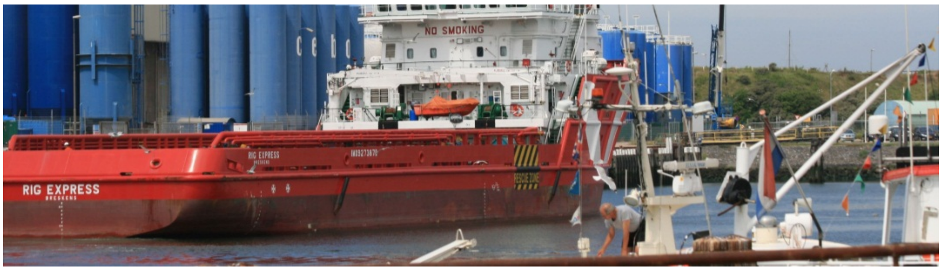
The development of a full standard methodology for testing ballast water discharges for gross non-compliance of the IMO's Ballast Water Management Convention (EMSA/NEG/12/2012)

S.M. Bierman, P. de Vries and N.H.B.M. Kaag

Report number C124/12



IMARES Wageningen UR

Institute for Marine Resources & Ecosystem Studies

Client:	European Maritime Safety Agency NEG122012@emsa.europe.eu Cais do Sodré 1249-206 LISBOA Portugal

Publication date: 30 October 2012

IMARES is:

- an independent, objective and authoritative institute that provides knowledge necessary for an integrated sustainable protection, exploitation and spatial use of the sea and coastal zones;
- a key, proactive player in national and international marine networks (including ICES and EFARO).

P.O. Box 68
1970 AB IJmuiden
Phone: +31 (0)317 48 09 00
Fax: +31 (0)317 48 73 26
E-Mail: imares@wur.nl
www.imares.wur.nl

P.O. Box 77
4400 AB Yerseke
Phone: +31 (0)317 48 09 00
Fax: +31 (0)317 48 73 59
E-Mail: imares@wur.nl
www.imares.wur.nl

P.O. Box 57
1780 AB Den Helder
Phone: +31 (0)317 48 09 00
Fax: +31 (0)223 63 06 87
E-Mail: imares@wur.nl
www.imares.wur.nl

P.O. Box 167
1790 AD Den Burg Texel
Phone: +31 (0)317 48 09 00
Fax: +31 (0)317 48 73 62
E-Mail: imares@wur.nl
www.imares.wur.nl

IMARES, institute of Stichting DLO
is registered in the Dutch trade
record nr. 09098104,
BTW nr. NL 806511618

The Management of IMARES is not responsible for resulting damage, as well as for damage resulting from the application of results or research obtained by IMARES, its clients or any claims related to the application of information found within its research. This report has been made on the request of the client and is wholly the client's property. This report may not be reproduced and/or published partially or in its entirety without the express written consent of the client.

A_4_3_2-V12.4

Executive summary

What is Gross Non-Compliance?

The International Convention for the Control and Management of Ship's Ballast Water and Sediments, as adopted by the International Maritime Organization (IMO), is targeted at reducing the risk of the introduction of invasive species and pathogens through ballast water discharges. To this end, the D-2 regulation of the convention specifies water quality requirements, in the form of upper limits on the concentrations of living organisms in ballast water discharges, to which all ships have to comply with.

In practical situations it may be difficult to determine the organism concentrations in discharged ballast water with high accuracy. Taking into account known variability in organism counts from samples of ballast water discharges taken during a set sampling protocol means that a 'Gross Non-Compliance (GNC)' threshold can be derived. If this is then used during Port State Control, Administration can be 99.9% sure that a vessel is in non-compliance with the D-2 standards in the Ballast Water Management (BWM) Convention when the GNC threshold is exceeded. This concept relies on the fact that the ballast water system will either work to the low D-2 standard, producing a compliant discharge, or not, producing a discharge that is equivalent to seawater - 10, 100 or 1000 times the D-2 Standard. As these non-compliant discharges provide the greatest risk, then, in the absence of practical protocols that can accurately and economically test to the D-2 standard, a gross non-compliance test is the next best option. This is not weakening the BWM Convention, it provides a practical interim test that can be used until sampling methodologies can be refined to account for the D-2 Standard.

Aim

To develop a full standard methodology for testing for gross non-compliance, complete with sampling protocol, analysis methodology and confidence limits, based on the existing EMSA research, so that when used by Port State Control, an Administration can be 99.9% sure that a vessel in in non-compliance with the standards in the BWM Convention.

Which approach has been used?

A good methodology for sampling a ballast water discharge would be to specify the procedure and minimum requirements that, if adhered to, would lead to a predictable level of reliability of the estimate of the mean concentration (on the basis of counts in the samples) of living organisms in the ballast water. The level of reliability for such a well-defined sampling procedure can be used to define a Gross Non-Compliance threshold (GNC threshold), such that if a mean concentration, as estimated from a sample, is above this threshold it can be taken as evidence of gross non-compliance. The GNC threshold is set at a concentration above the D-2 standard. The distance between the D-2 standard and the GNC threshold can be thought of as a 'guard band' to guard against false positives due to the possible imprecisions and biases in the sampling and analysis methodologies. A false positive would be a situation where a vessel which is in truth compliant is incorrectly classified as non-compliant due to imprecise measurements. In statistical language this is called a 'type I error'. The opposite of

this, where a vessel which is in truth non-compliant but incorrectly classified as compliant due to imprecise measurements is called a ‘type II error’
We note that by using this approach for testing for non-compliance, an implicit trade-off between type I and type II errors is made: the lower the pre-specified chance of incorrectly classifying a vessel which is in truth compliant, as non-compliant (“compliant but fined”; type I error) - the higher the chance of incorrectly classifying a vessel which is in truth non-compliant as compliant (“non-compliant but not fined”; type II error).

The width of the guard band is determined by the reliability of the estimate and by the desired level of protection against false positives. The width of the guard band was estimated from data on untreated ballast water from Gollasch and David (2010) (as the data availability for treated water was limited and had (near) zero counts of organisms). Therefore, it was assumed, that data for treated ballast water will show a similar, or lower variance. Similar variances are used in the calculations. Using lower variances (if future data sets support such), will result in a smaller guard band. The desired rate of false positives does not have a scientific basis, but needs to be set by Administrations. For example, in case the guard band is chosen in accordance with a type I error of $\alpha = 0.1\%$, an administration can be $100 - \alpha = 99.9\%$ sure that the observed counts in the samples could not have arisen by chance-like processes if the vessel is in truth compliance with the standards in the BWM Convention.

Estimated Gross Non-Compliance thresholds

The variability in counts of living organisms in samples from ballast water discharges is quantified using existing data from Gollasch and David (2010) on untreated ballast water discharges (as the data availability for treated water was limited and had (near) zero counts of organisms). Jørgensen et al. (2010) and Miller et al. (2011) proposed to use the Poisson distribution to quantify variability in counts of living organisms in samples taken from ballast water discharge. The Poisson distribution assumes that organisms are randomly distributed during the ballast water discharge, and that no errors are made in measurements of sampling volumes. The data from Gollasch and David (2010) show that the variability in counts of living organisms in samples is larger than can be explained by the chance-like processes of the Poisson distribution. A refined statistical model (Negative Binomial) has therefore been used and calibrated, based on the additional variability (over and above that explained by the Poisson distribution) that is observed in the data. The expected variability of counts from samples with concentrations of living organisms equal to the D-2 standard is estimated by extrapolation, using a scenario where the variability in counts is a multiple of the expected count in the samples. By recognising this greater variability, this method will provide gross non-compliance thresholds above that of those expected using the Poisson Distribution Model.

Using the estimated expected variability in counts in samples from ballast water discharges with concentrations equal to the D-2 standard, GNC threshold values can be calculated for specific sampling strategies. The GNC threshold indicates that a count higher than or equal to the threshold is highly unlikely (less than 0.1%) to occur by chance-like processes if the true concentration in the ballast water is equal to, or lower than, the D-2 standard. When a single subsample is analysed for a multitude of main samples (taken during discharge). The figures in Table 1 and Table 2 provide GNC thresholds for increasing numbers of main samples.

Table 1 and Table 2 only apply to a specific sampling scheme. For organisms ≥10 and <50 µm, the tables only apply when a 0.810 mL subsample is taken and analysed from one or more main samples. For organisms ≥50 µm the tables only apply when a 6 mL subsample is taken and analysed from 100 mL concentrate of one or more 500 L main samples.

Table 1 Gross non-compliance threshold values (α=0.1%) for direct counts in a subsample, estimated by the Negative Binomial model under the assumption that the true concentration in the entire discharge is equal to the D-2 standard (10 organisms per cm^3 or m^3).

Number of main samples	Number of subsamples	≥10 and <50 µm		≥50 µm	
		Sampled volume (cm^3)	Gross Non-Compliance threshold	Volume that samples represent (dm^3)	Gross Non-Compliance threshold
1	1	0.81	94	0.06*500	11
2	1	1.62	119	0.06*1000	13
3	1	2.43	139	0.06*1500	14
4	1	3.24	158	0.06*2000	15
5	1	4.05	175	0.06*2500	17

Table 2 Gross non-compliance threshold values (α=0.1%) for raised concentrations, expressed as concentrations per cm^3 or m^3. Results derived from the counts as presented in Table 5, with counts multiplied by the raising factors R.

Number of main samples	Number of subsamples	≥10 and <50 µm		≥50 µm	
		Sampled volume (cm^3)	Gross Non-Compliance threshold (#/cm^3)	Volume that samples represent (dm^3)	Gross Non-Compliance threshold (#/m^3)
1	1	0.81	116.0	0.06*500	366.7
2	1	1.62	73.5	0.06*1000	216.7
3	1	2.43	57.2	0.06*1500	155.6
4	1	3.24	48.8	0.06*2000	125.0
5	1	4.05	43.2	0.06*2500	113.3

We note that the derived GNC thresholds are a function of the sampling volume and sampling strategy, since the reliability of estimates of concentrations of living organisms will increase with increasing sampling volumes. Which sampling strategy and sampling effort should be selected is an Administration's decision. This decision may be based upon the height of the GNC value that is being considered appropriate/acceptable with respect to the D-2 standard, as well as the costs and practicability associated with particular sampling strategies. In other cases, sampling strategy may be determined by what is possible at a certain vessel at a certain moment. In the accompanying concept protocols, taking two main samples was taken as starting point as recommended in Gollasch, S. and David, M. (2010), but threshold values are given for 1 to 5 main samples.

Under which assumptions are the derived Gross Non-Compliance thresholds valid?

The derived GNC thresholds apply only if certain assumptions are valid. The overall assumption is that the random errors that are observed in the data of Gollasch and David (2010) using untreated ballast water, and the NB model that we use to describe it, are representative for the random errors that will be made by Port State Control (PSC) at inspections of treated ballast water. This assumption can be refined and split into separate specific assumption. These specific assumptions are discussed in more detail in appendix G.

Availability of data will aid the refinement of these thresholds over time. The understanding of variation and sampling/analysis error with respect to representativeness can be improved by including more data. This should preferably be data on treated ballast water and/or water with low (near D-2 standard) organism concentrations, which would help to get a more accurate estimate of over-dispersion and possible trends in time and of variance of organism concentrations. This may result in lower gross non-compliance thresholds.

In addition representativeness is also determined by procedural aspects. In other words PSC should use similar procedures for sampling and analysis as used by Gollasch and David (2010), as the GNC thresholds are based on those procedures. For this purpose protocols for sampling and analysis are proposed in appendices H. Also to make sure that quality of the sampling and analysis does not adversely affect the variance of PSC inspection, quality assurance should be integral part of the sampling protocol and analysis procedure.

An additional assumption that has to be made when applying a GNC threshold is that systematic errors in observed organism concentrations result in an underestimation of true concentrations. This error is difficult to quantify as true concentrations are usually unknown. This error can partly be reduced by assuring quality of, and providing clear protocols for sampling and analysis.

Conclusion

Gross Non-Compliance thresholds have been derived using the Negative Binomial distribution which has been parameterised using data of counts of living organisms in samples of untreated ballast water collected by Gollasch and David (2010).

When using the procedures described, counts of living organisms at or above the derived thresholds are observed, it can be stated with 99.9% certainty that the water is in gross non-compliance.

The derived GNC thresholds only apply if certain assumptions are valid. The validity of some assumptions need to be further assessed and discussed in the wider (scientific) community.

