Accepted Manuscript

A shipboard comparison of analytic methods for ballast water compliance monitoring

Johanna Bradie, Katja Broeg, Claudio Gianoli, Jianjun He, Susanne Heitmüller, Alberto Lo Curto, Akiko Nakata, Manfred Rolke, Lothar Schillak, Peter Stehouwer, Julie Vanden Byllaardt, Marcel Veldhuis, Nick Welschmeyer, Lawrence Younan, André Zaake, Sarah Bailey

PII: S1385-1101(16)30204-0
DOI: doi: 10.1016/j.seares.2017.01.006
Reference: SEARES 1503

To appear in: *Journal of Sea Research*

Received date: 15 September 2016
Revised date: 9 December 2016
Accepted date: 26 January 2017

Please cite this article as: Johanna Bradie, Katja Broeg, Claudio Gianoli, Jianjun He, Susanne Heitmüller, Alberto Lo Curto, Akiko Nakata, Manfred Rolke, Lothar Schillak, Peter Stehouwer, Julie Vanden Byllaardt, Marcel Veldhuis, Nick Welschmeyer, Lawrence Younan, André Zaake, Sarah Bailey , A shipboard comparison of analytic methods for ballast water compliance monitoring. The address for the corresponding author was captured as affiliation for all authors. Please check if appropriate. Seares(2017), doi: 10.1016/j.seares.2017.01.006

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Title: A shipboard comparison of analytic methods for ballast water compliance monitoring

Authors: Johanna Bradie[1], Katja Broeg[2], Claudio Gianoli[3], Jianjun He[4], Susanne Heitmüller[2], Alberto Lo Curto[5], Akiko Nakata[4,6], Manfred Rolke[2,7], Lothar Schillak[8,9], Peter Stehouwer[10], Julie Vanden Byllaardt[1], Marcel Veldhuis[11], Nick Welschmeyer[12], Lawrence Younan[13], André Zaake[14], Sarah Bailey[1]

Corresponding author: Johanna Bradie, Johanna.bradie@dfo-mpo.gc.ca

Address:

[1]Great Lakes Laboratory for Fisheries and Aquatic Sciences, Fisheries and Oceans Canada, 867 Lakeshore Rd., Burlington, Ontario L7S 1A1, Canada

[2]Federal Maritime and Hydrographic Agency (BSH), Bernhard-Nocht-Str.78, 20359 Hamburg, Germany

[3]SGS Italia Spa, Via Campodoro 25, 35010 Villafranca Padovana (PD), Italy

[4]Technical division, Satake Corporation, 2-30 Saijo Nishihonmachi, Higashi-Hiroshima-shi, Hiroshima-ken, 739-8602 Japan

[5]Aqua-tools, 26 Rue Charles Edouard Jeanneret 78300 Poissy France

[6]Current address: Test branch code S246, U.S. Naval Ship Repair Facility and Japan Regional Maintenance Center (SRF & JRMC), Detachment Sasebo, Tategami-cho, Sasebo-shi, Nagasaki-ken, 857-0063 Japan

[7]Permanent address: Fasanenweg 32, 21682 Stade, Germany, manfred.rolke@web.de

[8]Permanent address: TRITON Marine Science & Consult, MIMarEST, Oberdorfstrasse 12b, 68542, Heddesheim/Germany

[9]Temporary address: TRITON Marine Science & Consult, Andrea Papandreou 51, 16675 GLYFADA / Greece

[10]SGS Institut Fresenius GmbH, Rödingsmarkt 16, 22087 Hamburg, Germany

[11] Marine Eco Analytics, Havenkade 1A, 1779 GS, Den Oever, the Netherlands

[12]Moss Landing Marine Laboratories, 8272 Moss Landing Rd., Moss Landing, CA 95039, USA

[13]Turner Designs, Inc., 1995 N. 1st Street, San Jose, CA 95112

[14]bbe Moldaenke GmbH, Preetzer Chaussee 177, 24222 Schwentinental, Germany

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Abstract

Promising approaches for indicative analysis of ballast water samples have been developed that require study in the field to examine their utility for determining compliance with the International Convention for the Control and Management of Ships' Ballast Water and Sediments. To address this gap, a voyage was undertaken on board the RV Meteor, sailing the North Atlantic Ocean from Mindelo (Cape Verde) to Hamburg (Germany) during June 4-15, 2015. Trials were conducted on local sea water taken up by the ship's ballast system at multiple locations along the trip, including open ocean, North Sea, and coastal water, to evaluate a number of analytic methods that measure the numeric concentration or relative biomass of viable organisms according to two size categories (≥ 50 µm in minimum dimension: 7 techniques, ≥ 10 µm and < 50 µm: 9 techniques). Water samples were analyzed in parallel to determine whether results were similar between methods and whether rapid, indicative methods offer comparable results to standard, time- and labor-intensive detailed methods (e.g. microscopy) and high-end scientific approaches (e.g. flow cytometry). Several promising indicative methods were identified that showed high correlation with microscopy, but allow much quicker processing and require less expert knowledge. This study is the first to concurrently use a large number of analytic tools to examine a variety of ballast water samples on board an operational ship in the field. Results are useful to identify the merits of each method and can serve as a basis for further improvement and development of tools and methodologies for ballast water compliance monitoring.

Keywords: Ballast water management convention; detailed analysis; indicative analysis; invasive species; management; shipping; viable organisms

Highlights:

- Evaluated rapid, indicative analysis methods to monitor ballast water compliance
- Trials were conducted on a variety of samples sourced across the Atlantic Ocean
- Several rapid methods showed high correlation with traditional microscopy results

ACCEPTED MANUSCRIPT

1.0 Introduction

The International Maritime Organization adopted the International Convention for the Control and Management of Ships' Ballast Water and Sediments in 2004 to minimize the transfer of harmful aquatic organisms and pathogens in ships' ballast water (IMO 2004). Regulation D-2 restricts the concentration of viable organisms ≥ 50 micrometers in minimum dimension at discharge to < 10 viable organisms per cubic meter, and organisms < 50 micrometers and ≥ 10 micrometers in minimum dimension (hereafter, 10-50 µm) to < 10 per millilitre (IMO 2004). Now that the International Ballast Water Management Convention has been fully ratified, and will enter into force on September 8, 2017, there is a pressing need for ships to plan installations of ballast water treatment systems, and for regulators to plan the implementation of the Ballast Water Management Convention into their national legislation and Port State Control inspection programs.

Compliance monitoring and enforcement must be consistent, rigorous, and efficient (IMO, 2008); it can be divided into two main parts: ballast water sample collection and sample analysis, each of which is challenged by several difficulties (Gollasch & David, 2003; Gollasch & David, 2013; IMO 2013; Gollasch & David, 2015;). A number of tools and technologies are in development for both sampling and analysis, and recent studies have shown promising results for sampling devices (First et al., 2012; Bradie, 2016). For sample analysis, two types of methods may be employed: 'indicative' or 'detailed' analyses. 'Detailed' analyses, such as microscopy, provide a direct and precise measurement of the number of viable organisms in a sample that typically requires extensive scientific expertise, costly equipment, and a timeframe generally too long for a compliance enforcement scenario. In contrast, 'indicative' analysis methods

ACCEPTED MANUSCRIPT

should be rapid and easy to operate, typically measuring biological, physical, or chemical parameters that can be related to the number of viable organisms in a sample to provide an indication of potential non-compliance [gross exceedance] with Regulation D-2 (Bailey, 2015; Frazier et al., 2013; IMO, 2013). Indicative methods rely on various indicators to assess the viable biomass and/or viability of organisms in samples collected, including adenosine triphosphate (ATP) methods that detect cellular energy (Wright, 2012), fluorescence measurements that rely on the natural photosynthetic activity of chlorophyll-containing cells (phytoplankton) (Veldhuis *et al.*, 2006; Wright, 2012), and fluorescein diacetate (FDA) methods that detect enzymatic activity (non-specific intracellular esterases or enzymes; Welschmeyer & Maurer, 2011)(Gollasch & David, 2010). Some indicative methods use calibration curves to convert the measured parameter to an estimated organism concentration. Several promising approaches have been developed but require further study in the field to understand their methodological differences and to assess their comparability, accuracy, and precision (Gollasch and David, 2010; Gollasch *et al.*, 2012; Gollasch *et al.*, 2015).

