SORT 45 (1) January-June 2021, 19-32 DOI: 10.2436/20.8080.02.107

Verifying compliance with ballast water standards: a decision-theoretic approach

Eliardo G. Costa[1], Carlos Daniel Paulino[2] and Julio M. Singer[3]

Abstract

We construct credible intervals to estimate the mean organism (zooplankton and phytoplankton) concentration in ballast water via a decision-theoretic approach. To obtain the required optimal sample size, we use a total cost minimization criterion defined as the sum of the sampling cost and the Bayes risk either under a Poisson or a negative binomial model for organism counts, both with a gamma prior distribution. Such credible intervals may be employed to verify whether the ballast water discharged from a ship is in compliance with international standards. We also conduct a simulation study to evaluate the credible interval lengths associated with the proposed optimal sample sizes.

MSC: 62F15, 62P12.

Keywords: Optimal sample size, Bayes risk, Poisson distribution, negative binomial distribution.

1 Introduction

With the expansion of maritime traffic, ballast water has become the leading dispersing agent of invasive organisms with serious environmental, public health and economic consequences as indicated in Strayer (2010), McCarthy *et al.* (1992) and Marbuah, Gren and McKie (2014). In order to reduce the introduction of invasive species, specially zooplankton and phytoplankton, the international maritime community adopted the Ballast Water Management Convention (BWM Convention) in 2004, that has finally entered into force in 2017. Among other restrictions, the D-2 standard requires that deballasted water should contain no more than 10 viable organisms (referred to simply as organisms in the remainder) with maximum dimension between 10 μm and 50 μm per *mL* (IMO, 2004).

Given the large amount of ballast water carried by some vessels, it is impractical to analyze the whole water volume and an alternative is to rely on sampling methods that

[1] Departamento de Estatística, Universidade Federal do Rio Grande do Norte, Brazil.

[2] Departamento de Matemática, IST and CEAUL, FCUL Universidade de Lisboa, Portugal.

[3] Departamento de Estatística, Universidade de São Paulo, Brazil.

Received: July 2020

Accepted: February 2021

guarantee some acceptable error rates associated to the decision of whether a given deballasting process complies with the D-2 standard. Many authors (First *et al.*, 2013; Carney *et al.*, 2013; Gollasch and David, 2017; Casas-Monroy, Rajakaruna and Bailey, 2020) have addressed this issue, mentioning the quest for "representative" samples, without a consensus on a clear definition and examining samples obtained from a limited number of ship trips.

Very few articles deal with a more structured approach, in which a required sample size is computed to meet some maximum acceptable sampling error (Basurko and Mesbahi, 2011; Miller *et al.*, 2011; Frazier *et al.*, 2013). Costa, Lopes and Singer (2015, 2016), on the other hand, define "representative samples" as those that can be used to estimate the organism concentration in the ballast water tank with a pre-specified precision and use a frequentist approach to compute the optimal sample size with this characteristic. Costa, Paulino and Singer (2021) adopted a Bayesian approach to compute sample sizes required for estimating organism concentration obtained via two optimality criteria: the average coverage and the average length of credible intervals.

As many different tools or methods (*e.g.*, Niskin or Van Dorn bottles, plankton nets, pumps, or the in-line method) may be employed to collect samples from ballast water (Casas-Monroy *et al.*, 2020), it seems reasonable to include costs in the optimal sample size determination. With this in mind, we propose a Bayesian decision approach based on a criterion which minimizes the sum of the sampling method cost and the Bayes risk. An advantage of this approach is that the cost of collecting the sample is explicitly taken into account.

The proposed approach depends on an *ad hoc* loss function defined to accommodate the implications of using a credible interval for the organism concentration λ to decide for compliance or not with the D-2 standard. In a different setup, Etzioni and Kadane (1993) use a similar criterion with quadratic and logarithmic loss functions under a normal model. Sahu and Smith (2006) consider a loss function for the hypothesis testing problem of the parameter of a normal model. Islam (2011) and Islam and Pettit (2012, 2014) consider quadratic, linex and bounded linex loss functions for point estimation of the mean and the variance of a normal model with normal prior distributions, and also exponential and Poisson models both with a gamma prior distribution for point estimation of their respective parameters. Following a similar or the same approach, we may cite Pham-Gia and Turkkan (1992), Bernardo (1997), Lindley (1997), Parmigiani and Inoue (2009), De Santis and Gubbiotti (2017), among others.