Contents

Abbreviations, acronyms and symbols

c_{raised}	organism counts per volume unit in uptake or discharged water, raised from raw counts
c_{raw}	raw counts of organisms as determined in a subsample
D-2	regulation of the BWM Convention in which water quality requirements are specified
EMSA	European Maritime Safety Agency
G2	IMO’s guidelines on sampling of ballast water
G8	IMO’s guidelines for approval of ballast water management systems
GNC	Gross Non-Compliance
IMARES	Institute for Marine Resources and Ecosystem Services
IMO	International Maritime Organization
NB	Negative Binomial model
OET	sample taken ‘Over Entire Time’ of a discharge event, opposed to discrete samples labelled S1, S2 and S3
PSC	Port State Control
R	Raising factor used to convert raw counts (c_{raw}) in a subsample to counts per volume unit in the actual ballast water (c_{raised})
V_{conc}	Volume to which the main sample is concentrated
V_{samp}	Volume of the main sample taken during uptake or discharge
$V_{subsamp}$	Volume of the subsample take from either the main sample (V_{samp}) or the concentrate (V_{conc})

1 Introduction

One of the anthropogenic threats to maritime ecosystems worldwide is the introduction of invasive marine species into new environments through the discharge of the ballast water from vessels. The expanded volume of shipping traffic over the last decades has increased this pressure. Worldwide it is estimated that 3 to 5 billion tonnes of ballast water is discharged annually (Globallast, http://globallast.imo.org/), during which species can be transferred that may prove harmful to the recipient ecosystem.

The International Convention for the Control and Management of Ship's Ballast Water and Sediments, as adopted by the International Maritime Organization (IMO) in 2004, is targeted at reducing the risk of the introduction of invasive species and pathogens through ballast water discharges. To this end, regulation D-2 of the Convention specifies water quality requirements for ballast water which all ships have to comply with over time, once the Convention enters into force. These requirements are: less than 10 viable organisms per cubic metre greater than or equal to 50 micrometres in minimum dimension and less than 10 viable organisms per millilitre less than 50 micrometres in minimum dimension and greater than or equal to 10 micrometres in minimum dimension.

Article 9 of the Convention lays down the legal basis for inspection and sampling of the ballast water to collect data to determine whether a ship is in compliance with the Convention. The IMO's "Guidelines on Sampling of Ballast Water (G2)" introduces the concept that the sampling has to be representative of the entire discharge. This principle brings many statistical and practical issues into the development of an appropriate sampling protocol, resulting in sampling protocols that are prohibitively expensive, onerous and may cause undue delay to the vessel. Therefore, the concept of gross non-compliance has been raised and agreed within the IMO. EMSA Projects NEG/09/2010 (Gollasch & David, 2010) and NEG/10/2010 (Jørgensen *et al.*, 2010) identified a methodology based on testing for gross non-compliance of a system.

This concept relies on the fact that the ballast water system will either work to the low D-2 standard, producing a compliant discharge, or not, producing a discharge that is equivalent to 10, 100 or 1000 times the D-2 Standard. As these non-compliant discharges provide the greatest risk, then, in the absence of practical protocols that can accurately and economically test to the D-2 standard, a gross non-compliance test is the next best option. This is not weakening the Ballast Water Management (BWM) Convention, it provides a practical interim test that can be used until sampling methodologies can be refined to account for the D-2 Standard.

EMSA's research project "Testing sample representativeness of a ballast water discharge and developing methods for indicative analysis" (Gollasch & David, 2010), has provided a methodology for taking samples and analyse these using existing methodologies used for type approval. The methodology was, however, developed using untreated ballast water with clearly 'non-compliant' numbers of organisms.

EMSA's research project "The development of guidance on how to analyse a ballast water sample" (Jørgensen *et al.*, 2010) identified the 1-sample Poisson rate test as the most suitable statistical method to establish confidence limits when testing for gross non-compliance, using surrogate data.

2 Aim of the project

To develop a full standard methodology for testing for gross non-compliance, complete with sampling protocols, analysis methodologies and confidence limits, based on the existing EMSA research (i.e., data produced by Gollasch and David (2010)), so that when used by Port State Control, an Administration can be 99.9% sure that a vessel is in non-compliance with the standards in the BWM Convention when thresholds are exceeded.

3 Proposed general approach to testing for gross non-compliance

Our proposed approach to testing for gross non-compliance is to assume that a vessel is in compliance until it can be demonstrated, on the basis of data from samples of the ballast water, that the vessel is in non-compliance with a high level of certainty (99.9%).

A good sampling methodology would specify the procedure and minimum requirements that, if adhered to, would lead to a predictable level of reliability of the estimate of the mean concentration of living organisms in the ballast water. The level of reliability for such a well-defined sampling procedure can be used to define a Gross Non-Compliance threshold (GNC threshold) such that if a mean concentration, as estimated from a sample, is above this threshold, it can be taken as evidence of gross non-compliance (Figure 1). The compliance threshold is set at a concentration above the D-2 standard. The distance between the D-2 standard and the compliance threshold can be thought of as a 'guard band' to guard against false positives (Type I error) due to the possible imprecisions and biases in the sampling and analysis methodologies. A false positive would be a situation where a vessel which is in truth compliant is incorrectly classified as non-compliant due to imprecise measurements.

We note that by using this approach for testing for non-compliance, an implicit trade-off between type I and type II errors is made: the lower the pre-specified chance of incorrectly classifying a vessel which is in truth compliant as non-compliant ("compliant but fined"; type I error), the higher the chance of incorrectly classifying a vessel which is in truth non-compliant as compliant ("non-compliant but not fined"; type II error).

The width of the guard band is determined by the reliability of the estimate and by the desired level of protection against false positives. The width of the guard band had to be estimated from data on untreated ballast water from Gollasch and David (2010) (as the data availability for treated water was limited and had (near) zero counts of organisms). Therefore, it was assumed, that data for treated ballast water will show a similar, or lower variance. Similar variances are used in the calculations. Using lower variances (if future data sets support such), will result in a smaller guard band. The desired rate of false positives does not have a scientific basis, but needs to be set by Administrations. For example, in case the guard band is chosen in accordance with a type I error of $\alpha = 0.1\%$, an administration can be $100 - \alpha = 99.9\%$ sure that the observed count could not have arisen in the samples if the vessel is in truth compliance with the standards in the BWM Convention. The reliability is governed both by the (minimum requirements of the) sampling procedure and by the inherent variability of the system under study (as explained below). The width of the guard band may thus be different for different sampling procedures.

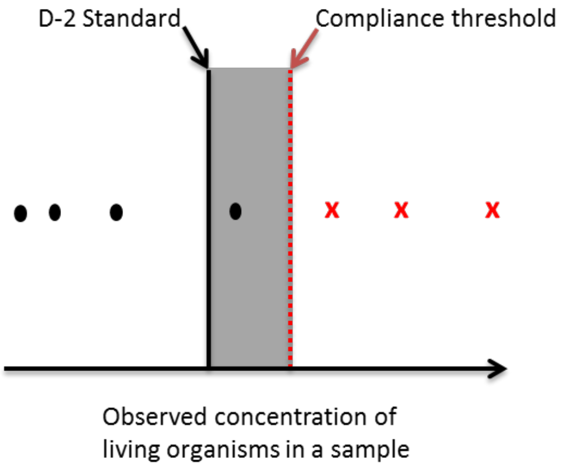


Figure 1 *Illustration of the use of a compliance threshold (red dotted vertical line) at a concentration above the standard (black solid vertical line) for assessing evidence of gross non-compliance of vessels. The distance at which the compliance threshold is set above the D-2 standard is the 'guard band' to guard against false positives (type I errors) due to the possible imprecisions and biases in the sampling methodology. Estimates of mean concentrations that fall above the compliance threshold give evidence of gross non-compliance (red crosses), whereas estimates below the threshold indicate no evidence of gross non-compliance (black dots).*

Based on the counts of numbers of living organisms in samples of the ballast water, a test statistic can be derived. The compliance threshold is set at the value below which $1 - \alpha$ % of possible test statistics (outcomes of the sampling procedure) are expected to fall given that the true (unobserved) concentration is equal to the D-2 standard. The probability that a sampling procedure yields a particular test statistic, given that the true concentration is equal to the D-2 standard, can be estimated using a probability distribution. Evidence of gross non-compliance would be proven if a point estimate, based on a sample, is higher than a concentration which can be expected to occur by chance-like processes if the true concentration is equal to the D-2 standard (Figure 2). In other words, evidence of gross non-compliance would be proven if observed organism numbers in a sample are significantly higher than that theoretically expected from a sample which is in compliance with the D-2 standard, but may have higher organism levels than the standard because it is influenced by the sampling and analytical errors inherent to biological sampling. Therefore, the threshold has been set to take this into account.

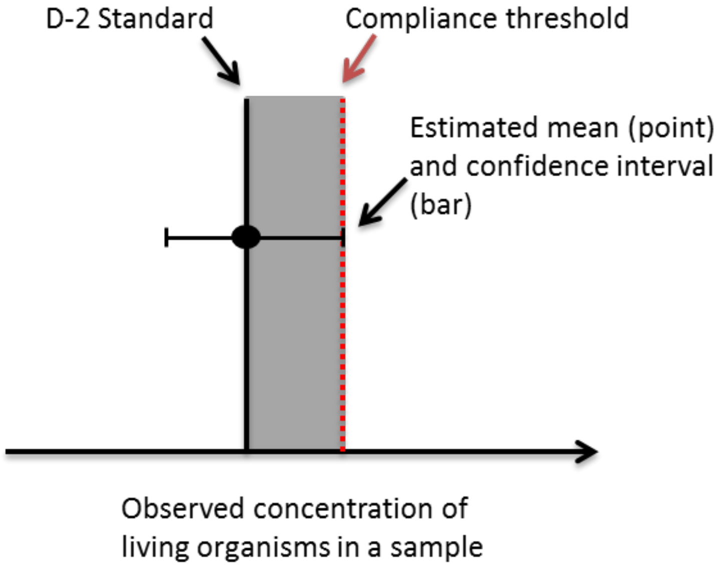

D-2 Standard
Compliance threshold
Estimated mean (point) and confidence interval (bar)
Observed concentration of living organisms in a sample

Figure 2 Illustration of the derivation of a compliance threshold from the estimated confidence interval of the estimate of a mean concentration of living organisms based on a sample.

The reliability of an estimate consists of two parts: accuracy and precision.

Precision refers to the size of random errors due to:

- inherent variability in concentrations of living organisms in the ballast water system;
- to sampling errors; and,
- to analytical errors.

An example of inherent variability in the ballast water system is variability in the mean concentration of living organisms in different parts of the discharge from the same water tank or between water tanks. This stratification can happen even when the ballast water has been pre-treated, however it can be reduced if treatment is undertaken on discharged. This is because successful treatment should homogenise the discharge to organism levels below the D-2 Standard. Sampling errors refer to the fact that in practice only a small sample (relative to the total volume of a discharge) will be taken leading to an element of chance in the parts of discharges which are included in the sample. Analytical errors are the various errors which are inherent to the methods used to estimate the concentrations of living organisms in the sample, including: errors in the determination of the sampled volume, counting errors, and the classification of organisms as belonging to a certain size group or as dead or alive. Some of these aspects can, under specific condition, also lead to a bias as described in the following section on accuracy. Inherent variability and analytical errors may in principle be estimated using experimental data, whereas sampling error can be controlled by the chosen (minimum requirements for) sample size and method. Gross non-compliance thresholds are based on models which consider the magnitude of random errors.

Accuracy refers to systematic errors, leading to biased estimates of mean concentrations. Accuracy is of primary concern, but is typically more difficult to quantify than precision. It is important to realise that the accuracy of an estimate based on a sample cannot be influenced by varying the sample size. It is therefore crucial to define the sampling procedure in such a way that a representative sample is obtained. However, full representativeness can in practice often not be guaranteed. Systematic errors are not addressed in the models used as true concentrations are unknown. We only have observed concentrations, whereas for quantification of systematic errors both are required. Assumptions with respect to systematic errors are made which are discussed in Chapter 9 and appendix G.

4 The Poisson rate test

The Poisson rate method, which makes use of the Poisson probability distribution (Jørgenson *et al.*, 2010; Miller *et al.*, 2011), should form the basis for estimating the mean concentrations, confidence intervals and compliance thresholds for ballast water discharges. However, the Poisson rate assumes that the living organisms are randomly distributed in the ballast water discharge, and that sample handling and analysis maintains this. Furthermore, the Poisson rate method takes only sampling error into account, which can be reduced by increasing the sample size. Counts can be more variable than predicted by the Poisson rate test when:

- sampling is not random, and/or the distribution of living organisms in the ballast water is not random and, therefore, not homogenously mixed;
- sample handling or analysis introduces additional variability. For example, random errors in the measurements of sample volumes introduce variability into estimates of concentrations of organisms per unit volume.

