



## Modeling sampling strategies for determination of zooplankton abundance in ballast water



Marco R. Hernandez<sup>a,\*</sup>, Mattias L. Johansson<sup>a,1</sup>, Yanyu Xiao<sup>b,c</sup>, Mark A. Lewis<sup>b</sup>, Hugh J. MacIsaac<sup>a</sup>

<sup>a</sup> Great Lakes Institute for Environmental Research, University of Windsor, Windsor, Ontario, Canada

<sup>b</sup> Department of Mathematical and Statistical Sciences, Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada

<sup>c</sup> Department of Mathematical Sciences, University of Cincinnati, Cincinnati, OH, United States

### ARTICLE INFO

#### Article history:

Received 15 August 2016

Received in revised form 16 November 2016

Accepted 19 November 2016

Available online 29 November 2016

#### Keywords:

Ballast water

Zooplankton

IMO D-2 standard

Modeling

Probability density function

Sampling effort

### ABSTRACT

Ballast water has been a major source of non-indigenous species introductions. The International Maritime Organization has proposed performance standard that will establish an upper limit for viable organisms in discharged ballast. Here we test different sampling efforts for zooplankton in ballast water on a commercial vessel. We fit different probability density functions to find the most representative and evaluated sampling efforts necessary to achieve error rates ( $\alpha$ ,  $\beta$ ) of  $<0.05$ . Our tests encompassed four seasonal trials and five sample volumes. To estimate error rates, we performed simulations which drew from 1 to 30 replicates of each volume (0.10–3.00 m<sup>3</sup>) for mean densities ranging between 1 and 20 organisms m<sup>-3</sup>. Fieldwork and simulations suggested that  $>0.5$  m<sup>3</sup> samples had the best accuracy and precision, and that the Poisson distribution fit these communities best. This study provides the first field test of a sampling strategy to assess compliance with the future IMO standard for large vessels.

© 2016 Elsevier Ltd. All rights reserved.

### 1. Introduction

Ballast water is one of the world's largest vectors for non-indigenous species (NIS) transfer (Molnar et al., 2008). Efforts to control this vector in the Great Lakes began in 1989 with voluntary mid-ocean ballast water exchange (BWE) for vessels entering with filled ballast-water tanks, which was followed by mandatory regulations in 1993. Regulations were extended to vessels with 'empty' ballast-water tanks in 2006 and 2008 in Canada and the USA, respectively. Ballast water management (BWM) has become a standard procedure worldwide, and is overseen by the International Maritime Organization (IMO). Current IMO best management practises request vessels with full ballast tanks conduct exchange on the open ocean to ensure that 95% of the ballast volume has been exchanged, to achieve an in-tank salinity of at least 30‰ (IMO, 2008a). While this procedure is effective in preventing the movement of NIS between freshwater ports that are connected by transoceanic routes (Bailey et al., 2011), it is less effective when both origin and destination ports are marine (Wonham et al., 2001). In 2004 the IMO proposed new performance standards (IMO D-2) (IMO, 2004). This agreement sets numerical limits on the density of two plankton size groups ( $<10$  viable organisms m<sup>-3</sup> for minimum

dimension  $>50$   $\mu\text{m}$  and  $<10$  viable cells mL<sup>-1</sup> for organisms between 10 and 50  $\mu\text{m}$ ) as well as for three bacteria indicators (IMO, 2004). The IMO D-2 convention has yet to be ratified and implemented (IMO, 2004).

Many companies and research groups are testing technology devices and processes to ensure compliance with IMO D-2 standards. Initial steps for approval include testing of devices by an independent third party at verification facilities designed to provide bench-scale estimations, usually referred to as land-based testing. Verification centers also must replicate treatment trials as part of the bench-scale evaluation. Sampling strategies and sampling effort are intended to be easily replicable (IMO, 2008b). Model ballast tanks must be  $\geq 200$  m<sup>3</sup>. For shipboard sampling, control and treated samples need to be collected in triplicate, that uptake and final densities be determined for control tanks, and that viable organism density be assessed before discharge of treated ballast water (IMO, 2008c). However, current guidelines provide no guidance on sample volumes or how they are collected.

Current technology devices have been tested primarily using land-based tests, though a subset has also used shipboard testing (Gollasch and David, 2010). However, no clear method exists for sampling on-board vessels, particularly for sampling directly from ballast tanks. Thus, an imbalance exists in the prescribed sampling process for land-based versus shipboard testing. Onboard sampling poses a major challenge as the IMO D-2 standard requires very low densities of zooplankton, and estimating live density of organisms requires large sample volumes, even under the best case (and unrealistic) scenario that

\* Corresponding author at: GLIER, University of Windsor, Windsor, Ontario N9B 3P4, Canada.

E-mail address: [hernand2@uwindsor.ca](mailto:hernand2@uwindsor.ca) (M.R. Hernandez).

<sup>1</sup> Lead authors.

organisms are randomly distributed (Lee et al., 2010; Miller et al., 2011; Frazier et al., 2013). Moreover, random dispersion of zooplankton in ballast tanks cannot be assumed, as organisms may aggregate and thus may exhibit a patchy distribution (Murphy et al., 2002; First et al., 2013).

