Journal of Environmental Management 180 (2016) 433–438

Contents lists available at ScienceDirect

Journal of Environmental Management

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

Research article

Sample size for estimating the mean concentration of organisms in ballast water

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, Brazil
[b] Departamento de Oceanografia Biológica, Instituto Oceanográfico, Universidade de São Paulo, Brazil

ARTICLE INFO

Article history:
Received 13 July 2015
Received in revised form
23 March 2016
Accepted 17 May 2016

Keywords:
Bounds on absolute and relative errors
Confidence interval
Coverage probability
Negative binomial distribution

ABSTRACT

We consider the computation of sample sizes for estimating the mean concentration of organisms in ballast water. Given the possible heterogeneity of their distribution in the tank, we adopt a negative binomial model to obtain confidence intervals for the mean concentration. We show that the results obtained by Chen and Chen (2012) in a different set-up hold for the proposed model and use them to develop algorithms to compute sample sizes both in cases where the mean concentration is known to lie in some bounded interval or where there is no information about its range. We also construct simple diagrams that may be easily employed to decide for compliance with the D-2 regulation of the International Maritime Organization (IMO).

1. Introduction

With the expansion of transoceanic trade routes and the increasing tonnage of commercial ships, ballast water discharges have become a major cause of unintentional introduction of invasive species to coastal waters worldwide (Ruiz et al., 2000). Harmful species released from ballast tanks may develop resident populations in recently colonized areas, leading to significant ecological and economic impacts in addition to public health concerns (Carlton, 2001; McCarthy et al., 1992). A convention on ballast water management was adopted by the International Maritime Organization (IMO, 2004) and is expected to enter into force soon. The D-2 regulation of the IMO convention requires that ballast water discharged by ships contain (i) fewer than 10 viable organisms with minimum dimension $\geq$50 μm per m^3, and (ii) fewer than 10 viable organisms with minimum dimension between 10 μm and 50 μm per mL.

The installation of on-board treatment units has been accepted by the international community as an effective strategy to attain the required concentrations and several systems have been approved by the IMO since the early years of the convention implementation (Tsolaki and Diamadopoulos, 2009). However, a significant challenge persists on how to evaluate compliance with the D-2 concentration standards because of the heterogeneous distribution of organisms within ballast tanks, their small size and the high discharge throughput. Given the enormous amount of ballast water carried in a great number of commercial ships, it is not practical to examine the entire volume and one must rely on sampling to verify compliance with the D-2 standards. For such purposes, sampling schemes must guarantee that the D-2 standards are satisfied up to a pre-specified margin of error. In this context, Miller et al. (2011) employed the Poisson distribution while Bierman et al. (2012), Frazier et al. (2013) and Costa et al. (2015) approached the problem by using the negative binomial distribution from a hypothesis testing point of view. Such an approach is convenient for decision making, but does not provide the magnitude of the mean concentration estimates. To incorporate this feature, we propose a solution based on confidence intervals.

Costa et al. (2015) indicated that under the negative binomial model, the sample volume v may consist of n aliquots with volume w so that an estimate of the mean concentration λ is $\widehat{\lambda} = (nw)^{-1}\sum_{i=1}^{n} X_i$ where X_i is the number of organisms detected in the i-th aliquot. Then, given an aliquot volume, from an estimation perspective, the problem may be viewed as one of determining the minimum number of aliquots required to obtain estimates $\widehat{\lambda}$ subject to a fixed upper bound on the estimation error. From the statistical point of view, the solution is similar for organisms with dimension

* Corresponding author.
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.jenvman.2016.05.043
0301-4797/

≥50 μm or in the range 10 μm–50 μm. In either case, we present our results in terms of organisms/unit volume so that we may use either mL or m^3 as the basis for specifying the mean concentration.

Chen (2008, 2011) and Chen and Chen (2015) proposed algorithms to compute sample sizes for the estimation of the mean of Poisson or binomial distributions with given bounds on the absolute or on the relative errors of maximum likelihood estimators. These authors showed that sample sizes may be obtained when the mean is known to lie within a bounded interval. Gamrot (2013) extended Chen's results and proposed a method for calculating the minimum sample size controlling the relative error in the binomial case. Chen and Chen (2012) proved that these results may be applied for estimating parameters of discrete distributions with the assumption that the coverage probability of the corresponding confidence interval considered as a function of the parameter of interest is continuous and unimodal. Our objectives are (i) to show that this assumption holds for the negative binomial distribution and (ii) to provide statistical procedures to obtain reliable mean concentration estimates, which may be used to verify compliance with the D-2 standards. We tested and empirically evaluated the proposed models under two different scenarios: (i) cases where the mean concentration is known to lie in a given interval and (ii) cases where no prior information on the mean concentration is available.

