ranges, as well as the detection limits, accuracy, and precision of the compliance monitoring device.

3.17
mineral matter
MM
mass of inorganic compounds present in water and separated by particle (≤0,7 µm) filtration

Note 1 to entry: MM is estimated as the mass difference between *total suspended solids* (3.22) and *particulate organic matter* (3.18).

3.18
particulate organic matter
POM
mass of organic matter present in water and separated by particle (≤0,7 µm) filtration

Note 1 to entry: Particulate organic carbon is a related quantity and composes a portion of the mass of POM. Although the two quantities are related, they are distinct and should be clearly identified.

3.19
precision
agreement between replicate measurements of a sample measured under the same conditions

Note 1 to entry: The same conditions include the same sample, the same instrument unit, and the same analyst, if applicable.

3.20
reagent-grade, purified water
water meeting the characteristics of Type I or II water, used as the basis for preparing challenge water for laboratory testing

Note 1 to entry: The characteristics of Type I or II water are defined in ASTM D1193-06.

3.21
reference method
analytical method that produces a value used as a benchmark

Note 1 to entry: Reference methods produce direct measurements of numerical concentrations that are comparable to the *discharge standard* (3.13).

Note 2 to entry: Reference methods are typically methods used in *ballast water management system* (3.5) *type approval testing* (3.25).

3.22
total suspended solids
TSS
mass of organic and inorganic matter present in water and separated by particle (≤0,7 µm) filtration

Note 1 to entry: TSS is composed of mineral matter and particulate organic matter.

3.23
trial
complete set of samples and sample analyses associated with a single test condition, such as water salinity

3.24
trueness
closeness of agreement between the average value obtained from a large series of test results and an accepted reference value

[SOURCE: ISO 5725-1:1994, 3.7, modified — Notes 1 and 2 to entry have been deleted.]

3.25
type approval testing
testing performed as part of a formal certification of a *ballast water management system* (3.5) for use aboard ships

3.26
uncertainty
measurement uncertainty
parameter, associated with the result of a measurement, which characterizes the dispersion of the values that can reasonably be attributed to the measurand

[SOURCE: ISO 21748:2017, 3.14, modified — "measurement uncertainty" has been added as a preferred term; Notes 1, 2 and 3 to entry have been deleted.]

3.27
viable
living and capable of reproduction

Note 1 to entry: Manufacturers shall indicate whether their *compliance monitoring device* (3.10) quantifies living or viable organisms, and the test should be designed to evaluate their claims using the appropriate *reference method* (3.21) for living or viable organisms.

4 General

4.1 Compliance with the discharge standard

A compliance monitoring device (CMD) determines whether a sample is likely to comply with or exceed the discharge standard (DS), such as the IMO Regulation D-2[4] which sets limits on the concentration of viable organisms in the following size or taxonomic groups:

— organisms ≥50 µm in minimum dimension;

— organisms ≥10 µm and <50 µm in minimum dimension;

— toxicogenic *Vibrio cholerae* (serotypes O1 and O139);

— *Escherichia coli;*

— intestinal enterococci.

NOTE National or regional authorities can define the same or similar categories and concentration limits.

The test methods described in this document are generic: the methods apply to any of the groups defined in the DS and the corresponding reference method used for each of those defined groups. This performance evaluation considers the claims of a CMD manufacturer, such as those defining the targeted group(s), the relevant DS e.g. References [4] and [5], and the stated limitations, such as limits on the salinity of the sample water.

4.2 Reference method

A reference method is an analytical approach used to quantify living or viable organisms in one of the categories of a DS (see 4.1). Typically, the reference method is used during type approval (TA) tests, such as those prescribed in the test protocols of the IMO[6] or the United States.[7] In general, a reference

method estimates numerical concentrations of living or viable organisms in a single category of DS. Consequently, each category examined requires a unique set of measurements using the appropriate reference method.

The numerical results from the reference method shall be reported with estimates of measurement uncertainty, which shall be determined in accordance with the approaches described in ISO 21748. Numerical values shall also be converted to one of three categorical values: "meets the DS", "exceeds the DS", or "indeterminant", which considers the ranges of values with a 95 % confidence interval of the measurement (see 5.1). This conversion permits a direct comparison between the CMD and the reference method.

4.3 Challenge water

Challenge water — as defined in this document — is similar to the "challenge water" used for TA testing.[6][7] Challenge water for CMD testing is only used in the subset of laboratory trials designed to evaluate the CMD performance under standardized and simplified conditions. The salinity of challenge water is either fresh (<1 g kg^{-1}),[1)] brackish (10 g kg^{-1} to 20 g kg^{-1}), or marine (28 g kg^{-1} to 36 g kg^{-1}), and it may be prepared from natural waters or with purified water amended with sea salts according to ASTM D1141-98.

Challenge water consists of abiotic characteristics, defined as minimum concentrations of dissolved organic matter (DOM), particulate organic matter (POM), and mineral matter (MM) and total suspended solids (TSS), as defined in TA test protocols.[6][7]

NOTE For other trials, including laboratory tests using ambient water, the salinity, DOM, POM, MM, and TSS are measured but not manipulated.

For the purpose evaluating the performance of CMD, abiotic challenge water components are as defined in TA test protocols,[6][7] but requirements for organism concentrations and diversity are specific to this test protocol and are defined in 4.4 and 6.2.

4.4 Test concentrations of organisms

The concentrations of living or viable organisms are defined relative to the DS. Concentrations below, approximately equal to, and above the DS are most relevant to this performance evaluation, and the target ranges ensure that samples meeting and exceeding the DS are included in the evaluation.

Target ranges are at a minimum defined below:

— below: >0 % and <50 % the DS;

— approximately equal to: ±50 % the DS;

— above: >150 % and <1 000 % the DS.

Concentrations below the DS shall be non-zero and measurable by the reference method (i.e. > limit of quantification).

