ICES AD HOC (WGBOSV) REPORT 2010

ICES ADVISORY COMMITTEE

ICES CM 2010/ACOM:65

Harmful Aquatic Organisms in Ballast Water - Overview of statistical methods that could be used to verify compliance with the D-2 Standard

By

ICES/IOC/IMO Working Group on Ballast and Other Ship Vectors (WGBOSV)

ICES
International Council for the Exploration of the Sea
CIEM
Conseil International pour l'Exploration de la Mer

International Council for the Exploration of the Sea
Conseil International pour l'Exploration de la Mer

H. C. Andersens Boulevard 44–46
DK-1553 Copenhagen V
Denmark
Telephone (+45) 33 38 67 00
Telefax (+45) 33 93 42 15
www.ices.dk
info@ices.dk

Recommended format for purposes of citation:

ICES. 2010. Harmful Aquatic Organisms in Ballast Water - Overview of statistical methods that could be used to verify compliance with the D-2 Standard. By ICES/IOC/IMO Working Group on Ballast and Other Ship Vectors (WGBOSV).. 7 pp.
https://doi.org/10.17895/ices.pub.19280759

Contents

Executive Summary

This document provides information on statistical methods and background information that could be used to verify compliance with the D-2 standard. The need for further data to verify the suggested approaches is highlighted

1 Introduction

The International Council for the Exploration of the Sea (ICES) ICES/IOC/IMO Working Group on Ballast and Other Ship Vectors (WGBOSV) met from 8–10 February, 2010 to discuss ongoing issues related to vectors of introduction of non native species. One Term of Reference was to "discuss and evaluate the sampling strategies under consideration at IMO and provide comment with the aim to prepare a written submission to relevant IMO committees in response to requested information. Statistical experts will be invited to attend and contribute at the meeting".

The WGBOSV reviewed several sampling approaches (including Pazouki *et al.*, 2009 and Miller *et al.*, submitted) and discussed the statistical theory behind obtaining samples representative of the whole discharge or portion of the discharge for each of the methods and the effect of sample size on the ability to detect non-compliance. The group continued this work intersessionally after the meeting and this document provides an overview of the results of this work.

The WGBOSV provides scientific opinion about three approaches to sampling ballast water for assessing compliance with the larger size category within D-2 (less than 10 viable organisms per $m^3 \geq 50$ µm in minimum dimension). These are flow integrated samples from the ballast discharge collected in three ways:

- One continuous sample collected over the entire discharge time;
- Multiple sequential (but non continuous) samples collected over the entire discharge time;
- A single sample collected over a portion of the discharge time

The continuous sample is collected throughout the entire discharge of a ballast tank or ship. Multiple sequential or single samples are where samples of known time duration are taken (e.g., 5, 10, 15 minutes). The timing of these samples should be determined by a recognised sampling protocol such as simple random sampling. Importantly, all sampling methods require that the instantaneous flow into the sample should be a constant proportion of the instantaneous flow in the ships' discharge line using an in-line sampling port following the G2 guidelines (IMO, 2008).

Basic assumptions

Based upon Miller *et al.*, the basic assumptions made when undertaking this work were that;

- all live and only live organisms within the sample were detected and counted;
- the entire sample is processed;
- an unbiased sampling method was used;
- the total discharge volume is large compared to the sample volume;
- interest lies in testing whether the mean density of organisms in the total discharge exceeds some threshold;
- only sampling error is considered; no observer error, for example;
- only the larger size category (≥ 50 µm) within D-2 was considered;

- sampling one organism does not affect the chance that another is sampled, e.g., organisms do not form chains or hook together.

It is acknowledged that these assumptions may need to be re-assessed once more data become available.

From the final assumption it follows that the discharge process can be treated as a non-homogeneous (or time varying) Poisson process (Cox and Isham, 1980).

