Marine Pollution Bulletin 91 (2015) 280–287

Contents lists available at ScienceDirect

Marine Pollution Bulletin

journal homepage: www.elsevier.com/locate/marpolbul

Implications of heterogeneous distributions of organisms on ballast water sampling

Eliardo G. Costa [a,*], Rubens M. Lopes [b], Julio M. Singer [a]

[a] Departamento de Estatística, Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, Brazil
[b] Instituto Oceanográfico, Universidade de São Paulo, São Paulo, Brazil

ARTICLE INFO

Article history:
Available online 12 December 2014

Keywords:
Negative binomial distribution
Non-homogeneous Poisson process
Poisson distribution
Stratified sampling

ABSTRACT

Ballast water sampling is one of the problems still needing investigation in order to enforce the D-2 Regulation of the International Convention for the Control and Management of Ship Ballast Water and Sediments. Although statistical "representativeness" of the sample is an issue usually discussed in the literature, neither a definition nor a clear description of its implications are presented. In this context, we relate it to the heterogeneity of the distribution of organisms in ballast water and show how to specify compliance tests under different models based on the Poisson and negative binomial distributions. We provide algorithms to obtain minimum sample volumes required to satisfy fixed limits on the probabilities of Type I and II errors. We show that when the sample consists of a large number of aliquots, the Poisson model may be employed even under moderate heterogeneity of the distribution of the organisms in the ballast water tank.

1. Introduction

The D-2 Regulation of the International Convention for the Control and Management of Ship Ballast Water and Sediments adopted by the International Maritime Organization (IMO) in 2004 sets upper limits on the concentrations of viable organisms in ballast water discharges to which ships must comply with (IMO, 2004). In particular, this regulation requires that ballast water discharged by ships contain (i) fewer than 10 viable organisms with minimum dimension $\geqslant 50\ \mu m$ per m^3, and (ii) fewer than 10 viable organisms with minimum dimension between 10 μm and 50 μm per mL.

Because of cost and time restrictions, examination of the entire volume of discharge is not feasible and sampling must be employed instead. In addition to the D-2 regulation, the G2 guideline from the IMO Guidelines on Sampling of Ballast Water (IMO, 2008) states that *"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."*

Although there has been some discussion regarding whether the D-2 regulation refers to any volume of water sampled from the discharged ballast water or to the entire discharge, authors like Gollasch et al. (2007) or Frazier et al. (2013) suggest that the latter interpretation seems more adequate. This is the approach we adopt. Issues concerning the implications of statistical "representativeness" of the sample are also extensively discussed in the literature. Experimental and simulated data to explore this aspect of the regulation was considered by Miller et al. (2011) and Carney et al. (2013), but also fall short of an explicit definition of a "representative sample". These two aspects of ballast water sampling seem to be the source of confusion and difficulty in setting the track required to enforce the IMO regulation.

A binomial model to compute the sample size "required to ensure that the sample proportion is representative of the population" was adopted by Basurko et al. (2011), but ground their solution on some rather restrictive assumptions. Furthermore, their results suggest that very large volumes of discharged ballast water should be sampled to assure compliance with reasonable accuracy. Poisson distribution models which lead to more realistic results was employed by Miller et al. (2011). Both approaches, however, are based on the assumption that the concentration of viable organisms (hereafter referred to as "concentration") is homogeneous throughout the ballast water tank. This may not be reasonable in practice (Miller et al., 2011; Carney et al., 2013). More recently, (Bierman et al. (2012), Costa (2013) and Frazier et al. (2013)), working independently, suggested negative binomial (NB) models that take the expected heterogeneity of the concentration into account.

* Corresponding author at: Caixa Postal 66281, São Paulo, SP 05314-970, Brazil. Tel.: +55 (11) 3091 6118; fax: +55 (11) 3091 6130.
E-mail addresses: eliardo@ime.usp.br (E.G. Costa), rubens@usp.br (R.M. Lopes), jmsinger@ime.usp.br (J.M. Singer).

http://dx.doi.org/10.1016/j.marpolbul.2014.11.030
0025-326X/

Table 1
Summary of the notation employed in the text.

X	Number of organisms detected in the sample
V	Total volume of the ballast water tank
V_h	Volume of stratum h of the ballast water tank
v	Sample volume
λ	Concentration in the ballast water tank
λ_h	Concentration in stratum h of the ballast water tank
$\delta,\ \gamma,\ \eta,\ \phi$	Parameters that define the form of the function that describes the variation of the concentration in the ballast water tank
α	Probability of Type I error (declaring non-compliant a ship that in reality complies with the D-2 regulation)
β	Probability of Type II error (declaring compliant a ship that in reality does not comply with the D-2 regulation)

A formal definition based on the common sense that "representative" samples are scaled down versions of the population and methods for selecting them was provided by Grafström and Schelin (2014). However, they mention that *"it is sometimes of more interest to have a small variance of an estimator than to have representative samples"*. As the D-2 regulation only requires that the overall concentration in the ballast water tank be smaller than a specified limit, from the statistical point of view, "representative" samples are those for which the analysis allows inference on this concentration with a specified accuracy.

Given that accuracy is directly related to the underlying statistical model and that different models are based on varying assumptions on the distribution of organisms in the tank, heterogeneity plays a major role in the quest for statistical representativeness. Based on empirical results, the Poisson model may be used even when the organisms are not randomly distributed within the ballast water tank, provided the sample consists of multiple independent aliquots (Frazier et al., 2013). We develop this issue further and propose specific models to accommodate different assumptions on the heterogeneity of the distribution and derive threshold values for determining whether a ship complies with the D-2 regulation under different sampling schemes and precision requirements.

In the second section we outline the technical details, compute the thresholds for each scenario and illustrate the procedures with data from Bierman et al. (2012). We conclude with the discussion of some practical issues and suggestions for future research in the third section.

2. Statistical methods and results

We develop our discussion referring to the regulation concerning viable organisms (hereafter referred to as "organisms") with minimum dimension $\geqslant 50\ \mu$m. The results are directly applicable to organisms with minimum dimension between $10\ \mu$m and $50\ \mu$m by changing the measurement units appropriately.

