Brazilian Journal of Probability and Statistics
2021, Vol. 35, No. 1, 158–171
https://doi.org/10.1214/20-BJPS470

Sample size for estimating organism concentration in ballast water: A Bayesian approach

Eliardo G. Costa[a], **Carlos Daniel Paulino**[b] **and Julio M. Singer**[c]

[a]*Universidade Federal do Rio Grande do Norte*
[b]*CEAUL, Universidade de Lisboa*
[c]*Universidade de São Paulo*

Abstract. Estimation of microorganism concentration in ballast water tanks is important to evaluate and possibly to prevent the introduction of invasive species in stable ecosystems. For such purpose, the number of organisms in ballast water aliquots must be counted and used to estimate their concentration with some precision requirement. Poisson and negative binomial models have been employed to describe the organism distribution in the tank, but determination of sample sizes required to generate estimates with pre-specified precision is still not well established. A Bayesian approach is a flexible alternative to accommodate adequate models that account for the heterogeneous distribution of the organisms and may provide a sequential way of enhancing the estimation procedure by updating the prior distribution along the ballast water discharging process. We adopt such an approach to compute sample sizes required to construct credible intervals obtained via two optimality criteria that have not been employed in this context. Such intervals may be used in the decision with respect to compliance with the D-2 standard of the Ballast Water Management Convention. We also conduct a simulation study to verify whether the credible intervals obtained with the proposed sample sizes satisfy the precision criteria.

1 Introduction

Evaluation of ballast water discharges from ships is a topic of current interest because the possible introduction of invasive species in stable ecosystems may bring serious environmental and economic consequences. Estimates of damage costs of invasive species may vary from 0.4 to 220 billion USD per year in 2008 prices (Marbuah, Gren and McKie, 2014, Table 1). Among other requirements, the D-2 standard of the Ballast Water Management (BWM) Convention requires that deballasted water should contain less than 10 viable organisms (referred to simply as organisms in the remainder) per mL, sized ≥ 10 and < 50 μm in minimum dimension (IMO, 2004). This class of organisms comprise zooplankton and phytoplankton. An overview of research in ballast water in the last thirty years is presented in Bailey (2015).

Given the large amount of ballast water (up to thousands of tons) transported by big vessels, one has to rely on sampling methods to verify whether the standard is satisfied. This is a difficult task, especially for organisms sized ≥ 10 and < 50 μm in minimum dimension. Recently, Jang, Kim and Choi (2019) mention that there is no established sampling methodology for such microorganisms. Although this topic has attracted the attention of many researchers, very few papers deal with sample sizes required to verify compliance with the D-2 norm. Exceptions are Basurko and Mesbahi (2011), Miller et al. (2011) and Frazier et al. (2013), who discuss the problem but do not provide a formal solution. Costa, Lopes and Singer (2015, 2016) attacked the problem with a more formal approach, defining the sample

Key words and phrases. Average coverage criterion, average length criterion, Poisson/gamma distribution, negative binomial/Pearson Type VI distribution.
Received September 2019; accepted March 2020.

as a set of n aliquots (sub-samples) in each of which the number of organisms is counted. The volume of each sample, say w, depends on the specific counting procedure (Casas-Monroy, Rajakaruna and Bailey, 2020). The sampling process is based on a probabilistic model and on a criterion according to which one must compute the number of aliquots of ballast water with volume w mL needed to decide with a certain margin of error whether the D-2 standard is complied with. One of the difficulties with this approach relates to the heterogeneous nature of the organism concentration in the ballast water tank (Murphy, Ritz and Hewitt, 2002, Carney et al., 2013, Casas-Monroy, Rajakaruna and Bailey, 2020). Based on frequentist methods, Costa, Lopes and Singer (2015, 2016) adopted models that take this heterogeneity into account. In particular, Costa, Lopes and Singer (2015) consider Poisson and negative binomial distributions and specify probabilities for Type I and II errors to test the hypothesis that the mean organism concentration in the tank is smaller than 10 organisms per mL. Costa, Lopes and Singer (2016), on the other hand, consider the same probability distributions and specify a lower bound to the probability that the difference between the mean concentration and its estimate be less than a fixed value. Costa (2017) suggests the adoption of more flexible models that may possibly incorporate knowledge acquired over time. Bayesian models are excellent candidates to incorporate such characteristics because that information may be considered in the prior distribution which may also be updated when more data is obtained.

Two criteria are widely used in the Bayesian literature, but not in the setup under investigation for sample size determination, namely, the average coverage and the average length of credible intervals (ACC and ALC, respectively). In both cases, we choose the smallest sample size that satisfies the condition imposed on some specified average characteristic of the posterior distribution of the parameter of interest.

For the ACC we compute the posterior probability of a highest posterior density (HPD) interval with fixed length for each sample $\boldsymbol{x}_n$ of size n and weigh it by the marginal distribution of the data. This average probability must be not smaller than a specified lower bound. For the ALC, on the other hand, we compute the length of an interval with fixed credible degree for each $(\boldsymbol{x}_n, n)$ and weigh it by the same marginal distribution. The average length must not be larger than a specified upper bound.

Adcock (1987, 1988) uses the ACC (with a different label) to determine sample sizes required to estimate multinomial probabilities under Dirichlet prior distributions as well as to estimate the mean and the variance of normal distributions with prior normal or chi-squared distributions for the cases where the variance is known or unknown, respectively. Joseph, Wolfson and Berger (1995) and Joseph, Berger and Bélisle (1997) use both the ACC and the ALC, among other Bayesian criteria for estimating the proportion and the difference between two proportions under binomial distributions with beta prior distributions.

