Proceedings of the 20th International Conference on Aquatic Invasive Species

Research Article

A test of the framework designed to evaluate compliance monitoring devices for ballast water discharge

Matthew R. First[1,*], Lisa A. Drake[2], Vanessa Molina[3], Cameron S. Moser[2], Stephanie H. Robbins-Wamsley[3], Scott C. Riley[3], Earle N. Buckley[4], Allegra A. Cangelosi[5], Katharine J. Carney[6], Thomas H. Johengen[7], Heidi Purcell[7], Euan D. Reavie[8], G. Jason Smith[9] and Mario N. Tamburri[10]

[1]*Naval Research Laboratory (NRL), Code 6137, Washington, DC, USA;* [2]*NRL, Code 6137, Key West, FL, USA;* [3]*Excet, Inc., Springfield, VA, USA;* [4]*Buckley Environmental, LLC, Mount Pleasant, SC, USA;* [5]*University of Wisconsin, Lake Superior Research Institute, Superior, WI, USA;* [6]*Smithsonian Environmental Research Center, Edgewater, MD, USA;* [7]*Cooperative Institute for Great Lakes Research, Ann Arbor, MI, USA;* [8]*University of Minnesota Duluth, Duluth, MN, USA;* [9]*Moss Landing Marine Laboratory, Moss Landing, CA, USA;* [10]*University of Maryland Center for Environmental Science, Cambridge, MD, USA*

*Corresponding author

E-mail: matthew.first@nrl.navy.mil

Co-Editors' Note:

This study was contributed in relation to the 20th International Conference on Aquatic Invasive Species held in Fort Lauderdale, Florida, USA, October 22–26, 2017 (http://www.icais.org/html/previous20.html). This conference has provided a venue for the exchange of information on various aspects of aquatic invasive species since its inception in 1990. The conference continues to provide an opportunity for dialog between academia, industry and environmental regulators.

Abstract

With the entry of the Ballast Water Management Convention into force, ballast water discharged from ships must meet standards limiting the concentrations of living organisms. Monitoring devices to confirm compliance with these standards would ideally be portable, easy-to-use instruments capable of rapid and accurate shipboard analysis of ballast water. Following a framework established for the validation of such potential devices, six devices were evaluated in a series of laboratory and field tests at three contrasting coastal locations. Devices were designed to quantify organisms in the ≥ 10 and < 50 μm size class. In all cases, the compliance monitoring devices were compared to the agreed-upon performance standard for quantifying living organisms: microscopy-based, vital fluorophore approach. Specific results from these validations are available elsewhere, although examples are shown to demonstrate the analytical and statistical approaches used for gauging—with data analysis and statistical approaches—each device's performance. Each metric used to evaluate devices (e.g., linearity, precision, and accuracy) was informative. However, linearity between the microscopy-based method and the compliance devices, especially along a large range of organism concentrations, would not be suitable for establishing performance criteria. Concentrations well below or above the limit for this size class (10 living organisms mL^{-1}) would be easily categorized as meeting or exceeding discharge standard and their values do not need to be well constrained and pinpointed. Precision, when measured as the coefficient of variation, was sensitive to the dimensions and scale of the devices' measurements, as certain devices calculated and reported cell concentrations, whereas other devices reported non-dimensional values along a wide dynamic range. Accuracy, defined as the agreement between the compliance device and the standard approach as to whether the sample met or exceeded the discharge standard, was measured by logistic regression analysis. Following this analysis, the likelihoods of detecting exceedances based upon cell concentration were calculated for each field site and cultured test organism. Accuracy was useful in defining the likelihood of correctly identifying an exceedance, and these likelihoods could be calculated for a range of cell concentrations. The concurrent testing of multiple compliance devices minimized the analysis burden as well as the logistical hurdles associated with field testing at multiple—for this study, three—locations. Eventually, the test procedures could be modified to measure variation among different units of the same device or applied to actual measurements of ballast water rather than communities of ambient organisms or cultured microalgae.

Key words: compliance device, ballast water management system, shipboard, protists, epifluorescence microscopy

Introduction

With the aim of reducing cross-ecosystem transport of aquatic organisms, the International Maritime Organization (IMO) set limits on the concentrations of living organisms permitted in discharged ballast water. Ships discharging ballast in international waters or territorial waters of nations signatory to the Ballast Water Management Convention must release fewer than 10 living organisms mL^{-1} within the ≥ 10 and < 50 µm size category and fewer than 10 living organisms m^{-3} for the ≥ 50 µm size category. Additionally, there are limits placed on certain indicator bacteria that, when present, would signal the presence of human pathogens. The United States Coast Guard (USCG) has enacted similar limits for vessels operating in US waters (USCG 2012).