Despite trying to overcome these issues through the development of a sampling protocol, it is plausible that both aspects may , to varying degrees, affect observed organism counts in ballast water. Therefore the bare Poisson rate test will underestimate the actual error. For example, clumping of organisms or random errors in the measurement of sample volumes are expected to lead to additional variation in the observed rate of organisms per volume of water, over and above the random sampling error. A refinement of the Poisson rate method is needed for the estimation of the reliability of estimates of mean concentrations. This refinement should take the other important sources of (both random and systematic) error as discussed above into account.

5 Sampling procedure

Understanding the sampling procedure is key in understanding the variance observed in the data. IMO has provided guidelines for ballast water sampling (the Guidelines (G-2), Annex 3 to resolution MEPC.173(58) (and the draft circular in development)) in assessing the compliance to regulation D-2. These documents mostly provide guidance for technical aspects, e.g. where to take a sample from the ballast water system. As each ship and ballast water system is different, there is no fixed sampling protocol. The “Guidelines on Sampling of Ballast Water (G2)”, however, does provide boundary conditions for such protocols:

- the sampling protocol should be in line with these Guidelines;
- 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;
- the sampling protocol should take account of the potential for a suspended sediment load in the discharge to affect sample results;
- the sampling protocol should provide for samples to be taken at appropriate discharge points;
- the quantity and quality of samples taken should be sufficient to demonstrate whether the ballast water being discharged meets with the relevant standard;
- sampling should be undertaken in a safe and practical manner;
- samples should be concentrated to a manageable size;
- samples should be taken, sealed and stored to ensure that they can be used to test for compliance with the Convention;
- samples should be fully analysed within test method holding time limit using an accredited laboratory; and
- samples should be transported, handled and stored with the consideration of the chain of custody.

5.1 Replicates and volumes sampled by Gollasch and David (2010)

As Gollasch and David (2010) found hardly any living organisms in treated ballast water, they focussed their effort on the analysis of uptake and discharge of untreated ballast water. The basic sampling procedure used, was set-up specifically to meet Port State Control (PSC) requirements. The outline of the sampling procedure is illustrated in Figure 3. Untreated water was sampled during the discharge on two different voyages, on two different ships. During the first voyage, there were 5 pumping events with untreated water (3 uptake and 2 discharge events). During the second voyage, there were 4 pumping events (2 uptake and 2 discharge events).

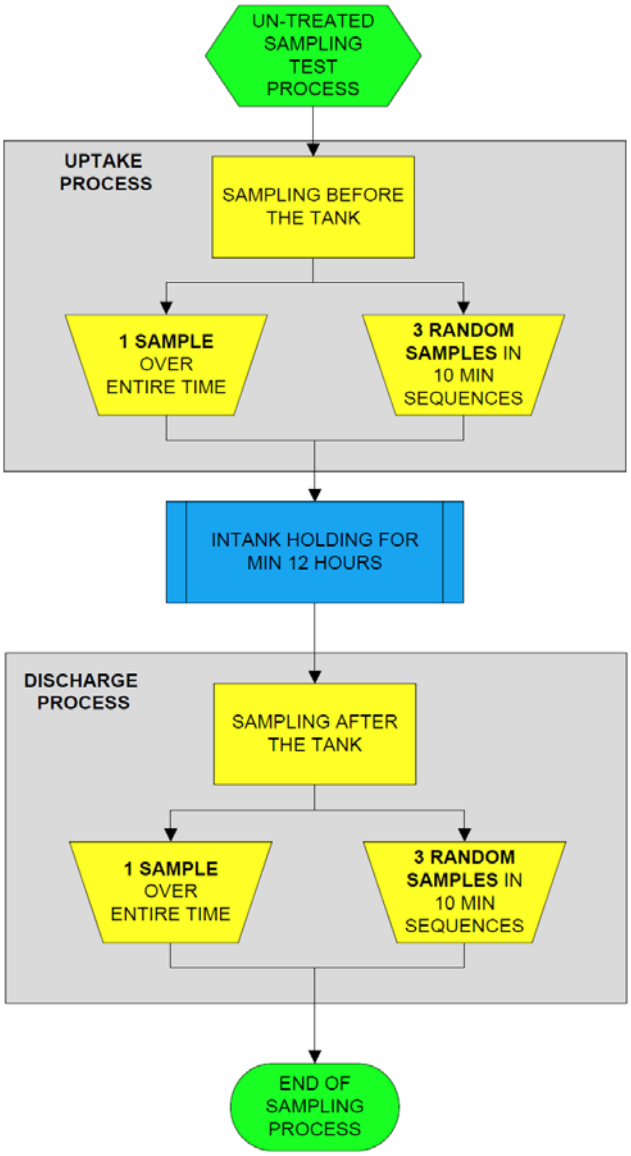

UN-TREATED SAMPLING TEST PROCESS
UPTAKE PROCESS
SAMPLING BEFORE THE TANK
1 SAMPLE OVER ENTIRE TIME
3 RANDOM SAMPLES IN 10 MIN SEQUENCES
INTANK HOLDING FOR MIN 12 HOURS
DISCHARGE PROCESS
SAMPLING AFTER THE TANK
1 SAMPLE OVER ENTIRE TIME
3 RANDOM SAMPLES IN 10 MIN SEQUENCES
END OF SAMPLING PROCESS

Figure 3 Sampling scheme for untreated ballast water (from Gollasch and David 2010)

During each of these pumping events 4 samples were taken: 1 continuous sample of the entire uptake/discharge volume (referred to as ‘Over Entire Time’ or OET) and 3 replicate discrete volumes intended at the start middle and end of the uptake/discharge (referred to as ‘S1’, ‘S2’ and ‘S3’ for beginning, middle and end respectively). By applying a flow splitter, the continuous and discrete samples could be taken in parallel. As the continuous samples were not replicated, there is no information on observed variability. Hence, these samples were not included in the analysis of the present study. The discrete samples were approximately 300-500 litres in volume and are subsampled for further analysis as described below.

We have grouped data of samples taken during the same cycle of discharge and uptake pumping events as belonging to the same ‘test’. This means that, in total, five tests were performed, each with ‘S1’, ‘S2’, ‘S3’ and ‘OET’ samples taken during both uptake and discharge. The tests are numbered 1,2,…5 and only during the 3rd test has a treatment been applied. The raw data are given in tables in appendix A.

5.1.1 Organisms ≥10 and <50 µm

A 1:1 subsample was taken from the discrete samples in a 10 litre bucket which after mixing was distributed over several 80 cm^3 bottles. These were sent to the lab for analysis. Organisms ≥10 and <50 µm were counted using flow cytometry. Such systems can either analyse a variable volume (stop analysing after a pre-set number of counts) or a fixed volume. For the analyses, 3 replicate subsamples of 270 mm^3 were counted by flow cytometry (pers. comm. Stephan Gollasch). The total volume of the subsample ($V_{subsamp}$) being analysed is hence $3 \times 0.27 = 0.81cm^3$. The three replicates came from a single 80 cm^3 bottle.

5.1.2 Organisms ≥50 µm

From the discrete sample a known volume (V_{sample}) between 292 and 540 litres was concentrated to a known volume (V_{conc}) of either 60, 80 or 100 cm^3. A 6 cm^3 subsample ($V_{subsamp}$) was taken from this concentrate and viable (living) organisms were counted under a binocular in equally shaped counting chambers. This resulted in a raw integer count number (c_{raw}). Counts are only available for the entire 6 cm^3, not per chamber.

6 Estimation of organism concentrations (≥50 µm) based on raw counting data

In the report by Gollasch and David (2010) only raised concentrations (c_{raised}) are presented. These raised concentrations in fact represent estimates, $\hat{\mu}$, of the (unobserved true) concentration μ of numbers of living organisms in the total ballast water discharge, based on an observed count in a sample of the discharge.

These raised concentrations are calculated from the raw counts as follows:

$$c_{raised}\left[\frac{\text{number of organisms}}{m^3}\right] = c_{raw}[\text{number of organisms}] \times R[m^{-3}], \qquad \textbf{Eqn. 1}$$

Where, for organisms ≥50 µm, $R[m^{-3}] = \frac{V_{conc}[cm^3]\times 1000[dm^3/m^3]}{V_{subsamp}[cm^3]\times V_{sample}[dm^3]}$,

which is the inverse of the volume of discharge which the subsample represents. We will refer to R as the raising factor.

For organisms ≥10 and <50 µm, $R[cm^{-3}] = \frac{1}{V_{subsamp}[cm^3]}$.

For example, suppose a raw count (c_{raw}) of 90 living organisms ≥50 µm was observed in a subsample of 6 cm^3 ($V_{subsamp}$), which originated from a sample volume of 500 dm^3 (V_{samp}), and then concentrated to a volume of 100 cm^3 (V_{conc}). Then, the raising factor is

$$R = \frac{100\times 1000}{6\times 500} = \frac{100}{3},$$

and the raised concentration (c_{raised}) is calculated as:

$$\hat{\mu} = c_{raised} = 90 \times \frac{100}{3} = 3000 \quad \text{organisms}/m^3$$

The reliability of estimates, expected of true concentrations $\hat{\mu}$ relies on the actual sampling volumes and counts which were made. This information has been retrieved in e-mail conversations with the authors of the Gollasch and David (2010) report, and the data has been included in appendix A of this report.

7 Estimation of the variance of estimated concentrations based on counts in samples

7.1 Theoretical results

This chapter attempts to explain the theoretical background in laymen's terms. A more accurate (but complex) background is given in appendix B.

7.1.1 Poisson rate test

The variability as defined by the Poisson rate test purely depends on raw counting data (c_{raw}) and is independent from the raising factor (R). If the Poisson rate test applies, the variability in raised concentrations (c_{raised}) can be simply obtained by multiplying the variability in the raw counts with the raising factor (R). However, in the dataset from Gollasch and David (2010) we observed significantly more variation than could be explained with the Poisson rate test. In other words: the Poisson distribution is over-dispersed. Hence, a different statistical model is required to compensate for this over-dispersion.

7.1.2 An over-dispersed Poisson rate test

The model proposed here to describe the over-dispersion is called the Negative Binomial (NB) model. Two different alternatives of the NB are studied here. In the first alternative, the variance is assumed to be equal to a fixed factor (φ) multiplied with the expected counts. In the second option the variance is assumed to increase quadratically with increasing expected counts and is parameterized with θ. Summarizing, the following models were considered:

- Model 1: The Poisson distribution in which the variance of the counts is equal to the mean (no over-dispersion):
$$\operatorname{var}(Y) = c$$
- Model 2: The 'quadratical' NB distribution with mean-variance relationship:
$$\operatorname{var}(Y) = c + {c^2}/{\theta}$$
- Model 3: The 'fixed factor' NB distribution with mean-variance relationship:
$$\operatorname{var}(Y) = \varphi \times c$$

A more detailed (mathematical) description is given in appendix B. These models were parameterized with data from Gollasch and David (2010) to derive a GNC threshold.

7.2 Parameterization of over-dispersed Poisson model

Estimates of the model parameters are given by the values of the parameters that maximize the likelihood of observing the data (from Gollasch and David, 2010) given the model. The parameters for models 2 and 3 were estimated at: $\varphi = 17.9$ and $\theta = 5.69$, for organisms ≥10 and <50 µm; and $\varphi = 3.17$ and $\theta = 40.8$, for organisms ≥50 µm. A more detailed derivation of these parameters is given in appendix C.

The observations are all made on discharges from untreated ballast water, with relatively high estimated mean concentrations compared to the D-2 standard. The estimates of parameters that describe the mean variance relationship ($\varphi = 17.9, 3.17$ and $\theta = 5.69, 40.8$) are themselves uncertain and are based upon only four discharge events with three replicate discrete samples. It is not possible to choose the correct model for the mean-variance relationship from this limited number of sample events in order to extrapolate to the mean-variance relationship around the D-2 standard, without further insight into the mechanisms underlying the over-dispersion. Also, it is not possible to derive a good estimate of the uncertainty of the over-dispersion parameters, based on this small number of sample events. Therefore, it is not possible to set a realistic upper bound on the dispersion parameter.

However, the model that provides to the widest distribution of possible counts in samples if the true mean concentration in the ballast water discharge is equal to the D-2 standard, is the model in which the variance is set at a multiple of the mean (model 3). This is because the raw counts in the subsamples ($V_{subsamp}$) are expected to be low, as visualized in Figure 4 and Figure 5.