NOTE 1 At concentrations 50 % of the DS (e.g. ≤4 organisms ml^{-1}), the probability that a random sample will have ≤9 organisms ml^{-1} is > 99 %. Likewise, at concentrations >150 % of DS, e.g. ≥16 organisms ml^{-1}, the probability that a random sample will have ≥ 11 organisms ml^{-1} is >95 %.

NOTE 2 Additional concentrations can be added, provided that both concentrations above and below the DS are included in the evaluation.

1) Salinity is reported here as g kg^{-1}, which for the purposes of this test protocol is approximately equivalent to Practical Salinity Units (PSU)[5].

5 Evaluation metrics

5.1 General

5.1.1 Overview

The evaluation considers three metrics: trueness, precision, and reliability, which are described in 5.2 to 5.4.

NOTE This document uses the terms "trueness" and "precision" to describe the accuracy of a measurement method, following the terminology in ISO 5725-1. This terminology differs from other common definitions, which define "accuracy" as the agreement between a measurement and the true quantity e.g. as shown in Reference [9].

5.1.2 Measurement protocols

The evaluation requires measurements from both the reference method and the CMD. Measurements shall be performed following standard, pre-established protocols. The pre-established protocols shall have guidance on whether the measurements meet data quality objectives and are acceptable.

EXAMPLE A data quality indicator can exceed the data quality objective, in which case, the protocol can require that the measurement is rejected and a new sample is analysed. It can be appropriate to reject a measurement, given measurements are not rejected without cause. It is expected that rejected measurements are recorded and reported.

The protocol for the CMD or the reference method can require multiple readings or repeated subsampling and analysis. In this case, the protocols shall indicate the process for reporting the mean and dispersion (e.g. range, standard deviation) of the set of numerical measurements, or for determining the overall sample disposition (e.g. meets or exceeds the DS).

5.1.3 Categorical outcomes

Readings from both the CMD and reference method are simplified into three categories:

— "meets the DS";

— "exceeds the DS";

— "indeterminate".

For the CMD, the categorical outcome is based upon guidance from the manufacturer. This guidance from manufacturers can include instructions for binning numerical measurements into one of the categories; the CMD may automatically classify the sample disposition and display terminology relating to one of the three categories.

EXAMPLE A CMD can report the outcome in terms of "risk", such as "low risk" or "high risk". These terms are comparable to "meets the DS" or "exceeds the DS", respectively.

For the reference method, the categories are assigned based upon the measured concentration and the uncertainty surrounding the estimate, considering, for example, the range of potential values within a 95 % confidence interval of the estimate. Estimates with confidence intervals spanning the DS are considered "indeterminate".

NOTE The reference method selected is well characterized, such that independent testing organizations have a historical record of usage and performance monitoring. These data support the estimates of measurement uncertainty and confidence intervals.

Though the evaluation metrics are based on categorical variables, it is important that all numerical measurements are collected and reported. For the CMD, all reported numerical information is recorded for each sample. The numerical results are important, as end-users may perform additional statistical tests using these measurements. For the reference method, all values used to calculate concentrations

(including sample volumes, reagent volumes, etc.) are recorded with the measurement. Additionally, the method detection limit and confidence intervals shall be reported for each measurement.

5.2 Trueness

5.2.1 Overview

Trueness is the agreement between measurements of the CMD and an accepted value from a reference standard or a reference method, which is a substitute for the "true" value. The uncertainty of measurements of the reference methods shall be determined in accordance with ISO 21748. Trueness is determined in:

— controlled laboratory tests with prepared challenge water (6.2);

— controlled laboratory tests with ambient water (6.3);

— field tests with water treated with a type-approved ballast water management system (BWMS) (6.4).

For laboratory tests, organism concentrations are manipulated to yield three samples for each trial (at each salinity), with concentrations below, approximately equal to, and above the DS, as defined in 4.4. For field tests, samples of ballast water shall be collected during deballasting — after BWMS treatment and neutralization processes (if performed) — in accordance with the requirements of ISO 11711-1 and ISO 11711-2. Although organism concentrations are expected to be below the DS, concentrations are not manipulated.

5.2.2 Measurement approach

CMDs can provide results as binary outcomes (e.g. "pass/fail"), numerical values, or both. All numerical data available shall be recorded, but the analysis of trueness is based on categorical outcomes, where a sample:

— exceeds the DS ("fail"): the sample exceeds the DS, has a high risk of exceeding the DS, or reports concentrations ≥10 ml^{-1} or ≥10 m^{-3}, for example;

— meets the DS ("pass"): the sample meets the DS, has low risks of exceeding the DS, or reports concentrations <10 ml^{-1} or <10 m^{-3}, for example;

— is indeterminate: the sample may exceed or meet the DS, but the determination cannot be made with confidence, for example, the measurement is not statistically different from the DS.

Statistical analysis is performed by placing all the test outcomes into an error matrix. Figure 1 shows a generic error matrix used to display measurement agreement between the reference method and the CMD. Based upon the analysis by both the reference method and the CMD, samples are assigned into one of nine test outcomes. Values in boxes show the tally of samples for each outcome, and the rows (R), columns (C), and the entire table (n) are used to calculate Cohen's kappa (κ) and other metrics, as described below (5.1.3).

Reference method	CMD: Exceeds DS	CMD: Meets DS	CMD: Indeterminate	
Exceeds DS	16	3	0	ΣR_1 19 (0,42)
Meets DS	4	12	1	ΣR_2 17 (0,38)
Indeterminate	0	3	6	ΣR_3 9 (0,20)
	ΣC_1 20 (0,44)	ΣC_2 18 (0,40)	ΣC_3 7 (0,16)	Total (n) 45 (1,00)

Key

DS discharge standard

$\sum R_i$ sum of row i, where i = 1, 2, or 3 (boxes with light grey shading)

$\sum C_i$ sum of column i , where i = 1, 2, or 3 (boxes with light grey shading)

n total number of samples (box with dark grey shading)

NOTE 1 Data are hypothetical and used only as an example for calculations. See Formulae (1) to (8).