Most ships will comply with these standards by using a Ballast Water Management System (BWMS), a shipboard treatment system that removes or kills organisms using filtration, UV radiation, heat, or a combination of these or other approaches (Tsolaki and Diamadopoulos 2010). These BWMS undergo verification testing, including land-based trials where organism concentrations in treated and control tanks are measured by microscope-based methods (US EPA 2010). A BWMS will demonstrate its efficacy in verification testing, but once installed and in operation, detailed examination of discharge water from operational vessels becomes challenging, and the ship's crew and local compliance officers will rely upon usage and maintenance logs to verify the BWMS operated as expected. However, periodic verification based upon organism concentrations in the discharge may be warranted.

Shipboard compliance-monitoring devices should be portable and easy to use instruments designed for rapid analysis of ballast water. These instruments must report directly on the regulated discharge standard for ballast water (through a surrogate parameter) or the probability of the discharge water exceeding the standard based upon one (or a few) samples. For the USCG, discharge limits are based upon the number of living organisms, and because some BWMS may kill organisms but not remove biomass, shipboard compliance devices must differentiate between living and dead organisms.

Current compliance devices typically target organisms in the ≥ 10 and < 50 µm size class, as sample volumes required are typically < 10 L, whereas the ≥ 50 µm size class may require several m^3 to obtain an accurate estimate of the concentration of a sparse community (Gollasch et al. 2015). Compliance devices have largely—although not exclusively (e.g., van Slooten et al. 2015)—used variable fluorescence fluorometry.

This variable fluorometry-based approach targets phototrophic microalgae, and it determines the relative abundance and physiological status of microalgae in a sample (Casas-Monroy et al. 2016; Bradie et al. 2018). Effective compliance devices will report results that correspond to concentrations of living organisms measured with the standard, microscopy-based techniques required in verification testing of BWMS.

This study follows on efforts to develop a framework for evaluating compliance devices (Drake et al. 2014) and the independent, technology evaluation process used by the Alliance for Coastal Technologies (ACT). Laboratory tests and field trials at multiple, contrasting coastal locations evaluated six compliance devices that used variable fluorescence fluorometry. The results of this effort are publically available (www.act-us.info/evaluations.php), but these particular set of results are not the focus of this report. Rather, this study focuses on the general approach used to evaluate compliance devices for ballast water based upon the comparison of the devices with standard, microscope-based counts of organisms, such as those described in land-based testing of BWMS (US EPA 2010). Simultaneous analysis of samples with cultured and ambient organisms was used to quantify the linearity, accuracy, and precision of the compliance devices relative to manual microscopy. This report describes the key test parameters, test quality procedures, and statistical analyses to provide a template for verification testing of newly emerging or currently existing (but evolving) compliance devices.

Materials and methods

Developing the test protocol

The test protocol was drafted by the test team, but prior to finalizing the document, a Technical Advisory Committee (TAC) conducted a critical review. The five members of the TAC reviewed the draft protocol prior to meeting with the testing team and the representatives of the technology vendors to develop consensus-based, finalized test protocols. Key test elements were: 1. Field-based testing using samples with mixed assemblages of ambient organisms at three contrasting locations, 2. Laboratory-based testing using samples with either of two cultured microalgae, and 3. Simultaneous measurements of all samples using the compliance devices and the standard, microscope-based, vital fluorophore procedure for quantifying organisms (US EPA 2010). These test elements are described in detail below.

Field tests

Evaluations of the instruments were conducted at three locations with contrasting water characteristics

and differing concentrations and compositions of microbiota: the Naval Research Laboratory (NRL) in Key West, Florida (latitude 24.58ºN; longitude: 81.79ºW) represented offshore, high salinity (36 practical salinity units [psu]); the Great Ships Initiative (GSI) land-based test facility (now the Great Waters Research Collaborative [GWRC]) in Superior, Wisconsin (46.71ºN; 92.05ºW) represented the freshwater Great Lakes (0 psu); and the Smithsonian Environmental Research Center (SERC) in Edgewater, Maryland (38.89ºN; 76.54ºW), located on the Chesapeake Bay, represented estuarine waters (9–13 psu). At each location, three independent trials were performed on separate (typically consecutive) days.