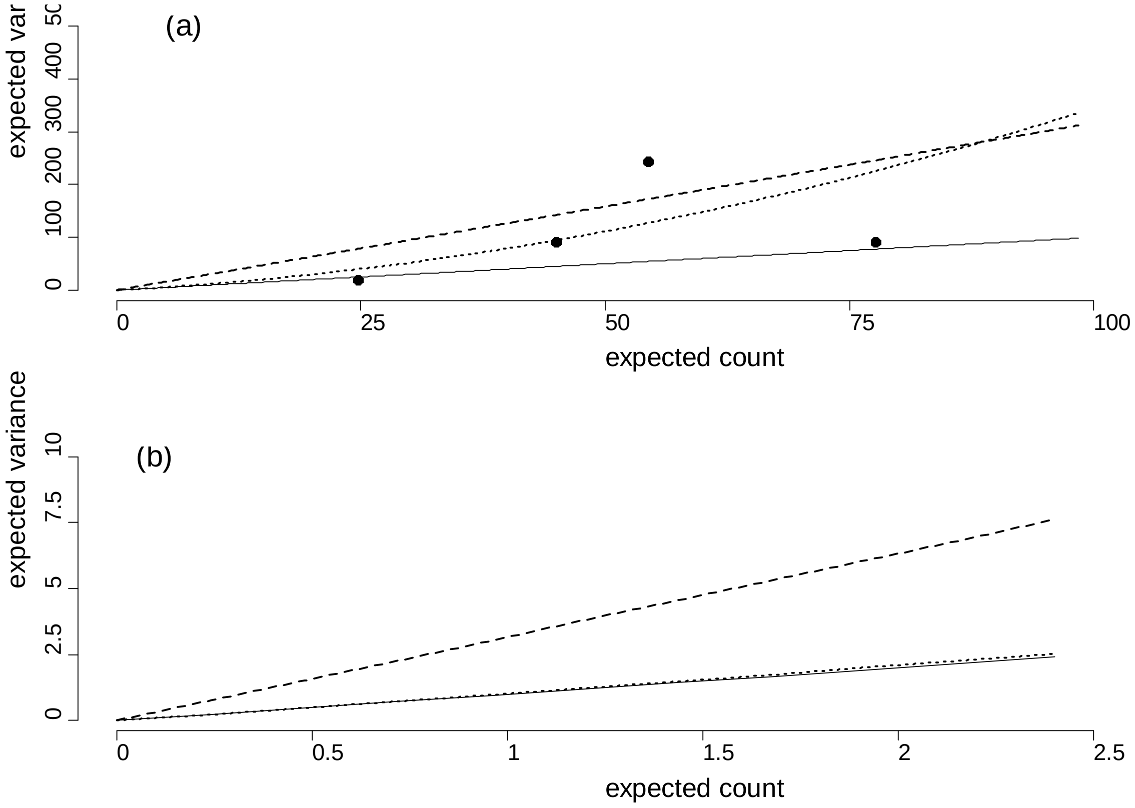


Figure 4 *Predicted mean-variance relationship for three models for the distribution of counts of organisms ≥50 µm in discrete samples of the ballast water (which is concentrated and subsequently sampled), given an unobserved true mean concentration in the entire discharge. Model 1 (solid line), Model 2 (dotted line), and Model 3 (dashed line). Panel (a) gives the predicted mean variance relationships for expected raw counts in samples ranging from 0 to 100, as well as the observed variance of the three replicates for each of the four discharge events. Panel (b) gives the predicted mean variance relationships for expected counts in samples ranging from 0 to 2.5. Expected counts are in the subsample, without raising of the data, hence variance in the raised data depends on the raising factor associated with a particular sampling scheme.*

Suppose that a discrete sample is taken of 500 dm^3 (V_{samp}) which is concentrated to 100 cm^3 (V_{conc}) and a single count is subsequently made in a subsample of 6 cm^3 ($V_{subsamp}$). For such a sample, the expected count in the subsample if the true concentration in the ballast water, μ, is equal to the D-2 standard is $\mathrm{E}(Y) = \frac{\mu}{R} = 0.3$. For such a small expected count, the model with $\mathrm{var}(Y) = 3.17\mu$ (Scenario 3, dashed line in Figure 4) gives the most largest estimate of the variance.

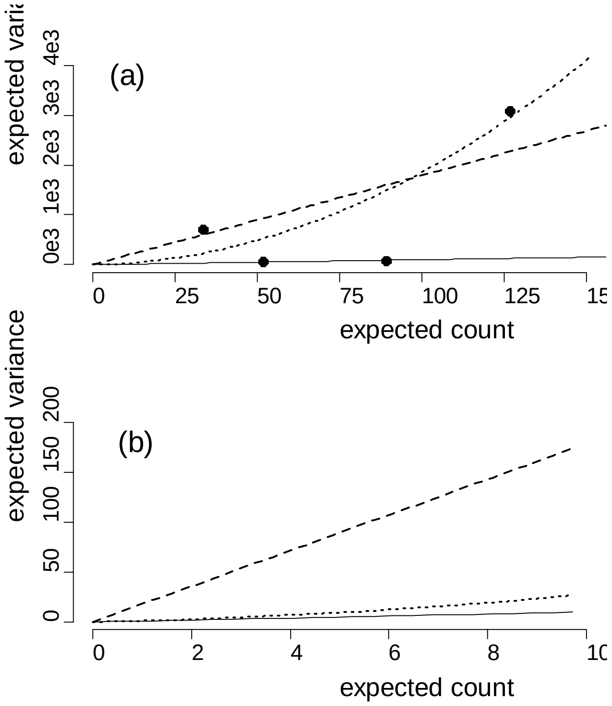


Figure 5 *Predicted mean-variance relationship for three models for the distribution of counts of organisms ≥10 and <50μm in discrete samples of the ballast water (which is concentrated and subsequently sampled), given an unobserved true mean concentration in the entire discharge. Model 1 (solid line), Model 2 (dotted line), and Model 3 (dashed line). Panel (a) gives the predicted mean variance relationships for expected raw counts in samples ranging from 0 to 150, as well as the observed variance of the three replicates for each of the four discharge events. Panel (b) gives the predicted mean variance relationships for expected counts in samples ranging from 0 to 10. Expected counts are in the subsample, without raising of the data, hence variance in the raised data depends on the raising factor associated with a particular sampling scheme.*

8 Derivation of gross non-compliance threshold

The gross non-compliance (GNC) threshold for a given sampling scheme can be derived from the NB distribution with mean-variance relationship given by Model 3 as described in Chapter 7. In order to be able to add confidence limits to the D-2 standard to create a gross non-compliance threshold, and refine the thresholds further, it is recommended that Model 3 should be tested with data from 'treated' ballast water discharges.

The GNC is derived by calculating for which raw count k, the probability of exceeding this count is smaller than 0.1%, given that the true concentration in the ballast water discharge being is equal to the D-2 standard.

For example, if counts are made in subsamples of 6 cm^3 each, from a 100 cm^3 concentrate which has been obtained by pouring a discrete discharge sample of 500 dm^3 through a sieve of 50 µm diagonal mesh, then:

- The true concentration at the D-2 standard is $\mu = 10$
- The raising factor $R = \frac{100}{3}$
- The expected count $c = \frac{3\mu}{100} = 0.3$

8.1 Sampling strategies and implications for the derivation of a GNC threshold

In practice, counts may be made in samples with varying volumes which may be taken from the discharge in a two-stage sampling design. In the first stage, discrete samples may be taken at different times during the ballast water discharge, and in the second stage multiple subsamples may be taken from these (concentrated) discrete samples.

Depending on the mechanisms underlying the over-dispersion in the counts, increasing the sample size at the second stage (increasing the volume of the subsampling of the concentrate of a discrete sample of the ballast water) may be expected to contribute less to the reliability of the estimate of the concentration in the entire discharge then increasing the number of discrete samples. The reason for this is that the variance between the 6 ml samples taken from the 100 ml concentrated sample is much less than the variance between the main samples taken from the discharge line. The mechanisms underlying the over-dispersion are not known, and the data do not allow further qualitative or quantitative investigation of potential sources of over-dispersion. Variation in concentrations during the discharge may for example be entirely random, follow some trend such as continuously increasing/decreasing concentrations, or may exhibit sudden stepwise changes in concentrations. We will assume a scenario in which the over-dispersion is caused entirely by variation in the concentration of living organisms over time during the discharge.

If over-dispersion is caused entirely by variation in the concentration during the discharge, the largest increase in precision can be expected to occur by increasing the number of discrete samples and spacing these samples regularly throughout the discharge from beginning to end (Cochran, 1963; unless concentrations in the discharge fluctuate with some period which coincides with the sampling frequency). This is illustrated by Table 3 versus Table 5 and Table 4 versus Table 6. For a more theoretical in-depth discussion we refer to appendix D

The GNC threshold that can be applied depends on the number of main samples and the number of subsamples drawn from each main sample and the volumes that are sampled. For organisms ≥10 and <50 µm, the GNC thresholds are derived where one or more 0.810 mL subsamples are taken and analysed from one or more main samples. For organisms ≥50 µm the GNC thresholds are derived where one or more 6 mL subsamples are taken and analysed from 100 mL concentrate of one or more 500 L main samples.

Table 3 versus Table 5 and Table 4 versus Table 6 show the effect (on the GNC threshold) of taking multiple subsamples from the same main sample versus taking a single subsample from multiple main samples. The difference between Table 3 versus Table 4 and Table 5 versus Table 6 is that in the latter tables the total raw counts in the subsamples is multiplied with the appropriate raising factor to obtain the raised counts in the discharged water.

The GNC threshold values that are derived using the Poisson distribution (Model 1) as proposed by Jørgensen *et al.* (2010) and Miller *et al.* (2011) are also given in the tables below for comparison.

A more detailed description of the derivation of the GNC threshold is given in appendices D and E.

Table 3 Gross noncompliance threshold values, estimated by two statistical models (Poisson distribution and NB distribution) that describe the distribution of the test statistic (counts of living organisms in samples of ballast water) under the assumption that the true concentration in the entire discharge is equal to the D-2 standard (10 organisms per cm^3 or m^3).

Number of main samples	Number of subsamples	≥10 and <50 µm			≥50 µm		
		sampled volume [*cm*^{3}]	Model 1: Poisson	Model 3: Negative Binomial	Volume that samples represent (*dm*^{3})	Model 1: Poisson	Model 3: Negative Binomial
1	1	0.81	19	94	0.06*500	4	11
1	2	1.62	31	184	0.12*500	5	19
1	3	2.43	42	274	0.18*500	6	27
1	4	3.24	52	364	0.24*500	7	35
1	5	4.05	63	455	0.30*500	7	43
1	6	4.86	73	545	0.36*500	8	51
1	7	5.67	82	635	0.42*500	9	59
1	8	6.48	92	725	0.48*500	9	67
1	9	7.29	102	816	0.54*500	10	75
1	10	8.10	111	906	0.60*500	11	83

Table 4 *Gross noncompliance threshold values (α=0.1%), expressed as concentrations per cm^3 or m^3. Results derived from the counts as presented in Table 3, with counts multiplied by the raising factors R, to obtain estimates of concentrations in the entire ballast water discharge, based upon the counts observed in the samples.*

Number of main samples	Number of subsamples	≥10 and <50 µm (#/cm^3)			≥50 µm (#/m^3)		
		sampled volume [*cm^3*]	Model 1: Poisson	Model 3: Negative Binomial	Volume that samples represent (*dm^3*)	Model 1: Poisson	Model 3: Negative Binomial
1	1	0.81	23.5	116.0	0.06*500	133.3	366.7
1	2	1.62	19.1	113.6	0.12*500	83.3	316.7
1	3	2.43	17.3	112.8	0.18*500	66.7	300.0
1	4	3.24	16.0	112.3	0.24*500	58.3	291.7
1	5	4.05	15.6	112.3	0.30*500	46.7	286.7
1	6	4.86	15.0	112.1	0.36*500	44.4	283.3
1	7	5.67	14.5	112.0	0.42*500	42.9	281.0
1	8	6.48	14.2	111.9	0.48*500	37.5	279.2
1	9	7.29	14.0	111.9	0.54*500	37.0	277.8
1	10	8.10	13.7	111.9	0.60*500	36.7	276.7

Table 5 *Gross noncompliance threshold values (α=0.1%). As Table 3 but based on a sampling strategy where one subsample has been obtained by from between 1 and 5 discrete main discharge samples over time.*

Number of main samples	Number of subsamples	≥10 and <50 µm			≥50 µm		
		Sampled volume (*cm^3*)	Model 1: Poisson	Model 3: Negative Binomial	Volume that samples represent (*dm^3*)	Model 1: Poisson	Model 3: Negative Binomial
1	1	0.81	19	94	0.06*500	4	11
2	1	1.62	31	119	0.06*1000	5	13
3	1	2.43	42	139	0.06*1500	6	14
4	1	3.24	52	158	0.06*2000	7	15
5	1	4.05	63	175	0.06*2500	7	17

Table 6 *Gross noncompliance threshold values (α=0.1%), expressed as concentrations per cm^3 or m^3. Results derived from the counts as presented in Table 5, with counts multiplied by the raising factors R, to obtain estimates of concentrations in the entire ballast water discharge, based upon the counts observed in the samples.*

Number of main samples	Number of subsamples	≥10 and <50 µm (#/cm^3)			≥50 µm (#/m^3)		
		Sampled volume (*cm^3*)	Model 1: Poisson	Model 3: Negative Binomial	Volume that samples represent (*dm^3*)	Model 1: Poisson	Model 3: Negative Binomial
1	1	0.81	23.5	116.0	0.06*500	133.3	366.7
2	1	1.62	19.1	73.5	0.06*1000	83.3	216.7
3	1	2.43	17.3	57.2	0.06*1500	66.7	155.6
4	1	3.24	16.0	48.8	0.06*2000	58.3	125.0
5	1	4.05	15.6	43.2	0.06*2500	46.7	113.3

9 Discussion

In Chapter 8 a set of Gross Non-Compliance (GNC) thresholds are derived, depending on specific sampling strategies. The compliance threshold is set at the count below which 99.9 % of possible counts (outcomes of the sampling procedure) are expected to fall given that the true concentration in the ballast water discharge is equal to the D-2 standard. The probability that the sampling procedure yields a particular count, given that the true concentration is equal to the D-2 standard, was estimated using the NB distribution. If a count higher than the given GNC threshold is observed, it can be concluded that this count is unlikely to have occurred if the true concentration in the ballast water is equal to (or lower than) the D-2 standard. It is important to note that in order to derive the GNC thresholds several assumptions had to be made. As a consequence, the GNC thresholds only apply when those assumptions are valid.