NOTE 2 Samples in agreement are in boxes along the diagonal, marked with thick borders.

Figure 1 — Generic error matrix with example data

5.2.3 Statistical calculations

Cohen's kappa (κ) is commonly used to determine measurement agreement, as it adjusts the overall observed agreement (P_O) to the expected agreement (P_E), which is the agreement due to chance.[10] Formulae (1) to (4) are used to calculate P_O, P_E, κ, the standard error of κ (σ_κ), respectively:

$$P_O = \frac{\sum A}{n} = \frac{(16+12+6)}{45} = \frac{34}{45} = 0{,}76 \tag{1}$$

$$P_E = \sum_{i=1}^{3} \frac{\left(\frac{\Sigma R_i \cdot \Sigma C_i}{n}\right)}{n} = \frac{\left(\frac{19 \cdot 20}{45}\right) + \left(\frac{17 \cdot 18}{45}\right) + \left(\frac{9 \cdot 7}{45}\right)}{45} = \frac{8{,}44+6{,}80+1{,}40}{45} = \frac{16{,}64}{45} = 0{,}37 \tag{2}$$

$$\kappa = \frac{P_O - P_E}{1-P_E} = \frac{0{,}76-0{,}37}{1-0{,}37} = \frac{0{,}39}{0{,}63} = 0{,}62 \tag{3}$$

$$\sigma_\kappa = \sqrt{\frac{P_O(1-P_O)}{n(1-P_E)^2}} = \sqrt{\frac{0{,}76(1-0{,}76)}{45(1-0{,}37)^2}} = \sqrt{\frac{0{,}18}{17{,}9}} = 0{,}10 \tag{4}$$

where

ΣA is the sum of samples where the reference method and the CMD are in agreement;

ΣR_i is the sum of samples in row i (where i = 1, 2, or 3);

ΣC_i is the sum of samples in column i;

n is the total number of samples.

The confidence interval (I) of κ is based upon the chosen significance level (α). Formula (5) is used to calculate I based on κ, σ_κ, and z, which is the quantile of a normal distribution corresponding to a chosen α.

$$I = \kappa \pm z \cdot \sigma_\kappa \tag{5}$$

Typical values of α (α = 0,1, 0,05, or 0,01) correspond to confidence intervals of 90 %, 95 %, or 99 % and corresponding z values of 1,28, 1,65, or 2,33, respectively. For these z values, confidence intervals (I_α) are calculated using Formulae (6) to (8), respectively.

$$I_{0,1} = \kappa \pm 1{,}28 \cdot \sigma_\kappa \tag{6}$$

$$I_{0,05} = \kappa \pm 1{,}65 \cdot \sigma_\kappa \tag{7}$$

$$I_{0,01} = \kappa \pm 2{,}33 \cdot \sigma_\kappa \tag{8}$$

NOTE The z values correspond to a one-tailed distribution, where z is determined from 1 – α. This is appropriate for testing the hypothesis that the observed data exceed a defined threshold. For a two-tailed distribution, z scores are determined at 1 - $\alpha/2$. For typical values of α (α = 0,1, 0,05, or 0,01), $\alpha/2$ equals 0,05, 0,025, or 0,005, and corresponding z values are 1,65, 1,96, or 2,58, respectively.

Values of $\kappa \geq 0{,}8$ indicate strong agreement between the two methods, but lower thresholds for κ (e.g. $\kappa \geq 0{,}6$) can be acceptable in some cases.[11,12] A set threshold allows a hypothesis test to determine whether the observed κ (κ_{obs}) is significantly greater than the threshold κ (κ_{thr}). The number of samples required to determine whether κ_{obs} exceeds κ_{thr} depend upon the expected difference between these two values, where large sample sizes are required to detect small differences between κ_{obs} and κ_{thr}. An approach for determining samples sizes is described in 7.4.

5.3 Precision

5.3.1 Overview

Precision is determined from multiple readings of a single sample. At a minimum, precision is measured using a simplified sample, such as the samples of challenge water (see 6.2).

NOTE 1 It is recognized that the end user can require samples of natural water or with a complex matrix, such as ambient water samples or treated ballast water from a ship or from type approval testing. These samples can be acceptable for precision, given concentrations remain stable throughout the analysis.

For each organism or organism size class, each trial consists of one sample below the DS and one sample above the DS, according to 4.4.

NOTE 2 Samples with concentrations approximately equal to the DS are not used to measure precision.

5.3.2 Measurement approach

Samples for precision shall be prepared with sufficient volume to allow ≥10 subsamples for analysis with the CMD and ≥2 subsamples for analysis using the reference method. Precision of the CMD is the agreement among ≥10 subsamples from a single sample.

NOTE 1 As all numerical data reported by the CMD are recorded (5.1.2), end-users can additionally calculate precision using numerical values, with the understanding that numerical comparisons of precision between the CMD and the reference method are not expected to be comparable.

Analyses using the reference method shall be performed on ≥2 subsamples at the start and the end of the set of CMD analyses to verify that:

— concentrations are within the target range (whether below or above the DS);

— concentrations remain stable throughout the trial.

A sample shall be well-mixed prior to collecting each subsample, which are analysed in succession to minimize changes in the organism concentrations over time. Subsamples for the CMD and separate subsamples for the reference method are collected synchronously, but analysis using the reference method is only required for the first and last subsample of the CMD analysis.

NOTE 2 Though not required, more frequent sampling and analysis using the reference method can be useful, for example, if the end-user requires a calculation of precision based upon numerical values. In this case, an equivalent number of subsamples and analyses (e.g. ≥10 for both the CMD and the reference method) would provide a direct comparison between the CMD and the reference method.

Organism loss or mortality can cause instability and drive concentrations below the DS (for ≥150 % DS) or below the limit of detection of the reference method (for the ≤50 % DS) samples. If the status of the sample changes, the trial shall be invalidated. If changes in organism concentrations invalidate a trial, a reduced number of subsamples shall be analysed for each sample, given the number of samples increases so that at ≥10 subsamples are collected and analysed for both concentration ranges.