Samples with mixed assemblages of ambient organisms were prepared by either diluting or concentrating natural water sampled at a depth of 1 m using a horizontal Van Dorn bottle. The volumes to be diluted or concentrated were based upon an initial count of organisms in the sample water. Dilution was performed by mixing the sample with 0.22-μm filtered sample water. Concentration was performed by screening sample water through a sieve with mesh netting to selectively retain organisms $\geq$ 10 μm. Following these procedures, four samples (each 10–15 L) were generated with different target concentration ranges: 0 living organisms (org.) mL^{-1} (the 0.22-μm filtered water to be used as a control or blank for fluorescence); 5–20 org. mL^{-1}, representing concentrations near the discharge standard (DS); 30–50 org. mL^{-1}, representing concentrations above the DS; and $\geq$ 50 org. mL^{-1}, representing concentrations well above the DS. Each of the four samples was mixed by gentle inversion and rotated a minimum of three times prior to distributing into three replicate subsamples. For each trial, the order of subsampling was randomized and each subsample was coded so that analysts were not aware of the source of the subsample. Aliquots of each subsample were provided for analysis by each compliance device and for the standard microscope-based analysis, as described below.

Laboratory tests

Each compliance device was used to analyze samples with one of two microalgae, both tested individually, at a range of concentrations. *Tetraslemis marina*, with minimum and maximum cell dimensions ranging from 8 to 15 μm, and *Prorocentrum micans*, ranging from 25 to 50 μm, represented organisms near the extremes of the $\geq$ 10 and < 50 μm size class, respectively. Strains were acquired from the National Center for Marine Algae and Microbiota (NCMA, Bigelow Laboratory, East Boothbay, Maine) and were

transferred every 3 to 5 days into nutrient-enriched seawater (Guillard and Ryther 1962) to achieve concentrations sufficient for testing (> 500 org. mL^{-1}). At the start of each of the three laboratory trials, an initial microscope count of both cultures provided an estimate of concentrations of the stock cultures. Stock cultures were then diluted with filter-sterilized seawater to yield 2-L samples with concentrations of 5, 10, 20, 50, and 100 org. mL^{-1} of either *T. marina* or *P. micans*. A filter-sterilized seawater control (i.e., 0 org. mL^{-1}) was also analyzed for each set of organisms. In total, this design generated six unique samples for each of the two organisms (n = 12). The samples were subsampled in a random order following the procedures described for the field trials.

Sample analysis: Microscopy

All participating vendors, the TAC, and the testing team agreed that the instruments would be compared to the standard, microscope-based, vital fluorophore technique for land-based verification testing of BWMS (US EPA 2010). The method was executed following standard operating procedures in place at the three test sites. Briefly, 500 mL of the sample aliquot was concentrated on a monofilament mesh sieve. Organisms concentrated on the mesh—rated to retain particles > 7 μm—were rinsed with filtered seawater to yield a 50-mL sample contained in a centrifuge tube. The sample was well mixed by inverting the tube five times prior to transferring 0.985 mL into a microcentrifuge tube. The sample was mixed with 5 μL of 1 mM fluorescein diacetate (FDA) and 10 μL of 250 μM 5-chloromethylfluorescein diacetate (CMFDA; Steinberg et al. 2011). The only substantial difference among the standard methods at the test sites was that GSI did not use CMFDA. Internal validation studies at GSI demonstrated that, for that location, analysis using FDA was equivalent to the combination of FDA and CMFDA. After incubating in the dark for 10 minutes, the entire volume was transferred onto a 1-mL, gridded Sedgewick-Rafter chamber. A concentrated suspension of 10-μm and 50-μm fluorescent microbeads was added to samples from the field tests as size reference.

Using an epifluorescence microscope with appropriate light filters for the fluorophores and set to 100× magnification, analysts manually scanned entire rows (each 50-μL) following pre-generated, random row assignments. Typically, 7 to 14 of the 20 rows were counted, depending on the time required to scan the rows, as analysis time was limited to within 30 minutes of the addition of the fluorophore labels to avoid background fluorescence obscuring organisms. For very sparse concentrations, the entire chamber

was counted. Organisms fluorescing were considered living. For samples from the field tests, organisms ≥ 10 and < 50 µm—judged by comparing the organisms to microbeads—were tallied. For samples from the laboratory tests, all living microalgae were tallied. Concentrations were calculated based upon the tally of living organisms, the volume scanned in the Sedgewick Rafter slide, which was dependent upon the organism concentrations and ranged from 0.35 to 1 mL, the volume of concentrated sample, and the total sample volume.

Sample analysis: Compliance devices

The six compliance devices tested (identified by letters A through F) were all based upon variable fluorescence fluorometry, an approach to detect the fluorescence yield of chlorophyll a in microalgae (ACT/MERC 2012). Of the six devices, five were capable of discrete sample analysis and were designed to be carried by hand aboard a ship. One flow-through device was engineered to be installed aboard a ship and integrated into the piping system to continuously monitor ballast water diverted from the main ballast line. One of the devices tested in the first round was updated and retested in the second round. Aliquots of 100 to 500 mL from each subsample were distributed to the five compliance devices, and this volume was sufficient for rinsing materials, purging the fluidics system (for the flow-through device), and collecting triplicate readings. The specific approaches for generating a measurement of organism concentrations (based on initial fluorescence yield of chlorophyll a) and physiological status (based on measurements of initial and maximum fluorescence yield) differed among the instruments. Other differences among instruments, including internal electronics, sample interrogation chambers, signal processing routines, and analytical algorithms, were considered proprietary information.