Wang and Gelfand (2002) use the same criteria to determine the sample size for the estimation of parameters of distributions belonging to the exponential family, of parameters in Weibull survival models as well as of parameters in logistic regression models. M'Lan, Joseph and Wolfson (2006) use the ACC and ALC criteria in the context of case-control studies; Stamey, Young and Bratcher (2006) also consider these criteria to estimate the parameters of Poisson distributions as well to estimate the difference or the ratio of the parameters of two Poisson distributions. We may also cite Joseph and Bélisle (1997), Joseph and Wolfson (1997), Rahme, Joseph and Gyorkos (2000), De Santis (2007), M'Lan, Joseph and Wolfson (2008) for related work.

We consider a Bayesian approach to compute minimum sample sizes required to obtain lower and upper limits of credible intervals for the mean organism concentration in a ballast water tank with specified average coverage or average length. Letting $\boldsymbol{x}_n$ denote a sample of size n determined according to the proposed approach, the credible intervals defined by the lower [say, $a(\boldsymbol{x}_n)$] and upper [say, $b(\boldsymbol{x}_n)$] limits will have in average, the specified coverage

or length. Given this interval, the ship from which the sample was collected is declared not compliant with the D-2 standard if $a(\boldsymbol{x}_n) \geq 10$ or compliant, if $b(\boldsymbol{x}_n) < 10$. Otherwise, if $a(\boldsymbol{x}_n) < 10 \leq b(\boldsymbol{x}_n)$, more data are needed to make a decision.

In Section 2, we describe the adopted Bayesian models. Sample size determination under both the ACC and the ALC criteria is discussed and implemented in Section 3. A simulation study to evaluate whether HPD intervals constructed with the proposed sample sizes satisfy the adopted optimality criteria is presented in Section 4. The last section discusses the results obtained from the computation of the minimum sample sizes as well as from the simulation study. Algorithms for sample size computations, written with the R language (R Core Team, 2016), are presented in the Supplementary Material (Costa, Paulino and Singer, 2020).

2 Bayesian models

2.1 Poisson model with a gamma prior distribution

Given that the expected organism concentration (per unit of volume) in the tank is λ, let X be the number of organisms in an aliquot of volume w; in this aliquot, we expect to find $\mathbb{E}[X|\lambda] = w\lambda$ organisms. Suppose that the organisms are homogeneously distributed in the ballast water tank, and that, given λ, X follows a Poisson distribution with mean $w\lambda$.

The natural (conjugate) choice for the prior distribution is a gamma distribution with parameters θ_0 and λ_0, namely $\lambda \sim \mathrm{G}(\theta_0, \theta_0/\lambda_0)$, for which the kernel of the probability density function is

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

This implies that $\mathbb{E}[\lambda] = \lambda_0$ and $\mathrm{Var}[\lambda] = \lambda_0^2/\theta_0$. In this context, λ_0 represents a prior expected concentration and θ_0 controls the variability of λ around λ_0. The gamma distribution provides ample flexibility to model the shape of the prior knowledge on the mean concentration λ. In Figure S1 of the Supplementary Material, we present gamma density functions with different shapes.

Consider a random sample $\boldsymbol{x}_n = (x_1, \ldots, x_n)$ of size n of $X|\lambda$ and a gamma prior distribution for λ. We may write the model hierarchically as follows

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

$$\lambda \sim \mathrm{G}(\theta_0, \theta_0/\lambda_0). \tag{2}$$

The posterior distribution of λ is also gamma, with parameters $\theta_0 + s_n$ and $nw + \theta_0/\lambda_0$, where $s_n = \sum_{i=1}^{n} x_i$, that is, $\lambda|\boldsymbol{x}_n \sim \mathrm{G}(\theta_0 + s_n, nw + \theta_0/\lambda_0)$. An example of prior and posterior densities is presented in Figure 1. The effect of the observed data is clearly observed to lead to a posterior distribution more concentrated than the prior distribution.

2.2 Negative binomial model with a Pearson type VI prior distribution

In contrast with the homogeneity assumption for the organism distribution in the tank inherent to the Poisson model, consider a more realistic situation where the organisms are distributed heterogeneously. A reasonable model for the organism distribution in this case is the negative binomial distribution, which may be motivated as follows.

Suppose that the organism concentration varies in the tank according to a gamma distribution with parameters ϕ and ϕ/λ. Consider n aliquots with volume w randomly selected from the tank and let λ_i denote the corresponding organism concentration, $i = 1, \ldots, n$. Assume that given λ_i, the corresponding number of organisms, X_i, follows a Poisson distribution with mean $w\lambda_i$. Thus, given λ and ϕ, X_i follows a negative binomial distribution with

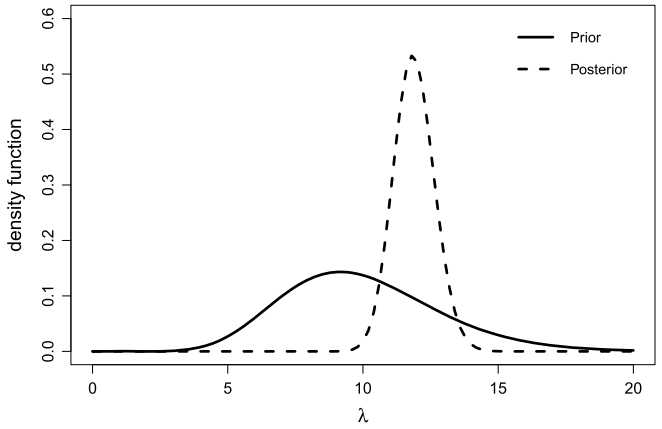

Figure 1 *Prior gamma distribution* $\mathrm{G}(\theta_0, \theta/\lambda_0)$ *and posterior gamma distribution* $\mathrm{G}(\theta_0 + s_n, nw + \theta_0/\lambda_0)$ *with* $w = 1$, $\lambda_0 = 10$, $\theta_0 = 12$, $n = 20$ *and* $s_n = 240$.