The parameter we are concerned with is the concentration λ per m^3 in the discharged ballast water. In statistical terms, deciding whether a ship complies or not with the D-2 regulation is equivalent to testing the null hypothesis $\lambda \leqslant 10$ *versus* the alternative hypothesis $\lambda > 10$ based on a sample of v m^3 of ballast water obtained from one or more aliquots collected during the discharge. The test procedure consists in determining a *compliance threshold* (c) in terms of the number of organisms observed in the sample, above which the ship is declared non-compliant. The compliance threshold is obtained by assuming an appropriate probability model and fixing the probability of Type I error (α) as well as the probability of Type II error [$\beta(\lambda_A)$] for a given value of the concentration $\lambda_A > 10$. Type I errors occur if we decide for non-compliance when, in fact, $\lambda \leqslant 10$; Type II errors occur if we decide for compliance when, in fact, $\lambda > 10$. The value $1-\beta(\lambda_A)$ is the power of the test to detect a non-compliant ship for which the concentration is $\lambda = \lambda_A > 10$. Ideally, α and β should be very small, but this may require large sample volumes and their choice must be based on cost, time and technical considerations.

A crucial point in this process relates to the choice of the probability model. We approach this problem by considering different assumptions on the distribution of organisms in the discharged ballast water. A summary of the notation employed in the text is presented in Table 1.

2.1. Sampling with homogeneous concentration under a Poisson model

We start by assuming an homogeneous distribution, in which case a Poisson model with mean $v\lambda$ may be adopted to compute the probability of observing the number X of organisms in a sample with volume v, *i.e.*,

$$P(X = k|v,\lambda) = \exp(-v\lambda)(v\lambda)^k/k!, \quad k = 0, 1, \ldots. \tag{1}$$

The decision rule is to declare non-compliance if $X > c$ where c is determined from

$$\alpha = P(X > c|v, \lambda = 10) \quad \text{and} \quad \beta(\lambda_A) = P(X \leqslant c|v, \lambda = \lambda_A), \tag{2}$$

for some specified $\lambda_A > 10$. If the sample volume v is composed of n independent ballast water aliquots with volumes w_i collected at different time intervals during the discharge and X_i denotes the number of organisms observed in the i-th aliquot, then $X = \sum_{i=1}^{n} X_i$ follows a Poisson distribution with mean $\sum_{i=1}^{n} w_i\lambda = v\lambda$ and the decision procedure is the same as described above. The reader is referred to Haight (1967) for details. An algorithm to compute the minimum sample volume v (or the number n of aliquots, each with volume w) required to satisfy pre-specified limits (α and β) on the probabilities of Type I and II errors under (1) and (2) is described as follows[1]

Step 1. Set initial values for $\alpha,\ \beta,\ w$ (or v_0) and $\lambda = \lambda_A > 10$;
Step 2. Take $n = 2$ (or $v = v_0$);
Step 3. Compute c through the first equation in (2). Then, with this value of c, compute $\beta(\lambda_A)$ through the second equation in (2);
Step 4. If $1-\beta(\lambda_A)\ \geqslant\ 1-\beta$, stop. The value n (or v) obtained in this step is the required value. Otherwise, set $n = n+1$ (or $v = v + \epsilon$, where ϵ is some convenient value depending on the dimension of the tank. For example, if the tank has a volume of 1000 m^3 we may set $\epsilon = 0.001$ m^3) and return to step 3.

Using this algorithm and specifying different limits for the probabilities of Type I error (α) and of Type II error for $\lambda_A = 12$, [$\beta(\lambda_A)$], we obtained the required sample volume (v), the corresponding compliance threshold (c) and the power of the procedure to detect different non-compliant concentrations (Table 2).

Setting $\alpha = 0.05$ and $\beta(12) = 0.05$, for example, the required sample volume is $v = 29.78$ m^3 and the decision rule is to declare compliant ships for which the number of organisms X in the sam-

[1] An R code for its practical implementation is presented in the Supporting Information.

Table 2
D-2 regulation required sample volumes (v) in m^3, compliance threshold (c) and power $[1-\beta(\lambda_A)]$ to detect non-compliant concentrations ($\lambda_A > 10$) based on a Poisson model.

α	β	Sample volume (v)	Compliance threshold (c)	Detection power for $\lambda_A =$			
				11.5	12	12.5	13
0.05	0.05	29.78	326	0.81	0.95	≈1	≈1
0.05	0.10	23.50	260	0.72	0.90	0.98	≈1
0.10	0.05	23.75	257	0.83	0.95	≈1	≈1
0.10	0.10	18.11	198	0.75	0.90	0.97	≈1

ple is $\leqslant$326 and non-compliant, ships for which $X > 326$. In this setup, if the actual concentration in the discharged ballast water is $\lambda_A = 11.5$, 12 or 12.5 org m^{-3}, the probabilities of decisions for non-compliance are, respectively, $1-\beta(11.5) = 0.81$, $1-\beta(12) = 0.95$ and $1-\beta(12.5) \approx 1$.

In practice, the homogeneity assumption underlying the Poisson model is seldom sustainable. The distribution of organisms in the discharged ballast water depends on tank configuration, discharge flow, uptake location, light and oxygen conditions, sediment ressuspension, duration of transoceanic voyage, types of organisms present, among other factors (Murphy et al., 2002; Frazier et al., 2013). The exact distribution is difficult, if not impossible, to determine a priori. Furthermore, in practice, overdispersion, *i.e.*, variance larger than the mean, is usually observed (Miller et al., 2011; Bierman et al., 2012); this invalidates the Poisson model. To bypass this problem, we identify alternative models and indicate the decision rule in each case.

2.2. Sampling with heterogeneous concentration under a stratified Poisson model

Consider the situation where the discharge may be divided into H strata with volumes V_h and concentrations $\lambda_h,\ h = 1,\ldots,H$ so that the overall volume is $V = \sum_{h=1}^{H} V_h$ and the concentration is $\lambda = \sum_{h=1}^{H} W_h\lambda_h$, where $W_h = V_h/V$. The strata may be defined according to different water levels or considering different tanks, for example. Note that we interpret the D-2 regulation as requiring that $\lambda \leqslant 10$ and not that each $\lambda_h \leqslant 10$.