The compliance devices designed for discrete analyses provided (1) a numerical measurement related to the abundance of organisms in the sample: total or live cell concentration, initial fluorescence yield, or other non-dimensional indices for abundance, and (2) a sample disposition, such as risk or likelihood of exceedance. The manufacturers' recommended protocols were followed for sample processing, analysis, and cleaning or rinsing following each reading (if required). Analysts manually recorded the key data from the instruments display, e.g., numerical measurements of abundance and sample disposition (i.e., low or high risk).

Data analysis

A linear regression analysis was used to determine the strength of the relationship between measurements of abundance by each compliance device (whether cell concentration or a non-dimensional variable such as fluorescence intensity) and cell concentrations determined by the standard technique. The regression coefficient (R^2) was used to compare the strength of the linear relationships of the field sites separately (n = 36 for each field site), the cultured organisms in the laboratory trials (n = 18 for each organism), and the combined field (n = 108) and laboratory (n = 36) data sets.

The measure of abundance was also used to determine the precision of the estimates. As each sample was subsampled and analyzed three times, the three measurements were used to calculate the coefficient of variation (CV) for each sample. The coefficient of variation—the standard deviation normalized to the mean of three values—is sensitive to the absolute value of the mean, and small variations among measurements are amplified as the mean approaches zero. To reduce the differences among instruments based upon their measurement scale, only subsamples where the mean value was > 10 (org. mL^{-1} or an arbitrary unit) were included in the analysis.

Accuracy was judged by the agreement between the compliance devices and the microscope analysis, in particular, whether the two approaches agreed or disagreed if a sample met or exceeded the discharge standard of ≤ 10 org. mL^{-1}. Compliance devices designed for discrete analyses provided either cell concentration or a measurement of likelihood of meeting or exceeding the discharge standard. These outcomes were converted into a binary variable, and samples were grouped as low risk of exceedance ("passing") or high risk of exceedance ("failing"). Outcomes were compared—using logistic regression—to the continuous range of microscope-based cell concentrations, centered on 10 org. mL^{-1}, the threshold for exceeding the discharge limit. Both linear and logistic regression analyses were performed using statistical software (SigmaPlot, V12.5; San Jose, CA). The constant (C) and coefficient (x) of the relationship between the binary variable and cell concentration were used to plot the likelihood of correctly predicting an exceedance (p) for a given population concentration of living organisms (Org.) using the logistical function (Hilbe 2009):

Eq. 1 $$p(Org.) = \left(\frac{1}{1 + e^{-(C + xOrg.)}}\right)$$

This relationship was used to calculate p for potential values ranging from 0 to 500 org. mL^{-1}.

Quality management procedures

All technical activities were conducted by personnel trained in the test procedures and operating within the Quality Management System (QMS) of their institution. The QMS outlined the policies, objectives, procedures, authorities, and accountability needed at the facilities conducting this work and for the testing personnel. Relevant to this evaluation, the key components of the QMS included the establishment of a test protocol, the use of standard operating procedures (SOP) for all critical operations (e.g., sampling, analysis, equipment operation), and a technical audit of the testing. Certain procedures to assure data quality were defined in the specific test protocol. For example, samples were blinded, but the key process of encoding the samples was overseen and verified by an affiliate who was familiar with the test procedures but not participating in sample analysis. Additionally, a technical system audit (TSA) was performed for laboratory trials and most field trials. The TSA verified that the test protocol, the associated SOP, and the QMS were followed while the experiments were underway.

As microscope-based analyses were performed by multiple analysts, for each trial, one subsample was randomly selected for analysis by a second analyst; the subsample was aliquoted, distributed, processed, and analyzed as other samples. The second analyst was not aware of the results of the first. Percent difference between the two analyses ≤ 25% were considered within the typical variation among analysts and among discrete samples, even when drawn from the same container.

Results and discussion

The test protocol and detailed results from the six compliance devices evaluated were posted on the ACT website (www.act-us.info/evaluations; available 10-Aug-2017). These reports include additional details on the test locations and methods. These reports allow end users to review individual performance data and determine which instrument best meets their needs. This report focuses on describing and evaluating the validation approach, and only examples of the test results are shown for demonstration, and therefore, the example data here does not identify the instruments. Example results are drawn from a set of four unique instruments (*A* through *D*), each capable of comparing the sample to the discharge standard.