For clarity, the notation employed on the text is summarized in Table 1.

2. Minimum number of aliquots for cases where the mean concentration is known to lie in a given interval

We assume a negative binomial distribution to compute the probability of observing the number X of organisms in an aliquot of fixed volume w obtained from a tank with mean concentration λ, *i.e.*,

$$\mathbb{P}[X = x|\lambda,\varphi] = \frac{\Gamma(\varphi+x)}{\Gamma(x+1)\Gamma(\varphi)}\left(\frac{w\lambda}{w\lambda+\varphi}\right)^{x}\left(\frac{\varphi}{w\lambda+\varphi}\right)^{\varphi}, \quad \lambda,\varphi>0, \tag{1}$$

for x = 0,1, …, where $\Gamma(x)=\int_0^\infty t^{x-1}e^{-t}\mathrm{d}t$ is the gamma function and φ is a shape parameter, also known in the literature as aggregation, clustering or heterogeneity parameter. In the ballast water set-up, small values of φ correspond to more heterogeneous distribution of the organisms in the tank. Large values of φ, on the other hand, correspond to homogeneous distributions and in this case, the negative binomial distribution may be approximated by a Poisson distribution.

Under the negative binomial distribution, the expected value of X is $w\lambda$ and the corresponding variance is $w\lambda + (w\lambda)^2/\varphi$, implying that this distribution may also model over-dispersion (variance larger than the mean). The parameter we are concerned with is the mean concentration λ in the ballast water tank. Suppose that n aliquots with fixed volume w have been randomly collected from the tank and that the corresponding number of organisms $X_1,\ldots,X_n$ follow independent negative binomial distributions. Then, we may use the maximum likelihood estimator $\widehat{\lambda} = (nw)^{-1}\sum_{i=1}^{n} X_i$, to estimate the mean concentration λ. Under this set-up, we want to determine the minimum number of aliquots such that the upper bounds on the absolute estimation error ($\left|\widehat{\lambda}-\lambda\right|$) or on the relative estimation error ($\left|\widehat{\lambda}-\lambda\right|/\lambda$) are $\varepsilon_a > 0$ and $\varepsilon_r \in (0,1)$, respectively, with minimum pre-specified probability (confidence level) $1-\rho$, $\rho\in(0,1)$, *i.e.*, to determine n such that for all λ

$$\mathbb{P}\left[\left|\widehat{\lambda}-\lambda\right| < \varepsilon_a\right] > 1-\rho \tag{2}$$

or

$$\mathbb{P}\left[\left|\widehat{\lambda}-\lambda\right| < \varepsilon_r\lambda\right] > 1-\rho. \tag{3}$$

Chen (2011) and Chen and Chen (2015) establish methods to compute the minimum sample size for constructing confidence intervals for the mean of a Poisson or of a binomial distribution with fixed confidence levels, controlling the absolute estimation error, the relative estimation error or both. The only requirement is the specification of a lower and an upper bound on the mean. In this setting, Chen (2011) and Chen and Chen (2015) show that the minimum confidence level, $1-\rho$, for $\mathbb{P}[\left|\widehat{\lambda}-\lambda\right| < \varepsilon_a]$ or for $\mathbb{P}[\left|\widehat{\lambda}-\lambda\right| < \varepsilon_r\lambda]$, considered as functions of λ, may be determined by computing these probabilities a finite number of times. We extend such results to the negative binomial distribution and show how they may be used to construct algorithms to compute the required sample sizes. The results are obtained in the form of a theorem stated and proven in the Supplementary material (Theorem 1) because of its technical nature.

In particular, an algorithm to compute the minimum number n of aliquots with volume w such that (2) is satisfied in cases where there is *a priori* information that the mean concentration λ belongs to the interval $[a,b]$ is described as follows.

Step 1. Set a value for ε_a;
Step 2. Set values for w, ρ, a, b, φ and take $n = 2$;
Step 3. Define the finite set of values of λ for which $\mathbb{P}[\left|\widehat{\lambda}-\lambda\right| < \varepsilon_a]$ must be computed as the union of the sets $\{a,b\}$, $\{\ell/(nw)+\varepsilon_a \in (a,b),\ \ell \text{ integer}\}$ and $\{\ell/(nw)-\varepsilon_a \in (a,b),\ \ell \text{ integer}\}$ which has at most $2n(b-a)+4$ elements;
Step 4. Compute $\mathbb{P}[\left|\widehat{\lambda}-\lambda\right| < \varepsilon_a]$ using (1) for all elements in the set defined in Step 3 and obtain the corresponding minimum;

Table 1
Summary of the notation employed in the text.

