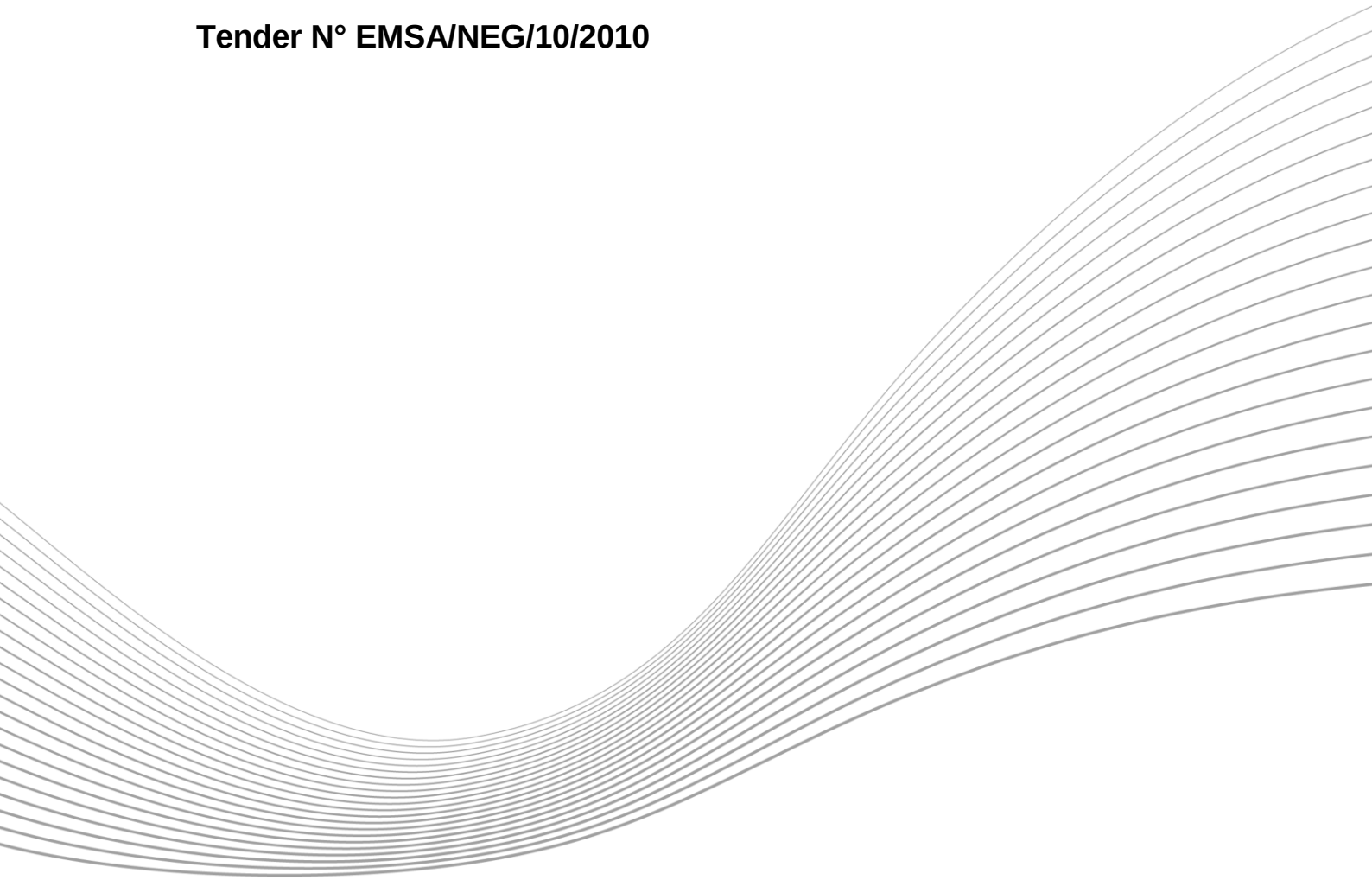
Final report to:
The European Maritime Safety Agency (EMSA)

Development of guidance on how to analyze a ballast water sample

Tender N° EMSA/NEG/10/2010



WATER • ENVIRONMENT • HEALTH

December 2010

Development of guidance on how to analyze a ballast water sample

Tender N° EMSA/NEG/10/2010

Agern Allé 5
DK-2970 Hørsholm, Denmark

Tel: +45 4516 9200
Fax: +45 4516 9292

e-mail: clj@dhigroup.com
Web: www.dhigroup.com

Client	Client's representative:
The European Maritime Safety Agency (EMSA)	Brian Elliott

Project	Project No
Development of guidance on how to analyze a ballast water sample	11808213
Authors Claus Jørgensen Kim Gustavson Jette Bjerre Hansen Thomas Hies	Date 2010.12.06 Approved by Torben Madsen

Revision	Description	By	Checked	Approved	Date
00	Final report				2010.12.06
00	Final draft report				2010.11.24
00	First draft report				2010.10.28
00	Interim report				2010.09.28

Key words	Classification
Ballast water Compliance monitoring Compliance assessment Analytical methods	☐ Open ☐ Internal ☒ Proprietary

Distribution	No of copies
EMSA	2 + pdf
DHI: CLJ/KIG/JBH/THI/ERA	1 + pdf

CONTENTS

APPENDICES

1 BACKGROUND

1.1 The IMO Ballast Water Convention

The introduction of invasive marine species into new environments by ballast water has been identified as one of the greatest threats to the sea ecosystems worldwide. The problem of invasive species is largely due to the expanded trade and traffic volume over the last few decades. It is estimated that ships transfer approximately 3 to 5 billion tonnes of ballast water internationally each year /30/, potentially transferring species of sea life that may prove ecologically harmful when released into a non-native environment, from one location to another.

The International Maritime Organization (IMO) has adopted the International Convention for the Control and Management of Ship's Ballast Water and Sediments /1/ to reduce the risk of spreading harmful aquatic organisms and pathogens released with ballast water. The convention requires that all ships' ballast water complies with the water quality requirements for discharges laid down in the D-2 regulation of the Convention.

Article 9 of the Convention lays down the legal basis for inspection for the purpose of determining whether a ship is in compliance with the Convention. Among others, it allows duly authorized officers to sample ships' ballast water in accordance with the IMO guidelines for ballast water sampling (G2) /4/.

On 26 April 2010, the European Maritime Safety Agency (EMSA) launched an invitation to tender regarding "the development of guidance on how to analyze a ballast water sample". DHI won the contract.

1.2 Objective

The Convention's D-2 Standards are expressed as numbers of viable organisms per volume, for different size classes of organisms (see Section 1.4). However, the Convention has not identified the methods to be used for compliance monitoring. The result of a measurement depends on the method used. Therefore, it is important to agree internationally on which analyses to use at a global level , in order to ensure a harmonised enforcement of the Convention. In addition, once an analytical result has been obtained, it must be determined whether the result complies with the Convention.

The objective of this project is to formalise a procedure for analyzing a ballast water sample for compliance purposes.

1.3 Aim

The aim of this project is the development of guidance on how to analyze and process the results from a representative ballast water sampling protocol.

1.4 Methods to be evaluated

The requirements to the quality of the discharge ballast water, which is defined in regulation D2 of the Convention, is shown in Table 1.1.

Table 1.1 The IMO ballast water performance standards (D-2)

Organism category	Standard
Organisms ≥ 50 µm	< 10 viable organisms/m^3
Organism size: ≥10 µm - < 50 µm	< 10 viable organisms /mL
Toxicogenic *Vibrio cholerae*	< 1 cfu/100 mL
Eschericha coli	< 250 cfu/100 mL
Intestinal Enterococci	< 100 cfu/100 mL

cfu: Colony forming units

Organisms ≥ 50 µm are mainly zooplankton. Organisms ≥10 µm and < 50 µm are mainly phytoplankton. However, both size groups contain both zooplankton and phytoplankton. The methods should thus be able to detect both the zooplankton and the phytoplankton and the size of the organisms. In addition, the unit of the D-2 standards for these size groups is viable organisms per volume, which means that the methods must be able to distinguish live organisms from dead organisms.

In addition to organisms ≥ 10 µm, the Convention sets standards for the feacal indicators, *Eschericha coli* and intestinal Enterococci and for toxicogenic *Vibrio cholerae*, which cause severe diarrhoea, dehydration and abdominal pain.

In this project, methods to determine zooplankton and phytoplankton from the scientific literature and from "land-based test facilities", as well as available internationally recognized methods, have been scrutinized and their suitability for compliance monitoring have been evaluated.

1.5 Review of statistical tools for analysis of data from "average" and "instantaneous" sampling protocols

The methods for compliance evaluation have been discussed. The "average" approach is, for the purposes of this document, defined as taking and analyzing more than two samples, and evaluating compliance by comparing the average and variation of the results to the standard. The "instantaneous" approach is for the purpose of this document defined as taking and analyzing one or more samples and comparing them to the standard on an individual basis.