In this paper, we report the results of a series of trials that were undertaken to compare methods for ballast water sample analysis by conducting replicated, comparative testing on marine water samples collected onboard the research vessel 'Meteor' in transit from Mindelo, Cape Verde to Hamburg, Germany. Water samples were collected from the ship's ballast water system and analyzed in parallel by multiple analytic methods for the ≥ 50 µm and 10-50 µm size classes (7 and 9 techniques, respectively) to assess comparability between methods, with particular reference to microscopic analysis as the standard method. In so doing, we evaluated the sensitivity and precision of the different methods and provide an overview of the benefits

ACCEPTED MANUSCRIPT

and drawbacks of each method (i.e. costs, training requirements, processing time, and interpretability of output). To our knowledge, this is the largest study to date to assess the comparability and reliability of analytic tools for ballast water compliance management under field conditions. Results offer insight into the benefits and limitations of each method, and support ongoing efforts to establish reliable uniform analytic methods for compliance monitoring under the International Ballast Water Management Convention.

2.0 Methods

2.1 Sample collection

Samples were collected during ballast water uptake of sea water while in transit (sea chest intake positioned at 2.5 m depth), except one trial where samples were collected during discharge of Mindelo harbour water that had been held in a ballast tank for three days. The main ballast line of the RV Meteor is equipped with multiple sampling points to allow simultaneous collection of paired samples of untreated sea water in the engine room. During the voyage, we used three different sample collection devices (plankton net, SGS Ballast Water Sampler 1 (BWS1), and Triton skid NP 6007 TG 18) to run 20 paired trials, collecting a total of 40 samples. The plankton net (50 µm diameter mesh) is the traditional method of concentrating ballast samples, whereas sampling skids are compact devices that have been developed to enable filtration and concentration of large volumes of water in a small space. During each trial, ~1000 L of water was concentrated for analysis of organisms ≥ 50 µm, using the 'cod' end (50 µm mesh, plankton net) or inbuilt filter (50 µm mesh, sampling skids) of each sample collection device, to a final volume of 1 L (some exceptions; see Table A2). The volume of water filtered

ACCEPTED MANUSCRIPT

was quantified using a magnetic flow meter (Seametrics WMP104-100) for net samples and built-in flow meters for the sampling skids. For each sample, between 10 and 16 L of water was taken for analysis of 10-50 µm organisms by collecting approximately 500 ml of the filtrate produced by each sampling device every minute.

All rinse water used during sample collection (and later analyses) was prepared by sequentially filtering local sea water taken through the ship's scientific sea water tap system through a series of meshes (1000, 500, 35, and 8 µm, nominal pore sizes) followed by filtration through a 0.2 µm passive (gravity-fed) filter cartridge (Whatman Polycap TC150). Rinse water was prepared prior to the start of each trial, so that the rinse water was sourced from the same geographic location as the samples being tested. Table A1 in Appendix A contains detailed trial information including salinity, temperature, sampling time and positions, sample collection devices used, ballast water flow rate, and total volume of water that passed through the ship's ballast system during the trial.

2.2 Sample preparation

All sample collection and further handling, like sample splitting and sieving, was completed in a uniform way, so that observed variability is more likely explained by analysis method rather than by sample handling. Water samples containing organisms and particles ≥ 50 µm were concentrated during sample collection, so post-collection processing was not required. Individual subsamples for each analysis method were taken by mixing each 1 L condensed sample by inversion five times, half-filling each subsample bottle (7 bottles, total volume 35-300 ml depending on analysis requirements), and repeating this procedure until bottles were topped up to the required volume. This splitting procedure (5x inversion of the

ACCEPTED MANUSCRIPT

sample bottle, half fill, 5x inversion, fill remainder) was used to fill all subsample bottles detailed below.

Water samples containing organisms and particles < 50 µm (10-16 L filtrate samples) were processed to generate the fractions required for the remaining analyses (< 50 µm for flow cytometry, 10-50 µm for all other methods). The subsample bottle for flow cytometry was filled first, and the remaining water was filtered on a 10 µm (pore diameter) Sterlitech polyester track etch (PETE) membrane filter. The retained particles were resuspended in filter-sterile sea water with a final concentration up to 16x times the original concentration (see Table A2). The concentrated sample was split into subsample bottles for analysis of the 10-50 µm size class (8 subsamples, volume 25-350 ml). For most analytic methods, there was no further assessment of the size of organisms in the size-fractionated samples (i.e. all organisms contained in a given sample were considered to be within the relevant size class). However, the Satake Pulse Counter uses pulse strength to estimate organism size (see Appendix B for details), and microscopists used photomicrographs of 50 µm calibration beads (≥ 50 size class) and Sedgewick-Rafter grid widths (10-50 size class) as size references.

After each trial, all sampling gear, sample carboys, and subsample bottles were cleaned in a dilute (100-200 ppm) or concentrated (2500 ppm) bleach solution (depending on equipment robustness) made using the ship's potable water supply to prevent cross-contamination of living organisms between tests. After bleaching, all equipment was rinsed with MilliQ water three times; plankton nets were rinsed with potable water three times before being rinsed once with MilliQ water and hung to dry. Prior to re-use, all sample carboys and

ACCEPTED MANUSCRIPT

subsample bottles were rinsed three times with 0.2 µm filtered sea water to remove any residues.

2.3 Subsample Analysis

Subsamples were analyzed in parallel using the analytic methods described in Table 1; detailed methodology for each analytic method is provided in Appendix B. The indicative methods fall into three major groups: those that detect chlorophyll fluorescence activity (CFA), those that rely on FDA as a viability probe, and those that measure ATP. CFA devices (Walz WATER-PAM, bbe 10cells, TD Ballast-Check2, and Hach BW680) measure baseline fluorescence under dark adaptation (F_0), and maximal fluorescence (F_m) under saturating light to estimate total active chlorophyll fluorescence (F_v) in the subsample ($F_v=F_m-F_o$) (Wright et al., 2015). The Hach BW680 and the bbe 10cells determine relative 'active' chlorophyll biomass estimates based on the F_v value, whereas the TD BallastCheck-2^{TM} estimates organism concentration using corrected fluorescence measurements (see Appendix B for details). The F_v measurements from the CFA devices can be converted into a cell number using an instrument-specific calibration value. In contrast, FDA methods rely on the conversion of FDA to fluorescein by viable cells, which makes them appear green when excited by blue light. Thus, FDA is not a 'traditional' stain, in the sense that it does not 'bind' or 'bond' with internal cellular compounds (as would a nuclear 'stain' like SYBR Green). In the case of the Satake Pulse Counter, the amount of fluorescence within each cell is used to estimate cell size, whereas the Moss Landing Marine Labs (MLML) bulk FDA method relies on measuring the absolute fluorescein production rate, as measured extra-cellularly after it diffuses out of the cell.

ACCEPTED MANUSCRIPT

From each subsample, three replicate measurements were analysed by each analytic method/device. Subsample bottles were inverted five times before withdrawing each replicate to ensure subsamples were well mixed, and all steps of each analysis were performed independently on each replicate. Due to time constraints and operational considerations (e.g. limited materials), some analytic methods could not be conducted for all trials and in some cases, the number of replicates was also limited (e.g. flow cytometry). The total number of replicates analyzed for a given comparison is shown in the text above each plot.