$\mathbb{E}[X_i|\lambda, \phi] = w\lambda$ and $\mathrm{Var}[X_i|\lambda, \phi] = w\lambda + (w\lambda)^2/\phi$, where ϕ is a shape (or agglomeration) parameter. This is denoted as $X_i|\lambda, \phi \sim \mathrm{NB}(w\lambda, \phi)$ and the probability function can be written as

$$f(x_i|\lambda, \phi) = \frac{\Gamma(\phi + x_i)}{\Gamma(x_i + 1)\Gamma(\phi)} \left(\frac{w}{\phi}\lambda\right)^{x_i} \left(1 + \frac{w}{\phi}\lambda\right)^{-\phi - x_i}.$$

For inferences on the parameter of interest λ, we consider ϕ fixed (taking on several values in the following analysis). A natural conjugate prior distribution for the parameter λ of the negative binomial distribution is a Pearson Type VI distribution (Johnson, Kotz and Balakrishnan, 1994a, 1994b), also known as a generalized beta prime distribution with a further scale factor ϕ/w, for which the kernel of the probability density function is

$$h(\lambda) \propto \frac{w}{\phi} \left(\frac{w}{\phi}\lambda\right)^{\theta_0 - 1} \left(1 + \frac{w}{\phi}\lambda\right)^{-\theta_0 - (\theta_0/\lambda_0 + 1)},$$

with location parameter 0, scale parameter ϕ/w and shape parameters θ_0 and $\theta_0/\lambda_0 + 1$, where λ_0 and θ_0 are known positive fixed constants (hyperparameters). We use the notation $\lambda \sim \mathrm{PVI}(0, \phi/w, \theta_0, \theta_0/\lambda_0 + 1)$. In this case, $\mathbb{E}[\lambda] = (\phi/w)\lambda_0$ and $\mathrm{Var}[\lambda] = (\lambda_0^2/\theta_0)[\phi^2(\lambda_0 + 1)/(w^2(1 - \lambda_0/\theta_0))]$, for $\lambda_0 < \theta_0$.

In the Poisson model with gamma prior distribution, we have $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|\lambda]] = \mathbb{E}[w\lambda] = w\lambda_0$, that is, the expected number of organisms when collecting an aliquot depends only on the hyperparameter λ_0. This makes sense since we are assuming homogeneity for the concentration, and regardless of the location where we collect an aliquot in the ballast water tank, we expect to find the same number of organisms. On the other hand, if we consider the negative binomial model with a Pearson Type VI prior distribution, we have $\mathbb{E}[X] = \phi\lambda_0$ so that the expected number of organisms in an aliquot depends on the parameter ϕ that controls the heterogeneity of the organisms in the tank. Note that ϕ is also a scale parameter for the prior distribution and the larger its value, the more spread out is the distribution with the other parameters fixed, indicating a vague prior knowledge about the parameter of interest. Furthermore, we can set the other parameters in such a way that the prior distribution may represent cases where there is high probability associated to an interval even when the value of ϕ increases. When λ_0 and θ_0 are fixed and ϕ increases, we have distributions representing cases with large variability. Examples are presented in Figures S2 and S3 of the Supplementary Material.

Consider a random sample of size n from $X|\lambda, \phi$, and a Pearson Type VI prior distribution for λ. We may write the model hierarchically as follows

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

$$\lambda \sim \text{PVI}(0, \phi/w, \theta_0, \theta_0/\lambda_0 + 1). \tag{4}$$

The posterior distribution of λ is Pearson Type VI, with the same location and scale parameters as the prior distribution, and shape parameters $\theta_0 + s_n$ and $\theta_0/\lambda_0 + n\phi + 1$, that is, $\lambda|\boldsymbol{x}_n \sim \text{PVI}(0, \phi/w, \theta_0 + s_n, \theta_0/\lambda_0 + n\phi + 1)$.

We must emphasize that ϕ plays two roles in model (3)–(4). In (3), it plays the role of a dispersion (or agglomeration) parameter. The larger is ϕ, the more homogeneous is the organism concentration in the tank. In the prior distribution (4), ϕ plays the role of scale parameter. Keeping the other parameters fixed, the larger is ϕ, the less precise is the prior knowledge about the parameter of interest (see Figure S2 of the Supplementary Material). This does not mean that if ϕ (previously known) is large we may only assign prior distributions with large variability, because we may specify the parameters λ_0 and θ_0 to adjust the precision of the prior knowledge even with large values of ϕ (see Figure S3 of the Supplementary Material).

3 Sample size determination

We consider two criteria to determine the minimum sample size, that is, minimum number of aliquots, required to estimate λ with a pre-specified precision.

3.1 Average coverage criterion (ACC)

The objective is to obtain the minimum sample size n such that the credible interval $R(\boldsymbol{x}_n)$ for λ has a pre-specified length with posterior probability at least equal to $1 - \rho$, that is,

$$\int_{R(\boldsymbol{x}_n)} h(\lambda|\boldsymbol{x}_n)\, d\lambda \geq 1 - \rho,$$

where $\boldsymbol{x}_n$ is a sample of size n and $R(\boldsymbol{x}_n)$ is a subset (an interval in our case) of the parameter space. Since the sample size determination precedes the actual sampling, we must consider all possible outcomes for $\boldsymbol{x}_n$ to achieve the objective. In this direction, we may weigh each outcome by its probability, that is,

$$\int_{\mathcal{X}^n} \left[\int_{R(\boldsymbol{x}_n)} h(\lambda|\boldsymbol{x}_n)\, d\lambda \right] g(\boldsymbol{x}_n)\, d\boldsymbol{x}_n \geq 1 - \rho,$$

where $\mathcal{X}^n$ is the sample space associated to $\boldsymbol{x}_n$ and $g(\boldsymbol{x}_n)$ is the marginal probability function of the outcomes.