1.6 Sampling

The call to tender stated explicitly that "Guidelines on how to take a sample – can be developed in house". Therefore, this is not a part of this project. However, sampling and analysis are both parts of the process of obtaining an analytical result and cannot be viewed separately. The sampling affects parameters such as detection limit and accuracy (precision + bias) and influences the evaluation of compliance.

2 CHALLENGES IN COMPLIANCE MONITORING

A number of challenges are associated with monitoring for compliance with the D-2 requirements of the Convention, challenges which are related to some of the basic assumptions of the D-2 requirements. First, the categorisation of organisms is based on size, not in biologically meaningful terms or on specified methods. Second, the organisms must be viable. The requirement for viability is logical and makes very good sense whereas the distinction between sizes is illogical.

The ideal method for compliance monitoring has a sufficient limit of detection, detects invasive species or indicator species, has a low variability, is cheap, quick and available in all parts of the world. However, the ideal method does not exist. Below, we will discuss some of the basic problems associated with the selection of methods for compliance monitoring and identify the basic requirements to the methods used for compliance monitoring. The criteria for ranking the analyses (see section 4.1) are based on the discussions below.

2.1 Viability

The D2-requirements concern – logically – viable organisms. In terms of the Convention, viability can ideally be seen as the ability to grow, proliferate or reproduce in the recipient where the ballast water is discharged. To determine viability in this sense would require an ecological study for each species in the sample and is not applicable for compliance monitoring.

The MEPC Guidelines for the approval of ballast water management systems G8 /3/ define viable organisms as "organisms and any life stages thereof that are living" (Section 3.12). Hence, the methods should be able to determine whether an organism is dead or alive. Since death criteria do not exist for the organisms in question, it is not possible to determine their exact time of death. Therefore, it is not a trivial task to determine the number of viable microorganisms in a sample at a given time.

A number of different death criteria may be applied as the principle of detection such as a disrupted cell membrane, absence of metabolic activity, absence of movement or non-culturability. Methods for counting number of viable organisms are being addressed in different fora such as the North Sea Ballast Water Opportunity Project.

The integrity of cell membranes can be determined by applying specific dyes that enter the cell if the cell membrane is disrupted, metabolic activity can be shown by determination of enzymatic-specific activities, movement can be visualised in a microscope and growth can be detected by observing the number of organisms or their activity over a period of time.

However, none of these criteria are ideal. For instance, organisms may have intact membranes but in reality be dead or dying as there may be a time lag from death to the disintegration of the cell membrane, leading to an overestimation. Metabolic activity may continue a while after the cell membrane has been disrupted, leading to an overestimation, and some organisms have dormant viable life stages with undetectable metabolic activity, leading to an underestimation. Mobility can only be used for mobile organisms. Using mobility may therefore result in an underestimation. Detection of

growth appears on the first glance to be the ideal criterion for phytoplankton. However, some algae have specific requirements, which may not be present during analysis, such as specific growth factors, temperature, light conditions and water movement. In addition, natural enemies such as grazers or viruses may repress growth, all resulting in underestimation of the number of organisms.

Furthermore, the different methods respond differently to different treatments. E.g., the cell may maintain integrity and metabolic activity several hours after UV-treatment, which means that analytical results on discharge after treatment in a BWMS may overestimate the number of live organisms.

2.2 Size

The D2 requirements are based on the size of the organisms, defined as the minimum dimension. The minimum dimension is defined in the MEPC Guidelines for ballast water sampling (G2) /4/ as:

".. based upon the dimensions of that organism's body, ignoring e.g., the size of spines, flagellae, or antenna. The minimum dimension should therefore be the smallest part of the "body", i.e. the smallest dimension between main body surfaces of an individual when looked at from all perspectives. For spherical shaped organisms, the minimum dimension should be the spherical diameter. For colony forming species, the individual should be measured as it is the smallest unit able to reproduce that needs to be tested in viability tests."

Although the definition is quite detailed and gives good guidance on where to measure and with regards to antenna etc., a number of problems still exist.

For instance, the minimum dimension is determined when looked at "from all perspectives". The size of an organism is often determined in a microscope or in a magnifier, which is two dimensional and therefore cannot look "from all perspectives".

For colony forming species, it is the individual cells that define the size. This means that each cell has to be measured (from all perspectives) and that some of the cells may belong to one size group whereas others of the same colony belong to another.

Another problem is that the size of the organisms is variable and depends on the nutritious status of the cells and where, in the cell cycle, they are (large before cell division). Therefore, organisms of the same species may belong to different size groups at different points of time.

And, in borderline cases, it can be very difficult to determine whether an organism is above or below the size limit.

2.3 Limit of detection (LoD) and limit of quantification (LoQ)

The limit of detection (LoD) is a characteristic of the analytical procedure and is defined in /5/. For chemical analyses, the LoD is the concentration, at which you can be 95% sure that the analyte is present in the sample.

However, for the analytical procedures used for determining the D2 Standard size groups, the limit of detection is 1 organism per volume sampled.

2.4 Analytical variability

The analytical variability (or accuracy) can be expressed as a combination of the precision and the bias. The precision is a measure of random errors and the bias is a measure of systematic errors. In practical terms, the accuracy can be determined by analyzing a number of identical samples and comparing the results with the true value. However, for biological samples, the true value is usually not known, and it can be difficult to determine the bias.

The sources of error are plentiful. Analysis of organisms is usually considered to follow the Poisson distribution (see Section 3.1), which means that the precision depends on the number or organisms counted. High numbers give lower relative variation. Other sources of variation are equipment, technicians, facilities and day-to-day variations.

Variation between laboratories is an important aspect, particularly when it comes to compliance monitoring since compliance should not depend on the laboratory. For chemical analyses, the variation between laboratories is evaluated by inter-calibrations and the performance of the laboratories can be tested in proficiency tests. Inter-calibrations are not easy to do on biological parameters as true values do not exist and the samples are not stable.

To ensure best possible analytical results, the laboratories should be well managed and required to determine the analytical variability by setting up uncertainty budgets for their analyses. This is best done by requiring that the analyses are performed as accredited analyses according to the ISO 17025 /6/.

2.5 Availability

The analyses for compliance monitoring must be broadly available in all parts of the world to allow monitoring globally. The availability depends mainly on the availability of the necessary skills, laboratory facilities and equipment. Many of the analyses in use today require expert knowledge or specialised equipment that are normally only available at universities and other research-based institutions.

2.6 Robustness

Robustness is the ability of the method to deal with a range of physical and chemical conditions such as temperature, salinity, range (high/low number of organisms), content of suspended solids and organic matter and types of organisms (autotrophic, mixotrophic, heterotrophic). In addition, the vulnerability during transport and handling is important as the organisms may die due to stress between sampling and analysis.

2.7 Practicability

A very important criterion for an analysis is the practicality, which includes issues such as simplicity, availability, cost, time to result and requirements to analytical skills and experience. Analytical methods should be as simple and cost and time effective as possible and must be available where the analysis is needed.

3 EVALUATION OF COMPLIANCE

Results from biological analyses of ballast water shall be used to decide whether the BWMSs show compliance or non-compliance with the limits of the D2 requirements /1/. The experience from land-based testing of BWMS is that they either work well, resulting in concentrations of organisms well below the D2 requirements or that they do not work well, resulting in concentrations of organisms well above the D2 requirements. Therefore, in most of the cases, the analytical results are expected to be either clearly above or below the D2 requirements. However, the BWMS may eventually become less efficient and other factors such as lack of maintenance may also influence the efficiency of the BWMS. It is therefore likely that there will be situations where it is difficult to determine compliance. Because non-compliance may have significant economical implications, it is important to have clear rules for assessing compliance.

In general, when limit values or standards are defined, they should be accompanied by instructions ("Evaluation rules") on how to evaluate the results in relation to the limit value /7, 8/. The absence of such evaluation rules invites a number of questions and ways of assessing compliance or non-compliance. Evaluation rules provide a prescription for the acceptance or rejection of a result and may take into account the uncertainty of the testing and the probability of making a wrong decision.

3.1 Principles of compliance assessment

If more than two test results are available from the laboratory testing, the average and variance of the results can be estimated using appropriate statistics[1] (see Section 2.4 for more information on analytical variability). This allows for an assessment of compliance or non-compliance, taking the uncertainty into account[2].

Figure 3.1 shows typical scenarios (1-4) arising when test results are used to assess compliance with an upper limit value. The dot shows the average result estimated from a number of test results and the horizontal line (wings) shows the estimated uncertainty.

[1] Normal distribution, logarithmic normal distribution, binomial distribution or the Poisson distribution

[2] Uncertainty related to test results consists of contributions from both sampling and testing, and it is well known that the sampling uncertainty in most situations constitutes a significant part of the total measurement uncertainty. However, very often only the uncertainty related to laboratory testing is taken into account, resulting in an underestimation of the total measurement uncertainty and thus providing an incorrect basis for an assessment of compliance. Guidance on appropriated methods of evaluating uncertainty is provided by /9/, /10/.