Given that access to tanks is often limited, one important question researchers seek to answer is the relationship between sampling method and sample representativeness (Gollasch and David, 2011). Zooplankton sampling in ballast tanks may be done using plankton nets via hatches (Briski et al., 2013; Simard et al., 2011) or, less commonly, by pumping a known volume from the tank into a plankton net (McCollin et al., 2008; Veldhuis et al., 2006; Gollasch and David, 2010). Sampling a ballast tank is complicated as access is limited while in port and very difficult while en route (Wright and Mackey, 2006). Samples must be representative of the entire population, easy to replicate, and unbiased. Another consideration is inherent stochasticity associated with low population densities, with concerns regarding both accuracy and precision (Lemieux et al., 2008). In addition, the sampling strategy must allow inferences to be made regarding densities of viable zooplankton in treated water.

Another important element is to determine the minimum water volume adequate for representative sampling (Gollasch and David, 2011). Several studies have addressed the effects of low organism density and sample volume on estimating the true density of zooplankton, using both Poisson and negative binomial distributions (Lee et al., 2010; Miller et al., 2011; Frazier et al., 2013; Costa et al., 2015). The validity of this theoretical approach has not yet been affirmed empirically. The Poisson distribution is suitable under the assumption of a centralized outflow that can be sampled entirely or in equal time intervals (First et al., 2013). A key challenge is access to the entire water column of a tank. Net tows likely introduce bias as only the upper portion of the tank is typically sampled.

In this study, we tested different sampling volumes using three in-tank sampling points to sample the full depth of a ballast tank on a working cargo vessel. Our goal was to identify the sampling efforts that will provide accurate density estimations of zooplankton at the very low abundances that the IMO D-2 standard requires for compliance. We also designed a simple model to contrast common distributions that have been examined theoretically to provide a sample volume that managers can utilize to verify compliance with the IMO D-2 standard.

## 2. Methods

Ballast samples were collected during voyages by the Federal Venture, between 2012 and 2013 [see Paolucci et al. (2015)]. The vessel transited from three ports (Saguenay, Trois Rivières, and Bécancour) in Quebec, Canada to two ports (Vila do Conde and Sao Luis) in Brazil. A single trial was conducted during each voyage where samples were taken and analyzed. Samples were collected from the largest ballast tank (Tank 2) on the starboard side, with 25 mm diameter inlet pipes (Alfagomma 266GL Water S&D PVC Standard Duty) installed at three depths (4.5, 14.5 and 16.0 m below top deck level) to account for vertical variation in organism distribution (Fig. 1). We selected those depths based on the geometry of the tank: 4.5 m is the middle section of the attached wing tank, 14.5 m is the highest open space in the double-bottom tank, and 16.0 m is just above the baffle line in the deepest portion of the tank. Each inlet pipe contributed one third of the total sample volume. To assess sampling effort, triplicate samples totalling 0.10, 0.25, 0.50, 1.00 or 3.00 m<sup>3</sup> were collected. Samples were collected two days after ballast-water exchange was performed in the North Atlantic region using a pneumatic, self-priming diaphragm pump. Ballast water was transferred from the tank to the forepeak of the vessel where it was filtered through a 35 µm plankton net. Water volume sampled was measured with a Seametrics flowmeter (WMP-Series Plastic-Bodied Magmeter). In-line valves were used to keep water flow rate

## Midship section

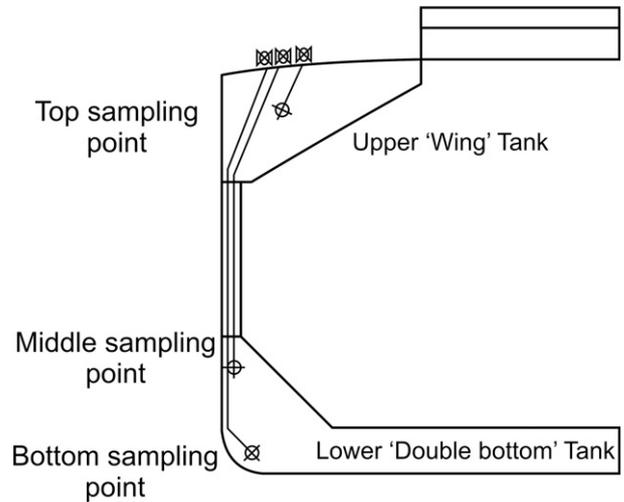


Fig. 1. Location of sampling ports inside the ballast tank.

to 40 L min<sup>-1</sup> in order to avoid mortality due to strong currents. Samples were then fixed in 95% ethanol for microscope counting. We assumed that all intact individuals encountered when processing under the microscope were alive at the time of capture. Each sample was counted entirely to assess population density. The order in which sample volumes were collected was randomized using a random number generator in Excel (Microsoft Inc.).