X	Number of organisms detected in a sample aliquot
w	Aliquot volume
λ	Mean concentration in the ballast water tank
φ	Shape parameter of the negative binomial distribution (considered known)
n	Number of aliquots
v	Total sample volume (=nw)
$1-\rho$	Confidence level
ε_a	Upper bound on the absolute estimation error
ε_r	Upper bound on the relative estimation error
a,b	Lower and upper bounds for the mean concentration
$\mathbb{P}[A\|B]$	Probability of the event A given the event B
Γ	Gamma function

Step 5. If this minimum coverage probability is greater than $1-\rho$, stop. The value n obtained in this step is the required value. Otherwise, set $n = n + 1$ and return to Step 3.

The algorithm for cases where an upper bound is fixed for the relative estimation error instead of the absolute estimation error is similar; it suffices to replace $\mathbb{P}[|\widehat{\lambda}-\lambda|<\varepsilon_a]$ for $\mathbb{P}[|\widehat{\lambda}-\lambda|<\varepsilon_r\lambda]$ and define the set for which this probability must be computed as the union of the sets $\{a,b\}$, $\{\ell/[nw(1-\varepsilon_r)]\in(a,b),\ \ell \text{ integer}\}$ and $\{\ell/[nw(1+\varepsilon_r)]\in(a,b),\ \ell \text{ integer}\}$ which has at most $2n(b-a)+4$ elements. Details are presented in the Supplementary material, and an R code for the practical implementation of both versions of the algorithm may be downloaded from www.ime.usp.br/~jmsinger/NBsamplesize.

In Tables 2 and 3, we present the minimum number of aliquots n computed via the algorithm described above for fixed bounds either on the absolute or on the relative estimation errors in different scenarios assuming that the mean concentration lies either in the interval [5,15] or in [2,25].

If for example, we have prior information that the mean concentration is between 5 and 15 organisms/unit volume and that the aggregation parameter is $\varphi = 10$, fixing the maximum absolute estimation error at $\varepsilon_a = 2$ and the minimum confidence level at 0.95, it follows from Table 2 that the required minimum numbers of aliquots n are 1476, 168, 52 and 37 for aliquot volumes 0.01, 0.1, 0.5 and 1, respectively. On the other hand, if we know that the concentration is between 2 and 25 organisms/unit volume, it follows from Table 3 that the required minimum numbers of aliquots n are 2476, 303, 109 and 85 for aliquot volumes 0.01, 0.1, 0.5 and 1, respectively. The corresponding decision is to declare compliance with the D-2 standards if the sample concentration $\widehat{\lambda}<8(=10-\varepsilon_a)$ or non-compliance, if $\widehat{\lambda}>12(=10+\varepsilon_a)$. In a similar context regarding the information on the aggregation parameter φ, knowing that the mean concentration lies in [5,15], fixing the maximum relative estimation error at $\varepsilon_r = 0.1$ and the minimum confidence level at 0.95, it follows from Table 2 that the required minimum numbers of aliquots n are 7891, 820, 196 and 117 for aliquot volumes 0.01, 0.1, 0.5 and 1, respectively. For mean concentrations in the interval [2,25], it follows from Table 3 that the required minimum number of aliquots n are 19,500, 1996, 430 and 236 for aliquot volumes 0.01, 0.1, 0.5 and 1, respectively. Here, compliance with the D-2 standards is characterized when $\widehat{\lambda}<9(=10(1-\varepsilon_r))$, and non-compliance when $\widehat{\lambda}>11(=10(1+\varepsilon_r))$.

Diagrams that may help the decision about compliance with the D-2 standards are displayed in Figs. 1 and 2. If the sample concentration falls in the green (red) interval we declare (non-) compliance with the D-2 standards; in the yellow interval, more aliquots should be obtained before a decision may be made with the prescribed confidence level. For any given sample concentration obtained with the required minimum number of aliquots, projections of the endpoints of vertical line segments comprised in the shaded region correspond to the limits of confidence intervals for the mean concentration satisfying the bounds on the absolute or relative estimation error, respectively.

2.1. Simulation study

To check our results in practice we carried out a simulation study fixing $\varepsilon_a = 1$, $\varepsilon_r = 0.05$, $\rho = 0.05$ and λ at 1, 3.5, 9, 13, 20 or 28. For each scenario, we drew 1000 samples from a negative binomial distribution with number of aliquots n determined from Tables 2 and 3, and computed the empirical coverage probability as the proportion of samples for which $|\widehat{\lambda}-\lambda|<\varepsilon_a$ in the absolute error case, or for which $|\widehat{\lambda}-\lambda|/\lambda<\varepsilon_r$ in the relative error case. For the cases where λ is in [5,15] or [2,25] we expect the estimated coverage probability to be at least 0.95. The results are displayed in Table 4 and in Tables 2–4 in the Supplementary material.