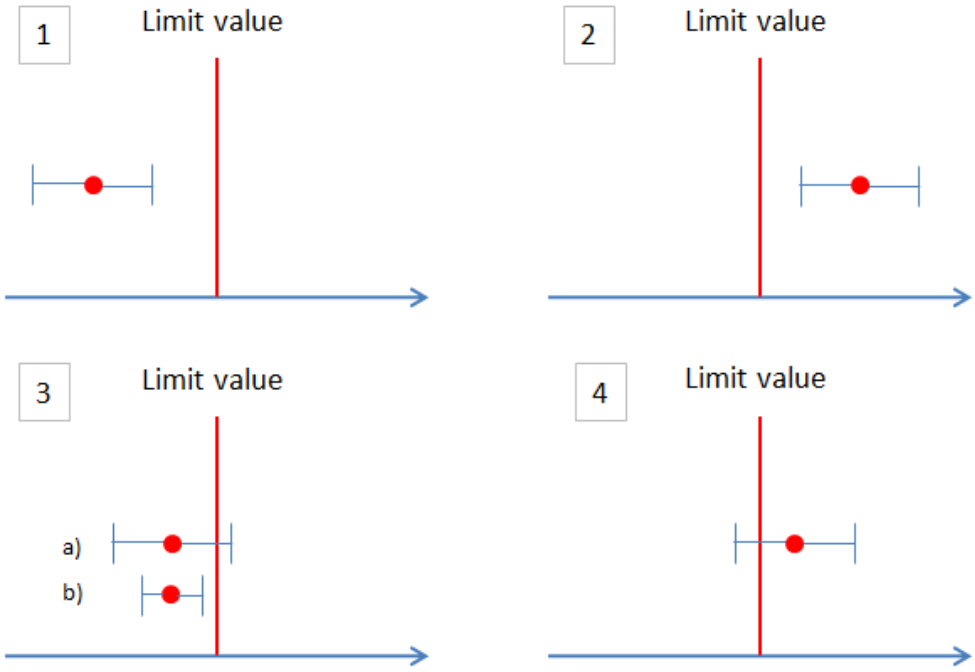


Figure 3.1 Assessment of compliance with an upper limit value

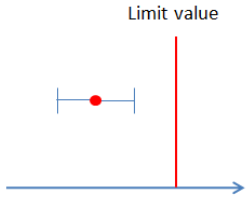
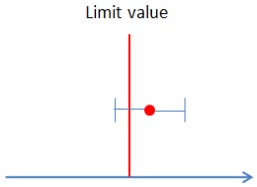
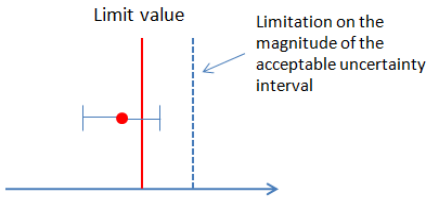
Evaluation of the results in scenarios 1 (a well performing BWMS) and 2 (a poor performing BWMS) is reasonably clear; the test results and their uncertainties provide good evidence that the true mean value of the test results is well above or well below the limit value, respectively. In scenarios 3-a) and 4, however, it is not possible to conclude (with the pre-defined certainty) that the true mean value is below the limit value. A possible remedy to this situation is to obtain more data, which provides a better estimate of the uncertainty and hopefully tightens the "wings" as shown in scenario 3b.

In general, the basic requirements for deciding whether the average test value complies with the standard or not are:

1. A specification giving the permitted upper limit value. For ballast water, this is the D2 requirements.
2. A specification of decision certainty (i.e. how certain must we be that we make the right decision)
3. A evaluation rule that describes how the uncertainty will be taken into account in the assessment of compliance.

Table 3.1 describes three different ways of formulating an evaluation rule that takes into account the un-certainty related to the estimation of the average test value.

Table 3.1 Three different ways of formulating an evaluation rule that takes into account the uncertainty related to the estimation of the average test value

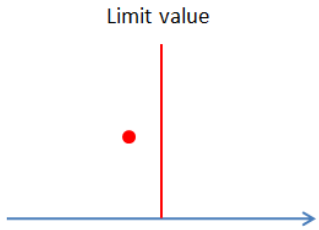
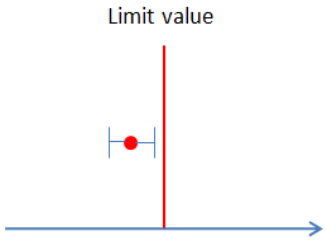
No	Relating test results to an upper limit value	Evaluation rule	Comments
1		The estimated average value plus the uncertainty must be below the limit value	A high degree of protection is achieved as the entire uncertainty range must be below the limit value. In some cases, this may lead to too strict criteria depending on the consequences of making a wrong decision
2		The estimated average value minus the uncertainty must not exceed the limit value	There is no incentive to reduce the uncertainty, quite the contrary - a high uncertainty will increase the chances of compliance. Such an evaluation rule should always be accompanied by a specification of the magnitude of the acceptable uncertainty interval
3		The estimated average value must be below the limit value and the average value plus the uncertainty must not exceed a specified value	It is most likely that the true value is below the limit value. However, with this evaluation rule it is accepted that some of the individual test results exceed the limit value up to a specified value set based upon restriction on the magnitude of the uncertainty interval

The selection of which rule to be used for compliance assessment of ballast water based on test results depends on the consequences of exceeding the limit values. If, for example, for one category of organisms, the consequences of a false compliance judgment are catastrophic, a high level of protection must be achieved. This would favour evaluation rule No. 1, which supports a high level of protection. On the other hand, if the organism category in question is relatively harmless, less strict requirements can be achieved through evaluation rules Nos. 2 or 3. Such choices should be made by legislators in cooperation with experts.

The situation is different if the assessment of compliance with a limit value is based on comparison of a single result with the limit value. Two scenarios are illustrated in Table 3.2 :

1. Assessment of a single test result not taking into account the uncertainty related to the test result
2. Assessment of a single test result considering uncertainty. The uncertainty can be determined by using the Poisson distribution as described in Section 3.1, using 1 sample. Alternatively, the performing accredited laboratory should provide an uncertainty estimate determined for the specific test method.

Table 3.2 Assessment of conformity of an instantaneous result

No	Relating test results to an upper limit value	Evaluation rule	Comments
A		To comply with the limit value, each single test result must be below the limit value.	Using this evaluation rule, the uncertainty is not taken into account and the magnitude of the uncertainty is unknown. Thus, the certainty of the decision of compliance or non-compliance is also unknown.
B		For this situation evaluation rules Nos. 1, 2 and 3 in Table 3.1 can be used.	The uncertainty introduced by sampling may be significant. Thus, using only the test uncertainty for conformity assessment may introduce a false decision certainty. Representative sampling is crucial for obtaining reliable results. The uncertainty (precision) can be determined on the basis of the Poisson distribution or an uncertainty based on the quality control carried out by the performing laboratory.

Applying evaluation rule A is simple but the certainty of the decision is unknown, which is undesirable. Using evaluation rule B allows for the same rules as the evaluation rules Nos. 1 to 3 in Table 3.1. However, the statistic evaluation can only be based on the Poisson distribution.

3.2 Evaluation by the "Average" and "Instantaneous" sampling approach

The selection of the appropriate statistical tool for assessment of compliance depends on the data and its distribution. A decision tree for selecting the appropriate test is shown in Figure 3.2 and described in more detail below.

3.2.1 The "average" approach

The "average" approach is for the context of this document defined as taking and analyzing more than two samples and evaluating compliance by comparing average and variation of the results to the standard.

According to the G2 guideline /4/, the sampling protocol should result in samples that are representative of the whole discharge of ballast water from any single tank or any combination of tanks being discharged. Whether this criterion is fulfilled depends on the size and number of samples taken, the span of time during which the individual samples are taken as well as the design and position of the sampling point(s). A low number of small samples and a short sampling time compared with the entire discharge event will give a low representativeness, because the probability of missing a high or low peak is high.

After obtaining the analytical results, the compliance evaluation should be performed according to an evaluation rule. The evaluation rule does not change or add to the D2-standard but it describes the principle of how to assess compliance. The evaluation rule could be either of the rules presented in Table 3.1, by application of an appropriate statistical test and selection of a level of decision certainty.

Once the evaluation rule is chosen, it can be formulated in statistical terms as a null hypothesis, which can then be tested statistically to determine compliance. Using evaluation rule 2 as an example, the null hypothesis H_0 and the alternative hypothesis H_A would be:

H_0: The average concentration of organisms equals the standard
H_A: The average concentration of organisms is smaller than the standard at a specified level of significance

Selection of an appropriate statistical method depends on the distribution of the data. If we assume that a measurement result can be represented by a continuous distribution (e.g. normal distribution, lognormal distribution etc), typical parameters like the mean, standard deviation or the median can be calculated and used for a statistical test. After successful testing of normality (using e.g. Anderson-Darling test or the Shapiro-Wilk test), the null hypothesis may be tested by 1-sample t-test at a selected significance level α and p-value (see examples in Appendix C).