The overall assumption is that the random errors that are observed in the data of Gollasch and David (2010) using untreated ballast water, and the NB model that we use to describe it, are representative for the random errors that will be made by PSC at inspections of treated ballast water. In general it is likely that if the step by step methodology is followed (see appendix H) then the random errors will be the similar.

This assumption can be refined and split into separate specific assumption. These specific assumptions are discussed in more detail in appendix G.

Part of the representativeness stems from the availability of data. The present study relies on only four discharge events of untreated ballast water during two separate trips. Availability of data will aid the refinement of these thresholds over time. The understanding of variation and sampling/analysis error with respect to representativeness can be improved by including more data. This should preferably be data on treated ballast water and/or water with low (near D-2 standard) organism concentrations, which would help to get a more accurate estimate of over-dispersion and possible trends in time and of variance of organism concentrations. This may result in lower gross non-compliance thresholds.

In addition representativeness is also determined by procedural aspects. In other words PSC should use similar procedures for sampling and analysis as used by Gollasch and David (2010), as the GNC thresholds are based on those procedures. For this purpose protocols for sampling and analysis are proposed in appendices H. Also to make sure that quality of the sampling and analysis does not adversely affect the variance of PSC inspection, quality assurance should be integral part of the sampling and analysis.

An additional assumption that has to be made when applying a GNC threshold is that systematic errors in observed organism concentrations result in an underestimation of true concentrations. This error is difficult to quantify as true concentrations are usually unknown. This error can partly be reduced by assuring quality of, and providing clear protocols for sampling and analysis. Nevertheless, as these errors usually result in the underestimation of the true organism counts, then the thresholds provided by the NB model are still valid.

10 Conclusion

Gross Non-Compliance thresholds have been derived using the Negative Binomial distribution which has been parameterised using data of counts of living organisms in samples of untreated ballast water collected by Gollasch and David (2010).

When using the procedures described, counts of living organisms at or above the derived thresholds are observed, it can be stated with 99.9% certainty that the water is in gross non-compliance.

The derived GNC thresholds only apply if certain assumptions are valid. The validity of some assumptions need to be further assessed and discussed in the wider (scientific) community.

11 Literature

Cochran, W. G. (1963) *Sampling Techniques.* John Wiley & Sons, 2nd edition

Dobson, A. J. (2002). An introduction to generalized linear models. Chapmann & Hall/CRC, Boca Raton, Florida, 2nd edition.

Gollasch, S. and David, M. (2010). Testing Sample Representativeness of a Ballast Water Discharge and Developing Methods for Indicative Analysis. Report No. 4. Research Study. European Maritime Safety Agency, Lisbon, Portugal. EMSA/NEG/09/2010.

Jørgensen, C., Gustavson, K., Hanse, J.B. & Hies, T. (2010) Development of guidance on how to analyze a ballast water sample. Final report to the European Maritime Safety Agency, EMSA (2010). EMSA, Lisboa, Portugal.

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

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

Miller, A.W., M. Frazier, G.E. Smith, E.S. Perry, G.M. Ruiz, and M.N. Tamburri, 2011. Enumerating Sparse Organisms in Ships' Ballast Water: Why Counting to 10 is Difficult? *Environ. Sci. Tech 45: 3530-3546.*

Niemelä, S. I. (2002). Uncertainty of quantitative determinations derived by cultivation of microorganisms. Publication J3/2002 Advisory commission for metrology chemistry section Expert Group for Microbiology.

Venables, W. N. & Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0

Quality Assurance

IMARES utilises an ISO 9001:2008 certified quality management system (certificate number: 57846-2009-AQ-NLD-RvA). This certificate is valid until 15 December 2012. The organisation has been certified since 27 February 2001. The certification was issued by DNV Certification B.V. Furthermore, the chemical laboratory of the Fish Division has NEN-AND-ISO/IEC 17025:2005 accreditation for test laboratories with number L097. This accreditation is valid until 27 March 2013 and was first issued on 27 March 1997. Accreditation was granted by the Council for Accreditation.

Justification

Rapport C124/12
Project Number: 430.51107.01

The scientific quality of this report has been peer reviewed by the a colleague scientist and the head of the department of IMARES.

Approved: Marcel Machiels
Research Scientist

Signature:

Date: ~date~

Approved: Floris Groenendijk
Department head

Signature:

Date: ~date~

Appendix A – Raw data from Gollasch and David (2010)

Table 7 Raw data ≥ 50 micrometre

Test number	Treatment	Uptake or Discharge	Sample type	Sample sequence	V_{samp} dm^3	V_{conc} cm^3	$V_{subsamp}$ cm^3	c_{raw}
1	untreated	UPT	S	1	450	80	6	172
1	untreated	UPT	S	2	450	80	6	152
1	untreated	UPT	S	3	450	80	6	143
1	untreated	UPT	OET		2595	100	6	312
1	untreated	DISCH	S	1	450	60	6	84
1	untreated	DISCH	S	2	450	80	6	66
1	untreated	DISCH	S	3	450	80	6	83
1	untreated	DISCH	OET		2869	80	6	248
2	untreated	UPT	S	1	300	80	6	38
2	untreated	UPT	S	2	300	80	6	43
2	untreated	UPT	S	3	300	80	6	24
2	untreated	UPT	OET		2324	80	6	65
3	untreated	UPT	S	1	380	80	6	11
3	untreated	UPT	S	2	380	80	6	21
3	untreated	UPT	S	3	380	80	6	28
3	untreated	UPT	OET		3287	80	6	66
3	treated	UPT	S	1	374	80	6	0
3	treated	UPT	S	2	364	80	6	0
3	treated	UPT	S	3	359	80	6	0
3	treated	UPT	OET		2717	80	6	0
2	untreated	DISCH	S	1	380	80	6	28
2	untreated	DISCH	S	2	380	80	6	39
2	untreated	DISCH	S	3	540	80	6	68
2	untreated	DISCH	OET		2562	80	6	144
3	treated	DISCH	S	1	380	80	6	0
3	treated	DISCH	S	2	380	80	6	0
3	treated	DISCH	S	3	493	80	6	0
3	treated	DISCH	OET		1924	80	6	0
4	untreated	UPT	S	1	350	100	6	51
4	untreated	UPT	S	2	350	100	6	58
4	untreated	UPT	S	3	350	100	6	45
4	untreated	UPT	OET		2381	100	6	243
4	untreated	DISCH	S	1	350	100	6	25
4	untreated	DISCH	S	2	350	100	6	32
4	untreated	DISCH	S	3	292	100	6	17
4	untreated	DISCH	OET		1763	100	6	113
5	untreated	UPT	S	1	450	100	6	226
5	untreated	UPT	S	2	450	100	6	124
5	untreated	UPT	S	3	450	100	6	201
5	untreated	UPT	OET		3243	100	6	523
5	untreated	DISCH	S	1	450	100	6	34
5	untreated	DISCH	S	2	450	100	6	57
5	untreated	DISCH	S	3	450	100	6	72
5	untreated	DISCH	OET		3105	100	6	227

Table 8 Raw data ≥10 <50 micrometre

Test number	Treatment	Uptake or Discharge	Sample type	Sample sequence	V_{samp} dm^3	$V_{subsamp}$ mm^3	c_{raw} Rep1	c_{raw} Rep2	c_{raw} Rep3
1	untreated	UPT	S	1	450	270	23	25	13
1	untreated	UPT	S	2	450	270	11	22	16
1	untreated	UPT	S	3	450	270	22	19	10
1	untreated	UPT	OET		2595	270	22	21	19
1	untreated	DISCH	S	1	450	270	11	26	20
1	untreated	DISCH	S	2	450	270	17	17	23
1	untreated	DISCH	S	3	450	270	17	6	18
1	untreated	DISCH	OET		2869	270	21	12	12
2	untreated	UPT	S	1	300	270	4	2	1
2	untreated	UPT	S	2	300	270	2	2	2
2	untreated	UPT	S	3	300	270	6	5	3
2	untreated	UPT	OET		2324	270	15	6	4
3	untreated	UPT	S	1	380	270	12	1	0
3	untreated	UPT	S	2	380	270	2	1	4
3	untreated	UPT	S	3	380	270	6	8	7
3	untreated	UPT	OET		3287	270	5	9	5
3	treated	UPT	S	1	374	270	0	0	0
3	treated	UPT	S	2	364	270	0	0	0
3	treated	UPT	S	3	359	270	0	0	0
3	treated	UPT	OET		2717	270	0	0	0
2	untreated	DISCH	S	1	380	270	8	6	2
2	untreated	DISCH	S	2	380	270	6	3	4
2	untreated	DISCH	S	3	540	270	29	17	25
2	untreated	DISCH	OET		2562	270	15	17	6
3	treated	DISCH	S	1	380	270	0	0	0
3	treated	DISCH	S	2	380	270	0	0	0
3	treated	DISCH	S	3	493	270	0	0	0
3	treated	DISCH	OET		1924	270	0	0	0
4	untreated	UPT	S	1	350	270	44	58	54
4	untreated	UPT	S	2	350	270	63	49	60
4	untreated	UPT	S	3	350	270	64	65	47
4	untreated	UPT	OET		2381	270	64	57	65
4	untreated	DISCH	S	1	350	270	15	21	28
4	untreated	DISCH	S	2	350	270	38	38	41
4	untreated	DISCH	S	3	292	270	77	68	54
4	untreated	DISCH	OET		1763	270	26	29	37
5	untreated	UPT	S	1	450	270	39	26	34
5	untreated	UPT	S	2	450	270	42	32	53
5	untreated	UPT	S	3	450	270	34	46	52
5	untreated	UPT	OET		3243	270	70	54	57
5	untreated	DISCH	S	1	450	270	26	29	44
5	untreated	DISCH	S	2	450	270	21	29	27
5	untreated	DISCH	S	3	450	270	37	26	28
5	untreated	DISCH	OET		3105	270	53	44	54

Appendix B – Theoretical background

Poisson rate test

If the living organisms are randomly distributed in the ballast water discharge, and sample handling and analysis does not introduce non-randomness in the distribution of organisms, the Poisson distribution may be used to estimate the reliability of estimates of μ̂ (the concentration of living organisms in the total discharge; see Chapter 6). Let Y be a random variable for the number of living organisms, and y the observed number of living organisms in a sample. Then, the Poisson probability distribution can be written as

$$f(Y = y; c) = \frac{e^{-c}c^{y}}{y!}, \qquad y = 0,1,2, \ldots \qquad \textbf{Eqn. 2}$$