Note that flow cytometric analysis was performed on subsamples preserved with formaldehyde (final concentration 2.3 %) and stored at –20°C until analysis; therefore the analysis of phytoplankton did not discriminate between viable and non-viable cells but reflects the total number of phytoplankton present.

2.4 Statistical Analysis

All analyses were performed in the R statistical programming environment (R Core Team 2015). Since there could be variability between samples collected at the same time using different sample collection devices in the ship's engine room, analytic results are only compared amongst measurements within the same sample (i.e. collected using the same sample collection device from the same sampling point in the same trial). The use of multiple sample collection devices is a potential advantage of our experimental design, since it reduces any bias that might be introduced by the choice of sample collection device (i.e. potential that health and/or diversity of sample may differ between collection methods, which may affect performance of analytic methods).

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

All measurement values were standardized to equivalent concentrations in the prepared subsample where necessary to account for any concentration steps performed for any analytic method (see Appendix B). Thus, measurements shown represent those in the subsample water after concentration during initial sample collection/preparation. For example, if 10 ml of the prepared subsample was concentrated to 1 ml for measurement by microscopy (i.e. 10x concentration) and 100 individuals were counted in this concentrated replicate (i.e. 100 individuals/ml), this value would be standardized to the concentration in the prepared subsample by dividing the measured value by the concentration factor (i.e. prepared subsample concentration = measured value in concentrated replicate/concentration factor). In this case, a value of 10 individuals/ml would be used for comparison against measurements by other analytic methods. One outlier measurement was dropped for SGS ATP (Aqua-tools) where a large organism was seen in the sample and a very high, concordant measurement was recorded; this replicate represents a rare event where one large organism was sampled, while this is of concern for ballast water sampling generally, it was excluded here as it is a real difference between samples, not a poor measurement, and thus not relevant for comparison between methods.

To compare results for each analytic method versus microscopy (i.e. dissecting and/or epifluorescence (with FDA) microscopy), pairwise scatterplots were made that compared all measurements taken for each subsample with replicates randomly paired between methods. A line of best fit was generated using Deming's regression which accounts for errors in both variables (Ripley and Thompson, 1987) and has been shown to give unbiased slope estimates for method comparisons (Linnet, 1993). Explicit error values were specified based on the

ACCEPTED MANUSCRIPT

standard deviation within each replicate. When this was not possible due to lack of replication (e.g. flow cytometry), it was assumed that the coefficient of variation was constant across all data values. The mean measurement (and standard deviation) for each sample for each analytic method is shown in Tables C.1 and C.2 for ≥ 50 µm and 10-50 µm samples respectively.

The various CFA devices were also compared to each other and to microscopy counts using the raw output of the variable fluorescence (Fv), with no manipulation or transformation of the raw data signals, to eliminate variation caused by device-specific conversion factors. This F_v was measured as the difference of chlorophyll fluorescence of dark-adapted phytoplankton (F_0), and the maximal fluorescence (F_m) under saturating light. As a result, the total active chlorophyll fluorescence (Fv) in the subsample is $F_v=F_m-F_o$ (Wright et al., 2015). These comparisons include the Pearson correlation coefficient *(r*) to quantify the strength of the relationship between each pair. *r* is useful to describe the total variability in the relationship between two methods even when methods are not measured in the same units, but note that it is dependent on the number of data points and the data range (Stockl et al., 1998), which limits comparability.

Importantly, variation in all pair-wise relationships explored is not only attributed to analytical differences and/or imprecision, but also sample-related effects (e.g. variation in the true number of individuals in each subsample)(Miller *et al.*, 2011).

3.0 Results

Natural variation in marine plankton communities throughout the voyage led to high variability in the concentration and composition of organisms observed in each size class across

ACCEPTED MANUSCRIPT

trials. The beginning of the cruise was spent mostly in oligotrophic (open ocean) water where biological variation was low and organism concentrations were at or below the limits specified in Regulation D-2, whereas very high concentrations were observed in the English Channel and the North Sea near the end of the cruise. Organism concentrations in the water ranged between ~1000 and ~800,000 individuals/m^3 for the ≥ 50 µm size class, and between 0.6 and 69.7 individuals/ml for the 10-50 µm size class (both as estimated by microscopy). Notably, the ≥ 50 µm size class was dominated by dinoflagellates (*Ceratium* sp. and *Protoperidinium* sp.) comprising, on average, 85% of individuals; remaining taxa typically consisted of copepods, rotifers, and tintinnids. The 10-50 µm size class was also dominated by phytoplankton, though occasionally small rotifers (e.g. Keratella) were observed in samples.

3.1 Results for ≥ 50 µm size class

Seven methods were used to analyze water samples containing organisms ≥ 50 µm: microscopy (visual inspection for motile organisms; FDA 'staining' for non-motile organisms), two FDA methods (Satake Pulse Counter, MLML bulk FDA), two CFA methods (measure only photosynthetic protists; Hach BW680, Walz WATER-PAM), and two ATP methods (MLML ATP, SGS ATP aqua-tools). In general, all analytic methods showed a positive correlation between recorded measurements and direct microscopy counts, and were sensitive enough to detect organisms in the water samples at a range of concentrations (i.e. all measurements for analytic methods were non-zero values when the microscope method detected live individuals; Figures 1; two exceptions/outliers: Walz WATER-PAM). As expected, variation between replicates increased with increasing abundance (Figure C.1).

ACCEPTED MANUSCRIPT

The Satake Pulse Counter is the only method that provided data outputs as estimated individuals/volume; this method tended to underestimate the concentration of individuals as compared to microscope counts (see Figure 1a, 1b; most data points fall below 1:1 line), but precision was higher than microscopy (i.e. less variation between replicates, Figure C.1). Interestingly, for low concentrations, both the Satake Pulse Counter and SGS ATP (aqua-tools) showed a strong relationship between microscope counts and measured value (ind/ml and ng.cATP/ml, respectively) with one group of outlier data (Figure 1b; outlier data points came from same trial). The Walz WATER-PAM and the Hach BW680 fluorometer also showed a fairly strong concordance with microscopy values at low concentrations (Figure 1b).

3.2 Results for 10-50 µm size class

Nine methods were used to analyze water samples containing organisms in the 10-50 µm size range: three FDA methods (epi-fluorescence microscopy using FDA stain, Satake Pulse Counter, MLML bulk FDA), four CFA methods (bbe 10cells, Hach BW680, Turner Designs' (TD) BallastCheck-2^{TM} , Walz WATER-PAM), flow cytometry (Beckman Coulter Epics-XL-MCL), and MLML ATP (Table 1). Five methods/devices were able to provide estimates as individuals/volume: Hach BW680, bbe10cells, TD BallastCheck-2^{TM}, flow cytometry (counts the sum of live and dead organisms), and the Satake Pulse Counter. Most methods were sensitive enough to detect organisms in the ballast water at a range of concentrations, including values less than 10 per mL (Figure 2a,b; but see Hach BW680 where several samples from open ocean oligotrophic waters were below the detection limit), and all methods showed a positive relationship between measured values and microscope estimates (Figure 2). The Satake Pulse Counter underestimated organism concentrations as compared to microscope estimates (see

ACCEPTED MANUSCRIPT

Figure 2; most data points fall below 1:1 line) and flow cytometry overestimated organism concentrations as compared to microscopy (see Figure 3a,3b; most data points fall above the 1:1 line). Variation between replicates generally increased with abundance (Figure C.2).