5.4 Reliability

Reliability is the ability to maintain integrity or stability of the CMD and data collection over time. The reliability of a CMD includes results of periodic calibration checks and analysis of a reference standard (if available), as required by the CMD manufacturer. The CMD shall also analyse both negative and positive control samples prior to and after each daily set of analyses.

Negative, positive, or calibration controls can be defined or addressed by the CMD manufacturer. If not specified, standardized approaches for obtaining negative, positive, and calibration controls may be defined by the test organization.

EXAMPLE 1 Negative controls are free of organisms. Purified water, artificial lake or seawater prepared without organisms, or filter-sterilized ambient water can all be used as negative controls. Positive controls contain target analytes or concentrations of organisms that are well above (>150 %) the DS and known to trigger a response from the CMD.

Reliability is assessed over the course of the performance tests, including both laboratory and field tests. Throughout the evaluation, described in 6.2 to 6.4, CMD use and operations — including preventative maintenance, cleaning and calibration — are performed according to manufacturer instructions and shall be documented. These observations, as well as results of daily analyses of negative and positive control samples, shall be logged.

Records shall include basic information, such as time and location, and qualitative descriptions to provide context to the events. Descriptions shall be recorded at the time of occurrence as narrative text that describes (at a minimum):

— instrument or assay failures;

— cleaning, battery changes, or other routine maintenance;

— challenges in the performance of the assay (including transport and set up of equipment);

— unscheduled maintenance, repair, or factory calibration required;

— problems occurring during calibration, analysis, or data reporting.

This log tracks the total number of events and allows calculations such as frequency of events based upon the number of uses or the time of use (e.g. mean time between non-scheduled maintenance events). Reliability is also determined by the rate of data recovery, which is the number of measurements produced relative to measurements expected (e.g. 95 of 100 measurements were obtained).

EXAMPLE 2 Examples of missing data can originate from broken or non-readable display screens or errors with the firmware or factory calibration.

CMD test personnel shall follow recommended troubleshooting and repair procedures. For any issues not easily resolved (e.g. issues that would require factory repair, unit replacement, or firmware modifications), the test personnel shall consider that the remaining samples for a trial in progress are missing. Future trials shall only be planned and executed when a functioning unit is available.

6 Evaluation test types

6.1 Overview

The minimal performance evaluation shall consist of three types of tests:

— laboratory tests using prepared challenge water with defined concentrations of cultured organisms and abiotic constituents (see 6.2);

— laboratory tests using ambient water with defined concentrations of ambient organisms, but without requirements for abiotic constituents (see 6.3); and

— field tests using treated ballast water without further manipulation (see 6.4).

Tests shall use sample water with temperatures between 4 °C and 35 °C or within a narrower temperature range, as specified by the CMD manufacturer.

Tests range in complexity to evaluate the performance of the CMD, from simple, well-characterized samples to ambient waters with mixed assemblages of organisms, and finally to treated ballast water from a ship. Nevertheless, it is acknowledged that instead of conducting laboratory tests with cultured organisms, the end user can require this laboratory testing to occur with ambient organisms that may be amended with cultured organisms to achieve the target concentrations. Additionally, it is recognized that the end user may also require laboratory tests described in 6.2 and 6.3, using water that is treated by a BWMS or has undergone a simulated ballast water treatment, in addition to being conducted using un-treated water as described in 6.2 and 6.3. Importantly, the concentrations of organisms in samples shall meet the requirements defined in 4.4, and the effects of treatment (such as residual oxidants or heat) shall not affect the survivability of organisms relative to untreated water. 6.2 to 6.4 describe the objectives and characteristics of each test type.

6.2 Laboratory tests with prepared challenge water

6.2.1 Challenge water

Laboratory tests with prepared challenge water examine the performance of the CMD using simplified water samples, with defined and controlled concentrations of abiotic and biotic constituents. Challenge water shall be prepared with ambient water or reagent-grade, purified water with ASTM-grade sea salts in accordance with ASTM D1141-98 or equivalent. Water shall be amended with DOM, POM, and MM to meet the requirements defined in a chosen BWMS test protocol. See References [6] or [7] for examples.

It is recognized that the end user may require laboratory tests conducted with water that is treated by a BWMS or has undergone a simulated ballast water treatment. If so, tests shall be conducted as described below. In this situation, water prepared for TA testing can be suitable for CMD testing when it meets the criteria defined in 4.4.

6.2.2 Cultured organisms

Challenge water shall be amended with cultured organisms, which are selected to meet the detection criteria for the CMD. For CMD that target organisms in a broad size class, the challenge water shall be amended with at least two dissimilar species (e.g. from two phyla) that fit within the size class.

NOTE As mentioned in 6.2.1, water prepared for TA testing that meets these requirements can be used.

The measurement approach used by the CMD guides the choices of which cultured organisms are used for laboratory tests with challenge water. For laboratory tests with prepared challenge water, the CMD shall be capable of detecting the chosen organisms. For example, if a CMD measures active fluorescence, then the cultured organisms shall have chlorophyll *a*. Likewise, for a CMD that detects motility, cultured organisms shall be motile.

The organisms selected shall be suited for the salinity and temperature of the test water. Table A.1 provides some potential organisms suitable for fresh, brackish, and marine waters, but other organisms may be used. Species identification and general characteristics (e.g. body dimensions, motility) of organisms used for testing shall be reported. For consistency among trials, cultures of organisms used shall be actively growing (e.g. phytoplankton cultures in exponential growth phase). All cultures shall contribute approximately equally to the overall concentration, so a single organism does not numerically dominate the sample.

EXAMPLE For two cultures, each comprises 40 % to 60 % of the total concentration of the mixture; for three cultures, each comprises 25 to 40 % of the total concentration of organisms. For CMD that target indicator microbes, water is amended with the specified microbe only.

