San Francisco Estuary Chlorophyll Sensor and Sample Analysis Intercomparison

Elizabeth Stumpner, U.S. Geological Survey California Water Science Center

Jamie Yin, Matthew Heberger, Jing Wu, and Adam Wong, San Francisco Estuary Institute

John Franco Saraceno, California Department of Water Resources

January 2022

SFEI Contribution Number 977

Executive Summary

This report presents an assessment of chlorophyll collection methods and anonymous results of field and laboratory comparisons in 2018 - 2019 by agencies in the San Francisco Estuary (SFE). The methods assessment and comparison exercises, with funding provided by the Delta Regional Monitoring Program and Bay Nutrient Management Strategy and in-kind contributions from participating agencies, are a first step to facilitate future comparisons and syntheses of data and inform best science practices in the region. *In situ* sonde comparison exercises found general agreement between two models of Yellow Springs Instrument (YSI) sensors, but the newer sensor (EXO v2 - total algae) measured higher chlorophyll fluorescence (fCHL) relative to the older YSI sensor (6-series 6025). Results may be attributed to the use of a two-point calibration and the fluorescence response of algal cultures in sensor development by the manufacturer. The laboratory comparison included participation by 12 distinct field - laboratory pairs (or groups), with one group analyzing filters using two analytical methods. Filters were collected in triplicate across three sampling events in 2018, and all sample results were pooled together. Results of statistical analyses indicated that nominal filter pore size, the grinding method associated with pigment extraction, and analytical methods do not introduce variability to the chlorophyll-*a* measurement (Chl-*a*). When Chl-*a* results were assessed by sample event, however, significant differences between nominal pore size and analytical methods existed; these differences could be attributed to the small sample size per event. Consistent reporting units and high-concentration calibration standards for field sensors among data collection agencies would improve the consistency and comparability of data collected in the SFE. More routine split sampling events, longer term sensor comparison exercises, and further processing and analytical comparisons that control for individual filterers may also enhance comparability in the region.

Suggested citation:

Stumpner E.B., J. Yin, M. Heberger, J. Wu, A. Wong, and Saraceno, J.F. 2022. San Francisco Estuary Chlorophyll Sensor and Sample Analysis Intercomparison. SFEI Contribution Number 977. Richmond, California: San Francisco Estuary Institute.

Table of Contents

List of Tables

List of Figures

Acknowledgments

The authors gratefully acknowledge the funding for this study from the Delta Regional Monitoring Program (Delta RMP) and the San Francisco Bay Nutrient Management Strategy.

This effort required the in-kind contribution, collaboration, and support of multiple scientists from the Bureau of Reclamation, California Department of Water Resources, California Department of Fish and Wildlife, Moss Landing Marine Laboratory, Bend Genetics, Estuary & Ocean Science Center, Regional San, San Francisco Estuary Institute, and the U.S. Geological Survey. Special thanks to Brian Bergamaschi, Bryan Downing, Phil Trowbridge, and the Delta RMP Technical Advisory Committee and Nutrients Subcommittee for their review of the Phase II development process and report.

We also wish to acknowledge the following individuals, as either members of the Chlorophyll Workgroup or direct participants in the sensor comparison, sample processing and laboratory comparison, field and laboratory methods assessments, or report review: Selina Cole, Janis Cooke, Laura Benninger, Melanie Lowe, Erwin Van Nieuwenhuyse, Tyler Salman, Jared Frantzich, Mike Dempsey, Scott Waller, Rachel Pisor, Shaun Phillipart, Ted Swift, Brittany Griffith, Sharon Gosselink, Sydney Berrios, Crystal Sturgeon, Alex Etheridge, Tamara Kraus, Tim Otten and Bend Genetics staff, Erica Nejad, Tara Schraga, Michelle Arias, Jodi Bloom, Mark

Marvin DiPasquale, Taylor Winchell, Erika King, Ariella Chelsky, Dave Senn, Gisella Cluster, Autumn Bonnema, Toni Ignoffo, Anne Slaughter, Anke Mueller-Solger, and Kurt Carpenter.

Introduction

Patterns and scales of phytoplankton abundance and species composition vary across water bodies worldwide (Cloern and Jassby 2010). In the San Francisco Estuary (SFE) and elsewhere, chlorophyll is commonly used as an indicator of phytoplankton abundance and food availability to the lower food web and is especially critical as the base of the food web for pelagic fish in the SFE (Jassby et al. 2003). Monitoring chlorophyll over time can be used to identify trends, understand drivers, and inform how management actions affect the pelagic food web. Chlorophyll concentrations may be measured in the laboratory as discrete samples (Chl-*a* as µg/L) or as *in situ* fluorescence measurements (fCHL as µg/L, mg/m^3, or relative fluorescence units). Different settings of sensors used for *in situ* measurements and sample processing, extraction, and analysis of samples in the laboratory may introduce variability to chlorophyll measurements within and across water bodies.

Field and laboratory measurements of chlorophyll in the SFE are reported by multiple federal, state, and non-governmental agencies. To synthesize *in situ* field and laboratory data and identify patterns and trends throughout time and space, data collected by different groups need to be comparable. Some examples of chlorophyll data repositories include the Interagency Ecological Program Data and Metadata (IEP Survey Data (ca.gov)), the U.S. Geological Survey National Water Information System (U.S. Geological Survey 2021), and the California Water Data Library (Map (ca.gov)). Chlorophyll measurements may differ across seasonal and sub-regional scales and among reporting agencies because instrument and sensor types, deployment settings, particle interference, filtering techniques, analysis methods can affect the measurements. Such variables identified in the conceptual model in Figure 1, make synthesis and trend analysis throughout time and space challenging.

The Chlorophyll Intercomparison Study was initiated in 2017 to measure and document the variability of chlorophyll measurements with the goal of improving the accuracy, precision, and comparability of discrete and *in situ* chlorophyll data collected in the region. The Phase I

planning involved identifying various methods used among and within agencies. Monitoring agencies within the SFE agreed that reducing this variability was important, and a plan was developed for additional work. In 2018, the Delta Regional Monitoring Program and San Francisco Bay Nutrient Management Strategy funded Phase II of the Chlorophyll Intercomparison Study to help understand sources of this variability and determine procedures to improve the accuracy of and reduce the variability in fCHL data obtained from *in situ* sensors and from Chl-*a* laboratory measurements. The Phase II work presented here represents the collaborative effort of field groups and laboratories representing multiple agencies and programs.

This Phase II report presents results for several tasks assigned to assess the variability of chlorophyll measurements made by multiple agencies and programs in the region. The tasks included: (1) assessing methods of different monitoring programs; (2) performing sensor comparisons in the field; and (3) organizing a sample processing and analytical laboratory comparison. This report also includes a literature review of past *in situ* fCHL sensor comparisons and laboratory comparisons. Finally, we provide conclusions that resource managers can use when evaluating data or making decisions about future studies and monitoring activities to improve the measurement of chlorophyll across the SFE.

The Task 1 laboratory methods assessments are included in the Methods section below. The methods assessment of *in situ* field measurements was presented to the Delta Regional Monitoring Program (Delta RMP) as the Task 1 report: *Assessment methods used to measure in situ chlorophyll fluorescence by different monitoring programs in the San Francisco Estuary*. For publication purposes, the Task 1 report is included here in Appendix 1. Briefly, differing field methods include sensor settings, calibration procedures, deployment and retrieval protocols, quality-assurance, and post-processing. The largest variability in methods between agencies and groups therein includes reporting units, calibration procedures, and sensor servicing and cleaning.

The Task 2 sensor comparisons involved side-by-side site deployments of fCHL sensors at two deployment locations. This comparison allowed for a targeted assessment of operational differences associated with *in situ* fCHL measurements that may introduce variability in measured values. These differences include instrument and sensor settings, deployment protocols, calibration, sensor servicing and cleaning, data transmission, data post-processing, and data reporting.