Representativeness

The term representative has a variety of interpretations (Kruskal and Mosteller, 1979a-c; 1980) including unbiasedness and good coverage. Some of the confusion about the performance of different testing procedures is due to ambiguity in the definition of representativeness. As long as the sample comes from a ballast pipe with a sample port that collects a sample representative of the water flowing through the pipe, a representative sample can be collected.

Sample size

Sample size requirements depend on the confidence level and power required for the purpose i.e., to prove that a vessel is in breach of the D-2 standard. This requires a high confidence, e.g., 95%, 99% or 99.9%, so that with high probability the vessel is in breach of the standard. If it is additionally required that ballast discharges with densities near to the standard are detected i.e., densities of 50, 20 or even 12 organisms per m³, then a high power is required to distinguish relatively low densities from the D-2 standard. These two issues are related. For a fixed sample size based on a random sample (see below), increasing the confidence level will mean higher sample counts will be required. To ensure that lower densities are detected when a continuous sample or multiple sequential samples are collected, the sample size i.e., the volume sampled, must be increased if the confidence limits and volume analyzed remain constant. The level of confidence and what density should be detectable are important considerations. Figure 1 in the appendix illustrates the relationship between the level of confidence, power, organism density and volume sampled.

One continuous sample collected over the entire discharge time

In this case, one sample is taken continuously throughout the entire discharge process of a ballast tank and would be collected and processed as the discharge was taking place. Owing to the assumption outlined in paragraph 5, sampling through time gives sample counts that are a sample from a Poisson distribution (Cox and Isham, 1980) with mean equal to the mean of the entire discharge.

This method is described and discussed in Miller et al. (submitted), which attempts to balance the analytical rigor and logistical reality in identifying what volumes of water can be reasonably sampled to yield accurate live counts of organisms ≥ 50 µm. They apply statistical modelling to a range of sample volumes and plankton concentrations and to calculate the statistical power of each combination to differentiate various zooplankton concentrations from the discharge standard of < 10 zooplankton per m³. Assuming that sampling and analytical errors or biases are negligible, the results demonstrate that continuous, time integrated sampling provides statistical confidence with manageable sample volumes. For example, taking a 7 m³ continuous sample from the discharge and enumerating all viable organisms the concentrated sample would allow densities of 14 organisms per m³ to be distinguished from 10 organisms per m^3 with a power of 80%. Further details are contained within the Miller et al. paper.

Multiple sequential samples collected over the entire discharge time

In this case, samples are taken at different points throughout the discharge process, but are not collected continuously, resulting in a number of observed counts and associated sample volumes. By sampling the discharge through time, sample counts have a lower variability than isolated grab samples, but more variability than the one continuous sample discussed above (since the entire discharge is not covered).

For the same reason a continuous sample is Poisson distributed, each individual multiple sequential sample will have a Poisson distribution with different mean. If these means follow a gamma distribution, then the counts from multiple sequential samples will have a negative binomial distribution (Evans *et al.*, 2000). This allows the calculation of appropriate limits for multiple sequential sampling. The variability of the means or the between sample variability will depend on the proportion of the ballast discharge sampled in each multiple sample: the greater the proportion, the more similar the sample means will be.

Three sampling approaches for collecting multiple sequential samples are:

- simple random sampling.
- stratified random sampling.
- systematic sampling.

For samples that cover a small proportion of the total discharge, stratified random sampling will increase sample coverage e.g., by sampling randomly from the beginning, middle and end of the discharge. Systematic sampling, where (from a random starting position) samples are taken every hour, might be a suitable practical approach.

The between sample variability may not be estimated well based on three 15-minute samples, for example. A solution is to use a data set collected using the same sampling protocol to estimate a value for the between sample variability, which could be an upper bound on the expected variability. The between sample variability depends on the proportion of the ballast that is being sampled, and the length of time over which sampling takes place, and these aspects would have to be standardised as much as possible to obtain consistent and comparable estimates of the between sample variability.