The CFA measurements had the strongest correlation with microscope results, both when comparing F_v values (r=0.84-0.94; Figure 3) and data converted to concentrations (Figure 2; Pearson correlation coefficient= 0.82-0.94; note concentrations could not be calculated for Walz WATER-PAM). The F_v values for devices measuring chlorophyll fluorescence were highly correlated with each other (range 0.96-0.97; TD BallastCheck-2^{TM} vs. Hach BW680 excepted; Figure 3), with TD BallastCheck-2^{TM} and bbe 10cells showing the highest similarity.

4.0 Discussion

Our study considered several types of 'indicative' analytic tools, including methods based on CFA, ATP, and FDA, and two detailed methods, flow cytometry and microscopy. Results indicate that measurements for all types of indicative methods showed concordance with microscopy results, whether those results were analyzed as raw biomass measurements or when converted to concentrations (where possible). There was variability between indicative methods in the strength of the relationship (which varied with numerical abundance of the organisms), and also differences in the sensitivity (i.e. minimum detection limit) of methods. It should be noted that some of the indicative methods are designed for ballast water monitoring where low organism concentrations are expected and therefore are not necessarily expected to give accurate results at high concentrations, but rather simply recognize non-compliance; in other words, unlike methods developed for standard marine research, the focus here is on

ACCEPTED MANUSCRIPT

accuracy at low concentrations near Regulation D-2 standards while accuracy at high concentrations often observed in nature is less important.

All analytic tools in this study aim to estimate abundance of organisms in ballast water samples, but the methods differ in terms of the taxa being measured and the units of their output. Both detailed analysis methods (i.e. microscopy and flow cytometry) are able to provide estimates in individuals/ml, but microscopy estimates account for both zooplankton and phytoplankton, whereas flow cytometry measurements only count phytoplankton (as used in this study, see Appendix B for methodology)(Veldhuis & Kraay, 2000). Further, since all flow cytometry samples were preserved before analysis in this experiment, measured concentrations are based on the cumulative number of organisms in samples (whether alive or dead before preservation). In comparison, microscopy (using motility and/or FDA 'staining') estimates are based only on living organisms. Likewise, FDA and ATP methods, which target metabolic activity and cellular adenylate content, respectively, quantify the biomass of both autotrophs and heterotrophs present in samples, albeit each measuring different properties of life. However, of these methods, only the Satake Pulse Counter (FDA method) provides estimates as individuals/ml, though other analytic methods (e.g. SGS ATP (aqua-tools)) may be able to estimate organism concentration for this size class in the future. In contrast, the CFA devices target only photosynthetic protists, but most can provide results in individuals/ml for the 10-50 µm size class through an empirical, instrument-specific conversion constant (i.e. bbe 10cells, TD BallastCheck-2^{TM}, Hach BW680).

Generally, analytic methods were quite sensitive to the detection of organisms in samples over a broad range of cell concentrations. Results indicated, however, that the various

ACCEPTED MANUSCRIPT

fluorometers (i.e. bbe 10cells, TD BallastCheck-2^{TM}, Walz WATER-PAM and Hach BW680), which had highly concordant results in agreement with previous work (e.g. Gollasch and David, 2012b), differed in their sensitivity (Figures 1-3). In particular, the Hach BW680 was less sensitive than the bbe 10cells and TD BallastCheck-2^{TM}. In total, 14 out of 32 samples were below the detection limit of the Hach BW680 in open ocean oligotrophic waters where viable organisms were detected by microscopy; five of these samples had concentrations > 10 individuals/ml. In contrast, bbe 10cells and TD BallastCheck-2^{TM} reported measured values for all samples analysed. The sensitivity differences between these instruments are likely explained by differences in their methodology. More specifically, both the Hach BW680 and the bbe 10cells determine relative 'active' chlorophyll biomass estimates based on the F_v value, which is calculated as the difference between the F_m and F_0 measurements. This can reduce the sensitivity of the instrument output, because the error in each measure is combined. The bbe 10cells measurement increases signal strength to avoid sensitivity loss by concentrating samples onto a selective filter.

For both size classes, the concentrations estimated by the Satake Pulse Counter were lower than measured by microscopy. There are multiple possible explanations for these differences. Firstly, the Satake Pulse Counter would likely register a group of colonial organisms as one pulse instead of many pulses; this could lead to discrepancies if the Satake Pulse Counter registers one large pulse (> 50 size class) when the microscopist counts multiple cells (likely 10-50 size class), or if one pulse is registered where the microscopist determines all cells within the colony are smaller than 10 µm. Alternatively, it is possible that differences in methods to estimate organism size could lead to discrepancies. The Satake Pulse Counter determines

ACCEPTED MANUSCRIPT

organism size based on pulse strength, with strong pulses attributed to large organisms and weak pulses attributed to small organisms, whereas the microscopists used photomicrographs of 50 µm calibration beads (> 50 size class) and Sedgewick-Rafter grid widths (10-50 size class) as size references. Inaccuracy by either method may lead to discrepancies in counts across both size classes, thus, it may be beneficial to examine the accuracy of size measurements for future comparisons of analytic methods. Further, discrepancies can arise due to human error, particularly when conducting microscopy (First & Drake, 2012) (though not limited to microscopy) due to inability to detect FDA 'staining' against background autofluorescence, fatigue, motion sickness, etc. Notably, Satake Pulse Counter and SGS ATP (aqua-tools) had similar outlier data points for the same sample, where measurements were quite high compared to microscope counts, which may indicate error by microscopy or poor mixing during the sample splitting process.

While this paper has compared the accuracy and precision of different analytic methods, factors such as cost, training requirements, and the time required to analyze and process a sample are also important considerations. Microscopy is the most demanding method, with requisite taxonomy skills, 20-60 minutes processing time per sample, and a substantial investment in equipment, making it an unlikely choice for compliance monitoring by non-scientific regulatory personnel. For the handheld CFA devices, including the bbe 10cells, TD BallastCheck-2^{TM} and Hach BW680, sample processing takes approximately 2 minutes and requires only minimal training; the devices range between $3,700-$4,800 and running costs are < $1 per sample (all values are approximate and in USD). The desktop Walz WATER-PAM, is a high-end scientific instrument, and has similar speed, per sample costs, and training

ACCEPTED MANUSCRIPT

requirements, but costs $16,700 to purchase. The Satake Pulse Counter has not yet been priced for market, but can be used with minimal training to analyze the ≥ 50 µm and 10-50 µm size classes; active processing time for both size classes is less than 5 minutes with results available for the ≥ 50 µm class in approximately 10 minutes and 30 minutes for the 10-50 µm class. The SGS ATP (aquatools) kit requires minimal training and processes samples for either the ≥ 50 µm or 10-50 µm size class in approximately 50 minutes; the initial start-up costs are $8,500 to purchase equipment and running costs are $35 per sample; some costs may decrease as devices mature from developmental technology to commercially-available units.

4.1 Caveats and Future Directions

In reviewing our results, it is important to keep in mind that all samples are subject to sampling effects and, due to random chance, it is not expected that every subsample or replicate measurement would contain identical numbers of organisms (Frazier *et al.*, 2013; i.e. if a 10 L ballast water sample containing 1000 organisms was split into ten 1 L subsamples, we would not expect that each subsample would contain exactly 100 organisms). Further, the variation between subsamples should increase with the mean (i.e. there is more variation between subsamples if the true concentration is 1000 individuals / L than 100 individuals / L). In fact, variation among subsamples is expected to equal the mean if all organisms have identical and independent chances of being sampled, but variation could be amplified by organism behaviour and/or the presence of colonies in samples. Therefore, even if all methods gave completely accurate results, variation in measurements between subsamples is expected (Miller *et al.*, 2011). Importantly, as a result, observed differences between measurements may

ACCEPTED MANUSCRIPT

represent a combination of true differences between subsamples and measurement error (Stockl, 1998). Herein, we visualized the precision of methods by quantifying the standard deviation among replicates for each trial (flow cytometry excluded, since measurements were not replicated). These graphs show that variability is low at low concentrations (i.e. the region of utmost concern for compliance monitoring), but they do not allow us to directly compare instruments quantitatively since most methods use different measurement units. Ideally, it would be beneficial if multiple replicates of multiple subsamples were analyzed by each tool to help parse out these sources of variation, but operational considerations limited the number of measurements taken during this voyage. Such an approach could be the focus of future empirical studies, or alternatively, lab-based studies may be useful to study the reliability and precision of these methods across a controlled range of values (see Vanden Byllaardt *et al.*, submitted).