For Task 3, we identified the combined variability that sample processing (e.g., filtering) and analytical methods may introduce to Chl-*a* results in the processing and laboratory

intercomparison exercise. The inclusion of individual laboratories' processing techniques in the laboratory methods assessment task was agreed upon by the Chlorophyll Workgroup because several field crews in the SFE do not complete their own Chl-*a* extraction and analysis.

Literature Review

Relevance to the San Francisco Estuary

Phytoplankton form the base of the pelagic food web in the SFE (Jassby et al. 2003). In particular, phytoplankton provide energy, carbon, and nutrients to zooplankton (Lehman 1992). Estuaries, such as the SFE, are important nursery habitats for juvenile fish and are home to other important species such as shellfish and shorebirds. The SFE is one of the most heavily modified estuaries in the world because of hydraulic mining, invasive species, changes in river flow and exports, loss of shallow wetland habitats, channelization and levee construction, and increased nutrient inputs (Nichols et al. 1986). These modifications have contributed to a 31-58% decline in Chl-*a* concentrations during the summer growing season from 1969-82 (Lehman 1992). Some of the potential causes for the decline in chlorophyll are grazing by invasive clams (Kimmerer and Thompson 2014), flow management and water residence time (Lucas et al. 2006, Monsen et al. 2007, Stumpner et al. 2020), decreases in seed stock from upstream sources (Parker et al. 2012), and increased nitrogen in the form of ammonium in some major tributaries (Dahm et al. 2016). Documenting this decline and obtaining real-time and accurate data throughout space and time are important for managers and decision makers to make informed recommendations.

Chlorophyll-*a* is the primary photosynthetic pigment among phytoplankton taxa and is used as a proxy for phytoplankton biomass (Latasa et al. 1996). Different photosynthetic pigments (Chl-*a*, Chl-*b*, phycocyanin, etc.) are used to identify the presence of different algal groups (Latasa et al. 1996). However, all photosynthetic organisms have Chl-*a* in common, leading to interest in its measurement. Many agencies report values in μg/L for the SFE, but collection and analytical methods are not consistent (e.g., solvent, use of grinding, analytical method). Ensuring these agencies are using comparable methods and obtaining similar results for the same parcel of water is important for understanding the ecosystem and for reporting on the status and trends of the lower food web in the SFE.

A 2017 study by the North Carolina Department of Water Resources provided unfiltered water sample splits to 17 different laboratories for analysis of Chl-*a* concentrations using their own standard operating procedures (North Carolina Department of Water Resources 2018). Laboratories had various protocols for analyzing these samples for Chl-*a* concentrations;

differences between the labs included filter type, volume filtered, storage time, and analytical method used. The study found inconsistencies with the quality of data and presented the data using mostly qualitative tables. There was no interpretation of best methods nor synthesis of the results. This was one of the only examples we found in the literature of an intercomparison study where multiple laboratories used a variety of methods —which in some cases were very similar — to obtain Chl-*a* values for the same water sample.

Methods for Measuring Chlorophyll

Chlorophyll concentrations can be measured using *in situ* sensors, collecting discrete water samples followed by laboratory analysis, or quantifying concentrations with remote sensing technology such as airborne or satellite spectral reflectance (Lorenzen and Jeffrey 1980, Lutz et al. 2006, Boyce et al. 2012). Each method offers strengths and weaknesses. Method continuity and intercomparability are important for comparing data and for the robustness of the several long-term monitoring projects in the SFE (Interagency Ecological Program Environmental Monitoring Program: IEP EMP, USGS Water Quality of San Francisco Bay; Triboli et al. 2003).

In situ fCHL can be measured with optical sensors in the field, which are commercially available tools that have been used in the SFE for decades. Sensors designed and manufactured by Yellow Springs Instruments (YSI), Turner Designs, WetLabs, and others are commonly used for continuous monitoring activities as well as short-term special studies. Manufacturers use the fluorescence response (e.g., excitation/emission pairs) of algal cultures to optimize sensor performance, and wavelength pairs may differ by manufacturer; the calibration of sensors may include cultures or rhodamine dyes. Sensor output may yield volts or millivolts, relative fluorescence units (RFU), or micrograms per liter (µg/L) that are typically calculated from fluorescence output. Measurements of 0-100 RFU range will typically line up with the 0-400 ug/L fCHL range, but the relationship is not always linear across various phytoplankton taxa with variable fluorescing pigments. These units of measurement represent different data streams (or channels) and should be calibrated individually. YSI suggests operators calibrate using one or both channels with their calibration solutions of choice (YSI, 2020); however, *in situ* measurements are compromised by changes in temperature and particle interference.

An assessment of the relationship between *in situ* fCHL and discrete Chl-*a* measurements as they relate to the phytoplankton community structure in the SFE was undertaken by Alpine and Cloern (1985), and more recently by Jassby et al. (2005). Alpine and Cloern (1985) found the *in situ* fluorescence per unit Chl-*a* response differed significantly between netplankton (>22 µm), nanoplankton (5–22 µm), and ultraplankton (<5 µm). Jassby et al. (2005) described the diel patterns observed in fCHL time series in the SFE resulting from diatom and cryptophyta

populations and also identified differences among discrete Chl-*a* concentrations along the same channel (San Joaquin River) due to these diel patterns and tidal forcings (Jassby et al. 2005).

Algal biomass may also be estimated through extraction of Chl-*a* from concentrated water samples and can be determined in the laboratory using fluorometric, spectrophotometric, or chromatographic methods. Fluorometry is recommended for freshwater systems with low concentrations of chlorophyll (but greater than 1 μg/L) or when pigment differentiation is not a concern (Berkman and Canova 2007). Spectrophotometry is recommended for freshwater systems with moderate to high concentrations of Chl-*a* greater than 1 mg/L (Berkman and Canova 2007). High performance liquid chromatography (HPLC) is the most precise method and is recommended for marine systems (Lutz et al. 2006). HPLC works by separating, identifying, and quantifying multiple phytoplankton pigments (Lutz et al. 2006) before analysis with a fluorometer or spectrophotometer.[1]

Latasa et al. (1996) conducted an intercalibration study among eight laboratories to determine the agreement between spectrophotometric and HPLC methods and found that the pigments showed more consistent results with spectrophotometric methods than with HPLC methods, even though HPLC has been shown to be more precise (Lutz et al. 2006). Drawbacks to the HPLC method are its high cost, complicated instrumentation, and long sample run time (~30 min for one sample compared to ~5 min for fluorometric methods).

The fluorometric method with acetone extraction is the most widely used (Holm-Hansen et al. 1965) because of its sensitivity and economy of time and materials.
Even if a laboratory uses the standard Holm-Hansen method, there can be variations with sample volume, sample storage and extraction solvents and procedures which can yield different results. Despite these differences, Axler and Owen (1994) found that over a wide range of concentrations variability could not be attributed to fluorescence and spectrophotometric methods or to storage and extraction techniques. In 1986, Australian researchers compared the performance of four methods for discrete Chl-*a* measurements based on a suite of marine samples (Murray et al. 1986). In their investigation, the HPLC method was determined as a 'convenient and accurate alternative to traditional methods for chlorophylls.'

[1] HPLC is often referred to as an analysis "method" for measuring Chl-*a*, but the high-performance liquid chromatography process itself just separates pigments. Analysis still requires a detector.