If the data (or transformed data) are not normal-distributed, the non-parametric 1-sample Wilcoxon test should be applied.

During analysis of organisms, the number of organisms is counted individually and it can be assumed that the collected data can be represented by a discrete distribution, e.g. Poisson distribution. In this case, the 1-sample Poisson rate test is applicable. In the 1-sample Poisson rate test, analytical results from each individual sample are pooled (i.e. the numbers of organisms over the entire volume are combined), which means that the variation between the individual samples are not taken into account.

Examples of the different statistical compliance evaluations are shown in Appendix C.

3.2.2 The "instantaneous" approach

The "instantaneous" approach is for the context of this document defined as taking and analyzing one or more samples and comparing them with the standard on an individual basis.

As for the "average" approach, the "instantaneous" sampling protocol must ensure that the sample or samples are representative of the discharge of ballast water. Representativeness can be ensured by taking one sample of sufficient size over the entire discharge event or by taking many samples, each of sufficient size, over a period of time, which is significant compared with the span of the discharge event.

After obtaining the analytical result(s), the compliance evaluation of each individual test result should be performed according to an evaluation rule. The rule may or may not take uncertainty into account as described in Table 3.2. However, not taking the uncer-

tainty into account results in an unknown certainty of decision and will not be considered further. Therefore, the evaluation rules could be either of the rules presented in Table 3.1, by application of an appropriate statistical test and selection of a level of decision certainty.

Once the evaluation rule is chosen, it can be formulated in statistical terms as a null hypothesis, which can then be tested statistically to determine compliance. Using evaluation rule 2 as an example, the null hypothesis would be:

H_0: The average concentration of organisms equals the standard
H_A: The average concentration of organisms is smaller than the standard at a specified level of significance

As mentioned above, the counting of organisms is usually considered to follow the Poisson distribution. In this case, the 1-sample Poisson rate test is applicable for testing the H_0 hypothesis.

Examples of the different statistical compliance evaluations are shown in Appendix C.

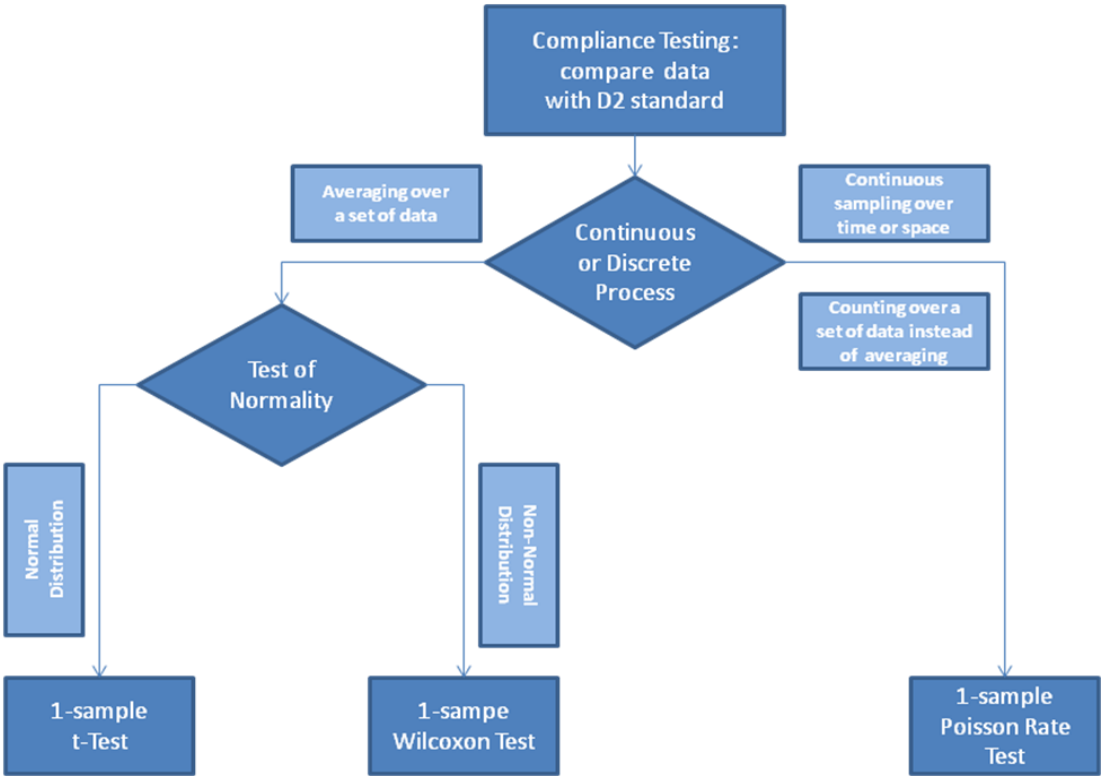


Figure 3.2 Decision tree for compliance testing data against the D2 standards

3.2.3 Pros and cons for "instantaneous" and "average" compliance assessment

As mentioned above, it is expected that the concentration of organisms in the discharge water is either well below or well above the D2 requirements. In most of the cases, compliance evaluation is thus expected to be obvious. In some cases, however, it will be unclear whether a ship is in compliance or not. It is in this type of cases that a statistical tool is required for evaluation. This section discusses the advantages and disadvantages of "instantaneous" and "average" compliance assessment.

The basic difference between the "instantaneous" and "average" approaches is that the "average" approach takes the variation of the concentration of organisms during the de-ballasting event into account whereas the "instantaneous" approach only uses 1 analytical result for the compliance assessment. Therefore, the "average" approach is normally preferred for compliance monitoring. However, the "average" approach has the disadvantage that a high variation between the analytical results, which is expected to occur, either leads to a very strict (evaluation rule 1) interpretation of the D2 standards, a very loose (evaluation rule 2) interpretation or yields indecisive results if the limit of variation is breached (evaluation rule 3). Any of these situations are undesirable.

The "instantaneous" approach has the inherent disadvantage that compliance is assessed for the individual samples. In the reality, this will mean that the requirement of G2 to take representative samples cannot be fulfilled unless all samples taken are representative. In addition, the probability on non-compliance will increase with the number of samples taken.

The specific case of the "instantaneous" approach, in which a flow-integrated sample is taken throughout the entire discharge event, does not have the problems described above if the sample size is sufficiently large. In addition, if the sample is taken properly, it will represent the average of the de-ballasted water and probably constitute a better representation than separate samples. However, if the discharge event takes place over a long span of time, it is likely that the sampling procedure will produce a negative bias because the initially sampled organisms may die prior to analysis due to the sample handling. In order to avoid the bias, subsamples of the flow-integrated sample may be analyzed across the sampling event or discrete samples may be taken and analyzed within the recommended maximum holding time and treated statistically by the Poisson distribution.

3.3 Recommendation

It is recommended to apply an evaluation rule that considers uncertainty.

It is also recommended to take either a flow-integrated sample or discrete samples of sufficient size and assess compliance by use of statistics based on the Poisson distribution. The flow-integrated sampling has the advantage of higher representativeness and, for long ballast water discharge events, the disadvantage of possible introduction of a negative analytical bias. Discrete sampling has the advantage of reducing the risk of negative bias and the disadvantage of a possible lower representativeness due to the risk of missing high or low peaks of organisms.

4 EVALUATION OF METHODS

4.1 Principles of evaluation

A search of peer-reviewed literature was carried out on the FINDit database, which is provided by the Technical Information Center of Denmark. The Technical Information Center of Denmark is the university library of the Technical University of Denmark. It is also the national centre for technical and scientific information. The database includes approx. 17,000 journals, primarily within technical, natural and medical sciences.

In addition, a letter was sent to all known land-based test facilities with a request to provide information on their methods.

A set of criteria (Table 4.1) was developed to evaluate the methods when applied to analysis for enforcement. Each method was given marks from 1 to 3. The mark 1 was given to a methodology, which is less useful for enforcement analysis. The mark 2 was given for methodology, which is useful for enforcement analysis. This mark is also given when the criterion is irrelevant for the specific method. The mark 3 was given for a methodology, which is suitable for enforcement analysis. A number of the evaluation criteria used are linked directly to the applicability for enforcement analysis. Therefore, this evaluation has no relevance for methods used for land-based testing.