Conducing our trials on board a ship in transit offered both advantages and disadvantages. While we were disadvantaged by the added complexity of conducting microscopy on a moving ship, we expect our results benefit from (i) organisms being subjected to the stresses of the ballast system during collection akin to real compliance scenarios and (ii) the diversity of communities that were sampled. Indeed, we expect that the composition of plankton communities is likely to affect the level of concordance observed between analytic methods. While our results showed that CFA devices performed well for the ≥ 50 µm size class, these results may be dramatically different if the samples for this size class had not been dominated by dinoflagellates, as was the case during this voyage. Thus, future work is needed to ensure that a variety of communities are considered when testing methods, including

ACCEPTED MANUSCRIPT

varying concentrations of zooplankton. The community composition may also affect how well our size-segregated samples conform to those defined in the D-2 convention (i.e. ≥ 50 µm and 10-50 µm in minimum dimension). As detailed in our methodology, samples were separated with a series of filters and all instruments except microscopy and the Satake Pulse Counter measured the sample without any further consideration of whether the organisms contained therein were of the expected size. Depending on the flexibility and shapes of organisms contained within our samples, it is possible that our samples did not exclusively contain organisms of the specified size since these properties can impact the efficiency of filtration. Any individuals excluded from microscopy counts based on size would thus not have been similarly excluded from measurement by the indicative tools; this could lead to higher estimates for the indicative tools versus microscopy.

In a related matter, it is interesting to note that previous ballast water counting protocols for the ≥ 50 µm size class have counted phytoplankton at the same time as motile zooplankton using a dissecting microscope, with the assumption that structurally intact phytoplankton cells are viable (e.g. First *et al.*, 2015). Since preliminary trials indicated that dinoflagellates were present in our samples at high abundance in the ≥ 50µm size class (on average 85% of individuals), we modified our analytic methods to include two sets of microscopy counts for this size class: (i) motile taxa counted on dissecting microscope, and (ii) non-motile phytoplankton counted with FDA 'stain' on the epifluorescence microscope, to avoid relying on the assumption that intact cells were viable. While we concede that the error rates (false positives and false negatives) of the FDA 'staining' were not quantified during our experiment, we assumed they would be similar to those reported in previous studies (e.g.

ACCEPTED MANUSCRIPT

Bentley-Mowat, 1982, Garvey *et al.*, 2007, Steinberg *et al.* 2011); conversely, the error associated with assuming all intact phytoplankton are viable was completely unknown (Paerl, 1978). Since dinoflagellates were so abundant in our samples, errors in determining viable vs. dead status could have significant impact on the results. While this issue lies outside of our analytic methods comparison, it may have significant implications for ballast water monitoring studies more generally, suggesting that examination of appropriate methods for enumerating viable phytoplankton in the ≥ 50µm size class could be considered further in future work.

Finally, it is important to highlight that each CFA device has a specific calibration/conversion coefficient that relates its results (F_v,F_m, or F_o) to cell concentrations. Consequently, different devices can generate different estimates even if they have the same fluorescence reading. In reality, the true relationship between fluorescence and cell concentration varies across phytoplankton communities depending on cellular chlorophyll content (can vary by a magnitude of 5x due to photo-adaption and an additional 5x due to nutrient limitation/saturation) and the size of cells in the community (range of cell volumes for 10-50 µm group varies 125x)(Veldhuis et al, 1997). Thus, chlorophyll content per cell is the major source of error for estimating cell numbers on the basis of bulk chlorophyll fluorescence and variation in calculated cell concentrations among instruments may depend on the similarity of the ballast sample community to that used to develop the calibration factor. Nonetheless, present results for these methods were promising for both size classes examined here. We expect that future studies to examine and quantify the variability in the relationship between fluorescence and cell concentration for ballast water communities would provide valuable insight.

ACCEPTED MANUSCRIPT

5.0 Conclusions

In conclusion, several promising indicative methods have been evaluated which should prove useful for ballast water compliance monitoring. These tools differ in their sensitivity, output, costs, training requirements, and processing time, all of which must be considered when selecting tools for ballast water compliance monitoring. Future work will be required to determine the accuracy of these methods at varying cell concentrations, to assess these methods in different types of plankton communities, and to determine if different processes for treatment of ballast water have effects on the analytical results of the instruments. A better understanding of the variability in concentration estimates due to the selection of calibration factors is crucial.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 1. Brief description of analytic methods used on the METEOR voyage. Full description and Standard Operating Procedures for each method are available in Appendix B. Method type is indicated as Det = Detailed analytic method or FDA, ATP or CFA for three types of indicative analytic methods where FDA = fluorescein diacetate, ATP = adenosine triphosphate, CFA = chlorophyll fluorescence activity. MLML = Moss Landing Marine Labs. BWI = Ballast Water index.

Analysis Method	Type	Description
Microscopy (movement)	Det	Samples were placed in a modified Bogorov chamber and live zooplankton (≥ 50 µm) were enumerated by observing movement.
Microscopy (FDA 'staining')	Det	Samples were 'stained' using FDA and fluorescing organisms were counted using a Sedgewick-Rafter counting chamber under an epifluorescence microscope equipped with a fluorescein isothiocyanate (FITC) narrow pass filter cube.
Flow cytometry	Det	Samples are inserted into flow cytometer for measurement. Phytoplankton cells are separated from other particles based on scatter and the red fluorescence of the chlorophyll present in the phytoplankton cell. The size range of phytoplankton is determined using spherical beads as an internal standard.
Satake Pulse Counter	FDA	Samples were stained using FDA and placed in portable stirring chamber that estimates the number of viable organisms based on the number of fluorescence pulses detected over a specific threshold.
MLML (streamlined) bulk FDA	FDA	Organisms were captured on a filter, placed into incubation, and tagged with FDA. Active organisms convert FDA to fluorescein, which is measured quantitatively in the bulk, whole-water incubation fluid.
MLML ATP	ATP	Organisms in sample were captured on a filter and immersed in a strong extraction fluid. Extraction tubes were mixed thoroughly after 1 hour extraction and measured for luminescence in the presence of luciferase enzyme; luminescence is linearly related to ATP concentration.
SGS ATP (aqua-tools)	ATP	Rapidly estimates living organisms through quantification of bioluminescent signal coming from the reaction of Luminase™ with intracellular Adenosine TriPhosphate (cATP).
Walz WATER-PAM	CFA	Desktop device to estimate phytoplankton biomass and photosynthetic activity. Measurements can be performed using whole water or size fractionated samples.
Turner Designs' BallastCheck-2™	CFA	Estimates abundance and assesses viability of phytoplankton based on fluorescence produced by organisms.
bbe 10cells	CFA	Estimates the number of living cells based on variable fluorescence (Fv) of chlorophyll of photosynthetically-active algae.
Hach BW680	CFA	Samples were deposited into a cuvette for measurement. Device displays BWI values and estimated risk based on average variable fluorescence response. Viable cell concentrations can be estimated based on BWI, which is proportional to variable fluorescence, F_v, in conventional PAM fluorometry.