If the CMD is designed to detect toxicogenic strains of *Vibrio cholerae* (O1, O139), laboratory trials with prepared challenge water shall include living or viable cultures of toxicogenic *V. cholerae* (O1, O139), as these pathogens may be rarely found in ambient waters or during shipboard testing. The laboratory and personnel performing this work shall be qualified to maintain, handle and dispose of toxicogenic *V. cholerae* safely.

6.2.3 Sample volumes and organism concentrations

Samples shall be prepared with sufficient volumes for the CMD, the reference method and ancillary analyses (7.7). This includes water required for sample processing (such as size-fractionation), replicate analysis, or quality control samples required in the CMD or reference method protocols. Typically, cultured organisms shall be maintained at concentrations that are several orders of magnitude more concentrated than the target concentrations. Stock concentrations are estimated using the reference method or another appropriate method, and these estimates are used to calculate the volume of the stock culture needed to meet the target concentrations and the required sample volume.

NOTE Stock organisms added to challenge water can dilute concentrations of DOM, MM, etc., so it can be necessary to amend large volumes of stock culture with abiotic constituents to meet requirements.

6.2.4 Sample handling and analysis

Samples with living organisms are dynamic and often sensitive, so procedures for handling, including processing, mixing, and sampling, shall not substantially change organism concentrations, viability status, or activity (e.g. motility). Samples, once prepared, shall be maintained within an acceptable temperature range and analysed soon after preparation.

Generally, prepared samples should be analysed within several hours (e.g. within 6 hours). However, it is recommended that independent testing organizations perform empirical tests to validate the maximum time that a prepared sample may be used.

Changes in expected concentrations of organisms (as measured by the reference method) indicate problems with handling and holding procedures.

NOTE A sample that no longer meets its intended target concentrations has undergone a substantial change in concentrations. This can occur rapidly, depending on the sample characteristics, so the sample hold time is adjusted according to observed changes in organism concentrations.

Analyses are performed following the protocols for CMD and the reference method. All results are logged in the test records, including numerical values of cell concentrations and other relevant parameters.

6.3 Laboratory tests using natural water with ambient organisms

6.3.1 Natural water

Laboratory tests with ambient water examine the performance of the CMD using natural water samples, with mixed assemblages of ambient organisms as well as natural abiotic constituents. Natural waters used for tests shall be within temperature and salinity ranges appropriate for the CMD. In general, natural waters are surface waters representative of those used for navigation by ships. Natural water shall be collected using common equipment and approaches, including discrete, manually collected surface water or pumped water from a piping system, granted the approach does not damage ambient organisms. Abiotic characteristics of the sample water (e.g. temperature, salinity, DOM, POM) shall be measured but are not manipulated (see 7.6).

6.3.2 Ambient organisms

Waters selected for testing shall include natural assemblages of organisms, including organisms outside the targeted size class or type of indicator microbe. Ambient water samples shall be diluted with filter-sterilized water from the sample location to achieve sample concentrations less than, approximately equal to, and greater than the DS, as defined in 4.4. For CMD that target organisms in a broad size class, concentrations of ambient organisms shall not be supplemented with cultured organisms. However, if ambient concentrations are insufficient for testing, ambient organisms shall be gently concentrated (e.g. using mesh netting) to achieve required concentrations.

For CMD that target indicator microbes, it is necessary to add the relevant bacteria from cultured stocks at concentrations less than, approximately equal to, and greater than the DS, as defined in 4.4.

For CMD that target organisms in broad size classes, a census of ambient organisms is required. Organisms shall be identified to the lowest possible taxonomic level possible, and both the overall concentration of ambient organisms and concentrations of major populations shall be reported. Recognizing that CMD can have taxa-specific biases, the purpose of this taxonomic characterization is to provide context to measurements of the CMD. For example, if the device is based upon the variable fluorescence of chlorophyll *a*, the relative concentration of photosynthetic organisms provides insight into the performance of the device and explains differences in the concentration measurements from the CMD and the reference method. Test personnel shall also note whether the ambient community includes colony-forming organisms, such as chain-forming diatoms or filamentous cyanobacteria. The method(s) used to determine the abundance and taxonomic composition of organisms can be manual microscopy or another widely used, acceptable approach that is independent of the measurement of the CMD.

6.3.3 Sample volume

As stated in 6.2.2, samples collected shall have sufficient volumes for the CMD, the reference method, and ancillary analyses (see 7.6), including the census of ambient organisms described in 6.3.2.

6.3.4 Sample handling and analysis

As specified in 6.2.4, samples shall be handled and processed to minimize substantial changes in organism concentrations. Within the sample, ambient organisms continue natural interactions, including growth and mortality due to predation. Consequently, it is particularly important to minimize

the time interval between sampling preparation and analysis, such that concentrations of organisms remain in the target ranges.

Analyses shall be performed following the protocols for CMD and the reference method. All results shall be logged in the test records, including numerical values of cell concentrations and other relevant parameters.

6.4 Field tests using treated water

6.4.1 Treated water

Field tests demonstrate the performance of the CMD under conditions of intended use (commonly shipboard, but possibly shoreside or other locations) by examining ballast water treated with a type-approved BWMS. All sample collection shall be performed in accordance with ISO 11711-2, which defines requirements for sample equipment and sampling procedures for ballast water samples. Sampling shall only occur at locations where ballast water is fully treated and neutralized (if applicable).

At a minimum, field tests shall examine three unique and independent deballasting events to provide a minimum of experimental replication, demonstrating the magnitude of variation in the performance of the CMD. Deballasting events may be from three separate ships or from a single ship, given the ballast water is collected and treated at separate times. Deballasting may occur at the port or while the ship is underway.

6.4.2 Organisms present post treatment

Treated ballast water may contain few living or viable organisms, given the BWMS operated as designed. Organism concentrations are not manipulated to meet set ranges, and it is possible that all trial and all samples meet the DS according to both the CMD and the reference method. In this scenario, the test approach can still capture false positives and is still appropriate to measure agreement and reliability.