Table4.1 Criteria for the evaluation of the analytical methods

1	Does the analysis have a sufficient limit of detection (LoD) and limit of quantification (LoQ)?
2	Interference
a	Does the analysis distinguish between size of the organisms?
b	Does the method distinguish correctly between viable and non-viable, i.e. is the death criterion acceptable when testing ballast water?
c	Are there other interferences?
3	Is the accuracy acceptable?
a	Does sample pre-treatment give rise to bias?
b	Does the analyst give rise to bias?
c	Precision?
d	Information on reproducibility and repeatability
4	Is the analysis robust?
a	Geographical variations
i	Temperature
ii	Salinity
b	Type of organisms: autotrophic, mixotrophic, heterotrophic
c	Range (high/low number of organisms)
d	TSS/DOC/POC levels
e	Transport/handling
5	Practicality
a	Complex/simple (many/few operations)
b	Availability
c	Cost (high or low)
d	Tie to result
e	Des the analysis require highly skilled analysts

The result of the evaluation of the methods is shown in Appendix A.

4.2 Organisms ≥ 50 µm

Organisms equal to or larger than 50 µm in minimum dimension are typically dominated by zooplankton, such like rotifers, crustaceans, molluscs, worms and jellyfish.

In the evaluated methods, sample assessments comprise determinations of the number of live and dead organisms present. In general, the evaluated methods include counting the numbers and a rough taxonomic classification of the organisms by microscopy examination at e.g. 20× and 50× magnification.

The viability of the organisms is determined by using movement and response to stimulus techniques such as poking the organisms and tapping the counting chamber; several methods define dead organisms as immobile organisms. Some methods use vital staining such as Neutral Red to determine viability. However, vital staining is not suitable for all groups of organisms, such as bivalves which are not coloured /28/. Others, such as polychaete barnacle nauplii, need to be inspected closely for stain uptake /29/. Therefore, the use of neutral red staining requires special attention. Staining methods should thus be used in combination with detection of movement in order to include moving but non stained organisms.

Prior to analysis, zooplankton are typically concentrated by use of plankton nets (50 µm in diagonal) and transferred to storage/transport-containers for further pre-treatment and analysis. The concentration procedure, transport and storing conditions as well as time from sampling to analysis may be critical due to damage of the organisms, subsequently an increase of the mortality.

Many studies indicated considerable mortality of zooplankton during storage. Maximum hold time should be validated for each analysis, e.g., so that the detectable zooplankton mortality over the hold time does not exceed, e.g., 5%. A holding time not exceeding 2 hours has been recommended /17/.

Most methods use simple counting with eye observation in microscope. In addition, one method uses digital video recording to increase analysis speed and reduce the problem of mortality during storage. One method combines flow cytometry and microscopy. This method, automatically counts, photographs and analyzes a discrete sample or a continuous flow. The technology is available on market. However, the equipment is expensive and use of the equipment requires specialized knowledge and experience. In flow cytometry, relatively small sample volumes can be analyzed and the samples may need to be further concentrated.

The methods received very similar scores except for the method combining flow cytometry and microscopy because of low score on practicality, i.e. high cost and specialised skills. Scores of the individual methods are shown in Appendix A.

Conclusions

1. Movement and response to stimulus techniques are simple techniques for examination of viable/dead zooplankton in the microscope. Laboratories that can perform counts of zooplankton using microscope are available in most parts of the world.

2. Vital staining may be a very useful supplement for examination of viable/dead zooplankton.
3. Distinction between size of the organisms is possibly in all of the evaluated methods.
4. Both storage condition and time before analysis may be critical as this may increase the mortality.
5. The use of video recording enhances the speed of analysis, thereby reducing the mortality and subsequently, the samples can be evaluated and documented by different persons.
6. Precision, reproducibility and repeatability should be determined and documented by the individual laboratories/methods.
7. All the methods are considered robust in relation to geographical variations, salinity, and types of organisms ≥ 50µm.

Recommendations
It is recommended to use simple microscopic examination of the organisms ≥ 50 µm for ballast water compliance monitoring. The microscopic examination of organisms is a robust, simple and cheap method, and laboratories for this analysis are available worldwide. Viability is determined by observing movement. Movement is induced by poking the individual animals and by tapping the counting chamber. In addition, it is recommended to validate the methods using vital staining and video recording for analysis of viable organism and to reduce the possibly bias for motility during transport and storing of samples.

4.3 Organisms ≥ 10 µm and < 50 µm

4.3.1 Phytoplankton and microzooplankton

The majority of organisms in the size range from ≥ 10 to < 50 µm are typically dominated by phytoplankton but usually also include microzooplankton and suspended benthic algae.

Several different methods have been used for determining the number of viable organisms present in ballast water (in land-based tests as well as in shipboard tests).

The different methods used for determining the number of viable organisms present in ballast water were evaluated according to the criteria described in Table 4.1. The scores of the individual methods are shown in Appendix B.

The majority of the methods use simple light microscopes, either upright or inverted microscopes or both. The initial costs of microscopes are relative high but light microscopes can be used for many years. Both upright and inverted microscopes are available all over the world and can be used for determining number, size, taxonomy, cell integrity and mobility, all of which are very useful parameters for determining the number of viable organisms in ballast water. High degrees of skill and experience are needed for identification of species and evaluation of the cell integrity. Identification of the organisms is not part of the compliance evaluation but taxonomic information on the organisms is often useful. The methods using microscopy for assessing number of viable organism have been given a high score in the evaluation.

More detailed microscopy can be made using an epi-fluorescence microscope. Epi-fluorescence microscopes are not as widespread as the simple light microscopes but they are often available at universities and research laboratories. Fluorescence generated by pigments in phytoplankton can be used for more detailed examination of cell integrity, viability by staining techniques and differentiation between heterotrophic and autotrophic organisms. Epi-fluorescence microscopy works best with live samples and samples preserved with formaldehyde or glutaraldehyde. A high degree of skill and experience is needed for successful application of epi-fluorescence microscopy. The methods using epi-fluorescence microscopy for assessing number of viable organism have been given a high score in the evaluation.

More advanced methods have been used to determine, size, fluorescence and viability in ballast water. In flow-cytometry, the individual cells or particles of the samples are channelled through a narrow tube to the detector. The technology is available on the market but it is relatively costly and requires specialized skills and experience. Some flow-cytometers use small sample volumes and may require a higher concentration of the sample. High particle loads (i.e. ballast water from rivers or estuaries) may therefore reduce the detection limit. Flow-cytometry reduces the manual labour required when compared to traditional microscopy techniques as it uses automatic quantification techniques. However, they may require a substantial manual handling of the recorded data. Flow-cytometers are relatively costly.

Techniques and standardized methods for the enumeration and viability analyses of microalgae and microzooplankton remain an active area of investigation. Fluorescein diacetate (FDA) and chloromethylfluorescein diacetate (CMFDA) have been used to determine viable cells/organisms. When non-specific esterase, which is an enzyme present in live cells, cleaves the stains, the resultant molecules fluoresce in green colour when illuminated with a blue light in epi-fluorescence microscopes. Highly skilled and experienced analysts are needed for identification of viable cells and organisms by epi-fluorescence. Highest score in the evaluation is given to methods using FDA/CMFDA stains in combinations with epi-fluorescence microscopy.

An immediate obvious elegant method to test viability of organism is to measure growth or metabolic activities in short-term or long-term incubation. Re-growth experiments, in which the increase in the number of cells was followed for days or quantified by Most Probable Number (MPN) techniques, have been used. A major drawback is that specific growth requirements may not be fulfilled during the incubation, e.g. temperature, light and movement and, finally, growth can be repressed by the presence of natural enemies like grazing zooplankton and viruses. Due to possibly bias and a large workload, re-growth experiments did not receive a high score in the evaluation.

Measurements of photosynthetic activity/metabolic activity by determining the fixation of $^{14}CO_2$ and PAM-fluorescence for assessing photosynthetic efficiency are both sensitive methods for analysis of viability of the microalgae. These methods, which are bulk parameters and not applicable for determining the number of viable organisms, only received low scores in the evaluation. However, bulk parameters may be useful as supporting parameters for evaluation of viability of the organism, e.g. detected by microscopy.

Cell densities in ballast may be low. Thus, samples must be concentrated to quantify accurately number of organisms. Planktonic organisms can be fragile and damaged by physical stress during concentration on plankton nets or filters and during storage. Planktonic organisms must be handled carefully to avoid damaging of the cells. The methods used for concentration of organisms between ≥10 - < 50 µm in ballast water samples include sedimentation and filtering.

Concentration of organisms can be achieved by a settlement method where a sample is poured into a settling cylinder and allowed sufficient time for cells to settle. After settling, the water from the upper portion of the sample is gently removed and the final volume noted. The settlement method can only be used on preserved samples and cannot be recommended as the viability can only be determined by the structural integrity. Another method involves concentration of the organisms by filtering the sample through e.g. a 10-µm mesh (i.e. plankton net). Following concentration, the organisms are resuspended and counted and evaluated by microscopy using counting chamber.

A more simple, rapid, flexible and cautious method for concentrating plankton cells is the use of transparent membrane filters (e.g. 10-µm pore size polycarbonate membrane filters). In this method, water samples are filtered directly onto a membrane, which subsequently can be placed directly on slides for examination of cells under the microscopy. All live, vital stained and fixed organisms can be evaluated on transparent membrane filters. The volume to be analyzed can be adjusted depending on the cell density.