ACCEPTED MANUSCRIPT

Figure Captions

Figure 1. Scatterplots showing the output from each analytic tool for the ≥ 50 µm size class compared to total microscopy estimates. Panel (a) shows full data while panel (b) shows the subset of data corresponding to microscopy estimates below 40 individuals per ml. All values are standardized to represent the concentration of organisms (or biomass indicator) in the prepared subsamples. The solid line indicates the line of best fit found using Deming regression (formula indicated above graph) and dashed lines (where applicable) indicate 1:1 line. Number of data points (n) is indicated above each plot. Note different x-axis limits in panel (a) corresponding to available data. Red dots are used for FDA methods; green dots indicate CFA methods, and blue dots indicate ATP methods.

Figure 2. Scatterplots showing the output from each analytic tool for the 10-50 µm size class compared to total microscopy estimates. Panel (a) shows full data while panel (b) shows the subset of data corresponding to microscopy estimates below 40 individuals per ml. All values are standardized to represent the concentration of organisms (or biomass indicator) in the prepared subsamples. The solid line indicates the line of best fit found using Deming regression (formula indicated above graph) and dashed lines (where applicable) indicate 1:1 line. Number of data points (n) is indicated above each plot. Note different x-axis limits in panel (a) corresponding to available data. Red dots are used for FDA methods; green dots indicate CFA methods, blue dots indicate ATP methods; purple dots indicate flow cytometry. Black dots represent points below detection limit of tool; points are plotted at minimum detection level (Hach BW680 only).

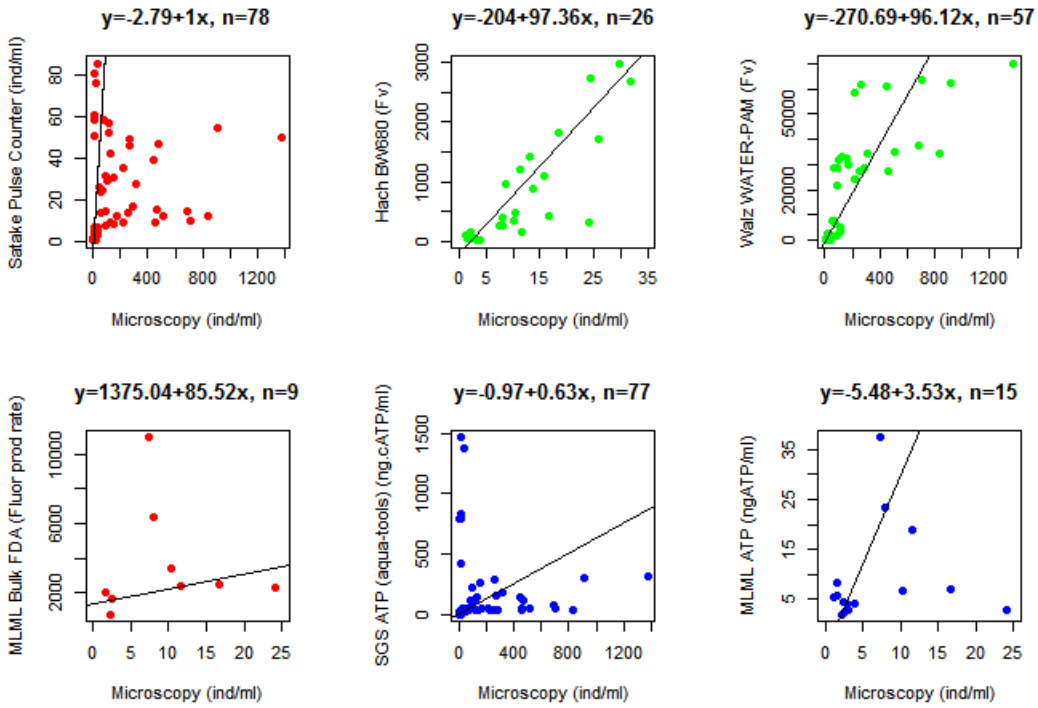
Figure 3. Scatterplots showing pairwise comparisons of raw Fv values for each CFA device and microscopy counts (ind/mL) for the 10-50 µm size class. The Pearson correlation coefficient (r) is indicated for each pair.

ACCEPTED MANUSCRIPT

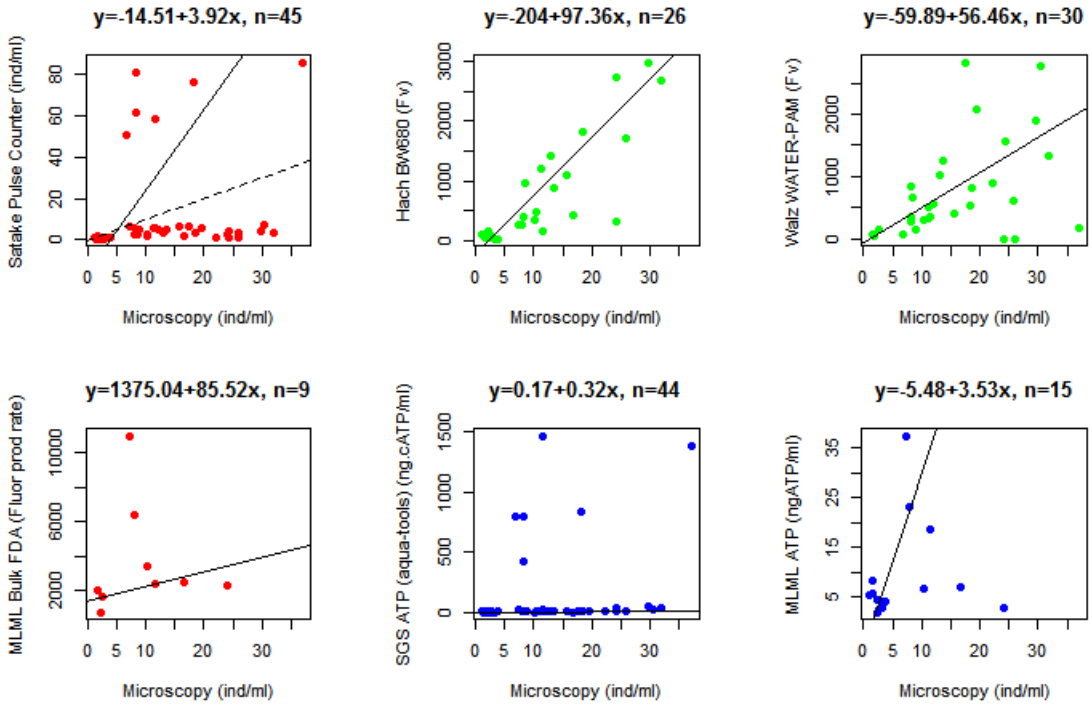
ACCEPTED MANUSCRIPT

Figure 1.