Assume now, that the sample volume v is obtained by collecting a ballast water aliquot with volume $v_h = n_h w$ in stratum $h, h = 1,\ldots,H$, where w is a pre-specified aliquot volume, n_h is the number of aliquots sampled in stratum h, and that the corresponding number of organisms in an aliquot with volume w, X_h, follows a Poisson model with mean λ_h. We consider the estimator $\widehat{\lambda}_h = n_h^{-1}\sum_{i=1}^{n_h} X_{hi}$ for $\lambda_h,\ h = 1,\ldots,H$, and $\widehat{\lambda} = \sum_{h=1}^{H} W_h\widehat{\lambda}_h$ for λ. Our objective is to determine the minimum sample volume such that $P(|\widehat{\lambda} - \lambda| < \epsilon) > 1-\alpha$ where ϵ denotes an acceptable difference between the estimator and the true value of λ. For such purpose, it suffices to establish n_h such that $P(|\widehat{\lambda}_h - \lambda_h| < \epsilon_h) > 1-\alpha_h$ with $\epsilon_h = \epsilon/W_h H$ and $\alpha = \sum_{h=1}^{H}\alpha_h$, provided that the minimum (a) and maximum (b) values for the concentration in stratum h are specified, *i.e.*, that $\lambda_h \in [a,b]$. This may be accomplished via the results of Chen (2007) outlined in Appendix A.

For example, consider the total discharge from a tank with volume 270 m^3 in four strata with volumes $V_1 = 135$ m^3, $V_2 = 75$ m^3, $V_3 = 40$ m^3 and $V_4 = 20$ m^3, a sample aliquot $w = 0.001$ m^3, an estimation error $\epsilon = 1$, and that $\lambda_1 \in [1,25]$, $\lambda_2 \in [1,40]$, $\lambda_3 \in [1,30]$ and $\lambda_4 \in [1,60]$, and $\alpha_1 = 0.02$, $\alpha_2 = \alpha_3 = \alpha_4 = 0.01$ which implies $\alpha = 0.05$. The required number of aliquots in strata 1–4 are, respectively, $n_1 = 544$, $n_2 = 329$, $n_3 = 71$, and $n_4 = 36$, which implies $v_1 = 0.544$ m^3, $v_2 = 0.329$ m^3, $v_3 = 0.071$ m^3, and $v_4 = 0.036$ m^3. We declare non-compliance if $\widehat{\lambda} > 11$ $(= 10 + \epsilon)$ with a probability of Type I error of $\alpha = 0.05$. [2]

[2] An R code for its practical implementation is presented in the Supporting Information.

2.3. Sampling with heterogeneous concentration under a non-homogeneous Poisson process model

If the concentration varies continuously along a given dimension (t) such as volume of discharged ballast water or discharge time, a non-homogeneous Poisson process (Basawa and Rao, 1980; Ross, 1996) may be employed to model the number of organisms observed in the sample. According to this model, the number of organisms observed over the continuous deballasting process follows a Poisson distribution with mean $\int_0^V \lambda(t|\boldsymbol{\theta})dt$, where $\lambda(t|\boldsymbol{\theta})$ is a function of known form, depending on parameters represented by the vector $\boldsymbol{\theta}$. The concentration in the tank is $\lambda = V^{-1}\int_0^V \lambda(t|\boldsymbol{\theta})dt$. Typical examples for the functions relating the concentration and the dimension along which we measure the deballasting process are

$$\lambda(t|\delta,\gamma) = \delta\gamma t^{\gamma-1} \quad \text{and} \quad \lambda(t|\delta,\gamma,\eta) = \eta + [(t-\gamma)/\delta]^2,$$

where δ, γ and η are parameters (elements of $\boldsymbol{\theta}$) to be estimated. Plots of such functions are displayed in Fig. 1. Here, we assume that the dimension along which the concentration varies is the deballasting volume. In the left panel, for example, the case $(\delta = 5, \gamma = 1)$ corresponds to a uniform concentration throughout the ballast water discharge, indicating that a Poisson model may be employed for analysis; the case $(\delta = 8, \gamma < 1)$, on the other hand, represents situations where the concentration decreases along the deballasting process. In the right panel, the function corresponds to a situation where there are large concentrations at the beginning of the discharge, small in the middle, and large again at the end. The parameter η may be viewed as the minimum concentration occurring in the deballasting process, γ is the deballasted volume at which the minimum concentration is reached and δ is associated with the rate at which the concentration decreases or increases.

Assuming that n aliquots with volumes $v_1,\ldots,v_n$ (negligible with respect to the total volume of discharged ballast water, V) are collected along the deballasted volumes $t_1,\ldots,t_n$ and that $x_1,\ldots,x_n$ represent the associated number of observed organisms, the corresponding probability function is

$$P(\boldsymbol{x}|\mathbf{t},\mathbf{v},\boldsymbol{\theta}) = \prod_{i=1}^{n}\exp[-v_i\lambda(t_i,\boldsymbol{\theta})][v_i\lambda(t_i,\boldsymbol{\theta})]^{x_i}/x_i!, \tag{3}$$

where $\mathbf{t} = (t_1,\ldots,t_n)$, $\mathbf{v} = (v_1,\ldots,v_n)$ and $\boldsymbol{x} = (x_1,\ldots,x_n)$. Maximum likelihood estimates (MLE) of $\boldsymbol{\theta}$ may be obtained by maximizing the likelihood corresponding to (3) and used to obtain an estimate of λ. The covariance matrix of the estimator of $\boldsymbol{\theta}$ may be obtained from the Fisher information matrix and the delta method may be employed to obtain the standard errors of its components (Sen et al., 2009).

Letting $\widehat{\lambda}$ denote the MLE of λ and $SE(\widehat{\lambda})$ the corresponding standard error, the decision rule is to declare non-compliance if the statistic $Z = (\widehat{\lambda} - 10)/SE(\widehat{\lambda}) > c$ with the compliance threshold c obtained from $\alpha = P(Z > c)$, where Z follows a standard normal distribution, and to declare compliance, otherwise. This procedure, however, requires a large number n of aliquots.