If the Poisson distribution applies, for a given true (unobserved) concentration of living organisms in the ballast water discharge (μ), the expected count ($c = \mathrm{E}(Y)$) and variance ($\mathrm{var}(Y)$) of the counts of living organisms in a sample of the ballast water are given by:

$$\mathrm{E}(Y) = \mathrm{var}(Y) = \frac{\mu}{R} = c \qquad \textbf{Eqn. 3}$$

The following shorthand notation is used to specify that observed counts Y are assumed to be Poisson distributed with mean rate c:

$$Y \sim \mathrm{Poisson}(c)$$

If the sample of the ballast water is representative of the discharge, an unbiased estimator of the true concentration of living organisms in the ballast water discharge, $\hat{\mu}$, is given by:

$$\hat{\mu} = c_{raised} \qquad \textbf{Eqn. 4}$$

In case the volumes are all measured without error, the variance of $\hat{\mu}$, $V(\hat{\mu})$, is then given by:

$$V(\hat{\mu}) = R^2 \times c_{raw} \qquad \textbf{Eqn. 5}$$

In practice, volumes will not be estimated without error. Random errors in volume measurements will lead to uncertainty in the raising factor R. A conservative (large) estimate of uncertainty in volume measurements is that the coefficient of variation is 5% of the measured volume. It is reasonable to assume that errors in volume measurements will be independent. Then, an approximation of the variance of the raising factor is given by:

$$V(R) = R^2 \times 3 \times 0.05^2 \qquad \textbf{Eqn. 6}$$

When the uncertainty in the raising factor is taken into account, the expected variance of $\hat{\mu}$ is approximated by (see Figure 6):

$$V(\hat{\mu}) = R^2 \times c_{raw} + R^2 \times {c_{raw}}^2 \times 3 \times 0.05^2 \qquad \textbf{Eqn. 7}$$

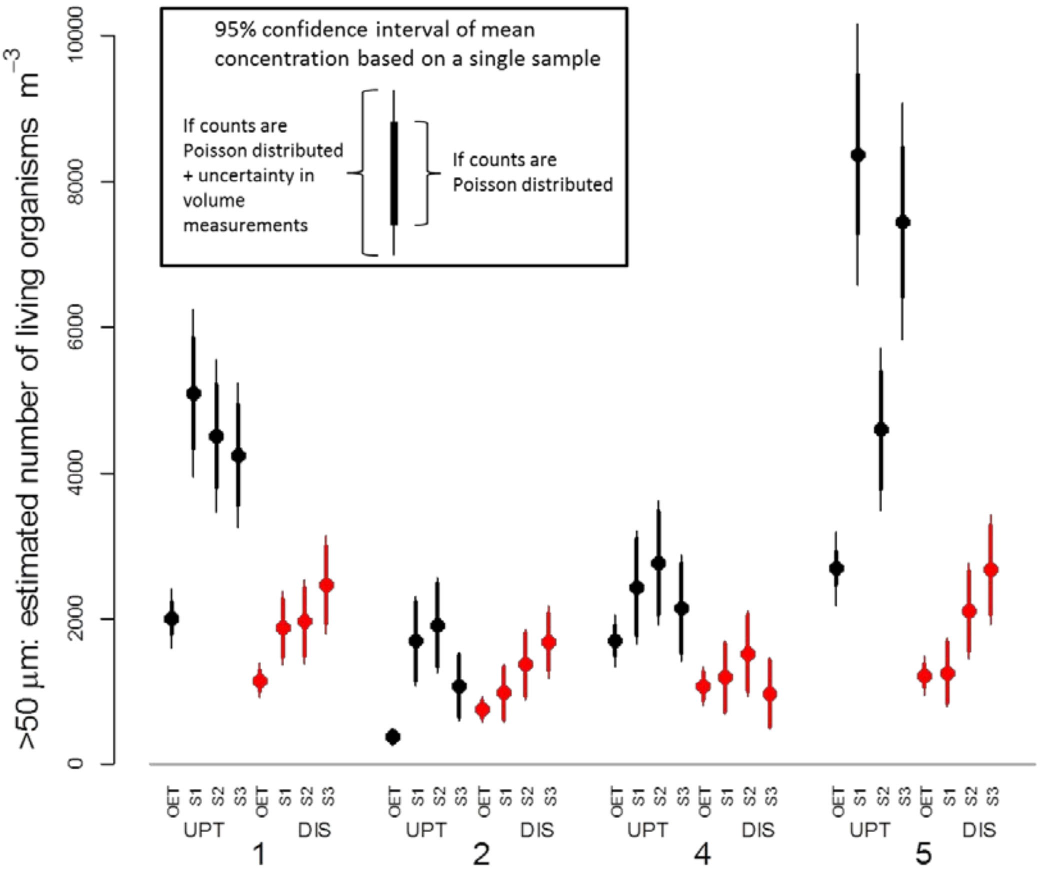


Figure 6 Variances under poisson distribution: indications that variance under Poisson not enough to capture variability in counts between repeated samples. Not even when uncertainty in volume measurements are taken into account. The numbers 1,2,4 and 5 refer to test events, 'UPT' to uptake and 'DISCH' to discharge, 'OET' to samples taken over the entire time of the discharge, and 'S1', 'S2' and 'S3' refer to discrete samples taken at the beginning, middle and end of the discharge.

An over-dispersed Poisson rate test

A variety of processes may result in counts in samples which are more variable than expected given the Poisson distribution. These processes include clumping of organisms, changes in concentrations of organisms over time during the discharge, and errors in the measurements of volumes. Such additional variability over and above the variability expected under the Poisson distribution, typically referred to as 'over-dispersion', is a common phenomenon in count data.

A more flexible model than the Poisson distribution would allow the variance to be a multiple, φ, of the mean:

$$\mathrm{var}(Y) = \varphi \mathrm{E}(Y) = \varphi \frac{\mu}{R} = \varphi c, \qquad \textbf{Eqn. 8}$$

with φ ($\varphi > 1$) a parameter that can be estimated from the data set of counts in the samples of the ballast water.

A model which allows the variance to be larger than the mean is the negative binomial probability distribution (NB distribution), given by

$$f(Y = y; c, \theta) = \left(\frac{\theta}{\theta+c}\right)^{\theta} \frac{\mathrm{G}(\theta+y)}{\mathrm{G}(y+1)\mathrm{G}(\theta)} \left(\frac{c}{\theta+c}\right)^{y} \qquad \textbf{Eqn. 9}$$

where G( ) is the Gamma function, $c = \frac{\mu}{R}$ is the expected raw count in the subsample, and the parameter θ determines the relationship between the variance and the mean (see Eqn. 10).

The NB distribution can be used as an alternative to the Poisson distribution to describe count data over an unbounded positive range whose sample variance is larger than the sample mean. For the NB distribution, the variance is a quadratic function of the mean:

With $\mathrm{E}(Y) = c$ and $\mathrm{var}(Y) = c + {c^2}/{\theta}$

For a given value of the expected count c, the variance can also be expressed as a multiple φ of the mean for the NB distribution:

$$\theta = \frac{c}{\varphi - 1}, \text{ since } \varphi c = c + {c^2}/{\theta} \qquad \textbf{Eqn. 10}$$

The following shorthand notation is used to specify that observed counts Y are assumed to be NB distributed with mean rate c and over-dispersion parameter θ:

$$Y \sim \mathrm{NB}(c, \theta)$$

Appendix C – Results: parameterization of over-dispersed Poisson model

The parameters for modeling over-dispersion in count data, φ and θ, may be estimated using the data set of counts in the samples of the ballast water. Let y_i be the observed count of living organisms in a sample from a concentrate of a discrete ballast water sample *i* (*i=1,2,…n*). The counts y_i may be modeled using the Poisson distribution or the NB distribution, given a predicted mean concentration in the total ballast water discharge event j:

$$y_i \sim \text{Poisson}(c_i)$$

or

$$y_i \sim \text{NB}(c_i, \theta)$$

With expected values c_i for the counts per sample predicted by the following model:

$$\log(c_i) = \log\left(\frac{1}{R_i}\right) + m_{j(i)} \qquad \textbf{Eqn. 11}$$

where $m_{j(i)}$ are the estimated mean concentrations of living organisms in the total discharge event j of which the discrete samples *i* were taken from.

Estimates of the parameters $m_{j(i)}$ and θ, $\widehat{m}_{j(i)}$ and $\hat{\theta}$, are given by the values of the parameters that maximize the likelihood of observing the data given the model (see for example Dobson 2002).

Table 9 *parameter estimates and log-likelihoods for the Poisson and NB models (Eqn. 11). The quantity $-2log(likelihood)$ is also commonly referred to as the deviance of the model, and is a measure of the relative goodness of fit the model.*

Parameter	≥10 and <50 µm		≥50 µm	
	Poisson (model 1)	NB (model 2)	Poisson (model 1)	NB (model 2)
$\widehat{m}_1$	4.16 (0.08)	4.16 (0.25)	7.64 (0.07)	7.64 (0.11)
$\widehat{m}_2$	3.72 (0.10)	3.72 (0.26)	7.23 (0.09)	7.22 (0.13)
$\widehat{m}_3$	5.05 (0.05)	5.05 (0.25)	7.13 (0.12)	7.12 (0.15)
$\widehat{m}_4$	4.70 (0.06)	4.70 (0.25)	7.61 (0.08)	7.61 (0.12)
θ	NA	5.69 (2.58)	NA	40.8 (31.2)
$-2\log(\text{likelihood})$	210.66	113.75	93.83	88.93

The model based on the Poisson distribution can be thought of as a special case of the model based on the NB distribution which has only one additional parameter (θ). The ratio of the likelihoods (the probability of the observed counts arising given the model) of the two models quantifies how many times more likely the data are under the NB model compared to the Poisson model. A statistical test for the improvement in likelihood due to the additional parameter is to compare twice the difference in the natural logarithm of the likelihood with a Chi-squared distribution with one degree of freedom (see for example Dobson 2002). This difference is given by 93.83 - 88.93 = 4.9 (for organisms ≥50), which has an associated p-value of 0.03 under the Chi-Squared distribution with 1 degree of freedom. For organisms ≥10 and <50 µm, this difference is even larger, with an associated p-value of <0.001. Thus, there is good evidence that the count data are over-dispersed, and that the NB distribution describes the distribution of counts in samples better than the Poisson distribution.

Next to the estimate of θ of the NB distribution (Table 9), another estimated of over-dispersion can be obtained by dividing the sum of the squared Pearson residuals $\sum\left(\frac{(y_i-\hat{m}_{j(i)})}{\hat{m}_{j(i)}}\right)^2$ of the Poisson model by the residual degrees of freedom (12 samples minus 4 estimated mean concentrations is 8 residual degrees of freedom). If the counts are indeed Poisson distributed around the predicted mean values $m_{j(i)}$, the ratio of the summed squared Pearson residuals to the degrees of freedom should be approximately equal to one. This ratio for the observed counts is $\frac{25.36}{8} = 3.17$ for organisms ≥50 µm and 17.9 for organisms ≥10 and <50 µm. This estimate of over-dispersion is equal to the parameter φ as estimated by the commonly used 'quasi-Poisson' model through quasi-likelihood estimation (Robert Wedderburn 1974). In the quasi-Poisson model, instead of specifying a full probability distribution for the observed counts, the variance is set as a multiple of the mean as specified in equation 8, and if the over-dispersion parameter is set equal to 1 the model reverts to the standard Poisson probability distribution.

Appendix D – Variability between sampling strategies and implications for the derivation of a GNC threshold

In case there are consistent trend in concentrations during the discharge events, there is a danger of obtained biased estimates of $\hat{\mu}$. There are indications in the data set that there are trend in concentrations over time (see figure 7), although it is possible that these trends are caused partly or wholly by sample handling. The implications of possible trends in concentrations are discussed in Appendix F.

Under the assumption that the over-dispersion in the count data is due to variability in rates during the discharge, the NB model may be derived as a mixture of the Gamma and Poisson distribution functions. Counts are NB distributed if they are drawn from the Poisson distribution with expected count c (rate parameter) which is itself drawn from a Gamma distribution with shape parameter β_1 and scale parameter β_2:

$$f(c;\beta_1,\beta_2) = \frac{1}{{\beta_2}^{\beta_1}}\frac{1}{\mathrm{G}(\beta_1)}c^{\beta_1-1}e^{-\frac{c}{\beta_2}}$$

$$y_i \sim \text{Poisson}(c)$$

The Gamma distribution function models the variability in concentrations during the discharge, whereas the counts in a discrete sample are expected to be Poisson distributed.

This Gamma-Poisson mixture is equal to a NB model with parameters θ and c, given the following parameterisation:
$\beta_1 = \theta$ and $\beta_2 = \Big(1 - \big(\theta/(c+\theta)\big)\Big)/\big(\theta/(c+\theta)\big)$

Using this parameterisation, the distribution of the sum of the counts in *N* systematically spaced discrete samples is given by:

$$\textstyle\sum_{i=1}^{N} y_i \sim \text{NB}(Nc, N\theta) \qquad \textbf{Eqn. 12}$$

Instead, the distribution of the sum of the counts in *M* subsamples from the concentrate of a single discrete samples is given by:

$$\textstyle\sum_{i=1}^{M} y_i \sim \text{NB}\left(Mc, \frac{c^2}{\beta_2^2\theta}\right) \qquad \textbf{Eqn. 13}$$

With this model, and with our estimate of $\text{var}(Y) = 3.17c$, the largest variance component is due to the variability in concentrations during the discharge (since under the Poisson distribution $\text{var}(Y) = c$). Therefore, increasing the sample volume by increasing the number of subsamples from the concentrate of a single discrete sample will increase the reliability of estimates less than when the sampling volume is increased by taking multiple discrete samples which are widely spaced during the discharge event.