(a)



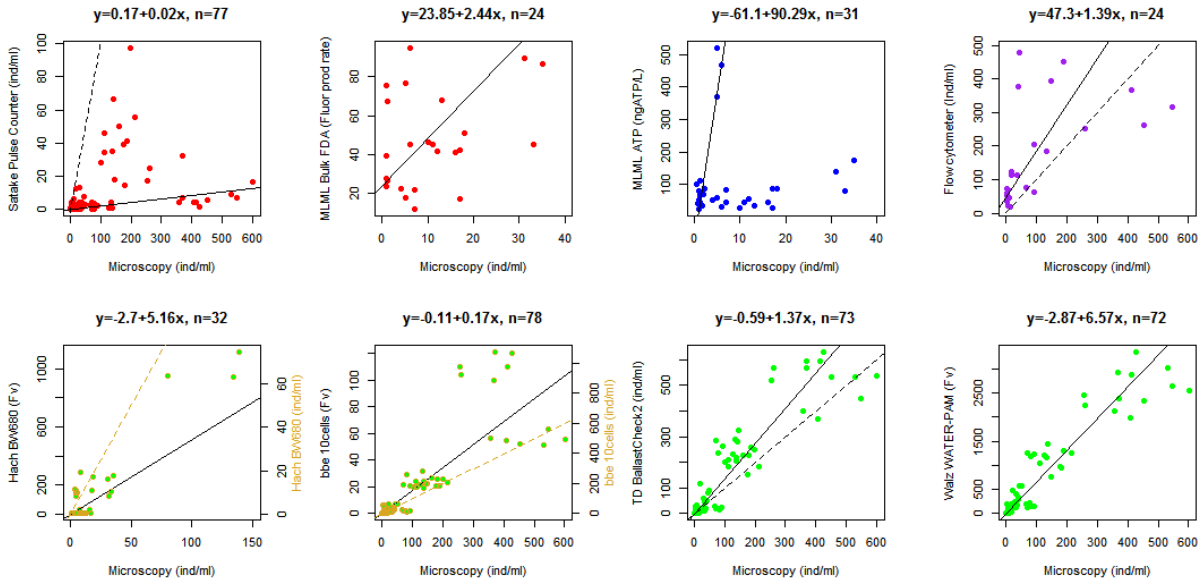
(b)



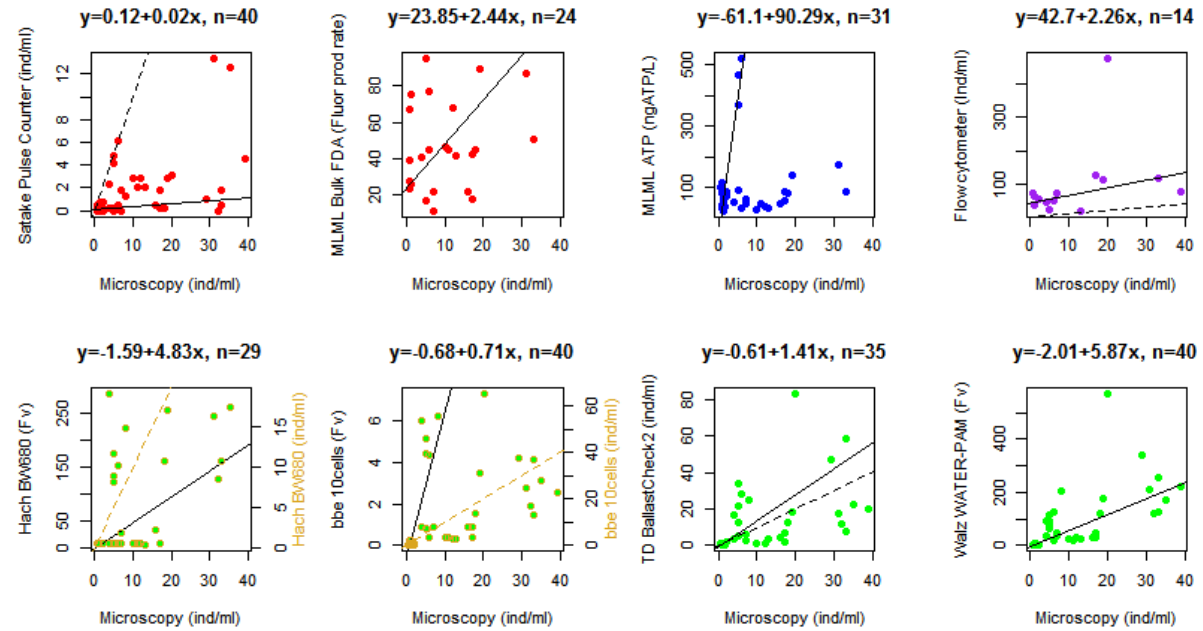
ACCEPTED MANUSCRIPT

Figure 2.

(a)

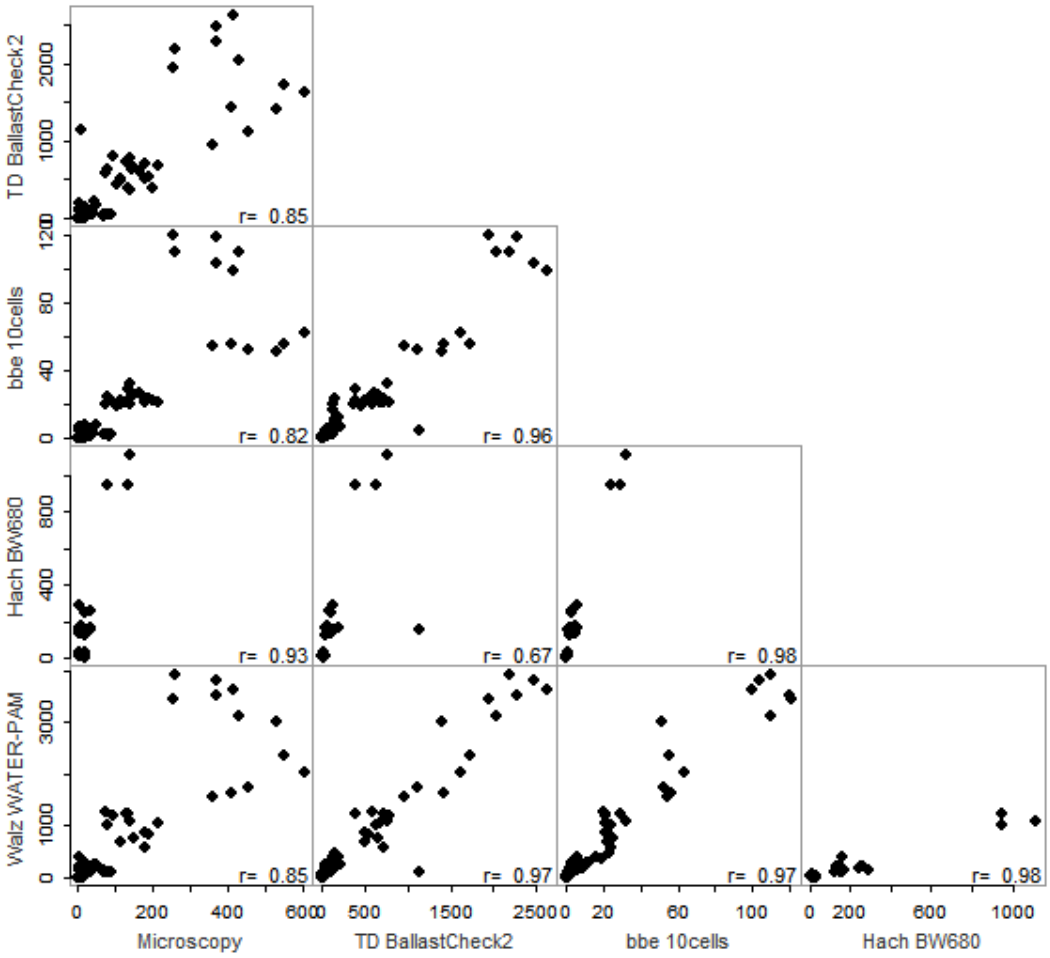


(b)



ACCEPTED MANUSCRIPT

Figure 3.



ACCEPTED MANUSCRIPT

Acknowledgements

This research benefitted from strategic planning and collaborative activities advanced through meetings of the ICES/IOC/IMO Working Group on Ballast and Other Ship Vectors. This paper presents the results of systematic investigative research, and neither endorses nor discourages the use of the technologies tested. We gratefully acknowledge the crew of the R/V METEOR for their collaboration and hospitality, Jarl Knutsvik for practical assistance on board, and many colleagues for helpful suggestions that contributed to our analysis and discussion. We thank Stephan Gollasch for the loan of various ballast sampling equipment and Wendy Allan (Carl Zeiss Canada) for the loan of a microscope.