Consider a sample $\boldsymbol{x}_n = (x_1, \dots, x_n)$ consisting of the counts of organisms in n aliquots (sub-samples) with a given volume w collected from a ballast water tank and a specified loss function L. The objective is to obtain the optimal sample size n_{o} that minimizes a total cost function consisting of the sum of a risk function r and a sampling cost $C(n)$. Once the required optimal sample size has been determined, the corresponding aliquots with volume w are collected (possibly on board or during the deballasting process), the organisms in these aliquots are counted and a credible interval with lower $a(\boldsymbol{x}_{n_{\mathrm{o}}})$ and upper $b(\boldsymbol{x}_{n_{\mathrm{o}}})$ limits for the mean organism concentration λ is computed. Considering that the D-2 standard requires $\lambda < 10$ for compliance, the ship is declared not compliant if $a(\boldsymbol{x}_{n_o}) \geq 10$ or compliant, if $b(\boldsymbol{x}_{n_o}) < 10$. Otherwise, if $a(\boldsymbol{x}_{n_o}) < 10 < b(\boldsymbol{x}_{n_o})$, more data are needed to make a decision.

In Section 2, we describe two Bayesian models required to compute the credible intervals. The first is appropriate for situations where the organisms are homogeneously distributed in the ballast water tank and the second may be needed for heterogeneous distributions. Sample size determination is presented in Section 3 in terms of a convenient loss function in a decision-theoretic approach. Additionally, we conduct a simulation study to evaluate the lengths of the credible intervals obtained for different combinations of the parameters governing the models and different sampling costs. We conclude, in Section 4, with a discussion of the results and of the difficulties associated to the establishment of the cost components.

2 Bayesian models

2.1 Poisson model with a gamma prior distribution

Let X be the number of organisms in an aliquot of volume w collected from a ballast tank with mean organism concentration λ. The expected number of organisms in this aliquot is $w\lambda$, *i.e.*, $\mathbb{E}[X|\lambda] = w\lambda$. Suppose that, given λ, X follows a Poisson distribution with mean $w\lambda$; this essentially corresponds to the assumption that the organisms are homogeneously distributed in the ballast tank. A possible and first natural choice for a prior distribution is the conjugate gamma distribution for which the density function is

$$h(\lambda) \propto \lambda^{\theta_0 - 1} \exp(-\theta_0 \lambda / \lambda_0),$$

where λ_0 and θ_0 are positive and known fixed constants, respectively interpreted as the prior mean and as a quantity inversely proportional to the prior variance. Thus, the larger (smaller) is θ_0, the smaller (larger) is the prior uncertainty about λ.

Considering a random sample of size n of $X|\lambda$ and a gamma prior distribution for λ, we may write the model hierarchically as follows

$$X_i|\lambda \overset{\text{iid}}{\sim} \text{Poisson}(w\lambda), \quad i = 1, 2, \ldots, n; \tag{1}$$

$$\lambda \sim \text{Gamma}(\theta_0, \theta_0/\lambda_0). \tag{2}$$

In this context, the posterior distribution of λ is also a gamma distribution with parameters $\theta_0 + s_n$ and $nw + \theta_0/\lambda_0$, where $s_n = \sum_{i=1}^n x_i$, *i.e.*, $\lambda|\boldsymbol{x}_n \sim \text{Gamma}(\theta_0 + s_n, nw + \theta_0/\lambda_0)$. Details are presented in the Supplementary Material.

2.2 Negative binomial model with a gamma prior distribution

Suppose that the organism concentration in the i-th aliquot is ℓ_i and the corresponding number of organisms is X_i, $i = 1,2,\ldots,n$. The expected number of organisms in the i-th aliquot is $\mathbb{E}[X_i|\ell_i] = w\ell_i$. For $i = 1,2,\ldots,n$, suppose that, given ℓ_i, X_i follows a Poisson distribution with mean $w\ell_i$ and that given a mean concentration λ in the tank, $\ell_i \sim \text{Gamma}(\phi,\phi/\lambda)$, so that $\mathbb{E}[\ell_i|\lambda|] = \lambda$ and $\text{Var}[\ell_i|\lambda] = \lambda^2/\phi$. Thus, given λ and ϕ, X_i follows a negative binomial distribution with $\mathbb{E}[X_i|\lambda,\phi] = w\lambda$ and $\text{Var}[X_i|\lambda,\phi] = w\lambda + (w\lambda)^2/\phi$, where ϕ is a shape (or agglomeration) parameter assumed known (see Amaral Turkman, Paulino and Müller, 2019, Appendix A on the Poisson-gamma mixture). We use the notation $X_i|\lambda,\phi \sim \text{NB}(w\lambda,\phi)$ and again assume a gamma prior distribution for λ.