3. Minimum number of aliquots for cases with no information on the mean concentration bounds

When there is no prior information about the mean concentration range, we propose a simple formula to compute the minimum number of aliquots required to estimate λ, adapting the results of Chen (2008) to the negative binomial distribution. In this case, however, computation of the minimum number of aliquots requires that upper bounds be set for both the absolute and relative errors simultaneously, *i.e.*, it relates to the determination of the minimum number of aliquots n such that

$$\mathbb{P}\left[|\widehat{\lambda}-\lambda|<\varepsilon_a \text{ or } |\widehat{\lambda}-\lambda|<\varepsilon_r\lambda\right]>1-\rho, \tag{4}$$

where $1-\rho$, $\rho\in(0,1)$ is the minimum confidence level.

The proposed approach does not require asymptotic approximations; it suffices to fix the confidence level, $1-\rho$, the aliquot

Table 2
Minimum number of aliquots n and total sample volume v (within parentheses) with prior information that the mean concentration belongs to the interval $[a,b] = [5,15]$ and minimum confidence level $1-\rho = 0.95$.

w			φ				
			0.5	10	50	100	1000
0.01	ε_a	1	7501 (76)	5851 (59)	5801 (59)	5801 (59)	5801 (59)
		2	1876 (19)	1476 (15)	1451 (15)	1451 (15)	1451 (15)
	ε_r	0.05	34,066 (341)	31,181 (312)	30,991 (310)	30,991 (310)	30,991 (310)
		0.1	8582 (86)	7891 (80)	7800 (78)	7800 (78)	7800 (78)
0.1	ε_a	1	2306 (231)	666 (67)	596 (60)	591 (60)	581 (59)
		2	578 (58)	168 (17)	151 (16)	148 (15)	148 (15)
	ε_r	0.05	6180 (618)	3260 (326)	3140 (314)	3119 (312)	3100 (310)
		0.1	1551 (156)	820 (82)	790 (79)	790 (79)	780 (78)
0.5	ε_a	1	1845 (923)	204 (102)	135 (68)	126 (63)	119 (60)
		2	461 (231)	52 (26)	35 (18)	32 (16)	30 (15)
	ε_r	0.05	3695 (1848)	775 (388)	652 (326)	636 (318)	620 (310)
		0.1	925 (463)	196 (98)	164 (82)	160 (80)	156 (78)
1	ε_a	1	1787 (1787)	145 (145)	76 (76)	68 (68)	60 (60)
		2	446 (446)	37 (37)	20 (20)	18 (18)	16 (16)
	ε_r	0.05	3384 (3384)	465 (465)	341 (341)	326 (326)	313 (313)
		0.1	847 (847)	117 (117)	86 (86)	82 (82)	79 (79)

Table 3
Minimum number of aliquots n and sample volume v (within parentheses) with prior information that the mean concentration belongs to the interval $[a,b] = [2,25]$ and minimum confidence level $1 - \rho = 0.95$.

w			φ				
			0.5	10	50	100	1000
0.01	ε_a	1	14,451 (145)	9851 (99)	9701 (98)	9651 (97)	9651 (97)
		2	3626 (37)	2476 (25)	2426 (25)	2426 (25)	2426 (25)
	ε_r	0.05	80,477 (805)	77,477 (775)	77,477 (775)	77,477 (775)	77,477 (775)
		0.1	20,228 (203)	19,500 (195)	19,500 (195)	19,500 (195)	19,500 (195)
0.1	ε_a	1	5766 (577)	1206 (121)	1011 (102)	991 (100)	966 (97)
		2	1443 (145)	303 (31)	256 (26)	248 (25)	243 (25)
	ε_r	0.05	10,848 (1085)	7901 (791)	7796 (780)	7748 (775)	7748 (775)
		0.1	2723 (273)	1996 (200)	1950 (195)	1950 (195)	1950 (195)
0.5	ε_a	1	4995 (2498)	435 (218)	243 (122)	219 (110)	197 (99)
		2	1249 (625)	109 (55)	61 (31)	55 (28)	50 (25)
	ε_r	0.05	4627 (2314)	1704 (852)	1581 (791)	1570 (785)	1550 (775)
		0.1	1161 (581)	430 (215)	400 (200)	395 (198)	390 (195)
1	ε_a	1	4898 (4898)	338 (338)	146 (146)	122 (122)	100 (100)
		2	1224 (1224)	85 (85)	37 (37)	31 (31)	26 (26)
	ε_r	0.05	3851 (3851)	931 (931)	805 (805)	791 (791)	775 (775)
		0.1	965 (965)	236 (236)	203 (203)	200 (200)	195 (195)