We conducted basic descriptive statistics (mean and standard deviation) for our four trials. Variance was grouped for fall and spring as those samples were not statistically different and mean densities were similar. Our first goal was to determine the best volume for sampling. Since the true density of organisms in the ballast tank was not known, we assumed that the mean density of organisms over all sample volumes in each trial was an accurate estimate of true density. Preliminary analysis of variance (ANOVA) revealed that volume sampled had a large impact on the density of organisms in the tank ( $p = 0.0056$ ). We estimated density based on the data points collected from the same volume. We assumed that if we sampled at the same volume repeatedly inside the tank, the density of organisms would follow a given probability distribution function (PDF). We performed the following analysis on each of five PDFs (Poisson, Weibull, Negative binomial, Gamma, and Log-normal) with respect to each volume individually. We estimated the parameters of each PDF by maximum likelihood estimation (MLE). Then, we created random number generators based on the estimated PDFs to sample more data points (i.e. one thousand data points) for the density of organisms for each volume, and calculated the mean square error (MSE) based on our assumption that the true density was the average of density estimates in all trials for each volume (Walther and Moore, 2005).

### 2.1. Modeling PDF for distribution of zooplankton

Our second goal was to determine how altering the spatial distribution of zooplankton would affect the sampling error rate. Specifically, our objective was to identify the number of samples of a particular volume that would be required to confidently state that a vessel was compliant with the IMO D-2 limit of <10 viable organisms m<sup>-3</sup> for zooplankton-sized organisms while keeping the rate of Type I and II errors below 5%. In other words, the cumulative sample number of each individual density (from 1 to 20 organisms m<sup>-3</sup>) required in each scenario was constrained to no more than a 0.05 error rate for both false positives and false negatives.

We modeled sampling from the ballast tank using a three-dimensional array in R (R Core Team, 2016). To simulate sampling from the tank, we defined each cell of the array as 1 L of water and the total volume of the array as approximately equal to the actual capacity of the tank used for our sampling (1,279,400 L in the actual tank, 1,300,000 L in our model  $100 \times 100 \times 130$  cell array). For each of 1000 replicates, we populated each cell in the array by drawing randomly from two commonly used PDFs (Poisson and Gamma) with mean densities from 1 to 20 organisms  $m^{-3}$ . For each PDF, we then sampled between 1 and 30 replicates using sampling points placed at particular heights in the array (to model our field design) but with randomly assigned length and width coordinates. The decision to cut off sampling at 30 replicates was somewhat arbitrary, but reflects the reality that it is impossible to collect and process large numbers of samples within a reasonable time in order to assess compliance. 30 represents a number of replicates somewhat above that which would normally be used in field sampling. In each case, we assessed the rate of false positives and false negatives (i.e. we tallied the number of cases where the true mean density was below 10, and the estimate was above 10, or where the true density was above 10 and the estimate was below 10) for all combinations of sample volume and replicate number and determined the minimum replicate number required to achieve error rates  $<5\%$  for each volume.

For the Poisson distribution, we also tested the effect on error rates of having organisms randomly but evenly distributed in the array (Even scenario) at the target density versus organisms preferring the upper wing tank (Uneven scenario: organisms randomly distributed in the 501,400 L upper section at a much higher density [up to  $\sim 500\times$  higher density] than the 778,000 L lower region while still achieving the same overall density as the even distribution). In addition, we modeled the effect of sampling only from the upper wing tank, as typically occurs in current working vessels. In an ideal Poisson situation with evenly distributed organisms, there should be no difference between sampling a given volume in a single large replicate versus a number of small replicates. However, because our simulations sampled randomly from a distribution, some variance between replicates occurred.

For the Gamma distribution, we simulated three different distribution shapes to test the effect of variance on our ability to accurately estimate the true density with different sample volumes and replicate numbers. In each simulation, we tested three levels of dispersion by setting the rate to 0.5, 1.0, and 2.0 to correspond with wide, medium, and narrow distributions, respectively, and then stepwise-adjusted the shape to achieve the desired mean, from 1 to 20 organisms  $m^{-3}$ .

### 3. Results

Although the vessel traversed essentially the same route from Canada to Brazil during all four trials, the geographic position of ballast-water exchange and subsequent location of sampling varied slightly from one trial to the next. Mean plankton density ranged from 285 to 1170 organisms  $m^{-3}$  (horizontal lines, Fig. 2), with a clear seasonal pattern: trial 1 (July) was highest, trial 3 (November) the lowest, and trials 2 and 4 (September and March) were similar and had intermediate densities (Fig. 2). From our field sampling, it was also evident that dispersion is larger in smaller volumes and that it is generally low at volumes  $>0.50 m^3$  (Fig. 2).

We observed no significant difference fitting the five distribution functions in our MLE of PDFs (Fig. 3), possibly owing to our small empirical dataset (12 data points from each sample volume). We did, however, note that the  $1.00 m^3$  sampling volume exhibited the lowest MSE term relative to other volumes tested (Table 1).