Unfortunately, there are not many studies designed to identify the form of the curves representing the variation in the concentration along the discharged volume; a study paving the way in that

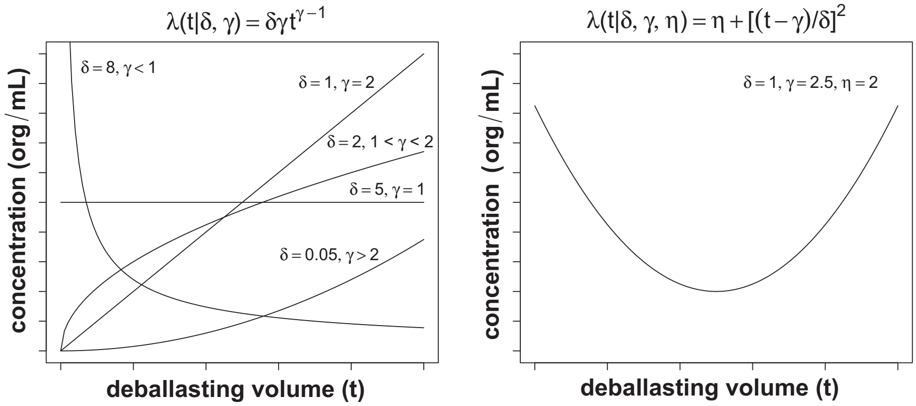

Fig. 1. Plots of the concentration function $\lambda(t|\boldsymbol{\theta})$.

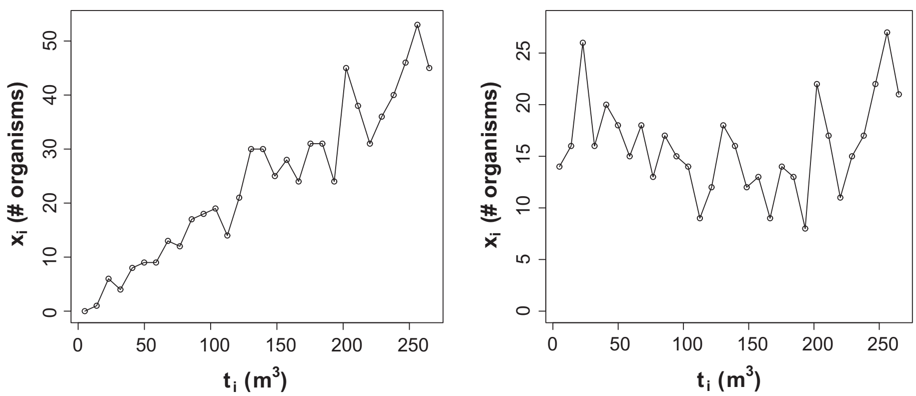

Fig. 2. Simulated counts of organisms for different volumes (t_i) of discharged ballast water using (3) with $\lambda(t|\delta,\gamma) = \delta\gamma t^{\gamma-1}$, $\delta = 0.05$ and $\gamma = 2.1$ ($\lambda = 23.63$) in the left panel and with $\lambda(t|\delta,\gamma,\eta) = \eta + [(t-\gamma)/\delta]^2$, $\delta = 40$, $\gamma = 130$ and $\eta = 12.5$ ($\lambda = 16.31$) in the right panel.

direction is presented in First et al. (2013). We consider two examples with simulated data using the functions represented in Fig. 1, letting t correspond to the discharged volume since the beginning of the deballasting process, assuming a total discharged volume of $V = 270\ \mathrm{m}^3$ and taking aliquots with volume $v_i = 1\ \mathrm{m}^3$. We set $\lambda(t|\delta,\gamma) = \delta\gamma t^{\gamma-1}$ with $\delta = 0.05$ and $\gamma = 2.1$, so $\lambda = \delta V^{\gamma-1} = 0.05 \times 270^{1.1} = 23.63$, and $\lambda(t|\delta,\gamma,\eta) = \eta + [(t-\gamma)/\delta]^2$ with $\delta = 40, \gamma = 130$ and $\eta = 12.5$, so $\lambda = \eta + [(V-\gamma)^3 + \gamma^3]/3V\delta^2 = 12.5 + (140^3 + 130^3)/(810 \times 40^2) = 16.31$. The data are presented in the Supporting Information and depicted in Fig. 2. The MLE of δ, γ, η and λ along with the corresponding standard errors are presented in Table 3. Setting $\alpha = 0.05$ it follows that the compliance threshold is $c = 1.64$ and since $Z = 13.6/0.89 = 15.34 > 1.64$ in the first case and $Z = 5.96/2.5 = 2.39 > 1.64$ in the second case, we declare non-compliance for both.

Table 3
Estimates obtained from simulated counts under model (3) with $\lambda(t|\delta,\gamma) = \delta\gamma t^{\gamma-1}$ with $\delta = 0.05$ and $\gamma = 2.1$ (Case 1), and $\lambda(t|\delta,\gamma,\eta) = \eta + [(t-\gamma)/\delta]^2$ with $\delta = 40$, $\gamma = 130$ and $\eta = 12.5$ (Case 2) along with standard errors and 95% confidence intervals (CI).

Case	Parameter	Estimate	Standard error	CI (95%)	
				Lower limit	Upper limit
1	δ	0.08	0.03	0.03	0.18
	γ	2.01	0.07	1.87	2.17
	λ	23.60	0.89	21.86	25.34
2	δ	45.45	6.67	36.10	69.08
	γ	133.67	10.33	108.38	156.04
	η	13.02	1.03	11.07	15.12
	λ	15.96	2.50	11.07	20.85

2.4. Sampling with heterogeneous concentration under a negative binomial model

When there is little (or no) information about how the concentration varies along the deballasting process, an alternative is to consider it as a random variable L following some convenient probability distribution. A possible approach is to assume that L follows a gamma distribution, for which the probability density function is

$$f(\ell|\lambda,\phi) = \frac{1}{\Gamma(\phi)}\left(\frac{\ell\phi}{\lambda}\right)^{\phi}\exp(-\phi\ell/\lambda)/\ell, \quad \ell > 0,$$

where $\Gamma(x) = \int_0^\infty t^{x-1}\exp(-t)dt$ is the gamma function. This implies that the concentration is $\mathbb{E}(L) = \lambda$ and that the corresponding variance is $\mathbb{V}(L) = \lambda^2/\phi$ (Johnson et al., 1994). This distribution has been used in many settings mainly because of its flexibility (see Johnson et al., 1994 and references therein, for example). Different values of its two parameters (λ,ϕ) correspond to different ways in which the concentration varies in the ballast tank. Plots of the gamma distribution for different values of the parameters λ and ϕ are represented in Fig. 3. For example, the cases $(\lambda = 5, \phi = 20)$ and $(\lambda = 20, \phi = 350)$ correspond to situations where the concentration in the ballast water discharge varies approximately between 2.5 and 7.5 org m^{-3}, and between 17 and 24 org m^{-3}, respectively. The case $(\lambda = 18, \phi = 3)$ corresponds to situations where the concentration in the ballast water discharge varies between 0 and more than 25 org m^{-3}.