Considering a random sample of size n from $X|(\lambda,\phi)$ and a gamma prior distribution for λ, we may write the model hierarchically as

$$X_i|\lambda,\phi \overset{\text{iid}}{\sim} \text{NB}(w\lambda,\phi), \quad i = 1,2,\ldots,n; \tag{3}$$

$$\lambda \sim \text{Gamma}(\theta_0,\theta_0/\lambda_0). \tag{4}$$

In this context, the posterior distribution of λ is not a known distribution and the computing of its summaries is analytically intractable. Thus, we rely on Markov chain Monte Carlo (MCMC) methods to generate random samples from the distribution of interest. In our case, we use the Metropolis-Hastings algorithm (Metropolis *et al.*, 1953; Hastings, 1970) based on a random walk to generate random samples from the posterior distribution of λ. With these samples we may compute related inference summaries. Details are presented in the Supplementary Material.

3 Sample size determination

An approach to the problem of sample size determination and credible interval estimation is to consider it as a decision problem (Lindley, 1997; Parmigiani and Inoue, 2009; Islam and Pettit, 2014). For this purpose, given that λ is the parameter of interest, it is necessary to specify a loss function $L(\lambda,d_n)$ based on a sample $\boldsymbol{X}_n = (X_1,X_2,\ldots,X_n)$ and a decision function $d_n \equiv d_n(\boldsymbol{X}_n)$. For a given n, the action $d_n(\boldsymbol{x}_n)$ consists of the specification of two quantities, the lower [say, $a(\boldsymbol{x}_n)$] and the upper [say, $b(\boldsymbol{x}_n)$] limits of a credible interval for λ.

Letting $f(\boldsymbol{x}_n|\lambda)$ be the sampling distribution for $\boldsymbol{X}_n$ and h be a prior distribution for the unknown parameter λ, the Bayes risk is (see Parmigiani and Inoue, 2009)

$$r(h,d_n) \;:=\; \int_\Lambda \int_{\mathcal{X}^n} L(\lambda,d_n) f(\boldsymbol{x}_n|\lambda) h(\lambda) d\boldsymbol{x}_n d\lambda, \tag{5}$$

where Λ is the parameter space and $\mathcal{X}^n$ is the sample space. The Bayes risk $r(h,d_n)$ may be viewed as the mean of the sampling expected loss expressed as a function of the parameter of interest weighted by the prior distribution; this summarizes the sampling expected loss over all possible values of the parameter of interest (here, the mean concentration λ).

The decision d_n^* that minimizes $r(h, d_n)$ among all possible decisions d_n is called a Bayes rule. Note that if the order of the integration may be inverted, we have

$$\begin{aligned} r(h, d_n) &= \int_{\mathcal{X}^n} \left[\int_{\Lambda} L(\lambda, d_n) h(\lambda | \boldsymbol{x}_n) d\lambda \right] f(\boldsymbol{x}_n) d\boldsymbol{x}_n \\ &= \int_{\mathcal{X}^n} \mathbb{E}\left[L(\lambda, d_n) \middle| \boldsymbol{x}_n \right] f(\boldsymbol{x}_n) d\boldsymbol{x}_n, \end{aligned} \tag{6}$$

where $f(\boldsymbol{x}_n)$ is the marginal distribution of the data, so that the decision d_n^* that minimizes $r(h, d_n)$ is the same that minimizes the posterior expected value of the loss function, namely $\mathbb{E}\left[L(\lambda, d_n) \middle| \boldsymbol{x}_n\right]$, for each $\boldsymbol{x}_n$. Given the specified action (the determination of the lower and upper limits of a credible interval for λ in our case), one must define a criterion to obtain an optimal sample size taking both the Bayes risk and the sampling cost into account. With this purpose, we minimize the total cost function $TC(n)$, customarily expressed by

$$TC(n) = r(h, d_n^*) + C(n),$$

where the function $C(n)$ needs to be specified. Here, we take $C(n) = cn$, with c being the cost of sampling an aliquot.

The additive structure of $TC(n)$ in terms of the cost of an action regarding the magnitude of λ and of the sample collection cost presupposes that they are measurable or scalable in some common unit (see Raiffa and Schlaifer, 1961, for example). In fact, we can view $C(n)$ as the relative cost of sampling expressed in terms of the cost associated to the Bayes risk.

Often it is not possible to compute $r(h, d_n^*)$ analytically. In such cases, we may use Monte Carlo simulations as an alternative. Since simulation methods are used, the estimates of $TC(n)$, denoted by $tc(n)$, may show a variation around its true value. We may reduce this variation by: (i) taking the number of Monte Carlo replicates as large as possible and/or, (ii) fitting a curve by least squares or some other method to a set of points $(n, tc(n))$. Müller and Parmigiani (1995) propose to fit the following curve to the estimates of $TC(n)$,

$$tc(n) = \frac{E}{(1 + Hn)^G} + cn,$$

where E, H and G are parameters to be estimated. The numerical methods required to estimate these parameters sometimes do not reach convergence depending on the initial values adopted to implement the corresponding algorithms. In order to simplify the fitting procedure and observing that the parameters H and G play similar roles and essentially represent the decreasing rate of the Bayes risk, we propose to fit the function