When organisms were evenly Poisson distributed in the ballast tank, simulations exhibited a clear relationship between sample volume, replicate number, and our ability to confidently state whether the ballast tank was compliant or not. As mean density of the sample approached the permissible limit of 10 organisms  $m^{-3}$ , the total volume of samples

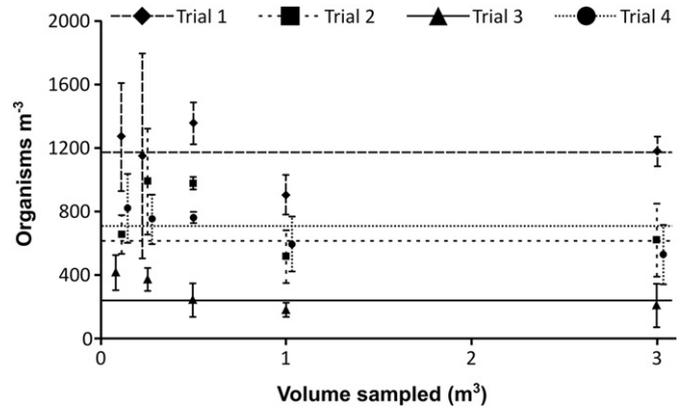


Fig. 2. Densities estimated from all four trials and five sampling efforts. Markers (diamonds – Trial 1, squares – Trial 2, triangles – Trial 3, and circles – Trial 4) indicate mean volume ( $n = 3$ )  $\pm$  one standard deviation.

required to assess compliance increased (Fig. 4, upper panel), and all sampling volumes eventually required  $>30$  samples to assess compliance. Smaller sampling volumes reached our arbitrary limit of 30 replicates earlier than did larger ones, leading to a larger window where sample sizes were insufficient to confidently assess compliance. For example, in our simulations a single  $0.10 m^3$  sample (purple line, Fig. 4 upper panel) could theoretically be sufficient to identify the sample as compliant (i.e.  $<10$  organisms  $m^{-3}$ ) if the true density was below 3 organisms  $m^{-3}$ . However, it would be impossible to confidently assess compliance of a sample with fewer than 30 replicate samples of  $0.10 m^3$  if true density were between 8 and 14 organisms  $m^{-3}$ . Overall, increasing the volume of samples improves our ability to confidently assess compliance as the true density approaches the 10 organisms  $m^{-3}$  limit (dotted vertical line, Fig. 4, upper panel).

In contrast to small volume samples, those of  $3.00 m^3$  required three or fewer replicate samples to confidently determine compliance when the true density was below 8 organisms  $m^{-3}$  or above 12 organisms  $m^{-3}$  (red line, Fig. 4 upper panel), and compliance could be assessed with 11–12 replicates if true density was very close to the maximum permissible limit (i.e. 9 or 11 organisms  $m^{-3}$ ). Intermediate sample sizes could be used to confidently assess compliance when the true density was  $<7$  or  $>13$  organisms  $m^{-3}$ , but as sample volume declined, the number of replicates required increased (Fig. 4, upper panel). As expected, across the range of densities tested, total sample volume seemed to be the key determinant of our ability to confidently

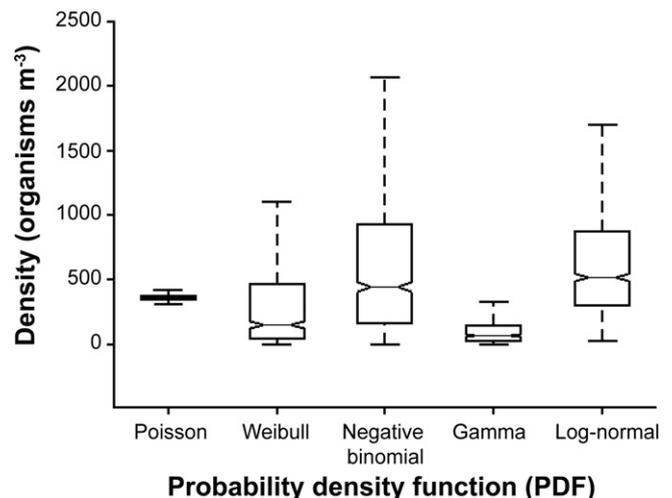


Fig. 3. Box and whisker plot for maximum likelihood of six probability density function testing  $1.00 m^3$  sample volumes.

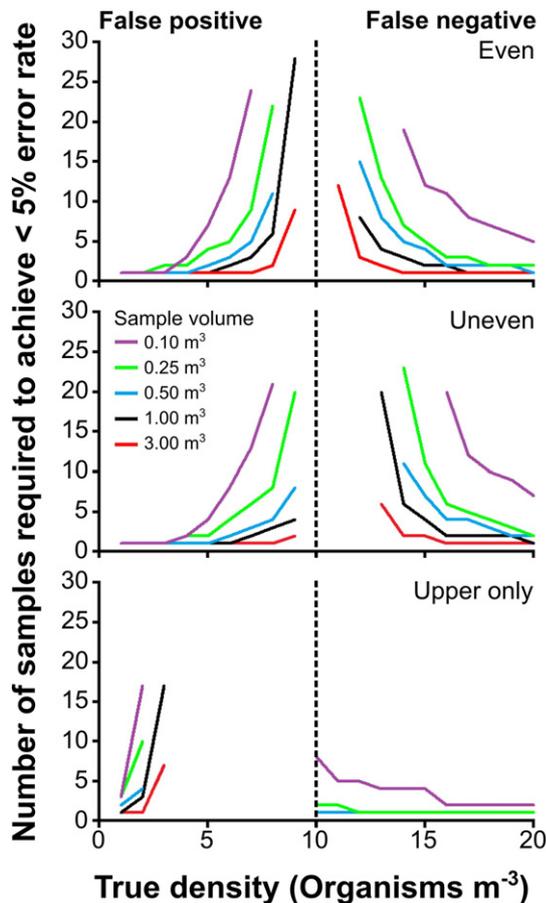
**Table 1**

Mean squared error (MSE \* 10<sup>-5</sup>) computed for each probability density function and each volume (m<sup>3</sup>). Lower values indicate less dispersion between data points and the distribution curve.