Previous Chlorophyll Intercalibration Studies

In 1980, Lorenzen and Jeffrey (1980) performed an intercalibration study to compare performance of laboratory fluorometric and spectrophotometric methods using research-grade pure pigments. They noted that sample origin is important when determining which method should be used. They concluded that samples collected in the surface of the euphotic zone are best analyzed using spectrophotometric or fluorometric methods. The additional Chl-*a* signal from the degradation products of senescent cells, detritus, and fecal pellets (e.g., phaeophytin-a) led to the introduction of the acidification technique so that these products do not contribute to the Chl-*a* values obtained from living phytoplankton (Lorenzen and Jeffrey 1980). The authors also recommended that *in situ* fluorometry should only be used to locate phytoplankton peaks for sampling but shouldn't be substituted for an accurate measurement of Chl-*a* concentration.

Over the past few decades, intercomparison or intercalibration exercises with field sensors and laboratory analyses have been done largely by oceanographers, with less focus in estuaries or freshwater. More recently, NASA and other space agencies have been major sponsors of chlorophyll intercalibration studies as they have launched satellites to reliably estimate phytoplankton biomass and primary productivity from orbit, based on ocean color (VanHeukelem 2002). To develop these algorithms, phytoplankton biomass must be ground-truthed. In 2006, a group of remote sensing specialists described the importance of *in situ* data (Lutz et al. 2006):

> "It is essential that the satellite coverage provided by the proposed network should be complemented by high quality and mutually-consistent *in situ* measurements of chlorophyll-a (chla), and if possible by a set of ancillary bio-optical measurements. Ideally these data could be used to establish regional remote-sensing algorithms for estimating chlorophyll-a from ocean-colour data."

Intercomparison studies are important for ensuring that measurements are comparable before and after a method changes, such as the advent of new technology or advanced scientific understanding. For example, over 20 years ago, a small intercalibration study helped scientists in the SFE understand the impact of changing one step in the laboratory method for Chl-*a* analysis (Triboli et al. 2003). This method was ultrasonication, which is the mechanical grinding and soaking of filters to aid extraction. In 1998, the IEP EMP program changed their method of chlorophyll extraction from the sonication method to a grinding method. Triboli et al. (2003) wanted to know whether there was a difference in the results from these methods which would necessitate a correction factor for new method data. To do this comparison, the authors collected replicated samples at four stations for one year and analyzed samples using both methods. Overall, the mean difference in methods was close to zero, and a data correction in

response to the 1998 method change was not needed. The results showed that the historical sonication method for Chl-*a* analysis was less precise than the current grinding method but was more sensitive for Chl-*a* concentrations below 10 µg/L.

Carstensen (2016) attempted to create an integrated dataset of Chl-*a* water grab measurements from monitoring programs in Norway, Denmark, and Sweden. This data synthesis was intended to assess differences in Chl-*a* values reported in different units to be able to combine datasets from the three national monitoring programs covering coastal water bodies. After analyzing data compiled by three different governments, the authors concluded that Chl-*a* is "a relatively robust proxy of phytoplankton biomass and relatively more precise than biovolumes/biomasses estimated from cell counts [which] frequently omit picoplankton specimens."

Sensor Intercalibration

In 2008, scientists from the University of the Pacific in Stockton, California, compared Chl-*a* values obtained through *in situ* fluorescence to values obtained through extraction and spectrophotometric determination (Burks et al. 2008). The authors demonstrated the difficulty and variability inherent in relating measured fluorescence to chlorophyll content. They concluded that some of the differences in the values obtained from the two different methods were due to the inability of the fluorometric method to differentiate between living Chl-*a* and phaeophytin. Another source of variability is "the consistent handling and analysis of highly unstable compounds such as Chl-*a*," which can occur from differences in sample handling and inconsistencies in analysis.

Methods

Field Sensor Intercomparison

For the 2018 field sensor comparison study, seven water-quality sondes with fCHL sensors were deployed at two continuous monitoring stations in the SFE: the San Joaquin River at Mossdale station, operated by the California Department of Water Resources, and the Liberty Island at Cache Slough (USGS station number 11455315; U.S. Geological Survey 2021), operated by the USGS (Figure 2). The sondes were deployed for approximately two weeks at each location; the period of deployment at the Mossdale location was August 16 –30, 2018, and from September 27 – October 10, 2018, at the Liberty Island location. Seven sondes, each owned by agencies making routine fCHL measurements in the system, were secured to a stainless-steel cage for side-by-side measurements. Sensors at both locations maintained an approximate 1-meter depth from the surface throughout the tidal cycle and were approximately 15 cm (6 inches) apart. Sensors

on the instruments measured water temperature, specific conductance, pH, dissolved oxygen, turbidity, fCHL, phycocyanin fluorescence, and dissolved organic matter fluorescence (fDOM). Five of the sondes were YSI model EXO v2 with Total Algae sensors (Xylem Inc, Rye Brook, NY) and two were previous generation YSI 6600 models with the 6025 fCHL sensor (hereafter 6-series sonde; Figure 3). The Total Algae sensor measures fCHL and phycocyanin (fPC, or PC) fluorescence simultaneously but the 6-series sensor does not have this capability. For the purpose of reporting, the sensor results are anonymous, and the agency is defined by a letter and the sensor type is retained in the results as -6S and -EXO2 (e.g., B-EXO2).

The deployment set-ups at the Mossdale and Liberty Island locations were similar but varied slightly (Figure 3). At Mossdale, the instrument cage was deployed on a taut-wire mooring approximately fifty feet upstream of the DWR operated continuous monitoring station. The instrument cage was attached to a ⅜-inch stainless-steel cable that attached to a surface marker and a concrete weight on the channel bottom. At the Liberty Island station, where currents are typically higher relative to Mossdale, the instrument cage was connected to the USGS continuous monitoring buoy in the middle of the channel with a stainless-steel cable and a six-foot aluminum pipe to maintain separation between buoys. For the Liberty Island deployment, a chain was attached at the bottom of the cage to maintain vertical placement in the water column.

Following the two-week deployment period, data performance verification to assess fouling and calibration drift of the sensors took place. Upon arrival at the station, a 'check sonde' measurement – a side by side comparison of an instrument with recently calibrated or checked sensors – was recorded to assess *in situ* conditions in the channel. Once the instrument cage was recovered, sondes were placed in a bucket of native water with a check sonde to account for sensor fouling over the deployment period (Figure 4). This exercise is generally referred to as a 'dirty bucket' test, and at this step, 'check sonde' measurements were recorded with individual instrument measurements. Following the 'dirty bucket' comparison, agencies opted to clean instruments in the field and re-deploy in buckets for the 'clean bucket' measurements or to clean instruments back in a laboratory setting.

Participants completed sensor calibration checks of their sondes either at the station or back at their respective offices or laboratories. Fouling and calibration drift corrections were completed in the data processing steps according to the quality-assurance and quality-control procedures of participating groups and agencies. These practices are outlined in the tables referenced in the Task 1 methods assessment report that is included as appendix 1 of this report.

Sample Collection, Processing, and Laboratory Methods Assessment

Participating agencies in the methods assessment include federal and state agencies, universities, and non-governmental organizations including the California Department of Water Resources' (DWR) Division of Environmental Services (DES) and Northern Central Regional Office (NCRO), California Department of Fish and Wildlife (CDFW), USGS, Reclamation, Moss Landing Marine Laboratory (MLML), Regional San, Bend Genetics, University of California at Davis (UC Davis), and the Estuary & Ocean Science Center at San Francisco State University. For this assessment, field crews and laboratories completed a survey to describe sample collection, processing, extraction, analysis, and standard curve techniques for Chl-*a* samples.

Whereas some laboratories collect, filter, and analyze their own samples, other laboratories receive samples (whole water or filters) from field crews. In an effort to capture filtering and laboratory methods across the SFE, 12 field-laboratory pairs, hereafter referred to as groups, were included in the laboratory intercomparison task. The agencies, field crews therein, and associated laboratories are identified in Table 1. Of the 12 groups, six groups typically complete their own analytical work. The study design used here was agreed upon by the Chlorophyll Workgroup and reflects current practices of many laboratories analyzing samples in the SFE – that is, certain field crews rely on a separate analytical laboratory. As identified in Table 1, nine analytical laboratories were surveyed under this task; four distinct field crew and laboratory pairs were affiliated with the USGS and three field crew-laboratory pairs were affiliated with the State of California (DWR and CDFW).