For models (1)–(2) and (3)–(4), the credible region in general is an interval and in this case, we consider the highest posterior density (HPD) interval to define $R(\boldsymbol{x}_n) = [a(\boldsymbol{x}_n), b(\boldsymbol{x}_n)]$. We fix the length $\ell > 0$ of the desired HPD intervals for λ, specify the minimum Bayesian coverage probability, $1 - \rho$ and determine the minimum sample size as well as the bounds $a(\boldsymbol{x}_n)$ and $b(\boldsymbol{x}_n) = a(\boldsymbol{x}_n) + \ell$ such that

$$\int_{\mathcal{X}^n} \left[\int_{a(\boldsymbol{x}_n)}^{b(\boldsymbol{x}_n)} h(\lambda|\boldsymbol{x}_n)\, d\lambda \right] g(\boldsymbol{x}_n)\, d\boldsymbol{x}_n \geq 1 - \rho. \tag{5}$$

Given $a(\boldsymbol{x}_n)$, $b(\boldsymbol{x}_n)$ and the parameters of the posterior distribution, the inner integral in (5) may be obtained computationally; the outer integral may be estimated via Monte Carlo simulation. An algorithm to obtain the minimum sample size satisfying the criterion is outlined in the Supplementary Material.

Table 1 *ACC (5) based minimum sample size (n) computed under the Poisson/gamma model (1)–(2) with $\lambda_0 = 10$, $\rho = 0.05$ and $\ell = 2$*

Aliquot volume (w)	Shape parameter (θ_0)				
	1.0	2.5	5.0	7.5	10.0
0.5	77	77	76	76	75
1.0	39	39	38	38	38

Table 2 *ACC (5) based minimum sample size (n) computed under the negative binomial/Pearson Type VI model (3)–(4) with $\lambda_0 = 10(w/\phi)$, $\rho = 0.05$ and $\ell = 2$*

Aliquot volume (w)	ϕ	Shape parameter (θ_0)			
		11	25	50	75
0.5	1	663	558	517	480
	8	134	126	118	111
	13	113	106	99	93
	22	97	92	86	80
	30	92	86	82	76
1.0	1	598	559	515	486
	8	97	92	86	81
	13	76	71	65	62
	22	61	56	53	50
	30	53	51	47	44

In Tables 1 and 2, we present sample sizes computed via ACC (5) setting the length of the required credible interval $\ell = 2$ for models (1)–(2) and (3)–(4), respectively. In Tables S1 and S2 of the Supplementary Material, we present the same scenario for $\ell = 4$. Note that for model (3)–(4) we consider $\lambda_0 = 10(w/\phi)$ to make the prior expected value equal to 10 in order to allow a comparison with model (1)–(2) for which we fixed $\lambda_0 = 10$. The values considered for the parameter ϕ were chosen to cover the range of estimates obtained from real data and reported in Casas-Monroy, Rajakaruna and Bailey (2020) as well as more extreme cases to mimic low and high aggregation of the organisms in the ballast water tank. For the Poisson/gamma model, the values considered for θ_0 were chosen to cover its parameter space in such a way that large and small prior variances were contemplated. For the negative binomial/Pearson Type VI model the values for θ_0 were chosen according to the constraint imposed by the prior variance, $\lambda_0 < \theta_0$.

3.2 Average length criterion (ALC)

An alternative criterion used to determine sample sizes is based on the average length of the posterior credible intervals. The rationale here is to set the minimum Bayesian coverage probability $1 - \rho$ and obtain the minimum sample size n by requiring that the length of the posterior credible region $\ell'(\boldsymbol{x}_n, n) = b(\boldsymbol{x}_n) - a(\boldsymbol{x}_n)$ be such that

$$\int_{\mathcal{X}^n} \ell'(\boldsymbol{x}_n, n) g(\boldsymbol{x}_n)\, d\boldsymbol{x}_n \leq \ell_{\max}, \tag{6}$$

where $\ell_{\max}$ is the maximum admissible length for the posterior credible region. The lower and upper bounds of the desired HPD interval may be obtained via numerical methods and

Table 3 *ALC (6) based minimum sample size (n) computed under the Poisson/gamma model (1)–(2) (and also using Theorem 1) with* $\lambda_0 = 10$, $\rho = 0.05$ *and* $\ell_{\max} = 2$

Aliquot volume (w)	Shape parameter (θ_0)				
	1.0	2.5	5.0	7.5	10.0
0.5	77 (61)	77 (70)	76 (73)	76 (73)	75 (73)
1.0	38 (31)	38 (35)	38 (37)	38 (37)	38 (37)

Table 4 *ALC (6) based minimum sample size (n) computed under the negative binomial/Pearson Type VI model (3)–(4) with* $\lambda_0 = 10(w/\phi)$, $\rho = 0.05$ *and* $\ell_{\max} = 2$

Aliquot volume (w)	ϕ	Shape parameter (θ_0)			
		11	25	50	75
0.5	1	434	437	441	432
	8	122	117	114	110
	13	101	100	95	91
	22	91	90	84	79
	30	87	85	79	75
1.0	1	358	383	403	397
	8	84	84	82	80
	13	66	65	62	61
	22	54	55	51	48
	30	50	48	47	44

the integral by Monte Carlo simulation. An algorithm to obtain the minimum sample size satisfying this criterion is outlined in the Supplementary Material.

Based on the ideas of M'Lan, Joseph and Wolfson (2008), who used a binomial model with a beta prior distribution, we may obtain the sample size using the ALC under the model (1)–(2) with no need for numerical methods via the following result.

Theorem 1. *Consider the Poisson/gamma model (1)–(2) and the average length criterion (6). Based on large sample approximation, the minimum n, to guarantee that the posterior credible interval average length is smaller than* $\ell_{\max}$, *is the smallest integer such that*

$$n \geq \frac{\theta_0}{w\lambda_0}\left\{\left[\frac{\lambda_0}{\theta_0}\frac{2z_{\rho/2}}{\ell_{\max}}\frac{\Gamma(\theta_0+1/2)}{\Gamma(\theta_0)}\right]^2 - 1\right\},$$

where $z_{\rho/2}$ *is the quantile of order* $1-\rho/2$ *of the standard normal distribution.*

The proof of Theorem 1 is presented in the Supplementary Material. In Tables 3 and 4 we present sample sizes computed using ALC (6) setting $\ell_{\max} = 2$ for models (1)–(2) and (3)–(4); in Table 3 (and S3 of the Supplementary Material) we present corresponding sample sizes (within parentheses) computed using Theorem 1. In Tables S3 and S4 of the Supplementary Material we present the same scenario for $\ell_{\max} = 4$.