Volume (m <sup>3</sup> )	Poisson	Weibull	Negative binomial	Gamma	Log-normal
0.10	1.2981	2.5946	2.5350	2.5364	2.7047
0.25	2.0119	3.9496	4.0674	4.0826	4.7422
0.50	1.6707	3.2963	4.0197	4.1046	6.3578
1.00	0.7853	1.5300	1.7222	1.7800	2.3707
3.00	1.4096	2.8947	3.2303	3.2271	5.5991

assess compliance when organisms were evenly Poisson distributed. For example, at a true density of 7 organisms m<sup>-3</sup>, compliance could be assessed with a minimum of 24, 9, 5, 3 or 1 sample(s) with corresponding volumes of 0.10, 0.25, 0.50, 1.00, or 3.00 m<sup>3</sup>, respectively. This reflects the expectation that, for Poisson-distributed populations, sampling a given volume in one large replicate or multiple small replicates should be mathematically equivalent. Here, differences likely reflect variation due to random sampling of our model tanks.

When organisms were unevenly distributed and were sampled from the full depth of the ballast tank (all three sampling ports), we saw a very similar pattern though the window of non-confidence (error rate > 0.05) moved toward false negatives (Fig. 4, lower panel). All



**Fig. 4.** Minimum sample numbers required at a given animal density and sample volume to achieve <5% false positive/false negative rate for Poisson-distributed organisms. False positives are shown to the left of the midline, false negatives to the right. The central gap in each line indicates that the minimum sample number required to achieve <5% false positive/false negative rate exceeds our arbitrary cutoff of 30 replicates at a given volume for those densities of organisms. The upper panel represents a case where organisms are evenly distributed throughout the tank. Middle panel shows the case where organisms favor the upper 1/3 of the tank and sampling is through three sampling ports (as in our field experiment). In the bottom panel, organisms are aggregated in the upper 1/3 of the tank and sampling is restricted to the upper portion of the tank.

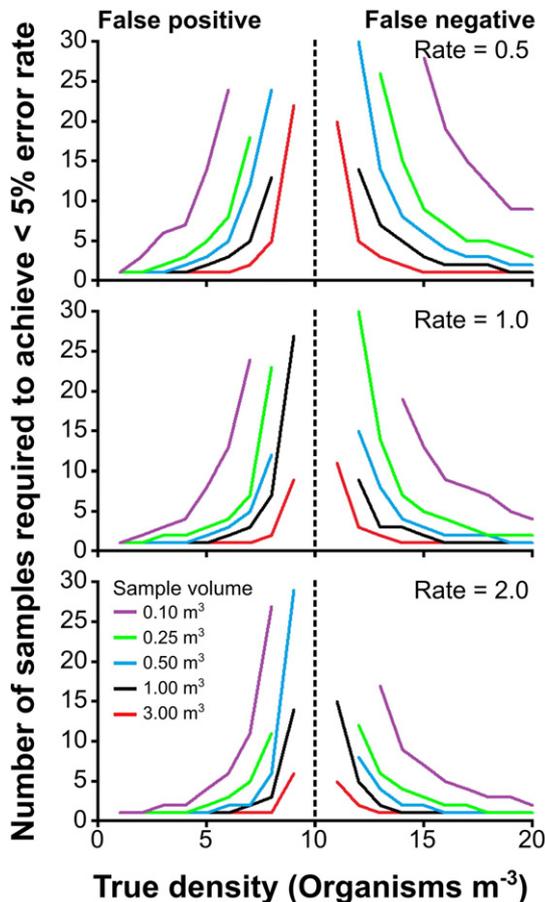
volumes except for 0.10 m<sup>3</sup> could be used to assess compliance when the true density of organisms was ≤9 organisms m<sup>-3</sup> (purple line, Fig. 4, lower panel); however, when the sample volume was low (e.g. 0.25 m<sup>3</sup>), a large (20) number of replicates was required (green line). The number of replicates required to confidently assess compliance dropped progressively from 8 to 4 to 2 replicates at 0.50, 1.00 and 3.00 m<sup>3</sup> (blue, black, red lines, respectively). The lower total volume required for samples of 1.00 m<sup>3</sup> (4 m<sup>3</sup>) versus 3.00 m<sup>3</sup> (6 m<sup>3</sup>) suggests that multiple 1.00 m<sup>3</sup> samples might be the most parsimonious sampling scheme given the time required to process samples under the microscope. The major difference between “uneven” and “even” scenarios is that there were more true densities above the compliance limit where we could not confidently assess compliance in the former scenarios. At a density of 13 organisms m<sup>-3</sup>, we could confidently assess compliance with sample volumes of 1.00 m<sup>3</sup> (black line) and 3.00 m<sup>3</sup> (red line), but both required sampling impractically large volumes of water: 20 m<sup>3</sup> (20 samples) for 1.00 m<sup>3</sup> and 18 m<sup>3</sup> (6 samples) for 3.00 m<sup>3</sup>.