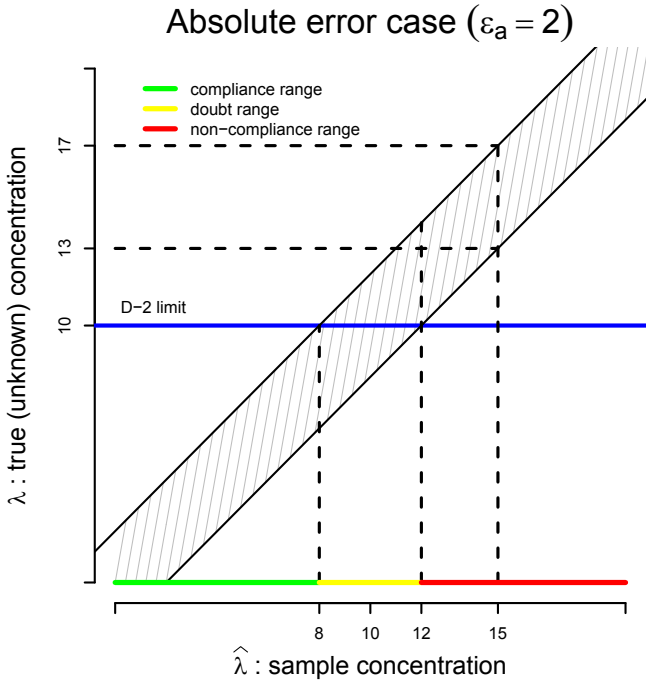

Fig. 1. Decision diagram for the absolute estimation error case with $\varepsilon_a = 2$.

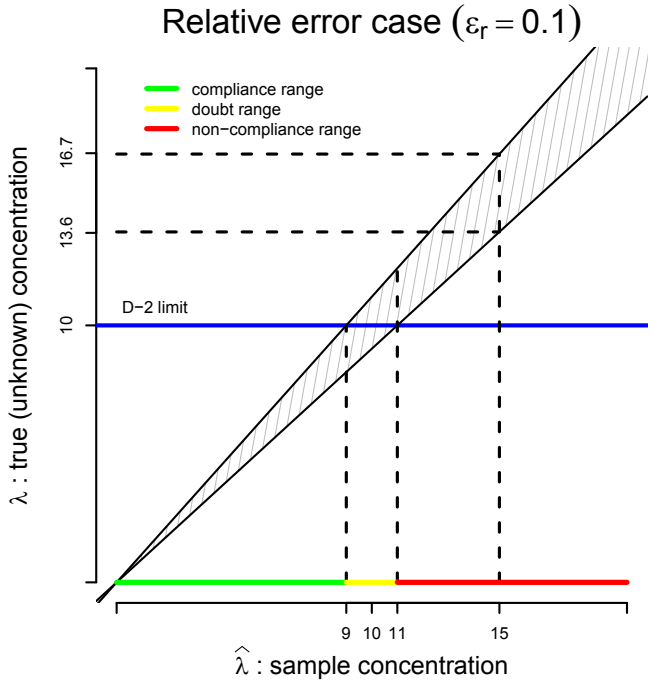

Fig. 2. Decision diagram for the relative error case with $\varepsilon_r = 0.1$.

volume w, and the upper bounds for the absolute and relative estimation errors. The basic result is provided in Theorem 2 stated and proven in the Supplementary material. The required number of aliquots is the smallest integer n satisfying

$$n > \frac{\varepsilon_r}{\varepsilon_a} \times \frac{\log(2/\rho)}{w(1+\varepsilon_r)\log(1+\varepsilon_r) - \left[\frac{\varepsilon_r}{\varepsilon_a}\varphi + w(1+\varepsilon_r)\right]\log\left(1+\frac{w\varepsilon_a\varepsilon_r}{\varphi\varepsilon_r + w\varepsilon_a}\right)}. \quad (5)$$

An R code to implement the computation of the minimum number of aliquots may be downloaded from www.ime.usp.br/~jmsinger/NBsamplesize.

The difference between (5) and the corresponding result in Chen (2008) lies in the denominator, where in the latter, the term ε_r is replaced by

$$\left[\frac{\varepsilon_r}{\varepsilon_a}\varphi + w(1+\varepsilon_r)\right]\log\left(1+\frac{w\varepsilon_a\varepsilon_r}{\varphi\varepsilon_r + w\varepsilon_a}\right). \quad (6)$$

Since, as expected, this term is greater than $w\varepsilon_r$, it is clear that the number of aliquots obtained under the negative binomial distribution are larger than those computed under the Poisson distribution with the same values for w, ε_a, ε_r and ρ. Note also that only the second term in the denominator of (5) depends on φ. The limit of (6) as φ approaches infinity *i.e.*, as the concentration distribution becomes more homogeneous, is $w\varepsilon_r$, which suggests that for large values of the shape parameter φ, the required number of aliquots may be obtained from the Poisson distribution. Results for given values of φ are displayed in Table 5.