$$tc(n) = \frac{E}{(1+n)^G} + cn,$$

that may be linearized as

$$\log[tc(n) - cn] = \log E - G \log(1+n), \tag{7}$$

where the term $-\log(1+n)$ may be interpreted as an explanatory variable and $\log[tc(n) - cn]$, as a dependent variable in a linear regression model. Assuming that an error is added, the estimates of E and G may be computed by least squares. Then, the optimal sample size n_o is the largest integer closest to

$$\left(\frac{\widehat{E}\widehat{G}}{c}\right)^{1/(\widehat{G}+1)} - 1, \tag{8}$$

where $\widehat{E}$ and $\widehat{G}$ are, respectively, the least squares estimates of E and G.

We adopt the loss function

$$L(\lambda, d_n) = \gamma\tau + (\lambda - m)^2/\tau,$$

where $\gamma > 0$ is a fixed constant, $\tau = (b-a)/2$ is the half-length and $m = (a+b)/2$ is the center of the credible interval (see Rice, Lumley and Szpiro, 2008). The first term involves the half-width of the interval which we may interpret as its precision. The second term, namely, the square of the distance between the parameter of interest (λ) and the center of the interval divided by the half-width to maintain the same measurement unit of the first term, may be interpreted as the bias divided by the precision. If the precision increases (τ decreases) the second term of the loss function increases. The weights attributed to each term are γ and 1, respectively. If $\gamma < 1$, we attribute the largest weight to the second term, prioritizing lower bias over precision; if $\gamma > 1$, the situation is reversed and if $\gamma = 1$, the two terms have the same weight.

For this loss function, the Bayes rule corresponds to the quantities which define the interval $[a^*(\boldsymbol{x}_n), b^*(\boldsymbol{x}_n)] = [m^* - \mathrm{SV}_\gamma, m^* + \mathrm{SV}_\gamma]$, where $m^* = \mathbb{E}\left[\lambda \middle| \boldsymbol{x}_n\right]$ and $\mathrm{SV}_\gamma = \gamma^{-1/2}(\mathrm{Var}\left[\lambda \middle| \boldsymbol{x}_n\right])^{1/2}$. For more details see Parmigiani and Inoue (2009), Rice *et al.* (2008) or Schervish (1995).

In a practical situation, once the required optimal sample size n_o has been determined along with the corresponding organism counts $\boldsymbol{x}_{n_o}$, the credible interval limits $a^*(\boldsymbol{x}_{n_o})$ and $b^*(\boldsymbol{x}_{n_o})$ are obtained via $\mathbb{E}\left[\lambda \middle| \boldsymbol{x}_{n_o}\right]$ and $\mathrm{Var}\left[\lambda \middle| \boldsymbol{x}_{n_o}\right]$, expressed in terms of the models described in the preceding section.

An algorithm to obtain the optimal sample size satisfying the total cost minimization criterion for the adopted loss function is outlined in the Supplementary Material and the corresponding R code is available in Costa, Paulino and Singer (2020). In Tables 1-2 we present optimal sample sizes computed for different values of the parameters defining the prior distributions for both models considered in Section 2. We set λ_0 and obtain the

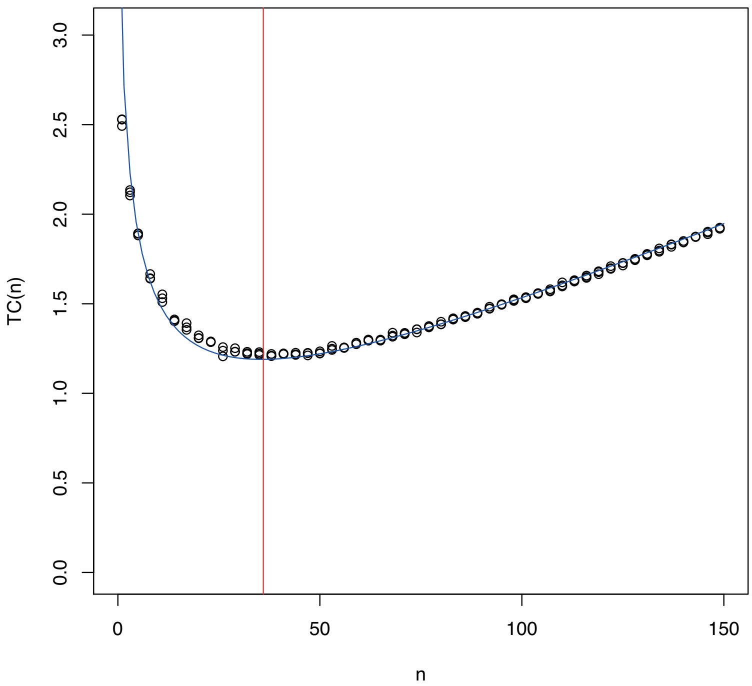

Figure 1: *Estimated total cost as a function of n for the negative binomial/gamma model with* $\gamma = 1/2$, $\phi = 22$, $w = 1$, $c = 0.01$, $\lambda_0 = 10$ *and prior variance equal to 4; the vertical line indicates the optimal sample size* $n_o = 36$.