Sample collection

The 12 groups that were surveyed collect water using a variety of methods that include using a Van Dorn sampler, filling a bucket, or using a peristaltic or impeller pump (Table 2). Most groups collect water at a depth of 1 m. One group typically samples at a depth of 2 m and 1 m off the bottom(USGS-Cloern). A second group samples at a depth of approximately 10 cm (USGS-MMD). Groups collect volumes between 60 mL and <4 L. Bend Genetics is strictly a contract laboratory and receives whole water or filter samples from clients.

Sample processing

Most groups filter samples in a laboratory setting within a few hours of collection following water sample storage on wet ice, but three groups routinely filter samples immediately on board their vessel (Table 3, Table 4). One group routinely filters samples within 24 hours of collection. Samples are typically inverted or mildly shaken to homogenize the samples, but one group routinely uses a churn splitter. The nominal pore size of filters used by all groups differ

between 0.3 – 1.5 μm and diameters are either 25 or 47 mm. Sample volume is measured with graduated cylinders or volumetric flasks, and lighting conditions are typically dim. Most groups filter using a light vacuum at about 5 psi (30 kPa). The elapsed time required to filter a sample is typically < 15 minutes, with one group reporting < 45 minutes. A few of the participating groups preserve their filters with magnesium carbonate, but most groups immediately freeze filters in aluminum foil in a freezer at –20°C or –80°C. The maximum hold time of most groups is 28 days, but one group stores filters for up to 90 days.

Sample extraction and analysis

Of the nine participating laboratories, three use Standard Method 10200H, five routinely use either EPA method 445.0 or a modified version of it, and one laboratory uses a spectrophotometric method by Parsons et al. (1984). Of the three laboratories using Standard Method 10200H, one laboratory uses a UV/Vis spectrophotometer; the remaining two laboratories use a fluorometer. The laboratory following the Parsons et al. (1984) method also uses a UV/Vis spectrophotometer. All fluorometers used by participating laboratories are the Turner Trilogy model (Turner Designs, San Jose, CA). The laboratory that has routinely used the method described by Parsons et al. (1984) has started following the modified EPA 445.0 method since the start time of this comparison (Table 5). Most laboratories use acetone as an extraction solvent, with the exception of one that uses ethanol. The steeping time of filters in solvent following extraction is <24 hours for all laboratories, three of the laboratories grind filters after extraction, and one uses sonication. Most laboratories use hydrochloric acid (0.1 N) for acidification to determine phaeophytin concentrations.

Of the nine laboratories, all using fluorometric methods build standard curves for analysis (Table 6). Participants either purchase liquid or solid reference material from Sigma Aldrich or Turner Designs and create serial dilutions of 5 – 8 points that include the range of 0 to 210 μg/L. The reference materials for the standard stocks are certified on a UV/Vis spectrophotometer by either the manufacturer or the laboratory according to the method by Jeffrey and Humphrey (1975).

Laboratory Intercomparison Methods

Variability in laboratory measurements may stem from sample processing (filtering technique), the method of grinding, extraction, and analysis followed, and type of optical instrumentation. Laboratories that analyze samples in the SFE rely on the EPA 445.0 Fluorometric method (Arar and Collins 1997), the Standard Method 10200H (Baird et al. 2017), and the Parsons et al. (1984)

spectrophotometric method (Table 5). These methods differ in their use of equations to correct for pheophytin (Arar and Collins 1997).

To capture variability of sample processing and analytical methods across the SFE, 13 sample splits were distributed to 12 groups. Of the 12 groups, seven typically complete their own analytical work in their laboratory (Table 1). As a direct comparison of two Chl-*a* analysis methods, one group analyzed filters using Parsons et al. (1984) and EPA Method 445 (Arar 1997), while the other groups used one analytical technique.

Water samples were collected in May 2019 at the DWR DES Grizzly Bay station and in July, and August 2019 at the NCRO continuous monitoring stations on the Toe Drain at Lisbon Weir (Figure 2). Upon arrival at the station, station measurements and 'check sonde' measurements were recorded. Water was collected at 1 m depth with a 4-L Van Dorn sampler, according to the recommendations of industry standards (Baird et al. 2017). Two 20-L carboys were filled with the contents of the Van Dorn sampler and immediately placed on ice in coolers. Samples were returned to the laboratory at the DWR West Sacramento office within three hours of sample collection.

Upon return to the laboratory, but prior to churn splitting, individual 1-L amber high-density polyethylene (HDPE) sample bottles were labeled for each crew-laboratory combination and assigned a random fill number. The contents of the carboy were shaken, and the 15-L churn was rinsed three times prior to refilling the churn (Figure 5). Slow and consistent churning of the sample occurred for several minutes prior to rinsing the sample bottles and caps three times. Before samples were split, the carboy was further agitated, and the churn was filled again using the second carboy if necessary. Churning of the sample continued while filling the 1-L bottles according to the randomized fill number. Because we were volume limited (the spigot of a churn splitter is set above the base of the churn and the entire contents cannot be dispensed) most of the bottles were filled to approximately 750 mL. Bottles to be filtered outside of the DWR laboratory space were filled to a volume of 1 L. All HDPE bottles were placed on ice until their contents were filtered.

Briefly, the participating groups used filters with nominal pore sizes ranging from 0.3 - 1.5 µm. All participants used relatively low vacuum pressure of 5 psi. Three groups preserved filters with magnesium carbonate prior freezing, while most participants immediately froze filters without adding preservatives. Because of logistical constraints, filtering was not always completed by direct representatives from groups. When this happened, participating staff followed the filtering procedures outlined in the groups operating procedures (Table 3, Table 4). When all sample splits were dispensed from the churn, seven were filtered immediately in the

laboratory. The remaining six whole-water samples were picked up or delivered for filtering that took place within 24 hours of sample collection. For each water sample, filters were collected in triplicate by each field crew-laboratory group and the volume of water passed through the filter was recorded. Prior to filtering, laboratory lights were dimmed. Filter types varied amongst groups and are further identified in Table 3.

All laboratories received Chl-*a* filters within two days of sample collection. For the majority of laboratories, extraction and analysis occurred within the 28-day holding time recommended by the Standard Method and EPA Method 445.0. One participating laboratory (Group 12-J) was unable to extract and analyze samples until October 2019, almost 5 months after the first round of sample collection and filtration. For the purpose of this report, results from the Chl-*a* analysis are anonymous. Survey results and group assignments are identified in the methods assessment tables in the Results and Discussion section.

Nutrient and phytoplankton enumeration samples were also collected from the churn to capture ambient nutrient and algae conditions in Grizzly Bay and in the Toe Drain at Lisbon Weir. Nutrient samples were collected by the USGS and CDFW for analysis by the USGS National Water Quality Laboratory and the Bryte Laboratory. Nutrient samples were collected in 1-L bottles and filtered within 2 hours following the churn split. USGS nutrient samples were filtered with a 0.45 µm high-capacity flow through filter and peristaltic pump. Filtrate was collected in a 125-mL amber HDPE bottle and immediately placed on ice. DWR nutrient samples were passed through glass fiber filters with a nominal pore size of 0.45 µm into 125-mL amber HDPE bottles. Phytoplankton enumeration samples were split directly into 125-mL amber HDPE bottles, immediately preserved with 4 mL of Lugol's solution, and sent for analysis at BSA Environmental Services laboratory.