In the uneven Poisson scenario, where organisms were concentrated in the top section of the tank and only that region was sampled (Fig. 4, lower panel), results were quite different. As organism density in the upper portion of the tank was much higher than the overall mean density, it was very easy to overestimate mean density; consequently, large sample volumes from tanks with low overall density (i.e. < 3 organisms m<sup>-3</sup>) were required to achieve an acceptable rate of false positives. In contrast, it took relatively small sample volumes (i.e. 1.00 m<sup>3</sup> total from any sample volume/replicate combination) to avoid false negatives, as few samples estimated densities lower than 10 organisms m<sup>-3</sup>.

Similar to the Poisson results sampled from throughout the tank, all sampling volumes with the Gamma PDF had a window of non-confidence for densities approaching the IMO D-2 standard of 10 organisms m<sup>-3</sup>. Overall, the relationships between different sample sizes were similar to that seen in the Poisson model, above. In all three dispersion scenarios, larger samples had narrower ranges where we failed to confidently assign compliance with reasonable replicate numbers (i.e. <30 replicates; Fig. 5). In the Gamma simulations, the key difference among the three different dispersion scenarios is that as dispersion decreased (rate increased), the range where we could not confidently assign compliance narrowed. This was most apparent in the smallest sample size (0.10 m<sup>3</sup>, Fig. 5, purple line). In the highest dispersion (rate = 0.5) model, we failed to confidently assign compliance for true densities from 7 to 15 organisms m<sup>-3</sup>, while for the intermediate dispersion (rate = 1.0) model the range is 8 to 14 organisms m<sup>-3</sup>, and for the more aggregated organisms (rate = 2.0) model the range is 9 to 12 organisms m<sup>-3</sup>. The other sample volumes tested exhibited a similar, if less pronounced, pattern. The other major difference was that the number of replicates for a given volume decreased with decreasing statistical dispersion. This was very pronounced in the 3.00 m<sup>3</sup> sample size, which maintained the same narrow range of non-confidence throughout all three rate scenarios, but required >20 replicates for confidence when dispersion was highest, 10–12 replicates at intermediate dispersion, and 5–6 replicates when dispersion was low (Fig. 5, red line). This pattern of a narrowing of the non-confidence range with decreasing dispersion, and a decrease in replicates required for confidence, was consistent across all five sample volumes. Consistent with the Poisson model, the largest sample sizes again returned the narrowest range of non-confidence for tractable sample numbers.

#### 4. Discussion

Even at very low densities, sampling volumes of 1.00 and 3.00 m<sup>3</sup> were able to accurately estimate zooplankton density in ballast tanks. However, the improvement in accuracy by adding additional samples was more practical for 1.00 m<sup>3</sup> than for 3.00 m<sup>3</sup> samples. The 1.00 m<sup>3</sup> samples had the lowest MSE scores in five out of six PDFs tested (all



**Fig. 5.** Minimum sample numbers required at a given animal density and sample volume to achieve <5% false positive/false negative rate for Gamma-distributed organisms. False positives are shown to the left of the midline, false negatives to the right. Panels represent high-dispersion (top, rate = 0.5), moderate-dispersion (middle, rate = 1), and low-dispersion (bottom, rate = 2) scenarios. Other figure details are as for Fig. 4.

except Log-normal), and were, therefore, the most accurate of all volumes tested (Table 1 and Fig. 3).

Sampling across the water column addresses problems inherent in sampling species with patchy distributions, and is required for testing IMO D-2 compliance (IMO, 2008b; Murphy et al., 2002). Zooplankton tend to aggregate in natural waters (First et al., 2013) and likely do so in ballast tanks as well. Our multiport sampling design allowed us to sample the entire water column, including the double-bottom portion, which is usually inaccessible. Thus, multiple sampling ports provide more accurate estimates of organism density than single ports or if researchers use deck-based plankton nets. Although we used an equal number of ports as Murphy et al. (2002), our design allowed us to collect water from the lower portion of the tank, something that their system was unable to carry out. This portion is also inaccessible to open hatch tow sampling. Our design also made it possible to take as many replicate samples as desired within a short period of time without affecting vessel operations.

The Poisson distribution had the lowest MSE scores in all volumes (Table 1). The results we obtained were similar for Gamma distribution in deriving the likelihood of over dispersion due to clumping. The Poisson distribution is commonly used for modeling zooplankton distributions in ballast tanks (Lee et al., 2010; Miller et al., 2011; Frazier et al., 2013; Costa et al., 2015), however, the Gamma distribution also has been used as a Poisson approximation. Gamma distribution estimates abundance distributions (Engen and Lande, 1996) and has been suggested for zooplankton in ballast water (Costa et al., 2015). A need exists to build data sets that allow identification of an appropriate PDF based on empirical data. Our attempt with a rather limited data set proved inconclusive.