value of θ_0 such that the prior variance is a constant, say, σ^2, *i.e.*, $\theta_0 = (\lambda_0/\sigma)^2$. See Figure S2 in the Supplementary Material. The values considered for ϕ were chosen to cover the range of estimates obtained from real data reported in Casas-Monroy *et al.* (2020). In Figure 1 we depict a curve fitted to the estimated total cost as a function of n for the negative binomial/gamma model with $\gamma = 1/2$, $\phi = 22$, $w = 1$, $c = 0.01$, $\lambda_0 = 10$ and prior variance equal to 4. The vertical line indicates the optimal sample size $n_{\mathrm{o}} = 36$.

We also carried out a simulation study to evaluate the lengths of the credible intervals and the respective Bayesian coverage probability computed from samples obtained with the proposed optimal sample sizes. For such purposes, we considered the optimal sample sizes obtained via either the Poisson/gamma or the negative binomial/gamma model for combinations of different values of c, θ_0 (and ϕ in the negative binomial/gamma model). For each scenario, we drew 1000 samples $\boldsymbol{x}_{n_{\mathrm{o}}}$ with the optimal sample size n_{o}, obtained the limits $a^*(\boldsymbol{x}_{n_{\mathrm{o}}})$ and $b^*(\boldsymbol{x}_{n_{\mathrm{o}}})$ of the corresponding credible intervals, computed the mean of their lengths and the mean of the Bayesian coverage probabilities (see Supplementary Material for more details). The results for the average lengths are displayed (within parentheses) in Tables 1-2. The average acceptance rates for the Metropolis-Hastings algorithm used in the negative binomial/gamma model ranged between 31% and 71%. The results for the Bayesian coverage probability are discussed in Section 4.

Table 1: *Optimal sample sizes n_o and estimated mean posterior credible interval lengths (within parentheses) under the Poisson/gamma model (1)-(2) with $w = 1$ and $\lambda_0 = 10$.*

γ	Aliquot cost (c)	Prior variance 1	2	4
1/2	0.001	145 (0.72)	157 (0.70)	164 (0.69)
	0.010	28 (1.45)	32 (1.50)	34 (1.48)
1	0.001	184 (0.45)	198 (0.44)	207 (0.44)
	0.010	35 (0.94)	40 (0.94)	43 (0.94)
2	0.001	237 (0.28)	252 (0.28)	263 (0.27)
	0.010	46 (0.60)	51 (0.60)	55 (0.59)

A simple algorithm with the steps required for the determination of n_o and for the decision with respect to D-2 standard follows.

Step 1. Set the values of λ_0 and θ_0 (prior distribution), ϕ (only for negative binomial model), w (aliquot volume), c (aliquot cost) and γ (loss function).

Step 2. Obtain the corresponding optimal sample size n_o using the algorithm provided in the Supplementary Material with the parameter values defined in Step 1.

Step 3. Sample n_o aliquots of water from the ballast tank of the ship and count the number of organisms in each aliquot. We denote these n_o organism counts as $\boldsymbol{x}_{n_\text{o}} = (x_1, \ldots, x_{n_\text{o}})$.

Step 4. With the organism counts $\boldsymbol{x}_{n_o}$ and γ compute the credible interval limits $a^*(\boldsymbol{x}_{n_\text{o}})$ and $b^*(\boldsymbol{x}_{n_\text{o}})$ via $\mathbb{E}\left[\lambda \middle| \boldsymbol{x}_{n_\text{o}}\right]$ and $\text{Var}\left[\lambda \middle| \boldsymbol{x}_{n_\text{o}}\right]$. If there is no closed form for these moments of the posterior distribution, compute estimates for these quantities simulating values from the posterior distribution (using MCMC or another simulation-based method) and taking the respective sample moments.

Step 5. Use the credible interval limits to decide for compliance with the D-2 standard as follows: declare compliance if $b^*(\boldsymbol{x}_{n_\text{o}}) < 10$, or non-compliance if $a^*(\boldsymbol{x}_{n_\text{o}}) \geq 10$. Otherwise, if $a^*(\boldsymbol{x}_{n_\text{o}}) < 10 < b^*(\boldsymbol{x}_{n_\text{o}})$, more data are required to make a decision.

4 Discussion

We propose a decision-theoretic approach to obtain an optimal number of aliquots required to estimate the organism concentration in ballast water and indicate how the results may be employed to verify compliance with the D-2 standard.