The results displayed in both subsections of Section 3 are commented in Section 5.

Table 5 *ACC based Bayesian coverage probability of HPD intervals estimated via simulation for some scenarios under the Poisson/gamma model (1)–(2) using sample sizes displayed in Table 1 for $\ell = 2$*

Aliquot volume (w)	θ_0	n	Probability quantile used to fix λ				
			1/6	2/6	3/6	4/6	5/6
0.5	1.0	77	1.00	1.00	0.98	0.94	0.85
	2.5	77	0.99	0.99	0.97	0.95	0.88
	5.0	76	0.99	0.98	0.95	0.94	0.92
	7.5	76	0.98	0.97	0.96	0.94	0.91
	10.0	75	0.97	0.98	0.95	0.95	0.93
1.0	1.0	39	1.00	1.00	0.98	0.95	0.86
	2.5	39	0.99	0.99	0.97	0.93	0.88
	5.0	38	0.99	0.98	0.97	0.94	0.90
	7.5	38	0.98	0.98	0.95	0.94	0.90
	10.0	38	0.98	0.96	0.95	0.96	0.92

4 Simulation study

We conduct a simulation study to verify whether the credible intervals obtained with the sample sizes proposed in Section 3 satisfy the precision criteria.

For each (prior distribution) scenario and sample size obtained via the ACC (5) displayed in Table 1 (and S1 of the Supplementary Material) we drew 1000 samples from a Poisson/gamma model (1)–(2) with values of λ fixed at the quantiles of order 1/6, 2/6, 3/6. 4/6 and 5/6 of the corresponding prior distribution. Then, for each sample we obtained the lower $[a(\boldsymbol{x}_n)]$ and upper $[b(\boldsymbol{x}_n)]$ limits of the HPD credible interval for the mean organism concentration in a ballast water tank with pre-specified average coverage probability $(1 - \rho = 0.95)$ and computed the proportion of intervals containing the fixed value of λ. The results are displayed in Tables 5 and S5. We expect that the estimates of the HPD Bayesian coverage probability to be at least 0.95.

Under the same model, but using sample sizes displayed in Tables 3 and S3, obtained via the ALC (6), we conducted a similar simulation study, the results of which are displayed in Tables 6 and S6. In this case, we expect that the estimates of length of the HPD intervals to be at most 2 (or 4).

The same strategy was conducted for data obtained via the negative binomial/Pearson Type VI model (3)–(4) using the sample sizes provided in Tables 2, 4, S2 and S4. The results are provided in Table 7, 8 and in Tables S7–S10 of the Supplementary Material.

For illustrative purposes, we consider two sets of hypothetical data and obtain the corresponding HPD intervals under both proposed models. The first set mimics a case with extreme heterogeneity in the organism concentration and the second, a case with where the organisms are homogeneously distributed. This kind of severe aggregation occurs when individuals of the same species are physically attached to each other forming colonies as indicated in Rajakaruna et al. (2018).

We first determined the sample size required to satisfy the ACC with $\ell = 2$ assuming a Poisson/gamma model with a prior distribution having $\lambda_0 = 10$ and $\theta_0 = 0.01$. This choice corresponds to a large prior variance. Setting $w = 1$ and $\rho = 0.01$, the required sample size is $n_{\mathrm{P}} = 104$. We then generated 104 observations from a negative binomial model with $\lambda = 9$, $\phi = 0.1$ and $w = 1$. The generated counts are displayed in Table 9 where the heterogeneity induced by the negative binomial model is evident.

The sum of the counts is $s_{n_{\mathrm{P}}} = s_{104} = 1173$ so that the corresponding HPD intervals (obtained via the algorithms described described in Sections 1.1.1 and 1.2.1 of the Supplementary Material) are, respectively, (10.3, 12.3) for the Poisson/gamma model setting $\lambda_0 = 10$ and

Table 6 *ALC based length of HPD intervals estimated via simulation for some scenarios under the Poisson/gamma model* (1)–(2) *using sample sizes displayed in Table* 3 *for* $\ell_{\max} = 2$

Aliquot volume (w)	θ_0	n	Probability quantile used to fix λ				
			1/6	2/6	3/6	4/6	5/6
0.5	1.0	77	0.85	1.27	1.66	2.09	2.66
	2.5	77	1.30	1.60	1.86	2.13	2.49
	5.0	76	1.52	1.75	1.93	2.12	2.37
	7.5	76	1.61	1.79	1.94	2.10	2.30
	10.0	75	1.67	1.83	1.96	2.10	2.27
1.0	1.0	38	0.86	1.28	1.67	2.10	2.68
	2.5	38	1.31	1.61	1.87	2.14	2.50
	5.0	38	1.52	1.74	1.93	2.12	2.37
	7.5	38	1.61	1.80	1.95	2.10	2.30
	10.0	38	1.66	1.82	1.95	2.08	2.26

Table 7 *ACC based Bayesian coverage probability of HPD intervals estimated via simulation for some scenarios under the negative binomial/Pearson Type VI model* (3)–(4) *using sample sizes displayed in Table* 2 *for* $w = 0.5$ *and* $\ell = 2$