Conditionally on a specific value $L = \ell$ for the concentration and on ϕ, assume that the number of organisms (X) in a ballast water aliquot of w m^3 follows a Poisson distribution with mean $w\ell$. This

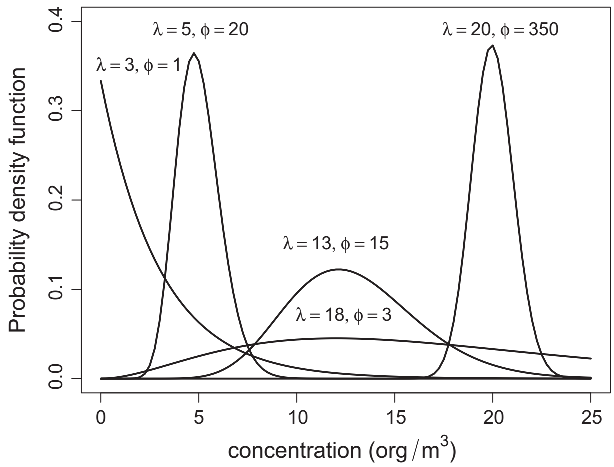

Fig. 3. Plots of the gamma distribution for different parameter values.

implies that (unconditionally) X follows a NB distribution with mean $w\lambda$ and dispersion parameter ϕ, *i.e.*, with probability function

$$P(X=k|\phi,w,\lambda)=\frac{\Gamma(\phi+k)}{\Gamma(k+1)\Gamma(\phi)}\left(\frac{w\lambda}{w\lambda+\phi}\right)^{k}\left(\frac{\phi}{w\lambda+\phi}\right)^{\phi},\quad \lambda,\phi>0, \tag{4}$$

for $k=0,1,2,\ldots$. In this context, $\mathbb{E}(X)=w\lambda$ and $\mathbb{V}(X)=w\lambda+(w\lambda)^2/\phi$, indicating that this model may accommodate a possible over-dispersion.

Suppose that n aliquots of w m^3 have been randomly collected from the discharged ballast water and that the corresponding number of organisms, $X_1,\ldots,X_n$ have independent Poisson distributions conditionally on their means, that, in turn, follow a gamma distribution. As a consequence, $X_i,\ i=1,\ldots,n$ follow NB distributions and $X=\sum_{i=1}^{n}X_i$ follows a NB distribution with mean $nw\lambda$ and dispersion parameter $n\phi$, *i.e.*, $\mathbb{E}(X)=nw\lambda=v\lambda$ and $\mathbb{V}(X)=nw\lambda+(nw\lambda)^2/n\phi=v\lambda+(v\lambda)^2/n\phi$. The reader is referred to Hilbe (2007) for details. In this case, the compliance threshold c is determined from

$$\alpha=P(X>c|\phi,w,\lambda=10)\quad\text{and}\quad\beta(\lambda_A)=P(X\leqslant c|\phi,w,\lambda=\lambda_A), \tag{5}$$

with

$$P(X=k|\phi,w,\lambda)=\frac{\Gamma(n\phi+k)}{\Gamma(k+1)\Gamma(n\phi)}\left(\frac{w\lambda}{w\lambda+\phi}\right)^{k}\left(\frac{\phi}{w\lambda+\phi}\right)^{n\phi},\quad \lambda,\phi>0, \tag{6}$$

for $k=0,1,2,\ldots$. For large ϕ or n the distribution of X may be aproximated by a Poisson distribution with mean $v\lambda$ (see Appendix B for technical details). This is in accordance with the empirical results mentioned in Frazier et al. (2013). In Fig. 4 we indicate plots of the NB distribution for different values of λ and ϕ along with the Poisson distribution with the same mean.

An algorithm to determine c and the number n of w m^3 aliquots such that the probability of Type I error and the probability of Type II error for a given value of $\lambda=\lambda_A>10$ are pre-specified (assuming ϕ is known) is essentially the same as that considered for the Poisson distribution with the substitution of (2) by (5). Using this algorithm and setting different limits on α and $\beta(12)$, we obtained the required sample volume ($v=nw$) with $w=0.001$ m^3 for different values of ϕ, the corresponding compliance threshold (c) as well as the power of the procedure to detect different non-compliant concentrations. The results are displayed in Table 4.

Setting $\alpha=0.05$ and $\beta(12)=0.05$, for example, the required minimum sample volume is $v=32.98$ m^3 and the compliance threshold is $c=361$ when $\phi=0.1$. The results presented in the last three rows of Table 4 are approximately equal to the results presented in Table 2, suggesting that for $\phi\geqslant 5$ we may use the Poisson distribution as an approximation to the NB distribution.

We illustrate the methods using data from Bierman et al. (2012), originally collected by Gollasch et al. (2010). The data, corresponding to organisms with minimum dimension between 10 μm and 50 μm per mL, are displayed in the Supporting Information. Three aliquots with volume $v=0.27$ mL were collected at the beginning, three at the middle and three at the end of both the uptake and the discharge process.

Assuming a Poisson model and letting $\alpha=0.05$, with a sample of $v=3\times3\times0.27$ mL $=2.43$ mL, the compliance threshold is $c=33$. Since the number of observed organisms in the uptake sample is $X=27$, we must declare compliance. For the discharge, we have $X=100$ and the decision would for be non-compliance. The corresponding power to detect a concentration $\lambda_A=12$ is $1-\beta(12)=0.21$. If additionally, we set $\beta(12)=0.10$, the required minimum sample volume is $v=23.76$ mL and the compliance threshold is $c=263$.