Table 4
Estimated coverage probability for some scenarios using the number of aliquots presented in Table 2 with $\varepsilon_a = 1$, $[a,b] = [5,15]$.

w	φ	n	λ					
			1	3.5	9	13	20	28
	0.5	7501	1	1	0.994	0.965	0.900	0.809
	10	5851	1	1	0.990	0.976	0.932	0.831
0.01	50	5801	1	1	0.985	0.961	0.905	0.851
	100	5801	1	1	0.991	0.973	0.911	0.850
	1000	5801	1	1	0.989	0.971	0.904	0.845
	0.5	2306	1	1	0.998	0.978	0.868	0.731
	10	666	1	1	0.989	0.965	0.883	0.815
0.1	50	596	1	1	0.988	0.968	0.913	0.851
	100	591	1	1	0.987	0.955	0.917	0.836
	1000	581	1	1	0.983	0.969	0.921	0.860
	0.5	1845	1	1	0.999	0.976	0.849	0.711
	10	204	1	1	0.992	0.963	0.904	0.786
0.5	50	135	1	1	0.991	0.966	0.918	0.830
	100	126	1	1	0.990	0.967	0.891	0.815
	1000	119	1	1	0.985	0.970	0.900	0.867
	0.5	1787	1	1	0.997	0.974	0.860	0.696
	10	145	1	1	0.998	0.975	0.882	0.754
1	50	76	1	1	0.989	0.963	0.901	0.811
	100	68	1	1	0.995	0.975	0.915	0.839
	1000	60	1	1	0.988	0.963	0.889	0.848

4. Discussion

We considered a procedure to obtain the minimum number of aliquots of a given volume to estimate the mean concentration of viable organisms in ballast water tanks with a pre-specified precision and indicated how the results may be employed to verify compliance to the D-2 standards.

As discussed in Costa et al. (2015), sample size determination in this set-up is highly dependent on the heterogeneity of the distribution of organisms in the ballast water tank. This heterogeneity may be summarized by the parameter φ determining the shape of the negative binomial distribution adopted as a model for the number of organisms in the sample. If we have prior information that the organism distribution in the tank is homogeneous, then the value of the parameter φ is large; otherwise, if we know that the corresponding distribution is heterogeneous, the aggregation parameter φ must be small. Information on φ must be obtained from historical or experimental data, such as in simulations of water sloshing patterns in ballast water tanks (Arai et al., 2002).

The results in Tables 2 and 3 indicate that when the aggregation parameter is large, *i.e.*, when the distribution of organisms in the tank is homogeneous, the total sample volume is constant irrespectively of the aliquot volume, suggesting that the choice of the aliquot volume is a matter of convenience. This favours the usual practice of concentrating a large volume of water and taking the concentrate as the sample. Consider the results displayed in Table 2 and suppose that the largest admissible error in estimating the mean concentration is $\varepsilon_a = 2$ organisms/m^3. When the distribution is homogeneous, it follows that the sample volume required to guarantee this margin of error is 15 m^3, irrespectively of the aliquot volume. Therefore, one may concentrate the 15 m^3 of ballast water and do a single analysis to identify the number of sampled organisms. If, for example, the organism count is 180, a simple calculation based on the Poisson distribution indicates that the lower and upper limits of an approximate 95% confidence interval are 10.25 and 13.75 for the mean concentration (leading to a non-compliance decision). If, on the other hand, a sample of 5 m^3 is considered and the organism count is 60, the corresponding lower and upper limits of an approximate 95% confidence interval are 8.96 and 15.04, and one would need to sample a larger volume to reach a decision (see Fig. 1). Details are presented in the Supplementary material.

When the aggregation parameter is small *i.e.*, when the distribution of organisms in the tank is heterogeneous, the total sample volume increases, suggesting that sampling more aliquots with small volume will capture the heterogeneity of the organism distribution with a smaller sample volume. Even if the sample ballast water is concentrated, a large number of aliquots must be collected to achieve the desired margin of error. This is in line with the flow integration sampling protocol (IMO, 2015), which considers obtaining multiple samples over a specified time scale repeatedly throughout the discharge.