The results in Table 1 obtained under the Poisson/gamma model indicate that the optimal sample size n_o increases as the prior uncertainty (variance) about λ increases, but the average interval length remains the same. For the negative binomial/gamma model we observe a similar behavior (see Table 2).

Table 2: *Optimal sample sizes n_o and estimated mean posterior credible interval lengths (within parentheses) under the negative binomial/gamma model (3)-(4) with $w = 1$ and $\lambda_0 = 10$.*

γ	Aliquot cost (c)	ϕ	Prior variance 1	Prior variance 2	Prior variance 4
1/2	0.001	1	276 (1.51)	345 (1.48)	347 (1.52)
		4	220 (1.05)	246 (1.03)	259 (1.03)
		8	162 (0.99)	185 (0.96)	206 (0.91)
		13	162 (0.89)	180 (0.87)	192 (0.85)
		22	156 (0.83)	172 (0.80)	182 (0.79)
	0.010	1	-	29 (3.25)	60 (3.18)
		4	21 (2.23)	37 (2.26)	42 (2.35)
		8	26 (1.93)	32 (2.05)	40 (1.99)
		13	25 (1.82)	33 (1.84)	38 (1.82)
		22	23 (1.75)	31 (1.74)	36 (1.71)
1	0.001	1	365 (0.96)	436 (0.95)	458 (0.96)
		4	232 (0.73)	267 (0.70)	292 (0.68)
		8	217 (0.61)	243 (0.59)	260 (0.58)
		13	208 (0.56)	229 (0.55)	244 (0.53)
		22	200 (0.52)	219 (0.51)	231 (0.50)
	0.010	1	-	48 (2.07)	78 (2.04)
		4	34 (1.42)	47 (1.47)	56 (1.47)
		8	36 (1.24)	45 (1.26)	51 (1.26)
		13	35 (1.16)	43 (1.17)	49 (1.15)
		22	35 (1.08)	42 (1.09)	47 (1.07)
2	0.001	1	478 (0.61)	510 (0.63)	678 (0.56)
		4	301 (0.46)	344 (0.44)	373 (0.43)
		8	281 (0.39)	310 (0.37)	331 (0.36)
		13	268 (0.35)	293 (0.34)	309 (0.34)
		22	257 (0.33)	279 (0.32)	292 (0.31)
	0.010	1	-	75 (1.31)	109 (1.27)
		4	61 (0.85)	70 (0.89)	70 (0.95)
		8	51 (0.78)	56 (0.82)	65 (0.80)
		13	51 (0.72)	54 (0.75)	62 (0.73)
		22	49 (0.68)	53 (0.70)	59 (0.68)

For both models and for all n_o in Tables 1-2 the average Bayesian coverage probabilities obtained in the simulation study were approximately 0.84, 0.68 and 0.52 for $\gamma = 1/2, 1$ and 2, respectively. These values are similar to the probabilities that a standard normal variable lies in the intervals $(-\sqrt{2}, \sqrt{2})$, $(-1, 1)$ and $(-1/\sqrt{2}, 1/\sqrt{2})$, respectively, and are consistent with the asymptotic normality of the corresponding posterior distributions. See Ferguson (1996, pg. 140), for example. However, we observe that this approximation also occurs for $n_o \approx 30$. To explain this, first, note that as $\theta_0 \to \infty$ the gamma distribution approaches a normal distribution (McCullagh and Nelder, 1989, pg. 287). For the Poisson/gamma model (1)-(2) the respective posterior distribution is also gamma with shape parameter $\theta_0 + s_n$, and the cases for which $n_o \approx 30$ are those where the prior variance is equal to 1 or 2 and correspond to θ_0 equal to 100 or 50, respectively. We may consider these values of θ_0 large enough to guarantee a reasonable approximation by the normal distribution.

The posterior distribution for the negative binomial/gamma model (3)-(4) is neither a gamma distribution nor a known distribution. To verify whether the normal approximation also holds in this case, we considered the smallest sample size in Table 2, namely $n_o = 21$ and generated 100 samples of size 100 from the posterior distribution of λ. We applied the Shapiro-Wilk test to each of these samples and observed that 90 out of 100 p-values were greater than 0.05, *i.e.*, that the normal approximation seems reasonable even for the smallest n_o in Table 2. This suggests, for example, that in order to obtain a Bayesian coverage probability of 0.95, we must have $\gamma = 1/1.96^2$. In general, if we want a Bayesian coverage probability approximately equal to $1-\rho$, we must set $\gamma = 1/[\Phi^{-1}(1-\rho/2)]^2$, where $\Phi^{-1}(\cdot)$ is the inverse probability function of the standard normal distribution. In other words, larger coverage probabilities requires smaller values for γ, which places more emphasis on the center than on the length of the corresponding credible interval.