Conclusions

1. Simple upright and inverted light microscopes are very useful for enumeration of morphologically healthy organisms and motile organisms as well as for measuring the size of the organisms. All parameters that are very useful for evaluating number of viable organisms in ballast water. High degrees of skill and experience are needed for identification of species and evaluation of the morphological health of cell. Both upright and inverted microscopes as well as skills and experience of analysis are available in most parts of the world.

2. Fluorescence generated from photosynthetic pigments and other pigments in phytoplankton can be used for more detailed examination of the morphological health of cells, for evaluation of stained organisms as well as for differentiation between heterotrophic and autotrophic organisms. Epi-fluorescence microscope, skills and experience of analysis are available at universities and other research laboratories worldwide.

3. Flow cytometers are advanced technologies available on the market. However, the equipment is expensive. Use of the equipment requires specialized knowledge and experience that are only found at some universities and other research laboratories worldwide. High particle loads (i.e. ballast water from rivers or estuaries) may therefore reduce the detection limit.

4. Fluorescein diacetate (FDA) and chloromethylfluorescein diacetate (CMFDA) have been used to determine viable cells/organisms. High degrees of skill and experience are needed for identification of viable cells and organisms by epi-fluorescence. The condition of organisms may affect staining effectiveness.

5. Re-growth experiments, in which the increase in the number of cells is followed for days, are slow and work intensive. In addition, a major drawback may be that specific growth factors during the incubation may not be fulfilled, giving a risk of bias.

6. Bulk parameter such as photosynthetic/photosynthetic activity can be used as supporting parameter for methods, in which the number of viable organisms is determined.

7. Planktonic organisms may be fragile and methods used for concentrating samples should receive great attention in selecting robust pre-treatment and analytical methods.

8. Currently, no single method exists that, in a simple manner, can be used for compliance monitoring for viable organisms between ≥ 10 - < 50 µm in ballast water. Generally, the present, available methods require high degrees of skill and experience. Standard Operating Procedures for microscopy and flow-cytometry are not easy to develop as several subjective judgments are made. Flow cytometers are expensive equipment and skill and experience of analysis are only available in some parts of the world.

Recommendations
Based on the evaluation made in the project, it is recommended to use FDA/CMFDA in combination with epi-fluorescence microscopy for determining the number of viable organisms in ballast water. However, as high degrees of skill and experience are needed and as the many subjective judgments increase the uncertainty, it is recommended to initiate development of a robust and more objective method(s) for compliance monitoring.

4.4 Bacteria

4.4.1 The faecal indicators: E. coli and Enterococci

The D2 requirements to the faecal indicators are identical to the requirements to excellent quality of coastal waters as defined in the European Bathing Water Directive /11/, which again is based partly on the WHO guidelines for recreational water /27/. The Directive requires the use of ISO 7899-1 or ISO 7899-2 for the determination of Enterococci and the use of ISO 9308-3 or ISO 9308-1 for the detection of *E. coli*. The requirements are based on a upper 95-percentile evaluation (Annex 2 of the Directive).

National regulations of bathing water in different parts of the world differ in the choice of analytical methods. E.g., in USA, *E. coli* is determined in accordance with Standard Method 9213D and Enterococci are determined in accordance with Standard Method 9230. Although there may not be equivalency between national methods used for bathing waters in different parts of the world and the ISO methods, it is suggested to allow the use of the nationally accepted methods to ensure a local availability. However, the methods must be quantitative (as opposed to presence/absence) and there should be a requirement for use of accredited analyses.

It is suggested to use a 95-percentile statistical evaluation because the standards of the Directive are related to the 95-percentile. The number of samples for the evaluation

should be sufficient to define the mean and standard deviations of the $\log_{10}$ bacterial enumerations.

The upper 95-percentile can be calculated as in /11/:

1. Take the $\log_{10}$ value of all bacterial enumerations in the data sequence to be evaluated (if a zero value is obtained, take the $\log_{10}$ value of the minimum detection limit of the applied analytical method instead)
2. Calculate the arithmetic mean of the $\log_{10}$ values (μ)
3. Calculate the standard deviation of the $\log_{10}$ values (σ)
4. The upper 95-percentile point of the data probability density function is derived from the following equation: upper 95-percentile = antilog ($\mu + 1{,}65\ \sigma$)

The 95-percentile statistical evaluation approach requires that a number of independent samples is taken over a period of time.

Alternatively, compliance can be assessed as described in Chapter 3. However, it should be kept in mind that the D2 standards for the indicator bacteria correspond to excellent water quality, and the risk related to moderate non-compliance can be considered small.

4.4.2 *Vibrio cholerae*

Vibrio cholerae is the causing agent of cholera, one of the organisms that have killed most people throughout history. Today, cholera is usually only a problem in emergency situations but it also occurs sporadically in areas with poor hygiene.

Analysis of *V. cholera* should be performed by traditional culturing methods such as ISO/TS 21872-1 /13/. Sampling can be carried out by filtering 100 mL of ballast water and incubating the filter according to the ISO/TS 21872-1. It is noted that work with confirmed *V. cholerae* requires specialised laboratories, approved to handle highly pathogenic microorganisms.

5 RECOMMENTATIONS

It is recommended to apply an evaluation rule that considers uncertainty.

It is also recommended to take either a flow-integrated sample or discrete samples of sufficient size and assess compliance by use of statistics based on the Poisson distribution. The flow-integrated sampling has the advantage of higher representativeness and, for long ballast water discharge events, the disadvantage of possible introduction of a negative analytical bias. Discrete sampling has the advantage of reducing the risk of negative bias and the disadvantage of a possible lower representativeness due to the risk of missing high or low peaks of organisms.

It is recommended to use simple microscopic examination of the organisms ≥ 50 µm for ballast water compliance monitoring. The microscopic examination of organisms is a robust, simple and cheap method, and laboratories for this analysis are available worldwide. Viability is determined by observing movement. Movement is induced by poking the individual animals and by tapping the counting chamber. In addition, it is recommended to validate the methods using vital staining and video recording for analysis of

viable organism and to reduce the possibly bias for motility during transport and storing of samples.

Based on the evaluation made in the project, it is recommended to use FDA/CMFDA in combination with epi-fluorescence microscopy for determining the number of viable organisms in ballast water. However, as high degrees of skill and experience are needed and as the many subjective judgments increase the uncertainty, it is recommended to initiate development of a robust and more objective method(s) for compliance monitoring.

For the indicator bacteria, it is recommended to use the national analyses used for bathing water analysis.

Analysis of *V. cholera* is recommended to be performed by traditional culturing methods such as ISO/TS 21872-1

In all cases, analysis should be carried out by accredited laboratories.

6 ACKNOWLEDGEMENT

The following test facilities and institutions are acknowledged for contributing with information on analytical methods:

Great Ships Initiative/Northeast-Midwest Institute, USA
NIVA, Norway
Naval Research Laboratory/SAIC, USA
Maritime Environmental Resource Center, USA
Korea Marine Equipment Research Institute (KOMERI)
Nippon Hakuyohin Kentei Kyokai, Japan
DHI, Denmark
Marine Scotland - Science
Newcastle University Marine International School of Marine Science and Technology, UK

7 REFERENCES

/1/ International Convention for the Control and Management of Ships Ballast Water and Sediments. London : International Maritime Organization, 2005.

/2/ Guidelines for safe recreational water environments Vol. 1 coastal and fresh waters. WHO, Geneva. 2003. ISBN 92 4 154580 1.

/3/ MEPC. Guidelines for approval of ballast water management systems (G8). Resolution MEPC.174(58). Adopted 10th October 2008.

/4/ MEPC. Guidelines for ballast water sampling (G2). Resolution MEPC.173(58). Adopted on 10 October 2008

/5/ JCGM 200:2008. International vocabulary of metrology — Basic and general concepts and associated terms (VIM). The Joint Committee for Guides in Metrology (JCGM/WG 2).

/6/ ISO/IEC 17025:2005. General requirements for the competence of testing and calibration laboratories.

/7/ EURACHEM/CITAC Guide (2007). Use of uncertainty information in compliance assessment.

/8/ King B. (1999). Assesment of compliance of analytical results with regulatory specification limits. Accred Qual Assur., 4: 27-30. Springer-Verlag 1999.

/9/ EURACHEM/CITAC Guide (2007). Measurement uncertainty arising from sampling – A guide to methods and approaches.

/10/ Nordtest TR 604 (2007). Uncertainty from sampling – a Nordtest handbook for sampling planners on sampling quality assurance and uncertainty estimation. NT technical report from Nordic Innovation Centre.

/11/ Directive 2006/7/EC of the European Parliament and of the Council of 15 February 2006 concerning the management of bathing water quality and repealing Directive 76/160/EEC.