The MLE of ϕ for the NB distribution obtained from the uptake data is $\widehat{\phi}=100$ and suggests that the Poisson distribution is an adequate option. For the discharge, on the other hand, $\widehat{\phi}=1.66$, suggesting that the Poisson model may not be acceptable.

Assuming a NB model with $\alpha=0.05$, with a sample of $v=3\times3\times0.27$ mL $=2.43$ mL, the compliance threshold is $c=39$ (in contrast with $c=33$ for the Poisson distribution). Since $X=100$ in the uptake sample, the decision would be for declaring non-compliance. Here, the corresponding power to detect a concentration $\lambda_A=12$ is $1-\beta(12)=0.13$. If we set $\beta(12)=0.10$, the required minimum sample volume is $v=65.88$ mL and the compliance threshold is $c=728$. To compute the MLE, we used the `fitdistr` function of the MASS package (R Development Core Team, 2013).

3. Discussion

The IMO Convention on ballast water management was approved in 2004 but has not been enforced after 10 years (one of the longest implementation phases in the organization's history) partly because many practical issues remain unresolved. By late-2014, 43 countries corresponding to 32.5% of the world's merchant fleet tonnage had ratified the Convention, but a minimum of 35% is required. One of the most critical dispute is towards a standardized, universally applicable sampling and analytical method for compliance assessment during port state control inspections (ICS, 2013). Once documented, a faulty performance of an installed ballast water treatment system may lead to serious legal penalties to crew and shipowners, meaning that inspection must be based on solid grounds. To assess the effective concentration in the ballast water discharge line, as required by the Convention, the inspection protocol must rely on sampling, as it is not practical to scan the entire ballast water volume within the time constraints of maritime operations. The IMO Convention provides generic sampling guidelines; for instance, the G8 guidelines for shipboard evaluation of treatment systems (IMO, 2004; Gollasch et al., 2007) requires sampling at three stages of the deballasting process (beginning, middle and end of the discharge), as an attempt to depict the heterogeneous distribution of the concentration within tanks. In Carney et al. (2013), the authors have shown that a three-stage sampling effort as described above is insufficient even for low-volume (1 m^3) test tanks, implying that the actual concentration in the ballast discharge may be quite different from that estimated via the proposed sampling protocols, including the one currently recommended by IMO.

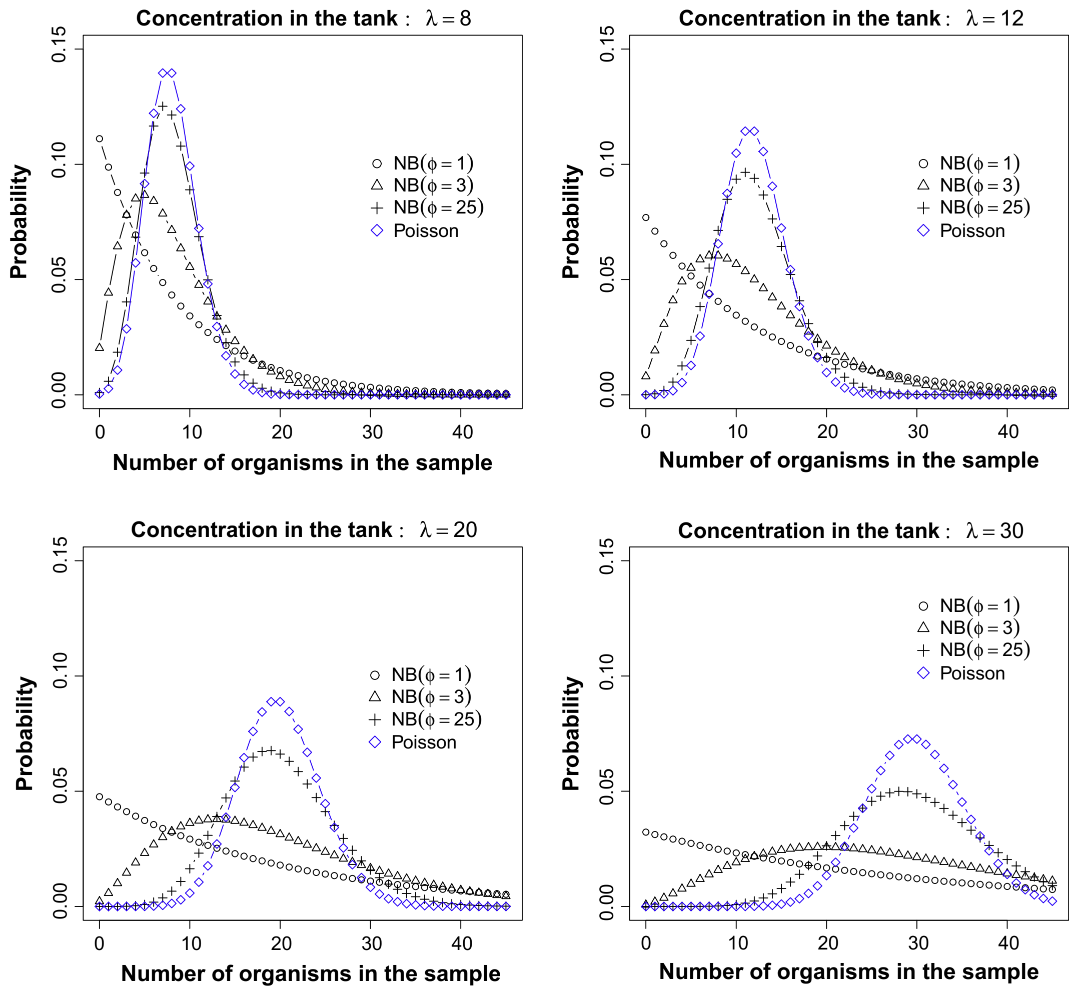

Fig. 4. Plots of the negative binomial distribution for different values of the parameter ϕ along with the Poisson distribution.

Table 4
D-2 regulation required sample volumes (v) in m^3, compliance thresholds (c) and power $[1-\beta(\lambda_A)]$ to detect non-compliant concentrations ($\lambda_A > 10$) based on a negative binomial model with $\phi = 0.01,\ 0.1$ and 5.