An example of a relationship between a measurement of abundance, in this case, a non-dimensional variable, and microscope-based cell concentrations includes results from three trials at three locations (Figure 1). Concentrations of living organisms ranged from 0 to ~ 3×10^2 org. mL^{-1}, so values were plotted on log-scale axes, however, an inset plot with linear-scale axes was overlaid so zero values could be displayed. Symbols were color and shape coded to differentiate among the field sites and indicate a second reported value, variable fluorescence, a non-dimensional value indicating the physiological status of microalgae within the sample. Finally, symbol outlines were color coded to indicate whether the sample disposition reported to be low (green outlines) or high (red outlines) risk of exceeding the discharge sample for all three subsamples. In some cases (especially for samples with concentrations near 10 org. mL^{-1}), the results of the three subsamples were not uniform; these cases were marked with orange outlines.

Linear regression analyses were performed for each field site separately (n = 36 for each field site), then all field sites combined (n = 108). Likewise, regression analyses were performed for *T. marina* and *P. micans* samples separately (n = 18), then together (n = 36). As R^2 is non-dimensional and independent of the scale used by the compliance devices—whether cell concentrations or other indices—it was suitable for comparisons among instruments. However, linearity over the large range of cell concentrations observed in these datasets may not be an appropriate metric for evaluating the compliance device. Compliance devices are designed to detect exceedances of the discharge limit; they are not necessarily optimized to display a linear response over several orders of magnitude of cell concentrations. In contrast to an analytical or research instrument, which may be judged by the extent and linearity of its dynamic range (Green 1996), compliance devices require high resolution within a limited range, while values outside that range do not need to be pinpointed. Therefore, R^2 would not necessarily be an appropriate metric for comparisons among compliance devices or establishing minimal requirements.

Precision is an important metric for compliance devices, as multiple readings with similar outcomes provide confidence in the result. Precision, measured as CV among subsample readings, was calculated among subsamples of each sample for all field (n = 36) and laboratory (n = 18) trials. However, CV was calculated only when the mean value was > 10 units (regardless of the measurement unit). Five of six instruments reported values < 10 in their dataset. For one instrument, as few as 12 of 36 analyses from the field trials reported mean values < 10 (data not shown).
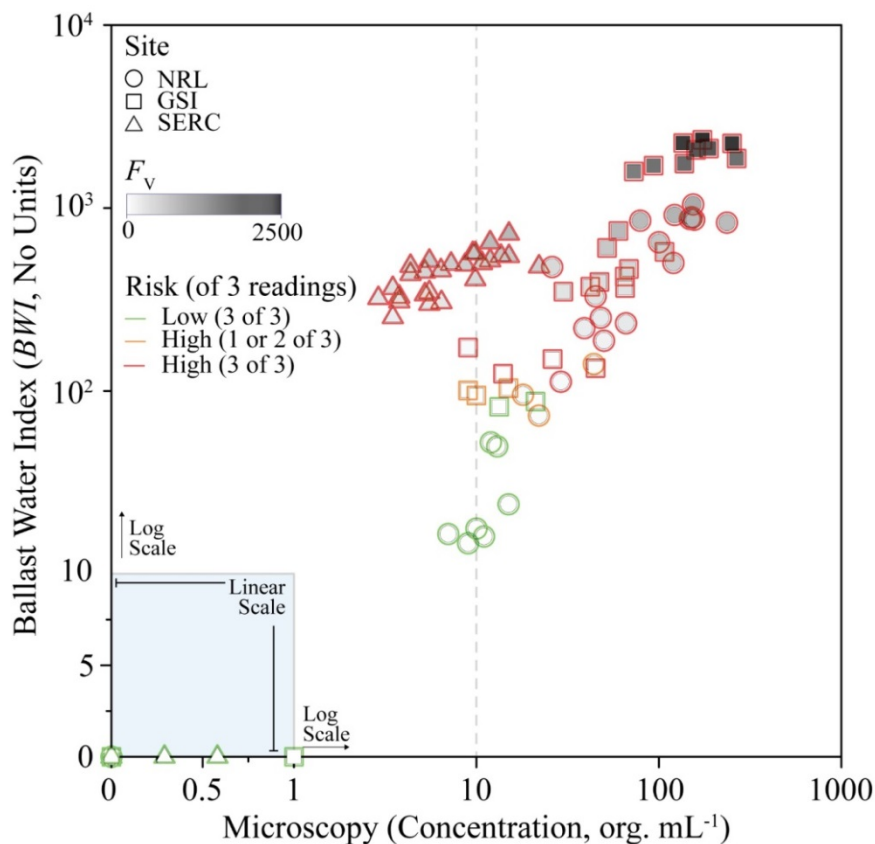
Figure 1. Comparisons of a non-dimension measure of organism concentration and microscope counts of organisms ≥ 10 and < 50 µm at all field sites. The relative concentration is based upon the fluorescence yield intensity of Instrument *C*. See the text for a description of the figure and legend definitions.