/12/ Bartram J, Rees G, ed. (2000) Monitoring bathing waters: a practical guide to the design and implementation of assessments and monitoring programmes. London, E & FN Spon. Published on behalf of the World Health Organization, Commission of the European Communities and US Environmental Protection Agency.

/13/ ISO/TS 21872-1:2007 Microbiology of food and animal feeding stuffs – Horizontal method for the detection of potentially enteropathogenic Vibrio spp. – Part 1: Detection of Vibrio parahaemolyticus and Vibrio cholerae.

/14/ UNESCO, Karlson B, Cusack C and Bresnan E. Microscopic and moleculars methods for quatitative phytoplankton analysis 2010. UNESCO 2010.

/15/ ETV. US Environmental Protection Agency Environmental Technology Verification Program Generic Protocol for the Verification of Ballast Water Treatment Technologies 2010.

/16/ Bresnan E. Identification and enumeration of toxin-producing phytoplankton in seawater samples in Scotland. Marine Scotland-Science 2010.

/17/ GSI. The Great Ships Great Ships Initiative (GSI) Research Protecols 2010. http://www.nemw.org/GSI/protocols.htm

/18/ McCollin T, Quilez-Badia G, Josefsen K.D, Gill M.E, Mesbahi E, Frid C.L.J. Ship board testing of a deoxygenation ballast water treatment. Marine Pollution Bulletin, 54:1170-1178, 2007.

/19/ David M, Gollasch S, Cabrini M, Perkovic M, Bosnjak D, Virgilio D. Results from the first ballast water sampling study in the Mediterranean Sea - the Port of Koper study. Mar Pollut Bull. 54(1). 2007.

/20/ Wright D A, R Dawson, V Caceres, C E Orano-Dawson, G E Kananen, S J Cutler and H G Cutler. Shipboard testing of the efficacy of SeaKleen as a ballast water treatment to eliminate non-indigenous species aboard a working tanker in Pacific waters. Env Tech 30(9):893-910, 2009.

/21/ Tamburri M. Test plan for the performance evaluation of the Severn Trent De Nora BalPuretm BP-2000 Ballast Water Management System. Maritime Environmental Resource Center, Maryland, US, 2009.

/22/ Carney K and Mesbahi E. Methods of determining viability used for the assessment of a ballast water treatment system. Newcastle University Marine International 2010.

/23/ Konishi H The ship Equipment Inspection Society of Japan, Konishi H. Standard Analytical Method recommended for quantitative analysis of organisms for approval of Ballast Water Management Systems 2010.

/24/ NIOZ 2009, Veldhuis M and Fuhr F. Final report of the land-based and shipboard testing of the SEDNA®-system.

/25/ DHI 2009. Gustavson K, Jørgensen C and Petersen G. DHI Methods for evaluation viable organism in Ballast water.

/26/ NIVA 2010. NIVA methods for evaluation viable organism in Ballast water.

/27/ WHO 2003. Guidelines for safe recreational water environments. Volume 1, Coastal and fresh waters. WHO Geneva. ISBN 92 4 154580 1.

/28/ Final report of the land-based and shipboard testing of the SEDNA®-system (Veldhuis M.J.W & Fuhr F, NIOZ 2008). http://www.bsh.de/de/Meeresdaten/Umweltschutz/Ballastwasser/Test_results.pdf

/29/ Elliott, D.T. and Tang, K.W. 2009. Simple staining method for differentiating live and dead marine zooplankton in field samples. Limnol. Oceanogr.: Methods 7, 2009, 585–594.

/30/ Globallast. http://globallast.imo.org/

A P P E N D I X A

Scores - Evaluation of methods for organisms ≥ 50 µm

	Score (0=ubrugelig, 1 dårlig, 2 middel eller irrelevant, 3 god, irrelevant)	Tamburri 2009	ETV 2010	McCollins et al 2007	David et al 2007	Wright 2009	DHI 2009	Niva 2010	NIOZ	Konishi 2(	GSI 2009	Carney and Mesbahi 2010	Unesco 2010
1.	**Does the analysis have a sufficient limit of detection (LoD) and limit of quantification (LoQ)**	3	3	3	3	3	3	3	3	3	3	3	**2**
2.	**Interference**												
a.	Does the analysis distinguish between size of the organisms?	3	3	3	3	3	3	3	3	3	3	3	3
b.	Does the method distinguish correctly between viable and non-viable, i.e. is the death criterion acceptable when testing ballast water	2	2	3	2	2	3	1	3	2	2	2	2
c.	Are there other interferences?	1	1	3	1	1	3	1	1	1	1	1	1
3.	**Is the accuracy acceptable**												
a.	Does sample pretreatment give rise to bias	2	2	2	3	2	2	2	2	2	2	2	1
b.	Does the analyst give rise to bias	2	2	3	2	2	3	2	3	2	2	2	3
c.	Precision?	2	2	3	2	2	2	2	2	2	2	2	2
d.	Information on reproducibility and repeatability	3	3	3	3	3	3	3	3	3	3	3	3
4.	**Is the analysis robust?**												
a.	Geographical variations	3	3	3	3	3	3	3	3	3	3	3	3
i.	temperature	3	3	3	3	3	3	3	3	3	3	3	3
ii.	salinity	3	3	2	3	3	3	3	3	3	3	3	3
b.	Type of organisms: autotrophic, mixotrophic, heterotrophic	3	3	3	3	3	3	3	3	3	3	3	3
c.	Range (high/low number of organisms)	3	3	3	3	3	3	3	3	3	3	3	1
d.	TSS/DOC/POC levels	2	2	2	2	2	2	2	2	2	2	2	1
e.	Transport/handling	1	1	1	1	1	1	1	1	1	1	1	1
5.	**Practicality**												
a.	complex/simple (many/few operations)	3	3	2	3	3	2	3	2	3	3	3	1
b.	availability	3	3	2	3	3	3	3	3	3	3	3	2
c.	cost (high or low)	3	3	3	3	3	3	3	3	3	3	3	1
d.	time to result	3	3	3	3	3	3	3	3	3	3	3	2
e.	does the analysis require highly skilled analysts	2	2	2	2	2	2	2	2	2	2	2	1
	Total score	**50**	**50**	**52**	**51**	**50**	**53**	**49**	**51**	**50**	**50**	**50**	**39**

A P P E N D I X B

Scores - Evaluation of methods for organisms ≥ 10 - < 50 µm

Score (0=ubrugelig, 1 dårlig, 2 middel eller irrelevant, 3 god, irrele		Konishi 2010	Konishi 2010	Konishi 2010	Konishi 2010	Wright 2009	Wright 2009	Wright 2009	Wright 2009	David et al 2007	Tamburri 2009
	Endpoint	Morphological	Mobility	Re-growth	Vital Calcein-AM	Morphological	Mobility	In vitro floroscen	24 regrowth	no of cellesand m	No of celles and morphol
1.	**Does the analysis have a sufficient limit of detection (LoD) and limit of quantification (LoQ)**	3	3	3	3	3	3	3	3	3	3
2.	**Interference**										
a.	Does the analysis distinguish between size of the organisms?	3	3	3	3	3	3	1	3	3	3
b.	Does the method distinguish correctly between viable and non-viable, i.e. is the death criterion acceptable when testing ballast water	1	1	3	3	1	1	1	2	1	1
c.	Are there other interferences?	2	2	2	2	2	2	2	2	2	2
3.	**Is the accuracy acceptable**										
a.	Does sample pretreatment give rise to bias	1	1	1	1	1	1	1	1	1	1
b.	Does the analyst give rise to bias	1	1	1	1	1	1	1	1	1	1
c.	Precision?	2	2	2	3	2	2	2	2	2	2
d.	Information on reproducibility and repeatability	2	2	2	2	2	2	2	2	2	2
4.	**Is the analysis robust?**										
a.	Geographical variations	3	3	3	3	3	3	3	3	3	3
i.	temperature	3	3	3	3	3	3	3	3	3	3
ii.	salinity	3	3	3	3	3	3	3	3	3	3
b.	Type of organisms: autotrophic, mixotrophic, heterotrophic	3	3	3	3	3	3	3	3	3	3
c.	Range (high/low number of organisms)	3	3	3	3	3	3	3	3	3	3
d.	TSS/DOC/POC levels	2	2	2	2	2	2	2	2	2	2
e.	Transport/handling	2	2	2	2	2	2	2	2	2	3
5.	**Practicality**										
a.	complex/simple (many/few operations)	3	3	1	1	3	3	3	2	3	3
b.	availability	3	3	3	1	3	3	3	2	3	3
c.	cost (high or low)	3	3	1	3	3	3	3	3	3	3
d.	time to result	3	3	1	3	3	3	3	1	3	3
e.	does the analysis require highly skilled analysts	1	1	1	1	1	1	2	1	1	1
	Total score	47	47	43	46	47	47	46	44	47	48