ϕ	θ_0	n	Probability quantile used to fix λ				
			1/6	2/6	3/6	4/6	5/6
1	11	663	1.000	0.999	0.997	0.985	0.893
	25	558	0.999	0.998	0.986	0.955	0.919
	50	517	0.994	0.988	0.974	0.958	0.913
	75	480	0.992	0.974	0.976	0.940	0.902
8	11	134	0.993	0.985	0.972	0.948	0.917
	25	126	0.977	0.976	0.952	0.942	0.928
	50	118	0.964	0.971	0.974	0.947	0.931
	75	111	0.959	0.978	0.954	0.967	0.924
13	11	113	0.987	0.982	0.974	0.949	0.909
	25	106	0.982	0.977	0.963	0.967	0.921
	50	99	0.978	0.961	0.963	0.953	0.929
	75	93	0.963	0.973	0.966	0.947	0.935
22	11	97	0.990	0.975	0.958	0.948	0.920
	25	92	0.977	0.961	0.969	0.953	0.924
	50	86	0.964	0.959	0.975	0.958	0.927
	75	80	0.965	0.973	0.974	0.967	0.942
30	11	92	0.980	0.968	0.975	0.950	0.900
	25	86	0.972	0.967	0.967	0.949	0.947
	50	82	0.967	0.963	0.965	0.961	0.936
	75	76	0.961	0.964	0.971	0.969	0.943

$\theta_0 = 0.01$, and (8.81, 10.81) for the negative binomial/Pearson Type VI model [for which we set $\phi \approx 0.1213$ (obtained via maximum likelihood), $\lambda_0 = 10(w/\phi) \approx 82.4$ and $\theta_0 = \lambda_0 + 1$, in order to obtain a large variance]. The first interval does not contain the organism concentration ($\lambda = 9$) used to generate the data and suggest (erroneously) non-compliance with the D-2 regulation. The second interval, on the other hand, contains the value $\lambda = 9$ (even with the

Table 8 *ALC based length of HPD intervals estimated via simulation for some scenarios under the negative binomial/Pearson Type VI model* (3)–(4) *using the sample sizes in Table* 4 *for* $w = 0.5$ *and* $\ell_{\max} = 2$

			Probability quantile used to fix λ				
ϕ	θ_0	n	1/6	2/6	3/6	4/6	5/6
1	11	434	0.934	1.232	1.570	2.048	2.931
	25	437	1.233	1.515	1.797	2.168	2.771
	50	441	1.438	1.668	1.897	2.158	2.578
	75	432	1.547	1.752	1.950	2.168	2.497
8	11	122	1.494	1.714	1.916	2.141	2.476
	25	117	1.686	1.841	1.983	2.135	2.339
	50	114	1.775	1.890	1.988	2.088	2.228
	75	110	1.818	1.908	1.983	2.067	2.175
13	11	101	1.582	1.780	1.960	2.160	2.434
	25	100	1.724	1.864	1.982	2.113	2.282
	50	95	1.820	1.917	1.999	2.083	2.199
	75	91	1.849	1.930	1.993	2.063	2.149
22	11	91	1.621	1.798	1.961	2.127	2.365
	25	90	1.740	1.864	1.970	2.079	2.226
	50	84	1.833	1.917	1.991	2.065	2.163
	75	79	1.874	1.938	1.994	2.055	2.125
30	11	87	1.632	1.808	1.959	2.118	2.330
	25	85	1.759	1.874	1.977	2.080	2.211
	50	79	1.850	1.932	1.997	2.066	2.155
	75	75	1.874	1.938	1.990	2.043	2.111

Table 9 *Simulated counts of the example via the negative binomial* (*NB*) *model* ($\phi = 0.1$, $\lambda = 9$ *and* $w = 1$) *and via Poisson* (*P*) *model* ($\lambda = 9$ *and* $w = 1$)

	0	0	2	65	0	0	0	19	0	6	0	1	2	0	10	27	4	0	41	0	53
	0	0	0	0	0	0	0	72	6	0	1	1	55	0	12	1	0	0	0	0	0
NB	4	0	3	0	0	248	20	1	0	5	0	0	16	0	0	7	208	0	94	0	0
	1	0	0	0	0	13	0	0	0	4	0	2	1	1	2	0	0	0	0	6	0
	75	0	50	6	0	0	5	0	8	1	0	0	4	0	0	8	0	2	0	0	
	5	10	10	10	12	10	3	7	10	9	10	9	7	13	7	12	7	7	6	7	7
	7	6	4	7	12	9	13	12	4	9	7	7	9	6	11	6	7	17	12	9	10
P	7	10	8	9	10	9	7	11	5	7	11	9	6	9	9	11	6	12	12	4	7
	3	7	11	7	9	4	9	6	13	3	11	5	9	8	5	7	10	14	9	6	9
	6	13	8	7	6	13	6	13	6	6	5	9	7	4	7	11	4	9	7	6	

sample size obtained under the Poisson/gamma model) and suggests (correctly) compliance with the D-2 regulation.

For the second scenario, we generated 104 observations via a Poisson model with $\lambda = 9$ and $w = 1$. The generated counts are displayed in Table 9 and shows the more homogeneous distribution of the organisms. The sum of the corresponding counts is $s_{n_P} = s_{104} = 859$ so that the associated HPD intervals are (7.29, 9.29) for both the Poisson/gamma model with parameters $\lambda_0 = 10$ and $\theta_0 = 0.01$ and the negative binomial/Pearson Type VI model with parameters $\phi \approx 233$ (obtained via maximum likelihood), $\lambda_0 = 10(w/\phi) \approx 0.043$ and $\theta_0 =$

$\lambda_0 + 0.01$. Both intervals contain the organism concentration $\lambda = 9$ suggesting (correctly) compliance with the D-2 regulation.

These examples suggest that the negative binomial/Pearson Type VI model accommodates both homogeneous and extreme heterogeneous situations while the Poisson/gamma model fails in the latter case. We must recognize, however, that both models behave quite similarly when heterogeneity is not extreme. This raises the question of eliciting prior information on the organism aggregation, but this depends on historical data which is not yet available.

5 Discussion

The results in Tables 1 and S1 obtained under the Poisson/gamma model indicate that the sample size does not decrease much when θ_0 increases, *i.e.*, when the variance of the prior distribution decreases. This may be explained by the homogeneity assumption for the expected concentration which is intrinsic to the adopted model. Unless we consider a precise prior distribution, the sample size required to satisfy the ACC will not change much. This feature is also visible when we compute the sample size under the same model using the ALC (see Tables 3 and S3).