We also observe that when the cost of sampling an aliquot c increases, the optimal sample size (and consequently, the average interval length) decreases (increases) under either model, but at the expense of an increase in the total cost (Tables 1-2). For example, if we set the prior variance equal to 4, $\gamma = 1/2$ and $\phi = 1$, from the results in Table 2, it follows that the optimal sample size for $c = 0.001$ is 347, generating a sampling cost of $C(347) = 0.001 \times 347 = 0.347$; the optimal sample size for $c = 0.010$, on the other hand, is 60, generating a sampling cost $C(60) = 0.010 \times 60 = 0.60$, an increase of $\approx 73\%$.

Although the aggregation parameter ϕ represents an important feature related to the heterogeneity of the organism distribution in the ballast water tank, under the total cost minimization approach, the optimal sample size is only slightly affected when ϕ increases (with the other parameters fixed) for $c = 0.01$. As displayed in Table 2, for $\phi \geq 4$ the optimal sample sizes are almost the same for different values of the prior variance. Also, note that for $c = 0.01$, $\phi = 1$ and prior variance equal to 1 we have no entry in Table 2 because there is no associated optimal sample size. This means that the cost of sampling outweighs the cost of decreasing the Bayes risk and it is not worth obtaining aliquots. This was also observed by Etzioni and Kadane (1993) and Islam and Pettit (2014, Table 1).

Such considerations point to a major difficulty of the proposed approach which is the quantification of the “costs” associated to the Bayes risk and to the sampling effort. Although the latter may be objectively calculated in terms of the technical aspects of the actual collection and analysis methods, the former certainly poses a complicated problem since it depends on quantitatively evaluating consequences of declaring a ship compliant or not based on a rule defined in terms of the credible interval. This is certainly a controversial and difficult problem; however, it permeates directly or indirectly, all methods of sample size determination and decision making.

Table 3: *Simulated organism counts obtained via the negative binomial model with $\phi = 8$, $w = 1$ and λ fixed as reported.*

λ	counts												
7	8	6	5	10	18	2	5	13	3	8	10	4	7
	7	8	11	6	3	3	4	12	3	1	6	2	7
	8	1	11	3	1	5	8	5	2	8	5	5	11
	4	4	3	11	6	2	10	6	6	7	7	8	
10	11	5	22	23	12	4	13	13	3	18	5	10	10
	15	20	27	3	15	4	5	11	11	21	7	3	6
	10	5	9	8	8	5	12	12	2	5	11	9	14
	10	9	10	15	15	10	11	8	7	7	8	6	
13	15	6	21	31	19	17	5	23	7	13	12	25	24
	6	24	12	3	11	7	23	13	5	3	9	16	9
	9	12	11	7	15	16	3	7	15	12	17	13	11
	13	17	20	9	11	8	9	11	8	12	3	13	

For illustrative purposes, we consider a set of hypothetical organism counts to obtain the associated credible interval based on the optimal sample size. Casas-Monroy *et al.* (2020) obtained estimates for ϕ varying from 8 to 22. For $\phi = 8$, the optimal sample size under the negative binomial/gamma model with $\lambda_0 = 10$, prior variance equal to 4 and $c = 0.010$ is $n_o = 51$ (Table 2). We generated 51 observations from a negative binomial model with $\lambda = 7$, $\phi = 8$ and $w = 1$ (see Table 3). Given the generated observations, we drew a sample of size 10,000 from the posterior distribution of λ with a burn-in of 1,000 iterations and a thinning of 10. The corresponding credible interval $[a^*, b^*]$ is $[6.10, 7.05]$. Now, if we generate 51 observations from a negative binomial model with $\lambda = 10$ (Table 3), the corresponding credible interval is $[9.61, 10.89]$. Finally, if we set $\lambda = 13$ to generate the 51 observations, the corresponding interval is $[11.58, 13.04]$. In all cases, the credible interval contains the value of the parameter of interest and lead to correct decisions relatively to compliance v.s. non-compliance with the D-2 standard.

Acknowledgements

The authors appreciate the enlightening comments of two anonymous referees which improved the paper considerably. This research received financial support from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, grants 153526/2014-9 and 304841/2019-6) and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, grant 2013/21728-2), Brazil. This research was also supported by the Fundação para a Ciência e Tecnologia (FCT), Portugal, under Projects UID/MAT/00006/2019 and UID/MAT/00006/2013.

References

Amaral Turkman, M. A., Paulino, C. D. and Müller, P. (2019). *Computational Bayesian Statistics: An Introduction*. Cambridge: Cambridge University Press.

Basurko, O. C. and 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, 183–190.

Bernardo, J. M. (1997). Statistical inference as a decision problem: the choice of sample size. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 46, 151–153.