Overall, this approach may be appropriate when the variability of the organisms in treated ballast water can be estimated based on suitable data (see paragraph 21) from ship-based tests, possibly augmented with land-based tests. Once a suitable estimate of variability is available then this approach may allow the confidence levels associated with the chosen volume of sample to be calculated and to determine the mean number of organisms that would indicate that D-2 has been exceeded. Initial calculations based on untreated water were carried out but without access to a large dataset with the numbers of organisms found in treated water the WGBOSV was unable to take this any further.

Single sample collected over a portion of the discharge time

This approach would be used to determine whether a given portion of water is not in compliance with the D-2 discharge standard. As with the other approaches, the sample collected would be representative of the water flowing through the discharge pipe. The data would be analyzed as in the one continuous sample (i.e., following a Poisson distribution).

Discussion

Initial attempts to verify the applicability of these methods has been hampered by the lack of available, peer-reviewed data on which to base the calculations of variability. In order to verify these methods data on the number of organisms present in treated water is required. Although these data will produce a lot of zeros (if the BWMS is working) they can be used for verification of the proposed methods. Availability of data was particularly a problem for the multiple sequential samples and an accurate estimation of between sample variability was not possible.

Concluding remarks

The scientific opinion of the WGBOSV is that one continuous sample yields results for measuring the density of organisms that are representative of the entire ballast water tank and are able to determine non-compliance at a lower threshold than multiple samples. However, this may not always be practical.

Multiple sequential samples can be applied to determine non-compliance if the sample variability is known. Uncertainty about the sample variability may be compensated for by adding a safety factor to the variability.

Collecting a single sample over a portion of the discharge time would allow a port state to determine if part of the discharge water was in non compliance.

The level of assumed between sample variability is crucial with multiple sequential samples. In the long term this could be estimated from a data set that should have been collected following a known sampling protocol, so that routine ballast water samples can contribute to this data set. In the meantime, a higher factor of variability may be assumed to compensate for existing uncertainty. Such a data set should cover the range of conditions thought to affect the variability of organism density in ballast discharge. These include:

- Temperature,
- Salinity,
- Geographical source of ballast,
- Residence time in tanks,
- Density of organisms in treated water,
- Sample duration,
- Vessel type.

The level of assumed variability should be reassessed periodically as more data becomes available.

References

Cox, D.R., and Isham, V. 1980. Point Processes. Monographs in Probability and Statistics. Chapman and Hall, London.

Evans, M., Hastings, N. and Peacock, B. 2000. Statistical Distributions: Third Edition. Wiley Series in Probability and Statistics. John Wiley and Sons, New York.

International Maritime Organization 2008. Resolution MEPC.173 (58) 'Guidelines for Ballast Water Sampling (G2).

Kruskal, W. Mosteller, F. 1979a. Representative Sampling, I: Non-Scientific Literature. International Statistical Review. 47(1): 13–24.

Kruskal, W. Mosteller, F. 1979b. Representative Sampling, II: Scientific Literature, Excluding Statistics. International Statistical Review. 47(2): 111–127.

Kruskal, W. Mosteller, F. 1979c. Representative Sampling, III: The Current Statistical Literature. International Statistical Review. 47(3): 245–265.

Kruskal, W. Mosteller, F. 1980. Representative Sampling, IV: The History of the Concept in Statistics. International Statistical Review. 48(2): 169–195.

Miller, W., Frazier, M. Smith, G. Perry, E. Ruiz, G. and Tamburri, M. (submitted). Enumerating Sparse Organisms in Ships' Ballast Water: Why Counting to 10 is not so Easy. Environmental Science and Technology.

Pazouki, K. Cabezas-Basurko, O. Mesbahi, E. Kemp, J. 2009. Part B: Ballast Water Sampling: Methods, Analysis, Representativeness and Legal Issues. Newcastle University. Project No. RP577.