Our simulation study suggests that when the absolute error is considered, the estimated coverage probability is greater than the minimum confidence level, even when the mean concentration λ is smaller than the lower bound. The situation is reversed when λ is greater than the upper bound. In the relative error case, the nominal confidence level is not attained only when λ is smaller than the lower bound. In the absolute error case it is important to specify the upper bound (b) correctly to guarantee the minimum coverage probability $1 - \rho$, whereas in the relative error case it is important to specify the lower bound (a). We also note that the coverage probabilities obtained in our simulation study are generally larger than the nominal level (adopted as 95%), suggesting that the results obtained from Theorems 1 and 2 are conservative, in the sense that

Table 5
Minimum number of aliquots n and total sample volume (within parentheses) required to satisfy (4) with minimum confidence level $1 - \rho = 0.95$.

w	ε_a	ε_r	Negative binomial (φ=)					Poisson
			0.5	10	50	100	1000	
0.01	1	0.05	21,099 (211)	15,305 (154)	15,061 (151)	15,030 (151)	15,003 (151)	15,000 (150)
		0.1	9194 (92)	7699 (77)	7636 (77)	7628 (77)	7621 (77)	7620 (77)
	2	0.05	13,599 (136)	7805 (79)	7561 (76)	7531 (76)	7503 (76)	7500 (75)
		0.1	5384 (54)	3889 (39)	3826 (39)	3818 (39)	3811 (39)	3810 (39)
0.1	1	0.05	7599 (760)	1805 (181)	1561 (157)	1531 (154)	1503 (151)	1500 (150)
		0.1	2336 (234)	841 (85)	778 (78)	770 (77)	763 (77)	762 (77)
	2	0.05	6849 (685)	1055 (106)	811 (82)	781 (79)	754 (76)	750 (75)
		0.1	1955 (196)	460 (46)	397 (40)	389 (39)	382 (39)	381 (39)
0.5	1	0.05	6399 (3200)	605 (303)	361 (181)	331 (166)	304 (152)	300 (150)
		0.1	1726 (863)	232 (116)	169 (85)	161 (81)	154 (77)	153 (77)
	2	0.05	6249 (3125)	455 (228)	211 (106)	181 (91)	154 (77)	150 (75)
		0.1	1650 (825)	155 (78)	92 (46)	85 (43)	77 (39)	77 (39)
1	1	0.05	6249 (6249)	455 (455)	211 (211)	181 (181)	154 (154)	150 (150)
		0.1	1650 (1650)	155 (155)	92 (92)	85 (85)	77 (77)	77 (77)
	2	0.05	6174 (6174)	380 (380)	136 (136)	106 (106)	79 (79)	75 (75)
		0.1	1612 (1612)	117 (117)	54 (54)	46 (46)	39 (39)	39 (39)

smaller sample volumes may be enough to ensure the required limits on the estimation errors are attained. This is the object of ongoing research.

According to the suggested sampling scheme, the decision with respect to compliance with the D-2 standards must be taken after the entire sample is collected. This may be a problem when the mean concentration is large, since non-compliance will possibly be detected only after the entire deballasting process is completed. A more sensible scheme should consider a sequential procedure in which non-compliance could be evaluated as the ballast water is discharged. A naive approach to this alternative based on the proposed sample size determination could be considered as follows. Suppose that n aliquots are required to guarantee that the estimation error falls within the prescribed limit. We may consider a decision based on the first n/K aliquots where K is a given constant. We give the ship owner the benefit of the doubt by supposing that no organisms are detected in the remaining $n - n/K$ aliquots and use these pseudo-observations along with those obtained in the first n/K aliquots to estimate the mean concentration. If the lower limit of the corresponding confidence interval for the mean concentration is larger than the D-2 standards, we declare non-compliance. Otherwise, we abandon the pseudo-observations, take an extra n/K aliquots assuming that no organisms are detected in the remaining $n - 2n/K$ aliquots and repeat the decision process. Consider, for example, a situation where $n = 100$ and $K = 10$; then we may count the organisms in the first 10 ($=100/10$) aliquots, assume that there are no organisms in the remaining 90 ($=100 - 100/10$) aliquots and apply the decision rule. If there is no evidence against compliance, we may sample 10 additional aliquots and count the organisms in the 20 available aliquots, assume that there are no organisms in the remaining 80 ($=100 - 2 \times 100/10$) aliquots and reapply the decision rule. Again if there is no evidence against compliance, we repeat the procedure until we obtain evidence against compliance or complete the sample of 100 aliquots. This approach may be useful to detect cases where the mean concentration is very large, reducing costs and preventing potential impacts.

Unfortunately, the stringent limits imposed by the D-2 concentration standards implies that large volumes of ballast water must be sampled to achieve reasonable margins of error, especially for organisms with dimension $\geq$50 μm. If, for example, the decision is based on a (more realistic) sample of 10 L and the mean concentration is 13 organisms/m^3, the probability of declaring non-compliance is approximately 12%. The scenario for organisms with dimension in the range 10 μm and 50 μm is more favourable, but still, a large number of aliquots must be examined when the distribution in the tank is heterogeneous.