True zooplankton densities were not known in our trials, thus we relied on a series of assumptions that justified using the mean of all sampling efforts per trial. Under these assumptions, large volume samples had higher precision and lower variability. Trials 1 and 3 also demonstrated that the largest volume (3.00 m<sup>3</sup>) estimated density better than smaller ones. However, in Trials 2 and 4 large volumes underestimated densities. While larger volumes - such as 3.00 m<sup>3</sup> - provided - in general - better estimates, they increased work load prohibitively and thus cannot be recommended (see Frazier et al. (2013)). We observed that 1.00 m<sup>3</sup> samples had the lowest MSE and provided a good estimation with a low rate of false positives when organism abundance was  $\leq 10$  individuals m<sup>-3</sup>, and a low false negative rate when density  $\geq 10$  individuals m<sup>-3</sup> for the two PDFs evaluated here. The error rate can be improved for estimates based on 1.00 m<sup>3</sup> samples by increasing the number of replicates (Figs. 4 and 5). Because our sampling technique was already an integration of three equal volumes, even a single replicate enhanced accuracy of the density estimate, and replicates at this volume are manageable.

There exists support for the argument that large volume samples offer better estimations assuming Poisson-based models (e.g. see Lee et al., 2010; Miller et al., 2011). However when the dispersion of organisms in the tank is unknown, there is a possibility to overestimate densities and wrongly conclude that vessels are not in compliance with the IMO D-2 standard (see Fig. 4). In our 'uneven' Poisson simulations, altering how animals are distributed in the tank modified not only the proportion of false positives and negatives, but the capability to accurately assess organism densities at all tested volumes. We agree with the aforementioned authors that larger volumes (e.g. 7.00 m<sup>3</sup>) provide a better estimator of density, though these volumes are impractical for organism enumeration at anything other than, and possibly including, a land-based testing facility. Our three sampling port design provides better opportunities to accurately quantify plankton present at low density. Theoretical minimum sampling volumes under our design slightly differed from those estimated by Frazier et al. (2013). We found that it would be theoretically possible to assess compliance with a single 0.1 m<sup>3</sup> sample, if true organism density was  $< 3$  individuals m<sup>-3</sup>, whereas Frazier et al. (2013) assert that a minimum of 0.4m<sup>3</sup> would need to be sampled to assess compliance. We argue that the differences between our findings reflect the different mathematical approaches used, rather than any significant disagreement in sampling recommendations. Differences may also stem from the composite nature of our samples, where every sample consisted of three 1/3 samples, taken from different parts of the ballast tank.

Our descriptive statistics highlighted that dispersion was larger on small sample volumes and decreased as volume increased (Fig. 2). Despite the non-significant difference among sampling volumes, we observed that sampling volumes below 0.50 m<sup>3</sup> are much more variable and thus less reliable (Fig. 2). Our comparison of MSE scores for all trials and volumes demonstrated that 1.00 m<sup>3</sup> had the smallest MSE and thus the best accuracy.

The two PDFs that we used to simulate sampling allow us to infer that when zooplankton populations are present at low densities, both 1.00 and 3.00 m<sup>3</sup> sample volumes provide good estimates of density with acceptable error rates ( $< 0.05$ ) versus smaller volumes.

Our study is limited by the number of trials and replicates within each sample volume, however it presents realistic working conditions and constraints likely to be encountered on ocean-going vessels. Validation procedures for IMO D-2 standard are in development. At present there exist no clear guidelines on sample volumes or sample number. We suggest 1.00 m<sup>3</sup> as a starting point and encourage collection of additional empirical data and assessment of sampling strategies.

Empirical data highlighted that integrative samples added precision to density estimations by reducing variance, and that large but practical volumes - such as 1.00 m<sup>3</sup> - benefit from it. MSE scores for 1.00 m<sup>3</sup> were lowest regardless of which PDF was used to fit our data, suggesting that this volume most accurately estimated true density. Finally, our

simulations revealed that increasing the size and number of samples improves confidence in compliance assessments, with the best tradeoff between accuracy, precision, and work load seemingly optimized with 1.00 m<sup>3</sup> samples.

### Author contributions

MRH, MLJ and HJM designed the study and wrote the paper, MLJ, YX and MAL conducted simulations, and all authors edited the manuscript.

### Acknowledgements

We are grateful to Fednav Inc. and the captains and crews of the Federal Venture for assistance in preparing and executing onboard experiments. Lab assistance was provided by Vishal Vara and Lucas Wilson. We acknowledge funding from Fednav Inc. and the NSERC CAISN network, and NSERC Discovery grants and Canada Research Chairs to M.A.L. and H.J.M. M.R.H. was supported by a scholarship from CONACYT Mexico. We gratefully acknowledge helpful comments of the reviewers.