Table 1. Example of ranges of CV values for two instruments, both measuring cell concentrations and both with ≥ 20 measurements (n) of CV.

Instrument	Mean CV	Median CV	Minimum CV	Maximum CV	n
A	24%	22%	6%	52%	20
B	21%	18%	9%	42%	22

Example ranges of CV values are shown in Table 1 for two instruments; both generated measurements of cell concentrations—and thus values on the same relative scale—and had at least 20 of 36 possible CV values. For these two instruments, both the mean and median of the set of CV values were <25%. Although subsamples were drawn from the same source, variation in subsampling, dispensing aliquots of subsamples, and transferring a portion of the aliquot into the sample vessel likely contributed to the variation among readings. An alternate approach for measuring variation—collecting multiple readings from the same sample aliquot (e.g., the same cuvette)—could be performed in some cases to track the variation among repeated readings. In these trials, however, that approach was not performed as some devices either destroyed or degraded the sample during analysis; for these instruments, repeated readings of a discrete sample were not possible.

Accuracy was measured by comparisons to the standard, microscope-based method performed for land-based validations of BWMS. The standard microscope-based method, like all analytical methods, is subject to uncertainty caused by variation among microscopes (and microscopists), reagents, fluorophore labeling efficiency, and interferences associated with the sample matrix. This uncertainty has been empirically measured (e.g., Reavie et al. 2010; Steinberg et al. 2011; MacIntyre and Cullen 2016). These and other investigations observed gaps between concentrations of living organisms reported by fluorophore-based assays and a reference method, e.g., growth assays (Gorokhova et al. 2012) or staining with Neutral Red (Zetsche and Meysman 2012). Measured quantitatively via flow cytometry, the fluorescence intensity of fluorophores varied among different species tested, and the difference in fluorescence between living and heat-killed organisms was not
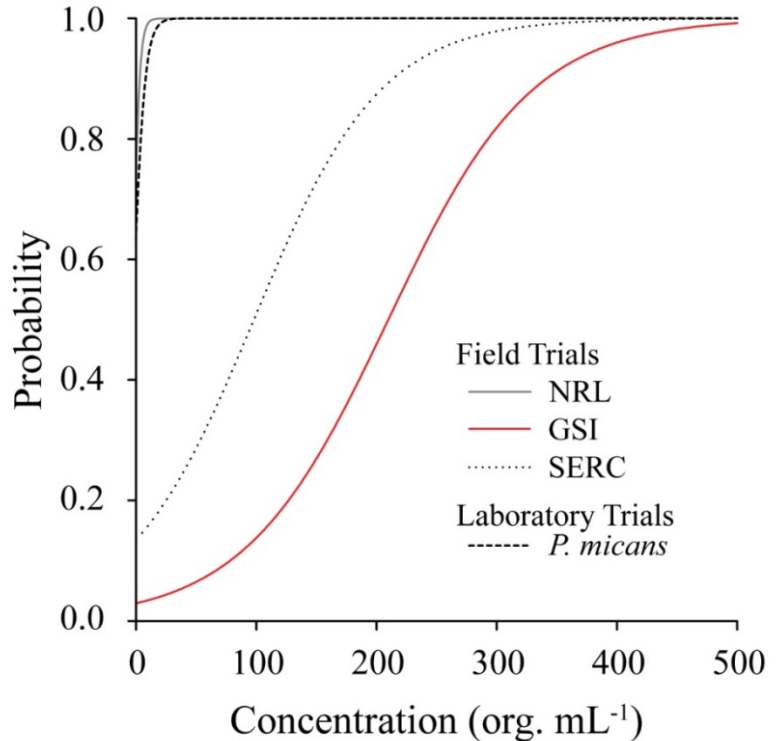
Figure 2. Probability of detecting an exceedance calculated along a range of concentrations using Eq. 1. The plots are based on field and laboratory results of Instrument *D*. See the text for a description of the figure and legend definitions.

always large enough to distinguish between the two populations (MacIntyre and Cullen 2016).