6.4.3 Sample volume

As stated in 6.2.3, samples collected have sufficient volumes to supply the water as required by the CMD, the reference method, and ancillary analyses (see 7.7). If ambient organisms are present, such that the reference method measures concentrations which are above the DS, a census of organisms remaining in ballast shall be conducted as described in 6.3.2.

6.4.4 Sample handling and analysis

As specified in 6.2.4, samples shall be handled and processed to minimize substantial changes in organism concentrations. For field samples, residuals of treatment, such as biocidal compounds, heat, and disinfection byproducts, can continue to affect the organisms present after the treatment. Minimizing the times when samples are held can mitigate these effects. Neutralizing active compounds also mitigates the effect of residuals of treatment on organisms remaining after the treatment.

Analyses shall be performed following the protocols for CMD and the reference method, as described in 6.2.4.

6.4.5 Test information and descriptions

Given the variety of BWMS approaches and the potential range of characteristics of natural waters used as ballast water, detailed sample and test descriptions are critical. Water characteristics importantly verify that the conditions are within the temperature and salinity ranges of the CMD, and that the treatment type of the BWMS is compatible with the CMD (see 7.2). The sample characteristics and

treatment descriptions also provide necessary context for findings. For all field tests, test personnel shall describe, at a minimum:

— Sample source information: date, time, and location of ballasting, ballast volume, tank location, BWMS description and operational details (treatment type, filtration details, treatment rate, residual biocides).

— Sampling information: collection method (e.g. installed sample port), sampling start and end times, sample flow rate and total volume, and sample chain-of-custody.

— Analysis information: Location(s) of analysis, start and end times, reagents used, calibration procedures, identification and role of personnel involved in the operation.

NOTE ISO 11711-2:2022, Annex B shows a shipboard sample collection worksheet, providing a template for the collection of relevant information.

7 Experimental design

7.1 General

Tests evaluate performance claims of the CMD, primarily the group(s) of organisms specified by DS. Performance claims also state whether the CMD operates in all water salinities or whether there are limitations in use, based upon salinity or other sample characteristics. Table 1 provides an overview of steps used to design experiments to evaluate a CMD, where common options and considerations are listed for most steps. These topics are described in 7.2 to 7.6.

Table 1 — Steps and considerations for designing experiments

Step 1. Identify the targeted DS group(s)
— organisms ≥50 µm in minimum dimension — organisms ≥10 and < 50 µm in minimum dimension — toxicogenic *Vibrio cholerae* (01 and 0139) — *Escherichia coli* — Intestinal enterococci
Step 2. Identify salinity ranges
— freshwater — brackish water — marine water
Step 3. Identify other characteristics and limitations
— portability, e.g. hand-held, benchtop, or in-line operation — water temperature limitations, e.g. sample water shall be be <35 °C — water characteristics limitations, e.g. turbidity, pH, etc. — organism characteristics limitations — treatment type limitations
Step 4. Determine the minimum set of tests

Table 1 *(continued)*

— laboratory tests using prepared challenge water — laboratory tests using natural water with ambient organisms — field tests using treated water
Step 5. Determine the minimum set of tests
— define the κ_{thr} value as described in 5.2.3 and 7.5 — define the significance level (α) — calculate n for laboratory tests
Step 6. Incorporate additional factors as required (optional)

7.2 CMD characteristics

The general characteristics of the CMD determine the set of tests required. For example, organizations designing the test shall first consider the DS group (or groups) targeted by the CMD and whether the CMD measures living or viable organisms, as well as the numerical limit(s) specified for the targeted DS group(s). If the CMD does not claim any limitations on water salinity, then the set of tests shall include trials using freshwater, brackish water, and marine water.

NOTE 1 When all three salinities are examined, laboratory tests with prepared challenge water are only conducted in freshwater and either brackish or marine water, not both.

Other factors of CMD design, such as portability and measurement type, determine, respectively, the test location and characteristics of the cultured organisms selected for laboratory tests with challenge water. Portable CMD are designed for discrete samples. As they are portable, these CMD can operate in a variety of locations, facilitating laboratory-based tests. For CMD intended for permanent installation aboard a ship, field tests ideally shall occur on a ship. A permanently installed, inline CMD operates in a unique and challenging environment, so testing in its operational environment is crucial. An inline CMD samples, processes, and analyses water within a piping system. As an inline CMD can operate continuously as water moves through the piping system, in-line CMD measurements shall be closely paired, in physical location and analysis time, with water samples collected for analysis via the reference method.

NOTE 2 While inline CMD are ideally tested on a ship as installed, tests located off ships are possible, given a customized test platform that can mimic the piping and flow characteristics specified for CMD.

7.3 Known CMD limitations

CMD manufacturers may specify known limitations on the performance of the CMD, such as those listed in Table 2. For laboratory tests with prepared challenge water, cultured organisms shall be chosen with considerations of the limitations of the CMD, such that selected organisms shall be detectable. For samples of natural waters and treated ballast water, any characteristics identified as performance limitations shall be measured and reported.

NOTE Additional factors, including water, organism, and treatment characteristics can be considered in additional tests (see 7.7).