Funding

We acknowledge the German Federal Maritime and Hydrographic Agency (BSH), Transport Canada, and Fisheries and Oceans Canada for funding.

References

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: 1–8.

Bentley-Mowat, J. 1982. Application of fluorescence microscopy to pollution studies on marine phytoplankton. Botanica Marina, 25, 203–204

ACCEPTED MANUSCRIPT

Bradie, J. 2016. METEOR Voyage M116/2: Report on performance of ballast water collection and analysis devices. Prepared for BSH (German Federal Maritime and Hydrographic Agency): 130 pages.

David, M, & Gollasch, S. n.d. Representative ballast water sampling for ballast water management compliance monitoring. Report. Available at http://www.geocities.ws/icts_papers/Papers/David,%20Gollasch%201.pdf

First, M. R., & Drake, L. A. (2012). Performance of the human "counting machine": evaluation of manual microscopy for enumerating plankton. Journal of plankton research, 34(12), 1028-1041.

First, M. R., Lemieux, E. J., Hyland, W. B., Grant, J. F., Moser, C. S., Riley, S. C.,Robbins-Wamsley, S., Stenberg, M. K., Wier, T. P., & Drake, L. A. (2012). Validation of a closed-housing filter skid for in-line sampling of aquatic organisms. Journal of plankton research, fbs007.

First, M. R., Robbins-Wamsley, S. H., Riley, S. C., & Drake, L. A. 2015. Towards minimizing transport of aquatic nuisance species in ballast water: Do organisms in different size classes respond uniformly to biocidal treatment?. Biological Invasions, 1-14.

Frazier, M., Whitman Miller, A., Lee II, H., Reusser, D. A. 2013. Counting at low concentrations: the statistical challenges of verifying ballast water discharge standards. Ecological Applications, 23(2), 339-351.

Garvey, M., Moriceau, B., Passow, U. 2007. Applicability of the FDA assay to determine the viability of marine phytoplankton under different environmental conditions. Marine Ecology Progress Series, 352, 17–26.

ACCEPTED MANUSCRIPT

Gollasch, S., Rosenthal, H., Botnen, H., Crncevic, M., Gilbert, M., Hamer, J., Hülsmann, N., Mauro, C., McCann, L., Minchin, D., Öztürk, B., Robertson, M., Sutton, C. & M. C. Villac. 2003. Species richness and invasion vectors: Sampling techniques and biases. Biological Invasions 5, 365-377

Gollasch S, David M. 2010. Testing Sample Representativeness of a Ballast Water Discharge and developing methods for Indicative Analysis. Final Report. EMSA Tender N°EMSA/NEG/09/2010. 124 pp.

Gollasch S, Stehouwer PP, David M, 2012. Technical Outline and Requirements for Organism Detection Systems for Establishing Compliance Enforcement with ballast water management requirements. Final report. Prepared for Interreg IVB North Sea Ballast Water Opportunity project: 68 pp.

Gollasch S, David M, 2012b. On board tests of the organism detection tools BallastCAM, FluidImaging, USA, Hach-PAM-fluorometer, USA, and Walz-Water-PAM-fluorometer. Results and findings. Prepared for Interreg IVB North Sea Ballast Water Opportunity project: 10 pp.

Gollasch S, David M. 2013. Recommendations for Representative Ballast Water Sampling. Final report of research study of the Bundesamt für Seeschifffahrt und Hydrographie (BSH), Hamburg, Germany. Order Number 4500025702.

Gollasch S, David M. 2015. Ballast water sampling and sample analysis for compliance control. In David M, Gollasch S (eds) Global Maritime Transport and Ballast Water Management

ACCEPTED MANUSCRIPT

– Issues and Solutions. Invading Nature. Springer Series in Invasion Ecology 8, Springer Science + Business Media, Dordrecht, The Netherlands. 306 pp.

Gollasch S, David M, France J, Mozetič P. 2015. Quantifying indicatively living phytoplankton cells in ballast water samples - Recommendations for port State control. Marine Pollution Bulletin, 101, 768-775. DOI: 10.1016/j.marpolbul.2015.09.037.

IMO (International Maritime Organization). 2004. International convention for the control and management of ships' ballast water and sediments. International Maritime Organization. Adopted on 13 February 2004. IMO, London, p. 36.

IMO (International Maritime Organization). 2008. Guidelines for Ballast Water Sampling (G2), IMO Resolution MEPC.173(58). International Maritime Organization, London, p. 14.

IMO (International Maritime Organization). 2013. Guidance on ballast water sampling and analysis for trial use in accordance with the BWM Convention and Guidelines (G2), BWM.2/Circ.42. IMO, London, p. 29.

IMO (International Maritime Organization). 2016. Status of multilateral conventions and instruments in respect of which the international maritime organization or its secretary-general performs depositary or other functions (as at 2 Aug 2016). IMO, London.

Linnet, K. 1993. Evaluation of regression procedures for methods comparison studies. Clinical Chemistry, 39(3), 424-432.

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. Environmental science & technology, 45(8), 3539-3546.

ACCEPTED MANUSCRIPT

Paerl, H. W. (1978). Effectiveness of various counting methods in detecting viable phytoplankton. New Zealand Journal of Marine and Freshwater Research, 12(1), 67-72.

R Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Richard, R.V., Grant, J.F., & Lemieux, E.J. 2008. Analysis of Ballast Water Sampling port Designs Using Computational Fluid Dynamics. Report No. CG-D-01-08, United States Coast Guard Research and Development Center, Groton, CT.

Ripley, B. & Thompson, M. 1987. Regression techniques for the detection of analytical bias, Analyst, 112, 377-383.

Steinberg, M. K., Lemieux, E. J., & Drake, L. A. 2011. Determining the viability of marine protists using a combination of vital, fluorescent stains. Marine biology, 158(6), 1431-1437.

Stockl, D., Dewitte, K., Thienpont, L. 1998. Validity of linear regression in method comparison studies: is it limited by the statistical model or the quality of the analytical input data? Clinical Chemistry 44, 2340-2346.

Veldhuis, M.J.W., Cucci, T., & Sieracki, M. 1997. Cellular DNA content of marine phytoplankton using two new fluorochromes: taxonomic and ecological implications. Journal of Phycology, 33, 527 - 541.

Veldhuis, M.J.W. & Kraay, G.W. 2000. Application of flow cytometry in marine phytoplankton research: current applications and future perspective. Sci Mar., 64: 121 – 134.

Veldhuis, M.J.W., F. Fuhr, J.P. Boon, C.T. Hallers-Tjabbers. 2006 Treatment of ballast water: how to test a system with a modular concept? Environmental Technology, 27, 909-921.

ACCEPTED MANUSCRIPT

Welschmeyer, N. A., & Maurer, B. (2011, October). A portable, sensitive plankton viability assay for IMO shipboard ballast water compliance testing. In *Proceedings of the IMO globallast conference: compliance monitoring and enforcement, Istanbul, Turkey* (pp. 127-139).

Wright, D. A. (2012). Logistics of compliance assessment and enforcement of the 2004 ballast water convention. Journal of Marine Engineering & Technology, 11(1), 17-23.

Wright, D. A., Welschmeyer, N. A., & Peperzak, L. (2015). Alternative, indirect measures of ballast water treatment efficacy during a shipboard trial: a case study. Journal of Marine Engineering & Technology, 14(1), 1-8.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Highlights

- Evaluated rapid, indicative analysis methods to monitor ballast water compliance
- Trials were conducted on a variety of samples sourced across the Atlantic Ocean
- Several rapid methods showed high correlation with traditional microscopy results

ACCEPTED MANUSCRIPT