Appendix

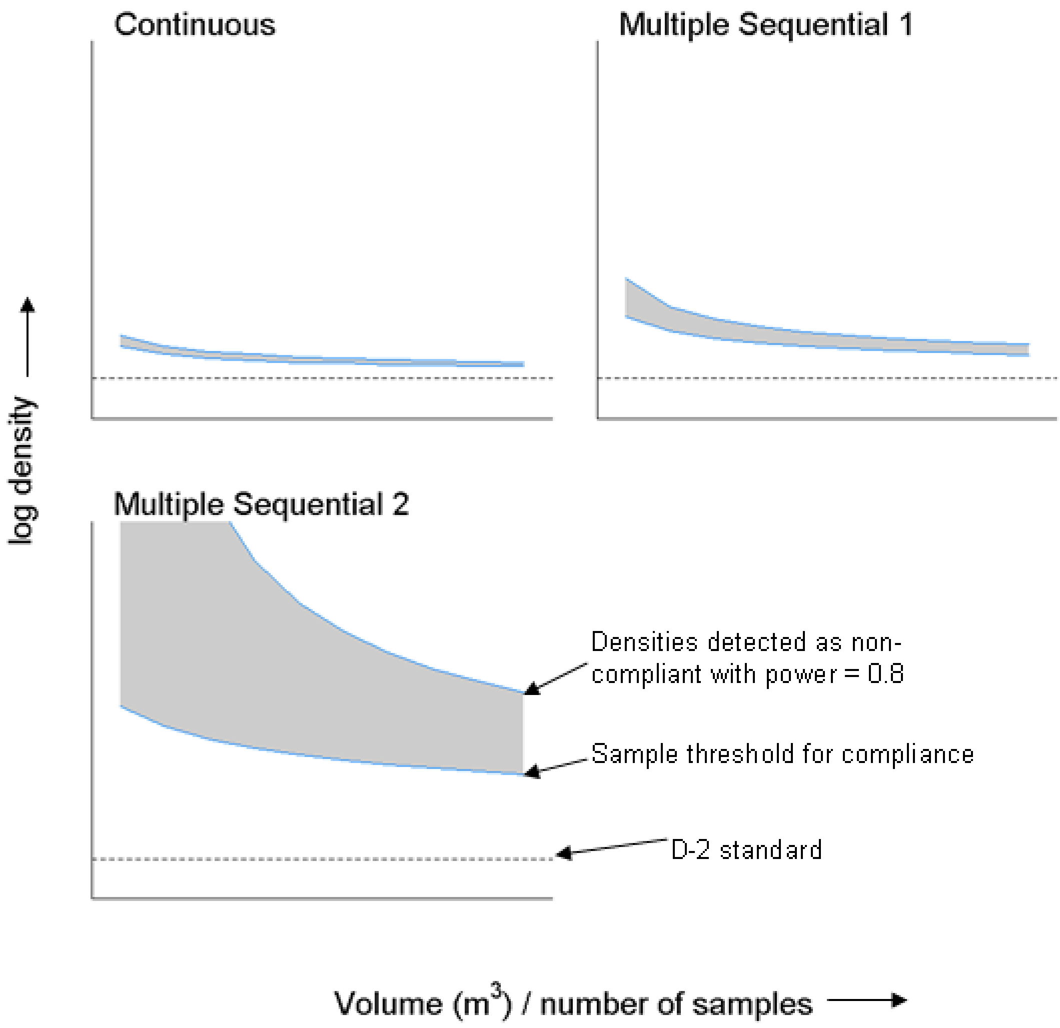

Figure 1

An illustration of the efficiency of two proposed sampling schemes. The figures for the multiple samples illustrate the effect of differences in sample variability with sample 1 having a low variability and sample 2 having a high variability. The D-2 standard is shown as a dotted line, the first line above this gives the sample threshold for compliance, if counts from a sample taken from the discharge are found to be above this line they will be non-compliant with 99% confidence. The next line relates to the density of organisms in ballast tanks, if organisms densities within the tank are above this line then non compliant counts in the samples will be detected with 80% power at 99% confidence, densities below this line are increasingly likely to give counts below the sample threshold.