On the other hand, under model (3)–(4) using either the ACC or the ALC with a fixed value for ϕ, the precision of the prior knowledge, controlled by θ_0 here, directly affects the required sample size. This also happens when we consider a fixed θ_0 and vary ϕ, that plays the role of a scale parameter in the prior distribution (see Tables 2, 4, S2 and S4).

Assuming that ϕ is known may be a disadvantage but we can circumvent this problem in a practical manner without considering a prior distribution for this parameter. The first and simpler way is to consider ϕ as small as possible, for example, $\phi = 1$. Since the sample size n decreases as ϕ increases, when we take ϕ as the minimum, we are being conservative, in the sense that the corresponding n is enough or more than enough to achieve the prespecified criteria settings. The second alternative is to consider a naive sequential procedure in which samples are selected one by one (or by lots). Observe that sample sizes obtained under a Poisson/gamma model (n_P) are always smaller than those obtained by a negative binomial/Pearson VI model (n_{NB}), with respective parameters fixed and write $n_{\mathrm{NB}} = n_{\mathrm{P}} + K$, where K is a positive integer. For fixed w, ℓ (or $\ell_{\max}$) and fixed hyperparameters, we may compute the sample size under a Poisson model, proceed with the sample collection obtaining n_{P} organism counts ($\boldsymbol{x}_{n_{\mathrm{P}}}$). Using these n_{P} organism counts we may compute an estimate for ϕ by maximum likelihood or by the method of moments (see Ludwig and Reynolds, 1988, eq. 3.5, for example) and with this estimate we may obtain n_{NB} and consequently K, which is the required number of additional aliquots. Since the prior distributions used in both models are different, we must choose the hyperparameters for the Pearson Type VI distribution which represent "equivalent prior knowledge" to those fixed in the gamma distribution. Given w, λ_0 and the estimate of ϕ, we may choose θ_0 such that the plot of the Pearson Type VI distribution is similar to the plot of the gamma distribution with previous hyperparameters used to obtain n_{P}.

The standard approach, on the other hand, would be to consider a prior distribution for ϕ which implies including at least an additional hyperparameter so that we must deal with another integral in order to obtain the marginal distribution of λ. This introduces further computational effort and is object of future research.

As in Inoue, Berry and Parmigiani (2005), we compare sample sizes obtained under different perspectives. Under the Bayesian approach fixing either ℓ or $\ell_{\max}$ (Tables 2, 4, S2 and S4), the sample sizes are, in general, smaller than those computed under a frequentist approach with ϵ_a (maximum absolute error estimation) equal to 1 or 2 (see Tables 2 and 3 in Costa, Lopes and Singer, 2016). This may be justified by the additional information provided

by the prior distribution relatively to that considered in the frequentist approach, where only lower and upper bounds for the parameter of interest are given.

For the ALC we also present a result (Theorem 1) which allows the computation of sample sizes under model (1)–(2) without the need for numerical and/or simulation methods. The corresponding sample sizes are consistently smaller than those obtained via Monte Carlo replicates, although the differences are not large. Note that since this theorem is based on large sample approximations, we expect a difference between the corresponding sample sizes and those obtained directly from the proposed criterion.

The simulation results (Tables 5–7 and Tables S5–S10 of the Supplementary Material) show similar results to those obtained under the simulation study presented in Costa, Lopes and Singer (2016). For smaller values of λ, the coverage criterion is attained well above the limit but the results are reversed for the larger values and the minimum fixed coverage is not attained. A similar conclusion holds when using the ALC. We also note that for values of λ smaller or equal to the median, the estimated coverage probability is larger than the proposed one. This is expected, but may not happen for values of λ greater than the median, mainly for the quantiles of order $5/6$ or higher, *i.e.*, in some cases the posterior interval does not contain λ, and this happens with estimated coverage probability smaller than the specified one. To justify this, note that for the Poisson/gamma model, given λ, S_n follows a sampling Poisson distribution with mean and variance $nw\lambda$. When the value of λ (*i.e.*, the prior quantile) increases, the variance of S_n also increases and consequently the variability of the posterior expected value increases, generating more HPD intervals that do not contain the true value of λ. The same happens for the negative binomial/Pearson Type VI model, where the sampling variance of S_n is $nw\lambda + n(w\lambda)^2/\phi$. When λ increases we observe the same behavior as in the Poisson/gamma model, but here, additionally, when ϕ increases, the sampling variance of S_n decreases. This may explain the increase (decrease) of the coverage probability (length) of HPD intervals when λ is fixed at the the $5/6$ quantile of the prior distribution. Also, note that ϕ is scale and shape parameter of the posterior distribution and this may explain the reversed behavior for the other quantiles (see Tables 7, 8, S7 and S8). This suggests that in practice, if the goal is a minimum coverage with probability $1 - \rho$, we should consider a sample size n corresponding to a minimum coverage probability greater than $1 - \rho$ in order to prevent or minimize this problem.

Although, for simplicity, we mention only one of the requirements of the D-2 standard, the proposed procedure is also valid for the requirement that deballasted water should contain less than 10 viable organisms per m^3, sized ≥ 50 μm in minimum dimension, provided the aliquot volume w is changed accordingly.

Practical issues related to the actual collection of the ballast water aliquots have been addressed by many authors (Carney et al., 2013, First et al., 2013, Gollasch and David, 2017). In a recent paper, Casas-Monroy, Rajakaruna and Bailey (2020) compares different methods for obtaining the sample, concluding that among three available competitors the in-line method may be the best one. Given that the BWM Convention requires ships to install a sampling port after their ballast water treatment system, the pipes within the required machinery may be used to collect aliquots along the entire deballasting process. The aliquots (with, say, 10 mL) are then integrated into a single volume from which the organisms are counted and the corresponding credible interval is computed.