### References

- Bailey, S.A., Deneau, M.G., Jean, L., Wiley, C.J., Leung, B., MacIsaac, H.J., 2011. Evaluating efficacy of an environmental policy to prevent biological invasions. *Environ. Sci. Technol.* 45, 2554–2561.
- Briski, E., Bailey, S.A., Casas-Monroy, O., DiBacco, C., Kaczmarska, I., Lawrence, Nasmith, L.E., 2013. Taxon- and vector-specific variation in species richness and abundance during the transport stage of biological invasions. *Limnol. Oceanogr.* 58, 1361–1372.
- Costa, E.G., Lopes, R.M., Singer, J.M., 2015. Implications of heterogeneous distributions of organisms on ballast water sampling. *Mar. Pollut. Bull.* 91, 280–287.
- Engen, S., Lande, R., 1996. Population dynamic models generating species abundance distributions of the gamma type. *J. Theor. Biol.* 178, 325–331.
- 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, 4442–4448.
- Frazier, M., Miller, A.W., Lee, H., Reusser, D.A., 2013. Counting at low concentrations: the statistical challenges of verifying ballast water discharge standards. *Ecol. Appl.* 23, 339–351.
- Gollasch, S., David, M., 2010. Testing sample representativeness of a ballast water discharge and developing methods for indicative analysis. Report No. 4. European Maritime Safety Association (EMSA), Hamburg.
- Gollasch, S., David, M., 2011. Sampling Methodologies and Approaches for Ballast Water Management Compliance Monitoring. vol. 23. PROMET-Traffic & Transportation, pp. 397–405.
- IMO, 2004. International Convention for the Control and Management of Ships' Ballast Water and Sediments London. International Maritime Organization, United Kingdom. [http://www.imo.org/en/About/Conventions/ListOfConventions/Pages/International-Convention-for-the-Control-and-Management-of-Ships'-Ballast-Water-and-Sediments-\(BWM\).aspx](http://www.imo.org/en/About/Conventions/ListOfConventions/Pages/International-Convention-for-the-Control-and-Management-of-Ships'-Ballast-Water-and-Sediments-(BWM).aspx) (online, accessed Dec 7, 2015).
- IMO, 2008a. Guidelines for Ballast Water Exchange (G6). International Maritime Organization, London, United Kingdom. <http://globallast.imo.org/wp-content/uploads/2015/01/G6-GUIDELINES-FOR-BALLAST-WATER-EXCHANGE.pdf> (online, accessed Dec 7, 2015).
- IMO, 2008b. Guidelines for Ballast Water Sampling (G2). International Maritime Organization, London, United Kingdom. <http://globallast.imo.org/wp-content/uploads/2015/01/G2-GUIDELINES-FOR-BALLAST-WATER-SAMPLING.pdf> (online, accessed Dec 7, 2015).
- IMO, 2008c. Guidelines for Approval of Ballast Water Management Systems (G8). International Maritime Organization, London, United Kingdom. <http://globallast.imo.org/wp-content/uploads/2015/01/G8-GUIDELINES-FOR-APPROVAL-OF-BALLAST-WATER-MANAGEMENT-SYSTEMS.pdf> (online, accessed Dec 7, 2015).
- Lee II, H., Reusser, D.A., Frazier, M., Ruiz, G., 2010. Density Matters: Review of Approaches to Setting Organism-based Ballast Water Discharge Standards. U.S. EPA, Office of Research and Development, National Health and Environmental Effects Research Laboratory, Western Ecology Division. EPA/600/R-10/031, Newport, OR.
- Lemieux, E.J., Robbins, S., Burns, K., Ratcliff, S., Herring, P., 2008. Evaluation of representative sampling for rare populations using microbeads. Report No. CG-D-03-08. U.S. Coast Guard Research & Development. Naval Research Laboratory, Key West, FL.
- McCollin, T., Shanks, A.M., Dunn, J., 2008. Changes in zooplankton abundance and diversity after ballast water exchange in regional seas. *Mar. Pollut. Bull.* 56, 834–844.
- 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 so easy. *Environ. Sci. Technol.* 45, 3539–3546.
- Molnar, J.L., Gamboa, R.L., Revenga, C., Spalding, M.D., 2008. Assessing the global threat of invasive species to marine biodiversity. *Front. Ecol. Environ.* 6, 485–492.
- Murphy, K.R., Ritz, D., Hewitt, C.L., 2002. Heterogeneous zooplankton distribution in a ship's ballast tanks. *J. Plankton Res.* 24, 729–734.
- Paolucci, E.M., Hernandez, M.R., Potapov, A., Lewis, M.A., MacIsaac, H.J., 2015. Hybrid system increases efficiency of ballast water treatment. *J. Appl. Ecol.* 52, 348–357.
- R. Core Team, 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
- Simard, N., Plourde, S., Gilbert, M., Gollasch, S., 2011. Net efficacy of open ocean ballast water exchange on plankton communities. *J. Plankton Res.* 33, 1378–1395.
- Veldhuis, M.J., Fuhr, F., Boon, J.P., Ten Hallers-Tjabbers, C.C., 2006. Treatment of ballast water; how to test a system with a modular concept? *Environ. Technol.* 27, 909–921.
- Walther, B.A., Moore, J.L., 2005. The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. *Ecography* 28, 815–829.
- Wonham, M.J., Walton, W.C., Ruiz, G.M., Frese, A.M., Galil, B.S., 2001. Going to the source: role of the invasion pathway in determining potential invaders. *Mar. Ecol. Prog. Ser.* 215, 1–12.
- Wright, D.A., Mackey, T.P., 2006. Shipboard and dockside trials of ballast water treatment technology. *Nav. Eng. J.* 118, 37–43.