ϕ	α	β	Sample volume (v)	Compliance threshold (c)	Detection power for $\lambda_A =$			
					11.5	12	12.5	13
0.01	0.05	0.05	62.36	682	0.81	0.95	≈1	≈1
	0.05	0.10	49.11	543	0.73	0.90	0.97	≈1
	0.10	0.05	49.67	537	0.83	0.95	≈1	≈1
	0.10	0.10	37.89	414	0.75	0.90	0.97	≈1
0.1	0.05	0.05	32.98	361	0.81	0.95	≈1	≈1
	0.05	0.10	26.03	288	0.72	0.90	0.97	≈1
	0.10	0.05	26.25	284	0.83	0.95	≈1	≈1
	0.10	0.10	20.03	219	0.75	0.90	0.97	≈1
5	0.05	0.05	29.78	326	0.80	0.95	≈1	≈1
	0.05	0.10	23.50	260	0.72	0.90	0.98	≈1
	0.10	0.05	23.66	256	0.83	0.95	≈1	≈1
	0.10	0.10	18.11	198	0.75	0.90	0.97	≈1
10	0.05	0.05	29.78	326	0.80	0.95	≈1	≈1
	0.05	0.10	23.49	260	0.72	0.90	0.98	≈1
	0.10	0.05	23.66	256	0.83	0.95	≈1	≈1
	0.10	0.10	18.11	198	0.75	0.90	0.97	≈1

Previous studies have suggested that increasing sample volumes provide greater precision and confidence in concentration estimates during ballast water testing for compliance (Basurko et al., 2011; Miller et al., 2011). The frequency of sample collection from the ballast water discharge is also likely to affect data accuracy, hence sampling at very frequent intervals (few minutes) should be required, raising concerns on ship operational delays (Carney et al., 2013). Such problems in ballast water sampling and analysis are connected to the heterogeneous distribution of organisms in ballast water tanks (Murphy et al., 2002). We provide evidence that appropriate statistical models may accommodate different assumptions on the heterogeneity of the concentration in the tank, paving the way to a more flexible, data-oriented sampling approach. In this context, we propose an algorithm to compute sample size, and specify the decision rule for each scenario. Threshold values may be computed for determining whether a ship complies with the D-2 regulation under different precision levels and sampling protocols. In particular, we suggest that, irrespectively of the distribution of organisms in the ballast water tank, a Poisson model may be employed to determine the sample volumes, provided it consists of a large number of aliquots collected along the deballasting process.

More information may be acquired once we gain more experience on how organisms are distributed in ballast water tanks and in the discharge (Frazier et al., 2013) and this may definitely lead to enhanced sampling schemes. The decision-making process on the ideal sampling protocol is further complicated by the combination of large discharge volumes during short time intervals with the low concentration expected when effective on board treatment systems are used. Novel monitoring approaches are urgently needed not only to enhance the actual process of counting organisms but also to decrease the risk of unintentional biological invasions through ship ballast water discharge without hampering routine maritime operations. Sampling and analytical techniques for evaluation of ballast water biology have so far relied on traditional methods derived from environmental assessment studies. Taxonomic composition, size determination and concentration estimates have been analyzed from samples collected with plankton nets, hydrographic bottles and small pumps, from restricted sampling points such as manholes and sounding pipes. Such sampling strategies suffer from several drawbacks including: (i) in-tank spatial and temporal variability may not be adequately evaluated with commonly-used sampling gear; (ii) sampling at the discharge line may hardly be accomplished on a routine basis; (iii) sampling and analysis are operator-dependent and time-consuming activities; (iv) no real-time information may be generated during ballast water operations, rendering monitoring practices a post-hoc verification practice; and (v) statistical constraints are not fully addressed when manual sampling strategies represent the only option available.

New sampling and analysis protocols for ballast water compliance assessment are expected to become available in the near future, and hopefully the concept of an in-line sampling method will become feasible as technology advances. Such system should be capable to record size and concentration data during the ballast water discharge, and use the data just collected as input to a statistical model similar to the ones proposed here. The model will then provide estimates of sampling volume and frequency for a subsequent data collection interval, in a continuous feedback loop until a decision on ship compliance may be achieved within acceptable confidence levels while keeping human supervision to a minimum. In other words, the future of ballast water monitoring for compliance is in standardized, ship-based automatic or semi-automatic sampling and analytical protocols, where reliable estimates and quality control are provided by robust, embedded statistical models.

The methods proposed here elucidate certain statistical aspects of ballast water sampling that have been constantly raised in the literature. The models are applicable to the same sampling schemes as employed at present, despite the amount of effort required to collect information on size and concentration at the necessary frequency and volume. Unfortunately we have no real world data at present to further assess the adequacy of the proposed models. However, ship-board tests are now under way to verify the behavior of the different models under contrasting organisms distributions. We expect that practical results will be available within one year.

Although other parametric mixtures, such as the Poisson-lognormal may be considered, the lack of knowledge about the distribution of organisms in the ballast water tank complicates the choice. We propose to use a nonparametric model based on a Dirichlet process mixture (see Müller and Quintana, (2004), for example) to allow for the required flexibility. We are also investigating sequential methods based on the proposed models to test for compliance with the D-2 norm during the deballasting process.

Acknowledgements

This research received financial support from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Grants 133211/2011-8, 308613/2011-2 and 307928/2010-1) and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Brazil. The authors appreciate the constructive suggestions of the referee and the editor. The authors thank Damian Matuszewski for the graphical abstract.

Appendix A. Technical details for the results in Section 2.2

For stratum h, X_{hi}, $i = 1, \ldots, n_h$, follow Poisson distributions with mean λ_h and an estimator is $\widehat{\lambda}_h = n_h^{-1}\sum_{i=1}^{n_h} X_{hi}$. The estimator of λ is $\widehat{\lambda} = \sum_{h=1}^{H} W_h \widehat{\lambda}_h$. Letting $\boldsymbol{x}$ denote the set of observed values, $x_{11}, \ldots, x_{Hn_H}$, it follows that

$$\{|\widehat{\lambda}(\boldsymbol{x}) - \lambda| > \epsilon; \boldsymbol{x} \in \Omega\} \subseteq \bigcup_{h=1}^{H} \left\{ W_h |\widehat{\lambda}_h(\boldsymbol{x}_h) - \lambda_h| > \epsilon/H; \boldsymbol{x}_h \in \Omega_h \right\},$$

where Ω_h is the sample space associated to the probability model for stratum h and $\Omega = \bigcup_{h=1}^{H} \Omega_h$. Then,

$$P(|\widehat{\lambda} - \lambda| > \epsilon) \leqslant \sum_{h=1}^{H} P(|\widehat{\lambda}_h - \lambda_h| > \epsilon / W_h H) \leqslant \sum_{h=1}^{H} \alpha_h,$$

and setting $\epsilon_h = \epsilon / W_h H$ and $\alpha = \sum_{h=1}^{H} \alpha_h$ the result follows.