A global set of standards or best practices for ballast water verification currently does not exist, and an intense debate persists within the IMO on the complex issue of sampling and analysis for compliance testing (IMO, 2015). One of the major challenges is to account for the uneven distributions of organisms at different depths of ballast tanks (Murphy et al., 2002; First et al., 2013), which is expected to result in variable concentrations along the discharge. Despite the IMO recommendations on taking repeated samples during the discharge to increase the ability to depict heterogeneous distributions, our results suggest that the number of samples generated by such an approach may be impractically large when using standard sampling and evaluation methods (*e.g.*, sequential filtering through plankton nets followed by lab-based microscopic analysis). However, instead of providing evidence that shipboard analysis for compliance testing is not feasible, this investigation emphasizes the need for developing innovative sampling and analysis methods to collect and process a large number of samples during the ballast water discharge. This could be achieved by high-volume particle imaging instruments as part of sample-in-flow monitoring systems (Matuszewski et al., 2015).

Acknowledgements

This research received financial support from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, grants 133211/2011-8, 308613/2011-2 and 311936/2013-0) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brazil. The authors are grateful to an anonymous reviewer for the enlightening comments.

Appendix A. Supplementary material

Supplementary material related to this article can be found at http://dx.doi.org/10.1016/j.jenvman.2016.05.043.

References

Arai, M., Makiyama, H.S., Cheng, L., 2002. Numerical simulation of sloshing of water in ship tanks during sequential ballast water exchange in seaways. In: Safety and Reliability, Ocean Space Utilization: Presented at the 21st International Conference on Offshore Mechanics and Arctic Engineering, June 23–28, 2002. American Society of Mechanical Engineers, Oslo, Norway, pp. 799–806.

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.

Carlton, J.T., 2001. Introduced Species in U.S. Coastal Waters: Environmental Impacts and Management Priorities. Pew Oceans Commission, Arlington, Virginia.

Chen, X., 2008. A Simple Sample Size Formula for Estimating Means of Poisson Random Variables. http://arxiv.org/abs/0804.3033 (accessed 25.08.14.).

Chen, X., 2011. Exact computation of minimum sample size for estimation of binomial parameters. J. Stat. Plan. Infer. 141, 2622–2632.

Chen, X., Chen, Z., 2012. Exact Sample Size Methods for Estimating Parameters of Discrete Distributions. http://arxiv.org/abs/1211.1912 (accessed 27.11.14.).

Chen, Z., Chen, X., 2015. Exact calculation of minimum sample size for estimating a Poisson parameter. Commun. Stat. Theory Methods. http://dx.doi.org/10.1080/03610926.2014.927497.

Costa, E.G., Lopes, R.M., Singer, J.M., 2015. Implications of heterogeneous distributions of organisms on ballast water sampling. Mar. Pollut. Bull. 91 (1), 280–287.

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., Lee II, H., Reusser, D.A., 2013. Counting at low concentrations: the statistical challenges of verifying ballast water discharge standards. Ecol. Appl. 23, 339–351.

Gamrot, W., 2013. On exact computation of minimum sample size for restricted estimation of a binomial parameter. J. Stat. Plan. Infer. 143, 852–866.

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

IMO, 2015. Guidance on Ballast Water Sampling and Analysis for Trial Use in Accordance with the BWM Convention and Guidelines (G2). IMO BWM.2/Circ.42/Rev.1, accessed on March, 17th 2016. http://www.imo.org/en/OurWork/Environment/BallastWaterManagement/Pages/Default.aspx.

Matuszewski, D.J., Cesar, R.M., Strickler, J.R., Baldasso, L.F., Lopes, R.M., 2015. Visual rhythm for particle analysis in sample-in-flow systems: application for continuous plankton monitoring. Limnol. Oceanogr. Methods 13, 687–696.

McCarthy, S.A., McPhearson, R.M., Guarino, A., Gaines, J., 1992. Toxigenic *Vibrio cholerae* 01 and cargo ships entering Gulf of Mexico. Lancet 339, 624–625.

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.

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

Ruiz, G.M., Rawlings, T.K., Dobbs, F.C., Drake, L.A., Mullady, T., Huq, A., Colwell, R.R., 2000. Global spread of microorganisms by ships: ballast water discharged from vessels harbours a cocktail of potential pathogens. Nature 408, 49–50.

Tsolaki, E., Diamadopoulos, E., 2009. Technologies for ballast water treatment: a review. J. Chem. Technol. Biotechnol. 85, 19–32.