Score (0=ubrugelig, 1 dårlig, 2 middel eller irrelevant, 3 god, irrelev		Tamburri 2009	Tamburri 2009	McCollins et al 20	McCollins et al 20	Carney and Mes	Carney and Mesb	ETV 2009	DHI 2009	DHI 2009	GSI 2009	NIOZ	NIOZ	NIOZ
ogy	Endpoint	CMFDA+FDA	Chlorphyll	No of celles and n	Chlorphyll	FDA FLOWCAM	FDA EPIfloroscen	FDA/CMFDA stain	FDA/CMFDA stain	Photosynthetic acti	FDA/CMFDA stain	Flow cytometry	PAM-fluorom	Morphological
1.	**Does the analysis have a sufficient limit of detection (LoD) and limit of quantification (LoQ)**	3	1	3	1	3	3	3	3	1	3	1	1	3
2.	**Interference**													
a.	Does the analysis distinguish between size of the organisms?	3	1	3	1	3	3	3	3	1	3	2	1	3
b.	Does the method distinguish correctly between viable and non-viable, i.e. is the death criterion acceptable when testing ballast water	3	1	1	1	3	3	3	3	3	3	2	1	1
c.	Are there other interferences?	2	2	2	2	3	3	2	2	2	2	2	2	2
3.	**Is the accuracy acceptable**													
a.	Does sample pretreatment give rise to bias	1	1	1	1	2	1	1	3	3	1	3	3	1
b.	Does the analyst give rise to bias	1	1	1	1	3	3	3	3	3	3	3	1	1
c.	Precision?	2	3	2	3	3	3	1	3	3	1	2	2	2
d.	Information on reproducibility and repeatability	2	3	2	3	2	2	2	2	3	2	2	2	2
4.	**Is the analysis robust?**													
a.	Geographical variations	3	3	3	3	3	3	3	3	3	3	3	3	3
i.	temperature	3	3	3	3	3	3	3	3	3	3	3	3	3
ii.	salinity	3	3	3	3	3	3	3	3	3	3	3	3	3
b.	Type of organisms: autotrophic, mixotrophic, heterotrophic	3	1	3	1	3	3	3	3	3	3	3	3	3
c.	Range (high/low number of organisms)	3	3	3	3	3	3	3	3	3	3	3	3	3
d.	TSS/DOC/POC levels	2	3	2	3	2	2	2	2	3	2	2	1	2
e.	Transport/handling	2	3	3	3	3	3	3	3	3	3	3	3	2
5.	**Practicality**													
a.	complex/simple (many/few operations)	2	3	3	3	1	1	2	2	3	2	2	3	3
b.	availability	1	3	3	3	1	2	3	3	1	3	1	1	3
c.	cost (high or low)	3	3	3	3	1	2	3	3	3	3	1	3	3
d.	time to result	3	3	3	3	3	3	3	3	3	3	3	3	3
e.	does the analysis require highly skilled analysts	1	3	1	3	1	1	1	1	1	1	1	1	1
	Total score	46	47	48	47	49	50	50	54	51	50	45	43	47

A P P E N D I X C

Examples of statistical analyses

The statistical tests should be carried out by trained statisticians to avoid erroneous conclusions.

It is recommended to use a statistical software package like Minitab (www.minitab.com) or R (http://www.r-project.org/).

The significance level, α, should be selected and the acceptable p-value defined before the tests are carried out.

(1 – α) is the confidence coefficient or level of confidence whereas the p-value determines the appropriateness of rejecting the null hypothesis in a hypothesis test. P-values range from 0 to 1. The smaller the p-value, the smaller the probability that the rejection of the null hypothesis is a mistake. Before conducting any analyses, determine your α level. A commonly used value is 0.05. If the p-value of a test statistic is less than your alpha, you reject the null hypothesis.

Because of their indispensable role in hypothesis testing, p-values are used in many areas of statistics including basic statistics, linear models, reliability, and multivariate analysis among many others. The key is to understand what the null and alternate hypotheses represent in each test and then use the p-value to aid in your decision to reject the null.

The smaller the p-value, the stronger the evidence against the null hypothesis and in favour of the alternative hypothesis.

Table C.1 Example data. Concentrations of zooplankton larger than 50 µm in 9 samples of 1 m^3 taken on 6 "sample events" (no/m^3)

	1	2	3	4	5	6
1 sample	14	0	0	10	1	9
2 sample	11	0	8	4	1	1
3 sample	4	1	2	3	3	1
4 sample	8	0	3	2	1	0
5 sample	6	1	2	10	1	2
6 sample	5	4	2	11	3	2
7 sample	3	4	0	12	17	3
8 sample	0	6	3	20	14	0
9 sample	0	4	0	10	15	1

1 sample t- test "Average approach":
Using the t-test at an α of 0.05 with the following hypothesis:

Null hypothesis: mean = 10.00 versus the alternative hypothesis: mean < 10.00

gives the following results:

	N	Mean	Standard deviation	95% upper bound	P-value
Event 1	9	5.67	4.72	8.59	0.000
Event 2	9	2.222	2.279	3.635	0.000
Event 3	9	2.222	2.489	3.756	0.000
Event 4	9	9.11	5.56	12.55	**0.322**
Event 5	9	6.22	6.92	10.51	**0.070**
Event 6	9	2.111	2.759	3.821	0.000

It is seen that the event samples 1, 2, 3 & 6 comply since the p-value is significantly below 0.05. However, for event 4 & 5, with a p-value of 0.322 and 0.07, the null hypothesis cannot be rejected and therefore these events do not comply with the D2 regulations and fails.

1 sample Wilcoxon test "Average approach":

Using the Wilcoxon test at an α of 0.05 with the following hypothesis:

Null hypothesis: median = 10.00 versus the alternative hypothesis: median < 10.00,

gives the following results:

	N	No for test	Wilcoxon statistic	Estimated median	P-value
Event 1	9	9	4.5	5.500	0.019
Event 2	9	9	0.0	2.000	0.005
Event 3	9	9	0.0	2.000	0.005
Event 4	9	6	9.0	10.00	**0.417**
Event 5	9	9	7.0	7.500	0.038
Event 6	9	9	0.0	1.500	0.005

It is seen that the event samples 1, 2, 3, 5 & 6 comply since the p-value is significantly below 0.05. However, for event 4, with a p-value of 0.417, the null hypothesis cannot be rejected and therefore this event does not comply with the D2 regulations and fails.

1-sample Poisson rate test "Average approach"

Using the 1-sample Poisson rate test at an α of 0.05 and applying the same hypothesis gives the following results:

	N	Total occurrences	Rate of occurrence	95% upper bound	P-value
Event 1	9	51	5.66667	7.1558	0.000
Event 2	9	20	2.22222	3.2291	0.000
Event 3	9	20	2.22222	3.2291	0.000
Event 4	9	82	9.11111	10.9480	**0.216**
Event 5	9	56	6.22222	7.7734	0.000
Event 6	9	19	2.11111	3.0977	0.000

It is seen that the event samples 1, 2, 3, 5 & 6 comply since the p-value is significantly below 0.05. However, for event 4, with a p-value of 0.216, the null hypothesis cannot be rejected and therefore this event does not comply with the D2 regulations and fails.

1-sample Poisson rate test "Instantaneous approach"

To illustrate as the 1-sample Poisson rate test we analyze the 9 single results of the event 4.

Again, we choose the same α level and hypothesis:

	Measurement	Rate of occurrence	95% upper bound	P-value
1	10	10	16.9622	**0.583**
2	4	4	9.15352	0.029
3	3	3	7.7536	0.010
4	2	2	6.29579	0.003
5	10	10	16.9622	**0.583**
6	11	11	18.2075	**0.697**
7	12	12	19.442	**0.792**
8	20	20	29.0620	**0.998**
9	10	10	16.9622	**0.583**

It is seen that the instantaneous data 2, 3 & 4 comply as the p-value is significantly below 0.05. However, for the data 1, 5-9, with a p-value of higher than 0.5, the null hypothesis cannot be rejected and therefore these instantaneous data do not comply with the D2 regulations and fail.

Test of normality of the of the 6 event data:
Using the Anderson Darling test at an α of 0.05 with the following hypothesis:

Null hypothesis: data are following a normal distribution versus the alternative hypothesis: data are not normal distributed

gives the following results:

	Anderson Darling value	P-Value
Event 1	0.22	0.776
Event 2	0.68	***0.048***
Event 3	0.75	**0.031**
Event 4	0.49	0.167
Event 5	1.15	**<0.005**
Event 6	1.10	**<0.005**

The test results clearly indicate for event 3, 5 & 6 that these data are **not** following a normal distribution (the p-value is significantly lower than 0.05, therefore the null hypothesis has to be rejected), whereas the data of the events 1 and 4 are following a normal distribution. The data set of the event 2 are just at the border with a p-value of 0.048, here additional testing would be required.

Since the event 5 is clearly not normal-distributed, the t-test as shown above is giving a wrong result compared to the Wilcoxon or Poisson test.