Table 2 — An example of CMD performance limitations

Performance limitations
— Measurement uncertainty, as reported by the CMD manufacturer, which may be a percentage of the measured value (e.g. ± 10 %) or an absolute value (e.g. ±2 organisms/ml).
Limitations based upon characteristics of sample water
— Water clarity (e.g. UV Transmittance, turbidity) — Other properties (e.g. pH, dissolved matter)
Limitations based upon characteristics of organisms
— Detects organisms with chlorophyll *a* only — Detects motile organisms only — Quantifies free-living, non-colonial organisms only — Other characteristics (e.g. requires culture-amenable organism, compatibility with specific probes, labels, or dyes)
Limitations based upon treatment type

— Compatibility (or incompatibility with treatment types)
  — oxidant-based treatment
  — ultraviolet light-based treatment
  — filtration treatment
  — other limitations based on treatments type

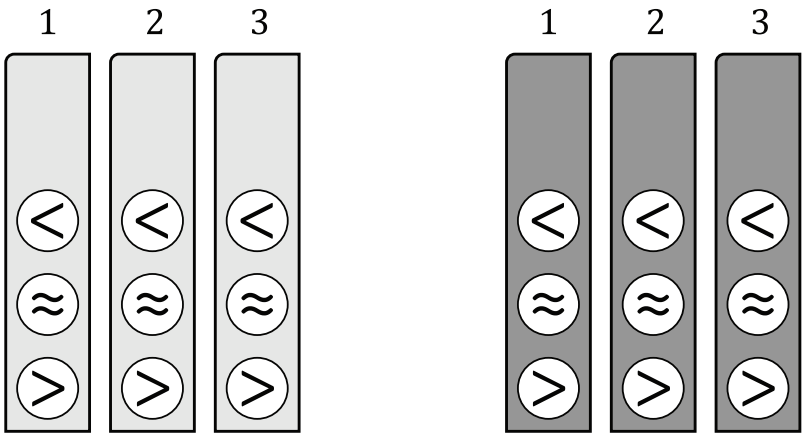
7.4 Basic evaluation requirements

Given the characteristics and limitations of a CMD, a minimum set of trials is determined as shown in Figures 2 to 4. Each size group defined in the DS requires a unique set of samples intended for the targeted size class or indicator microbe, and each requires an appropriate reference method. Therefore, if a CMD targets more than one group defined in the DS, each group is evaluated separately.

NOTE Multiple groups of the DS can be evaluated simultaneously, such that a sampling event for natural waters or treated ballast water can supply both sets of samples. In this case, the two sets of samples can share common sampling descriptions and ancillary analysis.

For trueness, a single trial shall include three samples with concentrations below, approximately equal to, and above the DS (as described in 4.4). For each salinity range, a unique set of trials for laboratory tests with prepared challenge water and laboratory trials with natural waters are required.

The simplest scenario considers a CMD designed for one organism group at one salinity. Three independent, replicate trials are performed using prepared challenge water with cultured organisms. An additional set of three independent, replicate trials are performed with ambient organisms from natural waters within the specified salinity range. Each trial has at least three samples, with organism concentrations below, approximately equal to the DS. A schematic is shown in Figure 2.



a) Challenge water with cultured organisms **b) Natural water with ambient organisms**

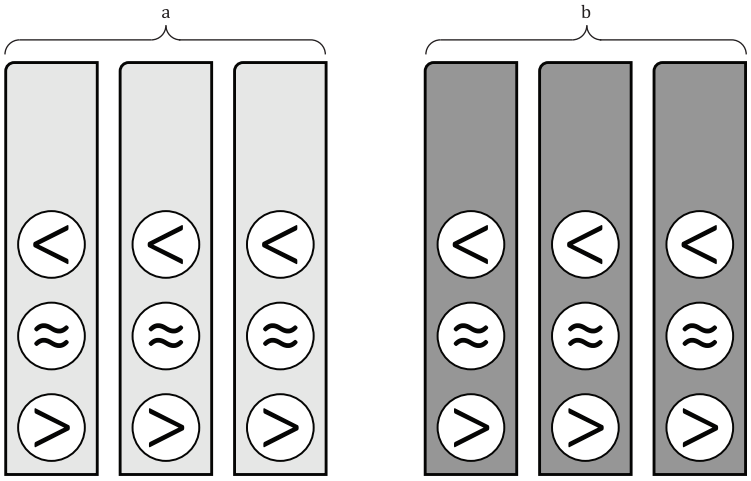
Key

1	independent trial 1
2	independent trial 2
3	independent trial 3
	light grey are trials with challenge water with cultured organisms
	dark grey are trials with natural water and ambient organisms
circles	samples for evaluation, prepared or manipulated to have organism concentrations:
<	less than the DS
≈	approximately equal to the DS
>	greater than the DS

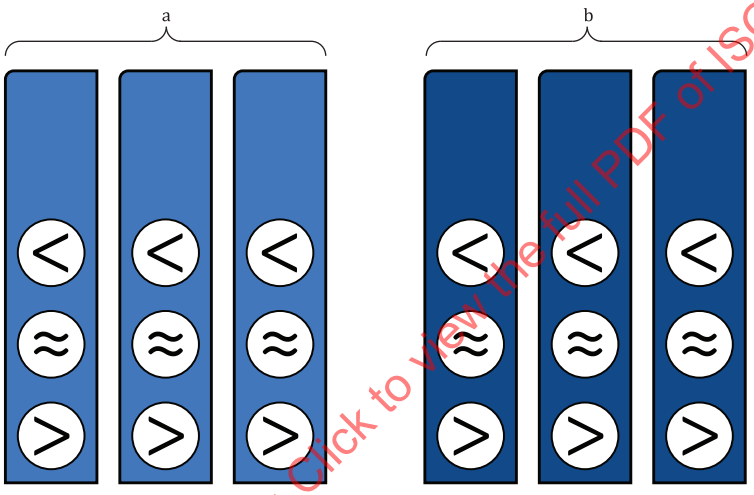
NOTE The schematic shows the minimum set of samples and trial to determine trueness for a CMD designed for a single water salinity.