Nitrite (NO_2^-) and nitrate plus nitrite ($NO_3^- + NO_2^-$) as nitrogen were determined by colorimetric analysis measured on an automated segmented flow analyzer according to methods in Fishman (1993) and Patton and Kryskalla (2011). Ammonium as nitrogen (NH_4^+) and orthophosphate as phosphate (PO_4^{3-} ; also referred to as soluble reactive phosphate) was also determined by colorimetric analysis by measurement on an automated-segmented flow analyzer (Fishman 1993). Total dissolved nitrogen (TDN) was determined by alkaline persulfate digestion (Patton and Kryskalla 2011).

Phytoplankton enumeration was completed using a counting and identification method for microplankton and nanoplankton in accordance with the American Public Health Association (APHA) Standard Method 10200 (Baird et al. 2017). Phytoplankton were enumerated to the lowest possible taxonomic level using membrane-filtered slides, as described in McNabb (1960),

and counted using a Leica DMLB compound microscope. A minimum of 400 natural units (colonies, filaments, and unicells) or a minimum of 50 fields were counted from each sample to provide accuracy within 90 percent confidence limits. To ensure complete enumeration, filters were viewed under high magnification (usually 630X) and a lower magnification (usually 400X). Cell biovolume of all taxa were estimated using formulae for solid geometric shapes by Hillebrand et al. (1999). Biovolume calculations were based on measurements of 10 organisms per taxon for each sample. The mean biovolume within each size class was used to calculate the total biovolume contributed by the taxon to its representative sample from Burkholder and Wetzel (1989).

Statistical Analysis

For the *in situ* sensor intercomparison measurements at the Mossdale and Liberty Island locations, we created time series plots and performed analysis with R statistical software (R Core Team 2018) using the 'psych' (Revelle, 2021) and 'ggplot ' (Wickham, 2016) packages. This generated time-series plots of the normalized data and distance from mean (z-score), scatter plot functions, Pearson correlation coefficients, density curves, and cumulative density functions.

Analysis of the Chl-*a* laboratory intercomparison results used the Mandel's *h* and *k* statistics to identify outliers according to recommendations by the International Standards Organization (ISO 1994) and described in Luping and Schouenborg (2000), to detect differences between samples obtained from different laboratories. The *h* statistic measured inter-laboratory consistency (reproducibility) by comparing means, and the *k* statistic measured within-laboratory consistency (repeatability) by comparing variance of replicates for all laboratories in the given study (Addinsoft 2020). The Grubb's test was used to detect outliers by eliminating one outlier at a time (Grubbs 1969). The Cochran C test is another outlier test and was used to check the dataset for homogeneity of variances (Cochran 1941).

In an effort to test whether sample processing and laboratory analytical methods contribute to extracted Chl-*a* variance, we ran an analysis of variance (ANOVA), and post hoc analyses were conducted using Tukey's pairwise honest significance difference (HSD) or Student's t-tests using JMP software version 14.2 (SAS Institute, Inc. 2007). Data for dependent variables (filter nominal pore size, grinding, and analytical method) were assigned as categorical variables, and replicate measurements were assigned as a random effect. Effects were considered significant if *p*-values were <0.05.

Results and Discussion

Sensor Intercomparison

The sensor intercomparison study took place in summer 2018, during a period of relatively low algal productivity across most of the freshwater extent of the SFE (Bergamaschi et al.in press). Both the Mossdale and Liberty Island locations are considered the freshwater, tidal extent of the SFE and were chosen to represent areas of high and low productivity for this study, respectively. High algal productivity on the San Joaquin River near Mossdale is common during the summer months (https://cdec.water.ca.gov/jspplot/jspPlotServlet.jsp?sensor_no=10333&end=&geom=small&interval=30&cookies=cdec01) and was measured during the August 2018 deployment (25–80 µg/L). Algal productivity at the southern breach of Liberty Island and throughout most of the Cache Slough Complex remained low throughout the summer and early Fall of 2018 (fCHL < 2 µg/L; https://waterdata.usgs.gov/ca/nwis/uv?site_no=11455315).

For reporting purposes, the results of the sensor comparison are anonymous and the -EXO2 and -6S following the agency letter identify the sensor used in the field exercise. A greater range of fCHL were recorded by the seven sondes at the Mossdale location relative to the Liberty Island location. The fCHL range for a given timestamp was more than ± 20 µg/L at Mossdale and ± 1 µg/L at Liberty Island (Figures 6, Figure 7); discrete samples collected at the end of the deployment confirmed the *in situ* measurements (Table 7). The fCHL time series at both locations shows a typical diel signal common to local algal populations (Jassby et al. 2005). The Mossdale time series indicates peaks of chlorophyll near the end of daylight and the lowest concentrations near mid-morning. The Liberty Island time series identifies peaks prior to sunrise and lowest concentrations near midnight.

All Mossdale time series data tracked one another well and their z-scores, a normalized difference of all measurements from the mean, adjusted with variance. The highest z-score occurred during a period of high fCHL, measured as high as 80 µg/L by select sensors (Figure 8). All seven-time series have strong correlations amongst each other (Pearson r > 0.9; Figure 9). Assessment of the Mossdale times-series by box and whisker plots convey that instruments B_EXO2 and E_EXO2 had a similar range whereas instrument C_6S had the lowest median value and interquartile range (Figure 9). The density curve (Figure 10) and cumulative density frequency (Figure 11) show that three prevailing distributions exist within the data; B_EXO2 and E_EXO2 data follow the same distribution and all remaining instruments, except C-6S, appear to be similarly distributed. The similar distributions seen in B_EXO2 and E_EXO2 were unaffected by differences in the meters' calibrations, because these meters were operated by different groups using different calibration protocols (Table A4).

Throughout the Liberty Island deployment, there are strong correlations (Pearson's r > 0.8) among five EXO2 datasets, with the exception of sonde F_EXO2, but correlations with data from two 6-series sondes are poor (Figure 12). Box plots of fCHL reveal many spikes, or outliers, in the time series (Figure 13). The 6-series sonde measurements contain several spikes, but the same was true for raw EXO v2 data prior to outlier deletion and data processing steps. Outlier deletion is defined by individual groups in their post-processing steps in Table 8A in Appendix 1; briefly, most groups delete data that are above a specified threshold or falls outside of 2 – 3 standard deviations for a specified period. The density curve and CDF show that none of the six datasets follows the same distribution (Figure 14 and Figure 15), and the different distributions appear to contribute to the discrepancy in measured data.

Negative values can occur in fCHL time series and may be an artifact of a poor or unnecessary sensor calibration. A poor calibration may occur as a result of a dirty sensor or compromised standard solution at the time of calibrations. Here we identify an unnecessary calibration as a calibration made when the sensor is performing within control limits (±0.5 turbidity unit or ±5%) of the measured value, whichever is greater (Wagner et al. 2006). Both occurrences may result in negative values (YSI 2019). Negative values were measured by sonde D_EXO2 during much of the Liberty time series.

One of the goals of the sensor comparison exercise was to find out if and to what extent different methods for deployment, calibration, and data collection affect the comparability of the fCHL datasets generated by different groups in the SFE. The assessment of methods in the Chlorophyll Task 1 report (Appendix 1) and time-series results of the sensor comparison here strongly suggest that historical fCHL data commonly collected by the older model YSI 6-series sonde will require careful review of instrument and sensor metadata prior to synthesis with modern fCHL data collected by EXO v2 series sondes. For the Mossdale deployment, the sensors on the EXO v2 generally recorded higher fCHL (~ 40 µg/L) than the 6-series sonde (~ 22 µg/L); the lowest concentrations in the time series were observed with the 6-series sonde "C-6S." During the Liberty Island deployment, the 6-series sonde data represented wider fCHL variability than EXO v2 data. Instrument A_6S had a number of spikes that are likely outliers; the two highest observations (> 100 µg/L) were removed from analysis. The likely cause of outliers in the time series may be floating vascular plants observed near the sensors both during deployment and retrieval at the Liberty Island station. That 6-series sondes appear to report lower fCHL compared to the EXO v2 is useful to know when reviewing past datasets and synthesizing historical data.