Appendix E – Derivation of gross non-compliance thresholds

The gross non-compliance (GNC) threshold for a given sampling scheme can be derived from the NB model with various values of the parameter $\hat{\theta}$. Here, we compute GNC thresholds for $\varphi = 1$ (Scenario 1) and $\varphi = 3.17$ or 17.9 (Scenario 3). The GNC is derived by calculating for which raw count k, the probability of exceeding this count is smaller than one in a thousand, given that the true concentration in the ballast water discharge being is equal to the D-2 standard. The count k is therefore the test statistic, and the GNC threshold is calculated by computing the distribution of the test statistic under the null hypothesis.

$$P(Y \leq k; c, \theta) = \sum_{y=0}^{k} \left(\frac{\theta}{\theta + c}\right)^{\theta} \frac{\mathrm{G}(\theta + y)}{\mathrm{G}(y+1)\mathrm{G}(\theta)} \left(\frac{c}{\theta + c}\right)^{y} \qquad \textbf{Eqn. 14}$$
$$P(Y > k; c, \theta) = 1 - P(Y \leq k; c, \theta),$$

where the expectation of the raw count $c = \frac{10}{R}$ depends on the sampled volumes that determine the raising factor R (equation 1), and $\theta = \frac{c}{3.17-1}$.

For example, if counts are made in subsamples of 6 cm^3 each, from a 100 cm^3 concentrate which has been obtained by pouring a discrete discharge sample of 500 dm^3 through a sieve of 50 µm diagonal mesh, then:

- The true concentration at the D-2 standard is $\mu = 10$
- The raising factor $R = \frac{100}{3}$
- The expected count $c = \frac{3\mu}{100} = 0.3$
- The over-dispersion parameter $\theta = \frac{0.3}{3.17-1}$

As explained in Appendix D and paragraph 8, the distribution of the sum of the random variables Y when counts are made in multiples of the sample volume depends upon the sampling strategy, with Eqn. 12 for repeated discrete samples and Eqn. 13 for repeated subsampling within a single discrete sample.

Appendix F – Some evidence of a trend in concentrations during discharge: implications for sampling

There are indications in the data set that there are trend in concentrations over time; since in nearly all discharge events estimates of concentrations tend to increase over time with lowest estimates in ‘S1’ and the highest in ‘S3’ samples taken at the beginning and end respectively (see figure 1).

If we extend the statistical model for the count data as used in appendix C (Eqn. 11), and include

$y_i \sim \text{Poisson}(c_i)$

With expected values c_i for the counts per sample predicted by the following model:

$$\log(c_i) = \log\left(\frac{1}{R_i}\right) + m_{j(i)} + as_i \qquad \textbf{Eqn. 15}$$

where s_i is the numerical sequence S1=-1, S2=0, and S3=1 and the parameter *a* models the slope (linearly increasing or decreasing) for sample sequence.

Table 10 parameter estimates and log-likelihoods for the Poisson model with trend in concentration with sample sequence (Eqn. 14). The quantity $-2log(likelihood)$ is also commonly referred to as the deviance of the model, and is a measure of the relative goodness of fit the model.

	Poisson
$\hat{m}_1$	7.64 (0.07)
$\hat{m}_2$	7.20 (0.09)
$\hat{m}_3$	7.12 (0.12)
$\hat{m}_4$	7.59 (0.08)
a	0.20 (0.05)
$-2\log(\text{likelihood})$	78.22

The model with a trend in concentration with sample sequence (Eqn. 14; Table 10) gives a significantly better fit to the data compared to the model without trend (Eqn. 11; table 8). The difference in deviance of 93.83 - 78.22= 15.61 has an associated p-value of <0.001 under the Chi-Squared distribution with 1 degree of freedom. Thus, there is strong evidence that concentrations increase with increasing sample sequence. The estimated over-dispersion from the model with trend for sample sequence, by dividing the sum of the squared Pearson residuals by the residual degrees of freedom (12 samples minus 5 estimated mean concentrations is 7 residual degrees of freedom) is $\frac{9.95}{7} = 1.42$, which is a value which can be expected to occur by chance if the counts are in truth Poisson distributed. Thus, if sample sequence is taken into account there is no evidence of over-dispersion.

It is unclear whether the trend has been caused by sample handling or by some aspect of the ballast water system. Holding time (the time between taking and analysis of the samples) may be expected to influence concentrations of living organisms, since death rates are expected to increase with increasing holding time. However, there is no indication that holding time decreased with increasing sample sequence. The S1 samples have been analysed directly after sampling, whereas the S2 and S3 samples were analysed after the S3 sample was taken. Thus, the S2 sample had the longest holding time (pers. comm. Matej). We also note that there is no evidence of a trend in the uptake samples.

In the presence of an increasing trend, basing estimates on a single sample taken at the beginning of the discharge is expected to lead to an underestimate of the true concentration in the total discharge. Furthermore, the variance in the counts in such a single sample taken at the beginning may be expected to be lower than the estimated variance with over-dispersion as estimated from the model without a trend.

Appendix G – Assumptions underpinning GNC and their plausibility

This section provides a detailed list of assumption underpinning the GNC threshold as derived in the present study. Their plausibility and implications are also discussed here. In addition, Table 11 summarizes which actions can be taken to improve the acceptance of the assumptions made.

Assumptions underpinning the GNC threshold values, discussion on their plausibility and implications

1. **Assumption**: The ships and Ballast Water Management Systems sampled by Gollasch and David (2010) are representative for the fleet that will be subjected to Port State Control.
   Plausibility: The analyses were based upon data collected as part of EMSA/NEG/09/2010 (Gollasch & David, 2010). As such they represent conditions in ballast water tanks with a relatively small capacity (max 257 m^3, voyage 1) and a short holding time (12, 24 and 48h). For the derived GNC thresholds to apply, it will have to be assumed that the estimated variability in counts of samples from this limited data set is representative of other types or sizes of ballast water systems. A longer holding time on the other hand might result in higher variance due to active (swimming) or passive (settling out) movement of the organisms. The data provided by Gollasch & David (2010), did not indicate increased variance from 12h to 24h, or 48h.
   Implications: Although larger tanks and larger volumes are likely to cause the large variance in the discharges of untreated water, this should be negated as the treatment of the ballast water homogenizes the water, reducing the variation as the Ballast Water Management Systems are designed to meet a set standard.

2. **Assumption**: The sampling procedure and analytical techniques used by Gollasch and David (2010) will be used as a basis for the procedures applied during State Port Control. E.g., errors in estimating sampling volume by Gollasch and David (2010) are similar to those to be made during Port State Control.
   Plausibility: The assumption is that the sampling procedure for PSC will be very comparable to the sampling procedures used for shipboard testing.
   Implications: None if the Sampling Protocol is followed.

3. **Assumption**: The sampling strategy meets the minimum requirements associated with the selected Gross Non Compliance threshold. More specifically:
   a. The specified volumes are used for taking discrete samples during the discharge event
   b. In case more than one discrete sample is taken, discrete samples are regularly spaced throughout the discharge for example taken at the beginning, middle and end of the discharge
   c. The sample was concentrated to an appropriate volume
   d. A sufficient volume is subsampled from the concentrate

   Plausibility: The sampling strategy (e.g. volumes of discrete and subsamples) dictates which GNC threshold would apply (see Table 3, Table 4, Table 5 and Table 6). Hence, once a particular GNC threshold is selected, minimum requirements for the sampling

should be associated with that particular threshold value. Which strategy and sampling effort should be selected is an Administration's decision. This decision may be based upon the height of the GNC value that is being considered appropriate/acceptable with respect to the D-2 standard, as well as the costs and practicability associated with particular sampling protocols.
Implications: None if the Sampling Protocol is followed.

4. **Assumption**: The variance of counts in samples as observed in untreated ballast water data of Gollasch and David (2010) will be the same or less in treated ballast water.
Plausibility: It is plausible that treated ballast water is more homogenised during the treatment than untreated water and that, hence, variance in treated ballast water is smaller than observed in that of untreated water. This assumption may require further study as longer holding times than applied in the study by Gollasch & David (2010) may also result in more variance due to settling out of organisms or swimming behaviour.
Implications: Treatment of the ballast water homogenises the water, as a result the variance in organism counts is expected to be reduced. The derived GNC would therefore still hold.

5. **Assumption**: The selected model (Model 3) and the data on which it was calibrated estimates the variance conservatively (an overestimation of variance).
Plausibility: Although a model which provides a more generous threshold (scenario 3) was selected, it was parameterised with only a small dataset. The variance is also extrapolated to the D-2 standard for which no data was included in the parameterisation step. On page 19 of Gollasch and David (2010) it is stated that "… more than 40 performance tests of different BWTS on different types of vessels showed that in most sampling events no living organisms were found at all". For purposes of estimating the variance of treated ballast water, such data are very valuable, even if they are (nearly) all zero counts.
Implications: Availability of data will aid the refinement of these thresholds over time. The understanding of variation and sampling/analysis error with respect to representativeness can be improved by including more data. This should preferably be data on treated ballast water and/or water with low (near D-2 standard) organism concentrations, which would help to get a more accurate estimate of over-dispersion and possible trends in time and of variance of organism concentrations. This may result in lower gross non-compliance thresholds.

6. **Assumption**: There is no trend over time during a discharge event in concentrations of living organisms.
Plausibility: When a trend over time (beginning, middle and end of the discharge) exists, there are implications for the GNC threshold value and its use. A positive trend is observed in the dataset of Gollasch and David (2010) for organisms ≥50 µm. As this is only a small dataset, this observation may be coincidental. However, the lack of a trend should be confirmed with additional data.
Implications: None if the Sampling Protocol is followed.

7. **Assumption**: Systematic errors (bias) in the counting of numbers of living organisms in samples will lead to an underestimation of true counts of numbers of living organisms (see also the note in the section below). More specifically:
    a. Dead animals are not counted as alive.
    b. Each individual organism is counted only once
    c. <50 µm organisms are not counted as ≥50 µm organisms
    d. <10 µm organisms are not counted as ≥10 µm organisms
    e. Samples are processed and analysed directly after sampling.

    Plausibility: For the purpose of deriving a threshold for Gross Non-Compliance as described in this report, systematic errors (bias) can be ignored as long as they result in an underestimation of the true number of organisms in the treated ballast water. More specifically on each sub-assumption:
    a. As no overestimation of living organisms should be made for the purpose of GNC, organisms should not be counted if there is doubt on the state of being of the organism (although registration of the category 'doubtful' should be encouraged).
    b. The same is true with regard to the possibility of counting organisms more than once. This error can be minimized by deploying qualified staff for the analysis. In addition, if there is any doubt an organism should not be counted.
    c. Again errors can be minimized here by deploying qualified staff. For ease of counting (avoiding the need to measure every individual), it may be considered to use coarser nets to concentrate the samples (e.g. 70µm mesh diameter).
    d. The issue is important with respect to the current guidelines. Many organisms with dimensions just below 10µm occur in similar (or higher) densities as the organisms in the 10-50 µm group. A careful assessment of the size is required to estimate the effect of false counts. For the upper boundary it is not really important, as the relevant volume is so much smaller.
    e. Holding time and processing of the sample may affect survival in the sample. This may therefore lead to underestimation of true living number of organisms in the sample. This is as explained before no issue for the purpose of Gross Non-Compliance.

    Implications: None if systematic errors are minimized by following the Sampling Protocol and result in underestimations of true concentrations.

A short note on systematic errors in the data of Gollasch and David (2010)

Systematic errors result in biased estimates of the mean (µ) which are very difficult to quantify. Their magnitude and direction should be analysed in specific controlled studies. Figure 7 shows that for organisms ≥50 µm the counts in the discrete sample are overestimates of those in the sample over the entire time or vice versa (the counts in the sample over the entire time are underestimates of those in the discrete samples). This is possibly an indication systematic errors in the analysis.