Figure 2 — Schematic of trials and samples to determine trueness at a single salinity

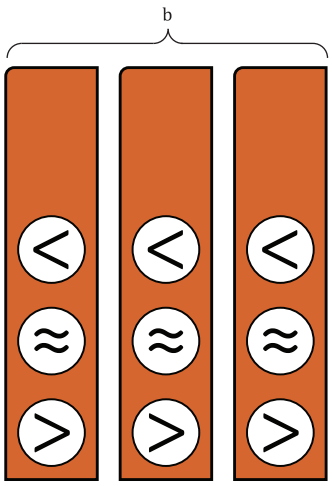
The next scenario considers a CMD designed for one organism group but at all three salinities. For this case, three independent, replicate trials are performed as shown in Figure 2, where the colour-coded boxes show independent trials unique to the water salinity and targeted DS group. If a CMD operates in all three salinity ranges, laboratory trials with prepared challenge water require only freshwater trials and brackish or marine water trials, not both. A schematic for this scenario is shown in Figure 3.



a) Laboratory trials for freshwater



b) Laboratory trials for brackish water



c) Laboratory trials for marine water

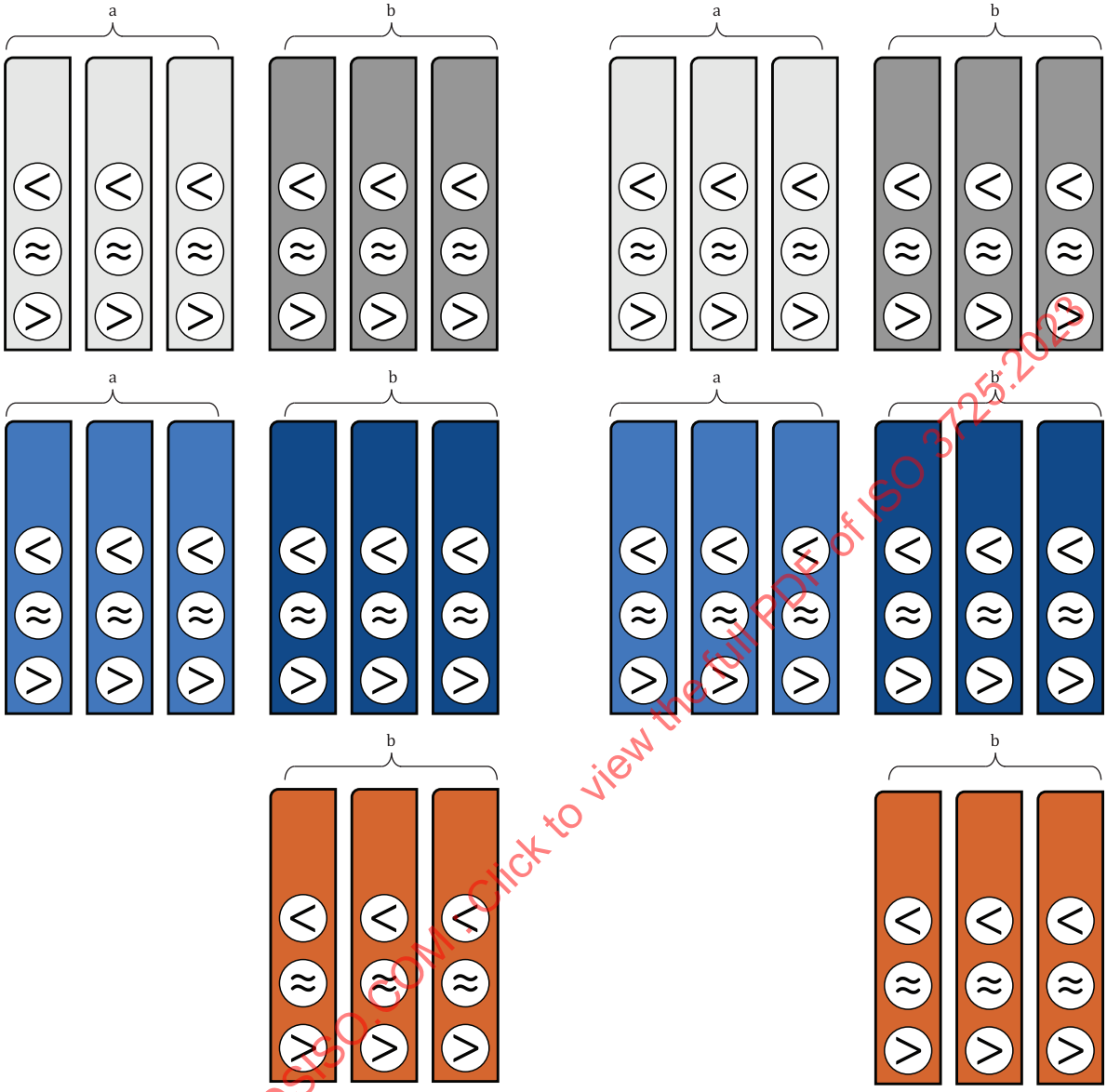
Key

[a]	Trials with challenge water and cultured organisms.
[b]	Trials with natural water and ambient organisms.
circles	samples for evaluation, prepared or manipulated to have organism concentrations:
<	less than the DS
≈	approximately equal to the DS
>	greater than the DS

NOTE The schematic shows the minimum set of samples and trial to determine trueness for a CMD designed for all three water salinities. When all three salinities are tested, trial of challenge water with cultured organisms are only required for brackish water or marine water (not both). In this schematic, trials with challenge water and cultured organisms are not performed.

Figure 3 — Schematic of trials and samples to determine trueness at three salinities

The final scenario considers a CMD designed for two organism groups, each at three salinities. Separate analyses are required for both organism groups, although a single prepared or collected sample may provide subsamples for both analyses, so that ≥ 50 µm and ≥ 10 µm and < 50 µm trials can occur in parallel. A schematic displaying trials and samples needed for this scenario is shown in Figure 4.



a) Laboratory trials for organisms ≥10 µm and <50 µm

b) Laboratory trials for organisms ≥50 µm

Key

a	Trials with challenge water and cultured organisms.
b	Trials with natural water and ambient organisms.
circles	samples for evaluation, prepared or manipulated to have organism concentrations:
<	less than the DS
≈	approximately equal to the DS
>	greater than the DS

NOTE The schematic shows the minimum set of samples and trial to determine trueness for a CMD designed for all three water salinities and for two groups of organisms defined in the DS: organisms ≥ 10 and < 50 µm and organisms ≥ 50 µm.

Figure 4 — Schematic of trials and samples to determine trueness for two size classes, each at three salinities