Notwithstanding the uncertainty associated with the standard method, it produces direct counts of living organisms. Included within these counts are heterotrophic organisms that are without chlorophyll and undetectable by the fluorometry-based compliance devices. The microscope-based technique is also an approved method for the certification of BWMS, so some level of consistency and comparability between certification testing and compliance monitoring would be important to the success of ballast water regulations. Agreement was based simply on the judgement of the compliance devices—indicating that the sample either met or exceeded the discharge standard—and the concentration measured by microscopy. Using empirically derived relationships, probabilities of detecting exceedance were calculated for cell concentrations ranging from 0 to 500 org. mL^{-1}, resulting in probability distributions for each field site and laboratory organism.

An example of the predicted probabilities for measuring exceedances is shown in Figure 2, where probabilities were based upon empirical relationships measured from each field site and for *P. micans*. For any actual cell concentration, the probability of correctly detecting an exceedance could be predicted

for each test site or cultured microalgae. In Figure 2, for example, the compliance device demonstrated high accuracy in field trials at NRL and in laboratory trials with *P. micans*: the probability of correctly predicting an exceedance when concentrations were exactly 10 org. mL^{-1} was 0.99 and 0.93, respectively. Other field sites, however, showed low predictability, and the probabilities for *T. marina* could not be calculated, as all *T. marina* samples were rated low risk of exceeding the discharge standard (logistical regression requires more than one outcome in the sample set).

Generally, compliance devices based upon variable fluorescence fluorometry produce results quickly (within minutes), operate without the need for reagents, and require only minimal sample processing. The fluorometry-based devices—at their core—require only optical components for illuminating chlorophyll *a* and measuring fluorescence emission. Therefore, the core components could be mounted into a small chassis, allowing for easy transport and hand-held operation. Fluorometry-based devices, including versions of devices evaluated in this study, have also been tested at other locations. In a test conducted with mixed assemblages of organisms sampled from the Adriatic Sea, the compliance devices demonstrated, in general, agreement with microscope counts

(Gollasch et al. 2015). Shipboard analyses of samples collected underway also revealed concurrence between microscope counts and the metrics of the compliance devices and—although differing in their detection limits—showed agreement among devices in their measurement responses across a range of organism concentrations (Bradie et al. 2018).

While this set of compliance devices was based upon variable fluorescence fluorometry, a standard method for validating compliance devices should be adaptable to other compliance devices, including those based upon ATP (van Slooten et al. 2015), bulk FDA hydrolysis (Akram et al. 2015), or another method. Because of this, simple and standardized metrics should gauge the performance of the device to the standard, microscope-based method. In this study, accuracy, which was measured via logistic regression analysis, provided parameters used to predict the likelihood of correctly identifying an exceedance, e.g., at 10, 30, 50, and 100 org. mL^{-1}. In addition to other considerations (e.g., analysis time, ease-of-use, and cost-per-sample), the likelihood estimates may be defined as a requirement by the device users, the compliance officers, and ship owners.

The other key metric to evaluate compliance devices is precision, which for compliance devices reflects the consistency among readings or agreement in outcomes. As compliance devices assist in decision-making, different outcomes among repeated readings weaken the justification of any enforcement actions. The approach used to measure precision, CV, is sensitive to the magnitude of the mean values. Thus, for typical reading-to-reading variability, instruments producing measurements of organism concentrations, especially when actual concentrations are < 10 org. mL^{-1}, will appear more variable than instruments producing measurements with a large dynamic scale.

The approaches for testing compliance devices described here provide an initial performance assessment and were designed to test multiple devices simultaneously. Concurrent evaluation of the compliance devices offers several advantages, in particular, comparing all devices to a single set of microscope counts reduces the analysis burden of comparing one device to one set of counts. Also, all devices analyzed the same samples, so variations among samples, e.g., due to the day-to-day variations in the assemblage of the ≥ 10 and < 50 µm community, are not a source of bias among tests. As more compliance devices become available (or as current devices are modified), concurrent analysis of multiple devices will be essential. As verification and validation of compliance devices progresses, additional testing should investigate the inter-unit

variability—i.e., differences among multiple units of the same compliance device—and the long-term stability of the device. Likewise, side-by-side microscopy and analysis with compliance devices using real ballast water would be necessary to verify that the device performs as expected. Ballast water is typically sequestered in the dark for long periods and exposed to suspended sediments and, potentially, dissolved metals. Such conditions could lead to the establishment of a microbial community obscured from detection by compliance devices or water samples with interferences. These sets of future evaluations, following the testing described herein, would provide high levels of confidence for compliance devices used for rapid, shipboard analyses.