Appendix B. Technical details on the approximation of the negative binomial distribution by a Poisson distribution

The characteristic function of the NB distribution in (4) is

$$\left[1 - w\lambda(e^{it} - 1)/\phi\right]^{-\phi}. \tag{B.1}$$

The limit of (B.1) as ϕ approaches infinity is $\exp[w\lambda(e^{it} - 1)]$, which is the characteristic function of a Poisson distribution with mean $w\lambda$. This implies that the NB distribution converges weakly to a Poisson distribution as ϕ approaches infinity (Sen et al., 2009), indicating that the NB distribution may be approximated by a Poisson distribution for large ϕ.

The characteristic function of the NB distribution in (6) is

$$\left[1 - nw\lambda(e^{it} - 1)/n\phi\right]^{-n\phi}. \tag{B.2}$$

If $n \to \infty$ and $w \to 0$ so that the value of nw remains equal to v, *i.e.*, sampling more and more aliquots with increasing less volume, the limit of (B.2) is $\exp[v\lambda(e^{it} - 1)]$, which is the characteristic function of a Poisson distribution with mean $v\lambda$. Therefore the NB distribution converges weakly to a Poisson distribution (Sen et al., 2009), indicating that the NB distribution may be approximated by a Poisson distribution when more aliquots with smaller volume are sampled.

Appendix C. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.marpolbul.2014.11.030.

References

Basawa, I.V., Rao, B.L.S., 1980. Statistical Inference for Stochastic Processes. Academic Press, London.

Basurko, O.C., Mesbahi, E., 2011. Statistical representativeness of ballast water sampling. In: Proceedings of the Institution of Mechanical Engineer, Part M, Journal of Engineering for the Maritime Environment, pp. 183–190.

Bierman, S.M., Vries, P., Kaag, N.H.B.M., 2012. 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). Tech. Rep. C124/12, IMARES Wageningen UR.

Carney, K.J., Basurko, O.C., Pazouki, K., Marsham, S., Delany, J.E., Desai, D.V., Anil, A.C., Mesbahi, E., 2013. Difficulties in obtaining representative samples for compliance with the Ballast Water Management Convention. Mar. Pollut. Bull. 68, 99–105.

Chen, X., 2007. Exact Computation of Minimum Sample Size for Estimation of Poisson Parameters. <http://arxiv.org/abs/0707.2116>.

Costa, E.G., 2013. Aspectos estatísticos da amostragem de água de lastro, Master's thesis, Departamento de Estatística, Universidade de Sao Paulo, Sao Paulo, in Portuguese. <http://www.teses.usp.br/teses/disponiveis/45/45133/tde-06052013-172123>.

First, M.R., Robbins-Wamsley, S.H., Riley, S.C., Moser, C.S., Smith, G.E., Tamburri, M.N., Drake, L.A., 2013. Stratification of living organisms in ballast tanks: how do organism concentrations vary as ballast water is discharged? Environ. Sci. Technol. 47 (9), 4442–4448.

Frazier, M., Miller, A.W., LeeII, H., Reusser, D.A., 2013. Counting at low concentrations: the statistical challenges of verifying ballast water discharge standards. Ecol. Appl. 23, 339–351.

Gollasch, S., David, M., Voigt, M., Dragsund, E., Hewitt, C., Fukuyo, Y., 2007. Critical review of the IMO international convention on the management of ship's ballast water and sediments. Harmful Algae 6, 585–600.

Gollasch, S., David, M., 2010. Testing Sampling Representativeness of a Ballast Water Discharge and Developing Methods for Indicative Analysis. Tech. Rep., GoConsult, Research Report to EMSA Tender EMSA/NEG/09/2010.

Grafström, A., Schelin, L., 2014. How to select representative samples. Scand. J. Statist. 41 (2), 277–290.

Haight, F.A., 1967. Handbook of the Poisson Distribution. John Wiley & Sons, New York.

Hilbe, J.M., 2007. Negative Binomial Regression. Cambridge University Press, Cambridge.

ICS, 2013. Annual Review. Tech. Rep., International Chamber of Shipping, Representing the Global Shipping Industry.

IMO, 2004. International Convention for the Control and Management of Ship Ballast Water and Sediments. <http://www.imo.org/>.

IMO, 2008. Report of the Marine Environment Protection Committee, Annex 3: Resolution MEPC, Guidelines for Ballast Water Sampling (G2) 173(58), International Maritime Organization.

Johnson, N.L., Kotz, S., Balakrishnan, N., 1994, . Continuous Univariate Distributions, second ed., vol. 1. John Wiley & Sons, New York.

Miller, A.W., Frazier, M., Smith, G.E., Perry, E.S., Ruiz, G.M., Tamburri, M.N., 2011. Enumerating sparse organisms in ships' ballast water: why counting to 10 is not easy. Environ. Sci. Technol. 45, 3539–3546.

Müller, P., Quintana, F.A., 2004. Nonparametric Bayesian data analysis. Statist. Sci. 19, 95–110.

Murphy, K.R., Ritz, D., Hewitt, C.L., 2002. Heterogeneous zooplankton distribution in a ship's ballast tanks. J. Plankton Res. 24, 729–734.

R Development Core Team, 2013. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>, ISBN 3-900051-07-0.

Ross, S.M., 1996. Stochastic Processes. John Wiley & Sons, New York.

Sen, P.K., Singer, J.M., de Lima, A.C.P., 2009. From Finite Sample to Asymptotic Methods in Statistics. Cambridge University Press, Cambridge.