The C_6S measured considerably lower concentrations relative to other instruments in the Mossdale deployment (Figure 6; Figure 10; Figure 11), but it is worth noting that the EXO2 data

distribution it most closely matched (C_EXO2) was operated by the same continuous monitoring group. Lower fCHL results, relative to other EXO2 time series, may be related to calibration practices and/or post collection data treatment, or sensor design. Differences in sensor design such as wavelength, filtering, and signal processing may affect data comparability (Foster et al., in press). For example, differences in the optical configuration, including excitation and/or emission wavelength, could result in different sensitivities to the same algal pigment or interferences such as dissolved organic matter or sediment. Differences in on-board data processing, including signal filtering such as smoothing, could also impact data comparability, especially in terms of data variance and "noisiness" and calibration drift.

For the purpose of this report, we present the time series in units of μg/L because all groups either calibrate sensors with fCHL values in μg/L or rely on factory calibration to this reporting unit. Given the survey results in Appendix 1, a few of the groups calibrate and report measurements in relative fluorescence units (RFU). The equation of the best linear fit between RFU and μg/L varies by dataset and may be site specific because of the presence of phytoplankton taxa. Errors associated with the conversion (calibration) between RFU and μg/L are not responsible for discrepancies in concentration data because sensors were all calibrated to μg/L for this exercise. However, the algorithms developed for the total algae sensor and 6-series sensors differ in which algal cultures they are based on - that is *Chlorella* and *Isochrysis* spp., respectively (Diego Davis and Melanie Poon, YSI, personal communication June 15, 2020).

The methods assessment identified that many groups (4 of the 7; Table A3 in Appendix 1) rely on the manufacturer to complete a factory calibration and calibrate sensors to a high fCHL standard. Although manufacturer calibrations may occur at an interval similar to groups completing their own calibration, using a check standard more frequently would provide a quality control step and help identify inaccurate or malfunctioning sondes. Given our reliance on these data to define high-biomass events, a standard operating procedure for the use of rhodamine or algal cultures to regularly check instruments would improve data fidelity and comparability.

Laboratory Intercomparison Assessment

The study consisted of 12 field-laboratory groups. Each group filtered and analyzed triplicate samples from three different collection events. We aimed to collect samples during low, medium, and high algal productivity conditions to capture variability across all conditions. Additionally, we factored staff availability and field conditions into consideration when determining sample collection date and locations.

In general, fCHL throughout the region in the summer of 2019 were relatively low (<5 µg/L) at many continuous monitoring stations (USGS station numbers 11455385, 11455508, 11455478, 11455095, 11455142, 11447650, 380631122032201, 11337190, 11455146, 11455315, 382996121401601, 11336790, 11455139, 11455140, 11447890, 11455143; U.S. Geological Survey, 2021).

The Chlorophyll Workgroup reviewed telemetered data to determine favorable conditions for sample collection. The first sample collected at Grizzly Bay on May 7, 2019, was chosen because fCHL concentrations were between 5 - 15 µg/L in the week leading up to the sample event; at the time of sample collection, the Grizzly station sonde recorded 9.9 µg/L (Table 8). The second and third sampling events, on July 9, 2019, and August 20, 2019, took place at the California Department of Water Resources (DWR) - NCRO station at Lisbon Weir - a location that typically has higher Chl-*a* concentrations relative to other locations in the region and is more easily accessible. The fCHL concentrations recorded at Lisbon Weir station were 7.8 and 19.9 µg/L at the time of the second and third sample collection, respectively (Table 8).

Based on the range of observations (Table 9), concentrations of fCHL were low - moderate for the May and July sampling events (7.1 ± 2.1 and 5.9 ± 2.7 µg/L, respectively), and moderate - high in August (20.9 ± 7.1 µg/L) (Table 10). Nutrient samples were also collected, and concentrations are reported in Table 8. These ancillary data are presented in this report for completeness; we have not performed any analyses using these data. The first sample event represented the highest ambient measured turbidities and dissolved organic matter concentrations, which may interfere with optical detection and compromise the fCHL measurement measurements (Foster et al., in press). Turbidity and dissolved organic matter interferences with optical sensors have previously been studied (Pellerin et al. 2009, Downing et al. 2012, Pellerin et al. 2013, Saraceno et al. 2017) and are current research topics in the SFE.

We followed the guidance of ISO 5725-1 that identifies appropriate statistical tests for laboratory consistency in intercomparisons (ISO 1994). Mandel's *k* statistics were used to detect differences among variances (repeatability) for entities included in the laboratory intercomparison. Mandel's k-statistics were calculated using the replicate filters across all sampling events and are presented graphically (Figure 17). Standard deviations are presented in Figure 18. Mandel's *h* statistics were used to detect differences among means (reproducibility) for entities included in the laboratory intercomparison (Figure 19). Summary statistics and estimates for repeatability standard deviations and reproducibility standard deviations were calculated both with all samples included (Table 10), as well as with outliers identified from entities 1-A and 7-D excluded (Table 11). Outliers were identified using the Cochran C test (Cochran 1941) and the Grubbs test (Grubbs 1969).

The Mandel's *k* and *h* statistics results show entities 1-A and 7-D were outliers because of their lack of repeatability and reproducibility, respectively. Entity 1-A had high within-laboratory variability (low repeatability) for all sampling events (Table 9, Figure 17, Figure 18).

The comparison of mean values for all events (including outliers) is shown in Figure 20. At the time of sample collection, field readings for the check sonde and instrument at the continuous monitoring station were recorded (Table 8). For all three events, the entity 7-D showed poor comparability to other entities (Figure 19), even though the within laboratory repeatability was good (Figure 17).

For events 1 and 2, entity 10-G had the second highest repeatability standard deviation (Figure 17), with a high Mandel's *k*-statistic (~5% significance level), indicating weakness within laboratory repeatability. Entity 9-F only submitted results for events 1 and 3 and their results showed excellent repeatability across all events relative to other pairs (Figure 17). Entities 12-J and 13-K had the lowest Mandel's *h*-statistic (the among-laboratory compatibility), meaning these entities reported values closest to the inter-laboratory mean values for each of the sampling events (Figure 19). Figure 20 shows the means for each replicate for each sampling event; entity 7-D has a noticeably higher mean result across all events.

To further evaluate Chl-*a* laboratory results, an ANOVA model tested whether the categorical variables of filter type, grinding, and/or analytical method introduce variance to Chl-*a* results (equation 1).

$$Chl-a \sim filter\ size + grinding + analytical\ method + filter\ size * grinding + filter\ size * analytical\ method + grinding * analytical\ method + filter\ size * grinding * analytical\ method \quad (equation\ 1)$$

Groups 1-A and 7-D that were identified as outliers were not included in the ANOVA models. For all pooled data, results did not indicate that Chl-*a* results differed significantly based on these three categorical variables. Both model runs using all Chl-*a* results and Chl-*a* result by event lost degrees of freedom when we attempted to evaluate interactions Table 12). Differences in Chl-*a* concentrations were significant when data were split up by sampling event. We discuss the implications of the interactions following a discussion of use of different nominal pore sizes, the grinding method, and types of analytical methods.

In the SFE, like all systems, phytoplankton sizes vary, and limitations exist with filtering techniques to assess biomass. Limitations include damage to large cells, cell geometry such that elongate cells pass through smaller pores, and filter clogging that retains smaller particles (Brewin et al. 2014). The laboratory methods assessment survey determined field-laboratory groups in the region use a variety of filter types with nominal pore sizes of 0.3, 0.7, 1.0, 1.2, and

1.5 µm (Table 3), but the ANOVA results do not suggest filter type alone is a significant factor of variance when all data are pooled together (Table 12). However, significant differences in Chl-*a* concentrations were observed for each of the three factors when data were split up by event, with events 1 and 2 both having significant differences ($p < 0.05$) across nominal pore size (Table 12.