Acknowledgements

References

ACT/MERC (2012) Alliance for Coastal Technologies/Maritime Environmental Resource Center. Workshop summary: Ballast water compliance monitoring using fluorometry. July 9; Annapolis, MD, Report: UMCES CBL 2012-029; http://www.act-us.info/Download/Workshops/2012/CBL_Fluorometry_Ballast_Compliance

Akram AC, Noman S, Moniri-Javid R, Gizicki JP, Reed EA, Singh SB, Basu AS, Banno F, Fujimoto M, Ram JL (2015) Development of an automated ballast water treatment verification system utilizing fluorescein diacetate hydrolysis as a measure of treatment efficacy. *Water Research* 70: 404–413, https://doi.org/10.1016/j.watres.2014.12.009

Bradie J, Broeg K, Gianoli C, He J, Heitmüller S, Curto AL, Nakata A, Rolke M, Schillak L, Stehouwer P Vanden Byllaardt J, Veldhuis M, Welschmeyer N, Younan L, Zaake A, Bailey S (2018) A shipboard comparison of analytic methods for ballast water compliance monitoring. *Journal of Sea Research* 133: 11–19, https://doi.org/10.1016/j.seares.2017.01.006

Casas-Monroy O, Chan PS, Linley RD, Vanden Byllaardt J, Kydd J, Bailey SA (2016) Comparison of three techniques to evaluate the number of viable phytoplankton cells in ballast water after ultraviolet irradiation treatment. *Journal of Applied Phycology* 28: 2821–2830, https://doi.org/10.1007/s10811-016-0798-3

Drake LA, Tamburri MN, First MR, Smith GJ, Johengen TH (2014) How many organisms are in ballast water discharge? A framework for validating and selecting compliance monitoring devices. *Marine Pollution Bulletin* 86: 122–128, https://doi.org/10.1016/j.marpolbul.2014.07.034

Gollasch S, David M, France J, Mozetic P (2015) Quantifying indicatively living phytoplankton cells in ballast water samples - recommendations for Port State Control. *Marine Pollution Bulletin* 101: 768–775, https://doi.org/10.1016/j.marpolbul.2015.09.037

Gorokhova E, Mattsson L, Sundström AM (2012) A comparison of TO-PRO-1 iodide and 5-CFDA-AM staining methods for assessing viability of planktonic algae with epifluorescence microscopy. *Journal of Microbiological Methods* 89: 216–221, https://doi.org/10.1016/j.mimet.2012.03.005

Green JM (1996) A practical guide to analytical method validation. *Analytical Chemistry* 68: A305–A309, https://doi.org/10.1021/ac961912f

Guillard RR, Ryther JH (1962) Studies of marine planktonic diatoms: I. *Cyclotella nana* Hustedt, and *Detonula conffervacea* (Cleve) Gran. *Canadian Journal of Microbiology* 8: 229–239, https://doi.org/10.1139/m62-029

Hilbe JM (2009) Logistic Regression Models. CRC Press, Boca Raton, Florida, USA, 656 pp

MacIntyre HL, Cullen JJ (2016) Classification of phytoplankton cells as live or dead using the vital stains fluorescein diacetate and 5-chloromethylfluorescein diacetate. *Journal of Phycology* 52: 572–589, https://doi.org/10.1111/jpy.12415

Reavie ED, Cangelosi AA, Allinger LE (2010) Assessing ballast water treatments: Evaluation of viability methods for ambient freshwater microplankton assemblages. *Journal of Great Lakes Research* 36: 540–547, https://doi.org/10.1016/j.jglr.2010.05.007

Steinberg MK, Lemieux EJ, Drake LA (2011) Determining the viability of marine protists using a combination of vital, fluorescent stains. *Marine Biology* 158: 1431–1437, https://doi.org/10.1007/s00227-011-1640-8

Tsolaki E, Diamadopoulos E (2010) Technologies for ballast water treatment: a review. *Journal of Chemical Technology and Biotechnology* 85:19–32, https://doi.org/10.1002/jctb.2276

U.S. Coast Guard (2012) Standards for living organisms in ships' ballast water discharged in U.S. waters. *Federal Register* 77, 17254–17320, https://www.federalregister.gov/documents/2012/03/23/2012-6579/standards-for-living-organisms-in-ships-ballast-water-discharged-in-us-waters

U.S. Environmental Protection Agency Environmental Technology Verification Program (2010) Generic Protocol for the Verification of Ballast Water Treatment Technology. Report number EPA/600/R–10/146, https://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=P10097A4.txt

van Slooten C, Wijers T, Buma AGJ, Peperzak L (2015) Development and testing of a rapid, sensitive ATP assay to detect living organisms in ballast water. *Journal of Applied Phycology* 27: 2299–2312, https://doi.org/10.1007/s10811-014-0518-9

Zetsche EM, Meysman FJR (2012) Dead or alive? Viability assessment of micro- and mesoplankton. *Journal of Plankton Research* 34: 493–509, https://doi.org/10.1093/plankt/fbs018