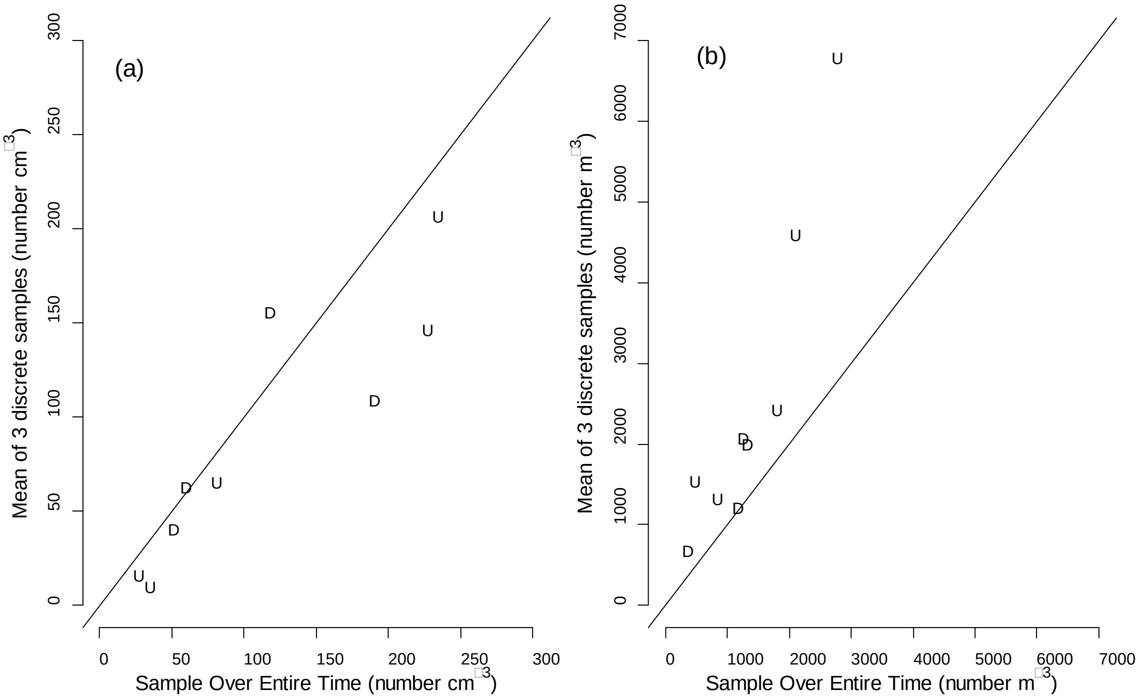


Figure 7 Observed raised counts in the sample taken over the entire time versus the raised mean count in the three discrete samples for (a) organisms ≥10 and <50 µm and (b) organisms ≥50 µm. Letters indicate specific pumping events, where 'U' are uptake events and 'D' are discharge events. The diagonal line is where y = x.

Table 11 Which actions can be taken to improve the acceptance of assumptions listed in the main text above

Assumption number	Collect more (representative) data	Use a protocol and quality assurance	Discuss in a wider community
1	X		
2		X	
3.a		X	
3.b		X	
3.c		X	
3.d		X	
4	X		X
5	X		X
6	X		X
7.a		X	X
7.b		X	X
7.c		X	X
7.d		X	X
7.e		X	X

Appendix H – A Draft Sampling Protocol for Determining Gross Non-Compliance of the D-2 Standard.

1.0 Aim

1.1 This document provides a Protocol for sampling and processing ballast water samples in such manner that the Administration can be 99.9% certain that a system is in gross non-compliance with the D-2 Standard. This sampling protocol takes into account the uncertainties arising from errors in sampling and analysis, incorporating the concept of representativeness, to provide the Administration with a test for gross non-compliance.

2.0 Background

2.1 In order to reduce the risk of ballast water spreading invasive species, the IMO adopted the “International Convention for the Control and Management of Ship’s Ballast Water and Sediments” in 2004. Ballast Water Management Systems (BWMS) are developed to treat ballast water in such a way that the ship’s discharge complies with the standards in Regulation D-2 of the Convention. Once a system is installed on board a vessel, port States and flag States are responsible for monitoring the performance of a BWMS. To do this, relevant officers may analyse the ballast water discharge to see whether the it complies to the D-2 Standards of the Convention. However, these officers have only limited options for sampling and analysis when compared to the options for testing at land-based test facilities. Therefore, the concept of gross non-compliance was developed. This is based on the premise that if a system works it is likely to produce a discharge with organism numbers at or under the D-2 Standard. If it does not, it is likely to produce a discharge with organism levels closer to those in the natural environment. Therefore, a test for gross non-compliance should set a threshold above which the number of organisms present in the discharge is so high that non-compliance can be assumed with a high level of certainty.

3.0 Limitations

3.1 The concept of gross non-compliance does not preclude any statistical substantiation of ‘full compliance’, however a different sampling protocol and data analysis will be required to test a discharge to the D-2 standard. This protocol provides an interim solution until full standardised, quick and effective methodologies for testing to the standards in the BWM Convention can be developed and refined. The Gross Non-Compliance threshold is only applicable when the presently described protocol is followed and underpinning assumptions (as described by Bierman *et al.*, 2012) are respected.

4.0 Scope

4.1 This protocol addresses sampling and analysis of organisms in both the size classes for plankton in the D-2 Standard during a discharge event. The protocol is an extension of the IMO guideline (G-2) and the practical way it has been applied during shipboard tests, as published by Gollasch & David (2010) as part of EMSA/NEG/09/2010. The statistical procedure to establish a raising factor to translate the actual counts into a probable number per m^3 (the unit of the D-2 standard) was developed by Bierman *et al.* (2012) as part of

EMSA/NEG/12/2012. This protocol is also based on the same sampling and analysis methodologies used in type approval testing.

5.0 Definitions

5.1 The Gross Non-Compliance (GNC) threshold is defined as the organism concentration below which 99.9 % of possible counts (outcomes of the sampling procedure) are expected to fall given that the true concentration in the ballast water discharge is equal to the D-2 standard.

6.0 Equipment Requirements

Sampling organisms in the discharge ≥ 50 micrometres in minimum dimension	**Sampling organisms in the discharge < 50 micrometres and ≥ to 10 micrometres in minimum. dimension**
Sampling net with cod-end and 50µm (diagonal) mesh size, flow meter and associated flexible piping to connect to the sampling point on the discharge.	
Large bucket to hold the sampling net (at least 75 l)	
Washing-bottle	Bucket to hold the main sample (approx.10 L)
20µm filter and filter apparatus	Dark storage sample bottles (100 ml)
Thermometer for registering sample temperature during sampling and holding	
100 ml graduated cylinder	For direct analysis -1 ml pipette and tips -1-2 ml vials to hold samples for flow cytometer analysis -flow cytometer
10 ml adjustable pipette and tips	
Microscope counting chamber	
Registration forms	
Dissecting microscope	For analysis at external laboratory -Polystyrene box -frozen cooling element

7.0 Step Wise Sampling Protocol

<table>
<tr><th>Sampling organisms in the discharge ≥ 50 micrometres in minimum dimension</th><th>Sampling organisms in the discharge < 50 micrometres and ≥ to 10 micrometres in minimum dimension</th></tr>
<tr><td colspan="2"></td></tr>
<tr><td colspan="2">The pipework should be set up so that the sampling point on the discharge is linked to a flow meter and then empties into a sampling net with a cod end attached, hung over the 75 l (or larger container, so that the cod end remains submerged all the time.</td></tr>
<tr><td colspan="2">A flow-rate of 3.00 m^3/h should be set from the sampling point on the discharge pipe.</td></tr>
<tr><td colspan="2">At least 2 samples of 500 litres of the discharge should be filtered; one at the beginning of the ballast water discharge and one near the end of the discharge.</td></tr>
<tr><td colspan="2">Ensure that any surplus filtered water is disposed of appropriately in line with the Ballast Water Management Plan and in accordance with the procedures on-board the ship.</td></tr>
<tr><td>Filter the discharge water during close to, but no more than, 10 minutes to obtain a maximum 500 l sample. Make a note of the sampled volume on the registration form.</td><td>Collect up to 5 litres of the filtered water over 10 minutes.</td></tr>
<tr><td>Fill a washing-bottle with the filtered water and use this to clean the inner surface of the net into the cod-end.</td><td>Carefully homogenise the sample by stirring and take a 50-100ml subsample for further analysis.</td></tr>
<tr><td>Finally flush the contents of the cod end into a sample bottle, or store directly in the cod-end (keeping the cod-end submerged in a bucket of filtrate).</td><td>Store the sample in the dark sample bottle until analysis.</td></tr>
<tr><td>Make sure that the storage container (either the cod-end or sample bottle) is of sufficient size and contains sufficient filtered water to prevent crowding effects.</td><td>If the sample needs to be stored for a longer period, pass the sample through a 50 µm mesh to ensure that grazing zooplankton are removed.</td></tr>
<tr><td colspan="2">Proceed to analyse the samples as described in Section 8</td></tr>
<tr><td colspan="2">This procedure assumes that the samples are analysed as soon as possible</td></tr>
<tr><td colspan="2">If the samples have to be send out to an analytical laboratory, pack them in a polystyrene box and protect them from being damaged.</td></tr>
<tr><td>The samples should be stored at approximately the same temperatures as the water they were taken from.</td><td>In order to prevent growth or decay, the samples should be stored in the dark at low temperatures (but not frozen) until analysis.</td></tr>
<tr><td>Some cooling may be necessary to prevent over-heating, but too much cooling can cause additional mortality. Generally the temperature should be kept within +/- 5°C of the original ballast water temperature.</td><td>Add a frozen icepack before closing the box.</td></tr>
<tr><td>Minimize exposure to daylight during storage</td><td></td></tr>
<tr><td colspan="2">No matter how samples are stored and transported, the conditions should always be measures using a thermometer, or thermologger and recorded appropriately. At least when the samples are stored and when they are analysed, and at beginning and end of any transport to a specific laboratory for analysis.</td></tr>
</table>

8.0 Analysis of the samples

Sampling organisms in the discharge ≥ 50 micrometres in minimum dimension using the sample collected from the filter (cod end) on board the ship.	**Sampling organisms in the discharge < 50 micrometres and ≥ to 10 micrometres in minimum dimension using the filtered water collected on board the ship.**
Concentrate the sample using a 20 µm filter	Homogenise the collected sample and take three 1ml subsamples for analysis.
Wash the residue from the filter into the graduated cylinder	Use a fluorescent stain to stain the sample for analysis of life cells.
Fill the cylinder to a final volume of 100ml with filtered water	Count living phytoplankton cells using a flow cytometer from 0.27 ml of each 1 ml subsample.
Carefully homogenise and take a 6 ml subsample and place it to 1 or more counting chambers	Register the counts for each subsample
Count the number of living organisms under a dissecting microscope at 60x magnification	Repeat these steps for the next sample
Register on a counting form and repeat these steps for the next sample	

9.0 Gross Non-Compliance Thresholds

9.1 The results obtained for each procedure should be added to obtain one result for organisms ≥ 50 micrometres in minimum dimension and one for organisms < 50 micrometres and ≥ to 10 micrometres in minimum dimension. These should then be then compared to the tables below. These tables contain extrapolated Gross Non-Compliance thresholds based on sample and subsample volumes and a statistically derived over-dispersion factor. If the accumulated counts from the subsamples are equal to or greater than the numbers in these tables, then it can be concluded with 99.9% certainty that the discharged water is in Gross Non-Compliance with respect to the D-2 standard. These numbers vary with the number of samples taken. Analysing an increasing number of subsamples within a sample, has less influence on the resulting thresholds.

9.2 Gross Non-Compliance thresholds for organisms greater than or equal to 50 micrometres in minimum dimension

Number of 500 L samples	Volume represented by analysing 6 ml subsample	Direct count[1]	GNC Threshold #/m^3
1	0.06 x 0.5 m^3	11	366.7
2	0.06 x 1.0 m^3	13	216.7
3	0.06 x 1.5 m^3	14	155.6
4	0.06 x 2.0 m^3	15	125.0
5	0.06 x 2.5 m^3	17	113.3

9.3 Gross Non-Compliance thresholds for organisms greater than or equal to 50 micrometres in minimum dimension less than 50 micrometres in minimum dimension and greater than or equal to 10 micrometres in minimum dimension.

Number of 5 L samples	Volume represented by analysing 0.81 ml subsample	Direct count[2]	GNC Threshold #/mL
1	1 x 5 L	94	116.0
2	2 x 5 L	119	73.5
3	3 x 5 L	139	57.2
4	4 x 5 L	158	48.8
5	5 x 5 L	175	43.2

10 References

Gollasch S. & M. David (2010) Testing sample representativeness of a ballast water discharge and developing methods for indicative analysis. GoConsult, research report to EMSA tender EMSA/NEG/09/2010.

Bierman S, P. de Vries and N.H.B.M. Kaag (2012) Testing ballast water discharges for gross non-compliance of the IMO's Ballast Water Management Convention. IMARES report to EMSA tender EMSA/NEG/12/2012.

Jørgensen, C., Gustavson, K., Hanse, J.B. & Hies, T. (2010) Development of guidance on how to analyze a ballast water sample. Final report to the European Maritime Safety Agency, EMSA (2010). EMSA, Lisboa, Portugal.

[1] Compare with sum of counts on registration form. If the sum of counts on the form is greater than or equal to the counts in this table, the sample is in gross non-compliance with the D-2 standard.