Carney, K. J., Basurko, O. C., Pazouki, K., Marsham, S., Delany, J. E., Desai, D. V., Anil, A. C. and Mesbahi, E. (2013). Difficulties in obtaining representative samples for compliance with the Ballast Water Management Convention. *Marine Pollution Bulletin*, 68, 99–105.

Casas-Monroy, O., Rajakaruna, H. and Bailey, S. A. (2020). Improving estimation of phytoplankton abundance and distribution in ballast water discharges. *Journal of Applied Phycology*, 32, 1185–1199.

Costa, E. G., Lopes, R. M. and Singer, J. M. (2015). Implications of heterogeneous distributions of organisms on ballast water sampling. *Marine Pollution Bulletin*, 91, 280–287.

Costa, E. G., Lopes, R. M. and Singer, J. M. (2016). Sample size for estimating the mean concentration of organisms in ballast water. *Journal of Environmental Management* 180, 433–438.

Costa, E. G., Paulino, C. D. and Singer, J. M. (2020). ssdet: Sample size determination in frequentist and Bayesian approaches. `http://www.github.com/eliardocosta/ssdet`. R package version 0.0.0.9400.

Costa, E. G., Paulino, C. D. and Singer, J. M. (2021). Sample size for estimating organism concentration in ballast water: A Bayesian approach. *Brazilian Journal of Probability and Statistics*, 35, 158–171.

De Santis, F. and Gubbiotti, S. (2017). A decision-theoretic approach to sample size determination under several priors. *Applied Stochastic Models in Business and Industry*, 33, 282–295.

Etzioni, R. and Kadane, J. B. (1993). Optimal experimental design for another's analysis. *Journal of the American Statistical Association*, 88, 1404–1411.

Ferguson, T. S. (1996). *A course in large sample theory*. London: Chapman and Hall.

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

Frazier, M., Miller, A. W., LeeII, H. and Reusser, D. A. (2013). Counting at low concentrations: the statistical challenges of verifying ballast water discharge standards. *Ecological Applications*, 23, 339–351.

Gollasch, S. and David, M. (2017). Recommendations for representative ballast water sampling. *Journal of Sea Research*, 123, 1–15.

Hastings, W. K. (1970). Monte Carlo sampling methods using markov chains and their applications. *Biometrika*, 57, 97–109.

IMO (2004). International convention for the control and management of ship ballast water and sediments. `http://www.imo.org/`.

Islam, A. F. M. (2011). *Loss functions, utility functions and Bayesian sample size determination.* Ph.D. thesis. Queen Mary, University of London.

Islam, A. F. M. S. and Pettit, L. I. (2012). Bayesian sample size determination using linex loss and linear cost. *Communications in Statistics-Theory and Methods*, 41, 223–240.

Islam, A. F. M. S. and Pettit, L. I. (2014). Bayesian sample size determination for the bounded linex loss function. *Journal of Statistical Computation and Simulation*, 84, 1644–1653.

Lindley, D. V. (1997). The choice of sample size. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 46, 129–138.

Marbuah, G., Gren, I.-M. and McKie, B. (2014). Economics of harmful invasive species: a review. *Diversity*, 6, 500–523.

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

McCullagh, P. and Nelder, J. (1989). *Generalized linear models*, 2nd ed. London: Chapman and Hall.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21, 1087–1092.

Miller, A. W., Frazier, M., Smith, G. E., Perry, E. S., Ruiz, G. M. and Tamburri, M. N. (2011). Enumerating sparse organisms in ships' ballast water: why counting to 10 is not easy. *Environmental Science and Technology*, 45, 3539–3546.

Müller, P. and Parmigiani, G. (1995). Optimal design via curve fitting of Monte Carlo experiments. *Journal of the American Statistical Association*, 90, 1322–1330.

Parmigiani, G. and Inoue, L. Y. T. (2009). *Decision theory: principles and approaches*. New York: John Wiley and Sons.

Pham-Gia, T. and Turkkan, N. (1992). Sample size determination in Bayesian analysis. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 41, 389–392.

Raiffa, H. and Schlaifer, R. (1961). *Applied statistical decision theory*. Boston: Wiley Cambridge.

Rice, K. M., Lumley, T. and Szpiro, A. A. (2008). Trading bias for precision: decision theory for intervals and sets. `http://www.bepress.com/uwbiostat/paper336`. Working Paper 336, UW Biostatistics.

Sahu, S. K. and Smith, T. M. F. (2006). A Bayesian method of sample size determination with practical applications. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 169, 235–253.

Schervish, M. (1995). *Theory of Statistics*. New York: Springer-Verlag.

Strayer, D. L. (2010). Alien species in fresh waters: ecological effects, interactions with other stressors, and prospects for the future. *Freshwater Biology*, 55, 152–174.