The Tukey pairwise (HSD) results for Chl-*a* concentrations based on nominal pore size indicate that filtering with the smallest pore size (0.3 µm) resulted in significantly different Chl-*a* concentrations compared to Chl-*a* concentrations measured following filtering with 0.7- and 1.0-µm pore sizes for event 1. Using the 1.0-µm filter resulted in significantly different chlorophyll concentrations compared to concentrations measured using other filter sizes in Event 2, and the small sample size (n = 3) could create a Type 1 error in our data (Table 12 Figure 21). The highest range in Chl-*a* concentrations was observed in the 0.7 µm pore size (n = 33; Figure 21*A*) and could also be attributed to different personnel completing filtering across all events for entities 6-C, 8-E, 11-H, and 13-K. We discuss the error introduced by individual filterers below.

Grinding was a less common practice in the extraction step among the laboratories that participated in this study. Of the two laboratories using grinding, one produced analytical data that were considered outliers, and therefore data from that laboratory were not included in the ANOVA. Chlorophyll-*a* concentrations measured by the other laboratory that used grinding were not significantly different from concentrations measured by other laboratories that did not use grinding when data from all events were pooled together (Table 12, Figure 22). A previous study in the SFE (Triboli et al. 2003) found that Chl-*a* concentrations did not differ significantly based on the use of grinding of samples with concentration > 10 ug/L.

Eight of the 12 groups use fluorescence techniques (Table 5) and the remaining groups use spectrophotometric techniques, but ANOVA results did not show significant differences in Chl-*a* concentrations using pooled data or by event. Measured ranges of Chl-*a* concentrations in the SFE were similar among analytical methods (Figure 23A). More samples were analyzed using the fluorescence EPA Method 445.0 (n = 60) compared to other methods ($9 \leq n \leq 18$). Samples analyzed using the fluorescence EPA Method 445.0 (n = 60) had a smaller interquartile range and wider overall range compared to samples analyzed using other methods (Figure 23A), but the tall upper whisker on the boxplot for samples analyzed using EPA Method 445.0 indicate that the observed range was influenced by a small number of high measured concentrations (Figure 23A).

In event 2, concentrations obtained using spectrophotometric *Standard Methods* (n = 6; Baird et al. 2017) were significantly different from results obtained using fluorometric methods ($3 \leq n \leq$

18) but were not significantly different from spectrophotometric-based concentrations (n = 3) obtained using methods by Parsons et al. (1984). However, the small sample sizes may have contributed to the significant differences, and the differences may not be meaningful in a practical or environmental context (Figure 23B). The results of 10-G and 11-H represent data that are a direct comparison of analysis on a spectrophotometer and fluorometer, respectively, by the same group—with the exception of the representative completing the filtering. Controlling for filterer is an important random effect as one laboratory with two trained filterers can introduce 3% to the relative percent difference (Tara Schraga, written communication, December 2020).

We designed our sample processing and laboratory comparison exercises with the goal of controlling for filter nominal pore size, grinding method, and analytical approach, but our design did not control for a variable that is subject to human error - the individual that is filtering any particular sample.

Standard deviations of replicates within each event were less than 1.0 for all groups except event 2 for 10-G and all events for 1-A (Table 9). Similarly, only event 2 for 10-G and all events for 1-A exceeded the 1 and 5% significance levels, respectively, for Mandel's k statistics, which is used to determine outliers based on within-laboratory comparisons (repeatability; Figure 17) The greater coefficient of variance among the groups (Reproducibility – COV (Sr) in Table 11) is evidence that sample handling prior to analysis to the analytical steps may be a likely reason for variation. Lastly, we observed that groups consistently reported lower or higher Chl-*a* values relative to each other and this effect may be explained by the laboratory's instrument calibration curve. The calibration information of participating laboratories is captured in Table 7, but the baseline of any individual instrument as well as all the steps leading up to analysis could contribute to this artifact in the dataset (Figure 20).

Of the standard operating procedure shared by laboratories for the methods assessment survey, entity 8-E has a rigorous training and filtering protocol that could be adopted by any participants that do not have such procedures documented. Having all groups and laboratories work together to identify a common set of data quality objectives and indicators would improve abilities to directly compare current Chl-*a* sample collection and analysis efforts. Redesigning experiments to control for "filterer" would also improve abilities to directly compare current Chl-*a* sample collection and analysis efforts.

Lastly, Chl-*a* concentrations are a bulk measurement that may not inform on the quality of algae present for the pelagic food web — that is, whether taxa are deemed beneficial or harmful. As such, we collected phytoplankton enumeration samples at the three events and determined that

cyanobacteria, specifically *Eucapsis*, were present (Table 13). The highest diversity of algae was present in the May sampling event in Grizzly Bay, where diatoms comprised 43% of the sample biovolume and cyanobacteria comprised 12%. The lowest diversity of algae was present in the Toe Drain during the August sampling event, where the highest temperature of 23.3°C was recorded, diatoms were not present in the sample, and cyanobacteria comprised 88% of the biovolume.

Conclusions

The tasks of the Phase II study include (1) assessing field and laboratory methods of different monitoring programs; (2) performing sensor comparisons in the field; and (3) organizing a sample processing and analytical laboratory comparison. The results presented here represent the first comprehensive effort to document variability in fCHL and Chl-*a* monitoring across the SFE. The discussion above reflects the first attempt by the Chlorophyll Workgroup to link current practices to observed chlorophyll variance in the field and laboratory. The conclusions that follow can be used by the Chlorophyll Workgroup to formalize longer-term collaboration involving standardization of equipment and staff training to monitor chlorophyll and evaluate collected chlorophyll data throughout the SFE. Below we describe targeted studies that would provide the data needed to develop uniform protocols for monitoring chlorophyll throughout the SFE. The Chlorophyll Workgroup and the intercomparison exercises were a collaborative effort that required funding from the Delta Regional Monitoring Program, San Francisco Bay Nutrient Management Strategy, and in-kind support by participating agencies.

In situ chlorophyll fluorescence sensors

- As described in an intercalibration study in Australia several decades ago, *in situ* fluorometry is best utilized in locating phytoplankton biomass peaks and is not a substitute for accurate measurements (Lorenzen and Jeffrey 1980). However, as demonstrated by an *in situ* nitrate study on the San Joaquin River, high-frequency measurements can remove time series aliasing that may occur by infrequent grab sampling (Pellerin et al. 2009).

- Both models of fCHL sensors in the sonde comparison exercises were generally well-correlated and tracked one another, but distributions of the time-series data varied between sensor models and among participants. Spikes in data (e.g., noise), particularly at lower concentrations, increased the relative percent difference between two of the sensor types. At the Mossdale deployment location, the older 6-series sensor measured

a lower concentration of algae relative to the Total Algae Sensors on the EXO v2. In addition, a higher mean and greater range of fCHL was measured at the Mossdale deployment location relative to the Liberty Island location. Longer-term datasets collected by the 6-series and total algae sensors could be used by scientists and resource managers to compare data collected by the sondes across seasonal conditions and develop correction factors between the sensor types to make historical data more directly comparable.