Acknowledgments

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 3304126/2015-2) 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.

Supplementary Material

Supplement to "Sample size for estimating organism concentration in ballast water: A Bayesian approach" (DOI: 10.1214/20-BJPS470SUPP; .pdf). Supplementary information.

References

Adcock, C. J. (1987). A Bayesian approach to calculating sample sizes for multinomial sampling. *Journal of the Royal Statistical Society Series D The Statistician* **36**, 155–159.

Adcock, C. J. (1988). A Bayesian approach to calculating sample sizes. *Journal of the Royal Statistical Society Series D The Statistician* **37**, 433–439.

Bailey, S. A. (2015). An overview of thirty years of research on ballast water as a vector for aquatic invasive species to freshwater and marine environments. *Aquatic Ecosystem Health & Management* **18**, 261–268.

Basurko, O. C. and Mesbahi, E. (2011). Statistical representativeness of ballast water sampling. *Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment*, 183–190.

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*. https://doi.org/10.1007/s10811-019-02034-x

Costa, E. G. (2017). Tamanho amostral para estimar a concentração de organismos em água de lastro: uma abordagem bayesiana. Ph.D. thesis, Departamento de Estatística, Universidade de São Paulo, São. Paulo. In Portuguese.

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). Supplement to "Sample size for estimating organism concentration in ballast water: A Bayesian approach." https://doi.org/10.1214/20-BJPS470SUPP

De Santis, F. (2007). Using historical data for Bayesian sample size determination. *Journal of the Royal Statistical Society Series A Statistics in Society* **170**, 95–113. MR2339100 https://doi.org/10.1111/j.1467-985X.2006.00438.x

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 & 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.

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

Inoue, L. Y., Berry, D. A. and Parmigiani, G. (2005). Relationship between Bayesian and frequentist sample size determination. *American Statistician* **59**, 79–87. MR2113200 https://doi.org/10.1198/000313005X21069

Jang, P. G., Kim, J. H. and Choi, K. H. (2019). Testing of organisms $\geq 10\ \mu m$ and $\leq 50\ \mu m$ for ballast water management systems. *Advances in Oceanography & Marine Biology* **1**.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994a). *Continuous Univariate Distributions, Vol. 1*, 2nd ed. New York: John Wiley & Sons. MR1299979

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994b). *Continuous Univariate Distributions, Vol. 2*, 2nd ed. New York: John Wiley & Sons. MR1326603

Joseph, L. and Bélisle, P. (1997). Bayesian sample size determination for normal means and differences between normal means. *Journal of the Royal Statistical Society Series D The Statistician* **46**, 209–226.

Joseph, L., Berger, R. D. and Bélisle, P. (1997). Bayesian and mixed Bayesian/likelihood criteria for sample size determination. *Statistics in Medicine* **16**, 769–781.

Joseph, L. and Wolfson, D. B. (1997). Interval-based versus decision theoretic criteria for the choice of sample size. *Journal of the Royal Statistical Society Series D The Statistician* **46**, 145–149.

Joseph, L., Wolfson, D. B. and Berger, R. D. (1995). Sample size calculations for binomial proportions via highest posterior density intervals. *Journal of the Royal Statistical Society Series D The Statistician* **44**, 143–154.

Ludwig, J. A. and Reynolds, J. F. (1988). *Statistical Ecology: A Primer on Methods and Computing, Vol. 1*, 1st ed. New York: John Wiley & Sons.

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

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 & Technology* **45**, 3539–3546.

M'Lan, C. E., Joseph, L. and Wolfson, D. B. (2006). Bayesian sample size determination for case-control studies. *Journal of the American Statistical Association* **101**, 760–772. MR2256186 https://doi.org/10.1198/016214505000001023

M'Lan, C. E., Joseph, L. and Wolfson, D. B. (2008). Bayesian sample size determination for binomial proportions. *Bayesian Analysis* **3**, 269–296. MR2407427 https://doi.org/10.1214/08-BA310

Murphy, K. R., Ritz, D. and Hewitt, C. L. (2002). Heterogeneous zooplankton distribution in a ship's ballast tanks. *Journal of Plankton Research* **24**, 729–734.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rahme, E., Joseph, L. and Gyorkos, T. W. (2000). Bayesian sample size determination for estimating binomial parameters from data subject to misclassification. *Journal of the Royal Statistical Society Series C Applied Statistics* **49**, 119–128. MR1817878 https://doi.org/10.1111/1467-9876.00182

Rajakaruna, H., VandenByllaardt, J., Kydd, J. and Bailey, S. (2018). Modeling the distribution of colonial species to improve estimation of plankton concentration in ballast water. *Journal of Sea Research* **133**, 166–176.

Stamey, J. D., Young, D. M. and Bratcher, T. L. (2006). Bayesian sample-size determination for one and two Poisson rate parameters with applications to quality control. *Journal of Applied Statistics* **33**, 583–594. MR2240928 https://doi.org/10.1080/02664760600679643

Wang, F. and Gelfand, A. E. (2002). A simulation-based approach to Bayesian sample size determination for performance under a given model and for separating models. *Statistical Science* **17**, 193–208. MR1925941 https://doi.org/10.1214/ss/1030550861

E. G. Costa
Departamento de Estatística
Centro de Ciências Exatas e da Terra
Universidade Federal do Rio Grande do Norte
Natal, RN 59078-970
Brazil
E-mail: eliardocosta@ccet.ufrn.br

C. D. Paulino
Departamento de Matemática
Instituto Superior Técnico
Universidade de Lisboa
Lisboa 1049-001
Portugal
E-mail: daniel.paulino@tecnico.ulisboa.pt

J. M. Singer
Departamento de Estatística
Instituto de Matemática e Estatística
Universidade de São Paulo
Caixa Postal 66281
São Paulo, SP 05314-970
Brazil
E-mail: jmsinger@ime.usp.br