- The calculation of fCHL by *in situ* sensors depends on a number of factors, including the manufacturer's algorithm development for the optical response of a particular algal class (Gregor and Maršálek 2004). The presence of inorganic and organic particles present at the sampling location is an environmental condition that may change across seasons and water years. A more rigorous sensor deployment exercise may inform agencies as to which sensors are most robust for the SFE.
- Most fCHL sensors can report data in RFU and in μg/L, but the precision of the measurement is based on the presence of algal classes in the water column, environmental conditions therein, and the calibration of the sensor using a primary or secondary standard. All groups identified in this intercomparison study use either factory calibration or rhodamine dye (which is considered a secondary standard) to calibrate to a high fCHL value. Algal-based cultures are considered primary standards (Berkman and Canova 2007), and an algal-based culture that is specific to an autochthonous algal class may standardize fCHL measurements across the SFE. Calibrating and reporting fCHL measurements in the same units would improve data comparability and data synthesis efforts. Additional fCHL data collection over longer periods at a common fixed continuous monitoring station would provide larger datasets for more rigorous statistical analysis.
- Another aspect of standardizing calibration methods is determining the type (e.g., two-point) and frequency of calibrations of instruments used to continuously measure chlorophyll in the SFE. Similar calibration solutions would further improve data comparability; the manufacturer could complete factory calibration for the high fCHL standard, or rhodamine dyes or algal cultures could be used as primary calibration standards.
- A uniform sampling period at a 15-minute timestamp would improve comparability of collected data. Setting all sondes to either 'normal' or 'burst' mode would standardize

median filtering of data, reduce variability across time series, and improve the comparability and synthesis of collected data.

- Avoiding calibration 'errors' or 'over-calibrations' at zero concentration that yield negative fCHL concentrations at low concentrations (which occurred during the Liberty Island deployment in this study) would further improve the comparability and accuracy of collected chlorophyll data.

- Development and adherence to a uniform set of detailed standard operating procedures (SOPs) by all agencies likely would improve the comparability of chlorophyll data collected in the SFE. A cooperative and comprehensive training of personnel from all participating agencies likely would improve the consistency of collection and processing of chlorophyll samples.

Sample processing prior to chlorophyll-*a* extraction and analysis

- Filtering is a processing step that may occur outside the jurisdiction of the analytical laboratory. Filtering can introduce variability to samples and has been shown to contribute to differences as high as 3% even between trained individuals (Tara Schraga, written communication, December 2019). Whereas many laboratories that routinely collect samples have rigorous operating procedures for filtering, many participating entities have staff members that collect and filter samples but do not complete the analysis. Irrespective of which entity completes which step, a rigorous training program for staff involved in filtering from all groups likely would improve repeatability within and among laboratories. In their SOP, entity 8E has set strict standards to ensure repeatability and reproducibility before an individual is allowed to filter independently. Routine comparisons between individuals of the same group could also be another quality assurance step added to current protocols.

- The effects of not adequately shaking a sample bottle were observed in the results reported by entity 1-A. The processing and laboratory methods assessment revealed that all other groups typically homogenize a sample bottle prior to filtration, but entity 1-A did not adequately homogenize their bottle in between pours during this comparison. Typically, an entity would collect three 1-L bottles for replicate samples, but for this exercise, the volume distributed to individual groups was limited to 1 L because we were constrained by the volume of the churn. Mimicking conditions that most groups routinely sample and filter would improve the comparability of chlorophyll data collected in the future.

- Holding samples in cold and dark conditions for up to 24 hours may not contribute to variability as previously considered, but the effects of holding times and conditions likely depend on the type of algae present in any given sample (Tim Otten, oral communication, April 2020). Most entities filtered within a few hours of collection, but entity 4-B filtered the day following sample collection. Entity 4-B uses the same filter and analytical laboratory as entity 5-B, but significant differences were not found between the sets of filters from these two entities.

Chlorophyll-*a* laboratory analysis

- Spectrophotometric vs. fluorometric methods did not stand out as a cause of variability, and both methods seem reasonable for conditions in the SFE. There was a significant difference between the SM10200H spectrophotometric methods and other analytical methods for samples in event 2. Previously, fluorometry has been identified as being better suited for lower concentrations in freshwater, but our ANOVA results do not suggest significant differences in analytical methods - specifically the spectrophotometric and fluorescence methods.
- A study design with more statistical power that controls for the individual filterer would improve our abilities to directly compare methods used to obtain chlorophyll data. In addition, rigorous training standards for individual field crews and laboratories likely would improve the repeatability and reproducibility among filterers.
- Participating groups used filters of nominal pore size from 0.3 - 1.5 μm, and significant differences in reproducibility were noted across individual sample events. Selecting a standard pore size (GF/F) for use by all filterers likely would reduce variability in measured Chl- *a* data. GF/F filters – originally produced by Whatman™ - have a nominal pore size of 0.7 μm; however, they are also considerably more expensive than other filters. Research groups such as universities may be reluctant to embrace a more expensive standard without a demonstrated benefit. GF/F filters also involve more glass material than GF/C (1.2 μm) or Pall A/E (1.0 μm) filters, which may cause challenges for laboratories that grind and filter. Studies by Saldanha-Corrêa (2004) and by Hillebrand et al. (1999) found little difference between retention rates of filter types typically used in this size range.

Additional Conclusions by the Chlorophyll Workgroup

- A long-term 'experimental' station where multiple algal fluorescence sensors can be deployed may improve understanding of what fluorescence sensors are measuring in the SFE. A long-term, experimental station could be used as a standard against which to compare new technologies that are developed to improve the accuracy of chlorophyll data measured *in situ*. Accurate, long-term fCHL data can be used to understand changes in the pelagic food web in the SFE.

- Routine sample processing and interlaboratory comparisons likely would improve comparability of collected chlorophyll data but would require significant commitment by agencies. Targeted field and sample processing and laboratory exercises like those used in this study provide quantitative results that managers can use to make science-based decisions for data synthesis efforts and to make changes in monitoring activities that inform ecological decisions. Providing participating entities with an algal culture for a standardized serial dilution and analysis likely could be used to further address the variability of chlorophyll results produced by each participating entity.

- The effort described in this report captured the widest variability across select groups and laboratories in the region. The Chlorophyll Workgroup decided against sending bottles of water to individual laboratories because the workgroup did not agree that sample design would capture the variance in results reported within our region. Instead, the workgroup chose to capture the combined error of filtering and analysis within the exercises described in this report. As mentioned previously, several groups sample for Chl-*a* and complete their own filtering prior to delivering or shipping filters to their selected analytical laboratory. The current dataset could not be used to determine how much the random effect variable of "filterer" may have on measured chlorophyll concentrations.

  Assessments of the relationship between fCHL and discrete Chl-*a* measurements as they relate to the phytoplankton community structure, as were conducted by Alpine and Cloern (1985) and more recently by Jassby et al. (2005), would provide additional information about chlorophyll in the SFE and variability in measured chlorophyll concentrations that was not covered by the study described in this report.

References

Addinsoft. 2020. XLSTAT: Mandel's h and k statistics for outliers. https://www.xlstat.com/en/solutions/features/mantel-statistics-for-outliers.

Alpine, Andrea E., and James E. Cloern. 1985. "Differences in in Vivo Fluorescence Yield between Three Phytoplankton Size Classes." *Journal of Plankton Research* 7 (3): 381–390. https://doi.org/10.1093/plankt/7.3.381.

Arar, E.J. and Collins, G.B. 1997. *Method 445.0: In vitro determination of chlorophyll a and pheophytin a in marine and freshwater algae by fluorescence* (pp. 1-22). Cincinnati: United States Environmental Protection Agency, Office of Research and Development, National Exposure Research Laboratory.. https://cfpub.epa.gov/si/si_public_record_report.cfm?dirEntryId=309415.

Axler, Richard P., and Christopher J. Owen. 1994. "Measuring Chlorophyll and Phaeophytin: Whom Should You Believe?" *Lake and Reservoir Management* 8 (2): 143–51. https://doi.org/10.1080/07438149409354466.

Baird, R.B., Eaton, A.D. and Rice, E.W., Eds. 2017. Standard Methods for the Examination of Water and Wastewater. 23rd Edition, American Public Health Association, American Water Works Association, Water Environment Federation, Washington D.C.

Bergamaschi, B.A., T.E.C. Kraus, B.D. Downing, E.B. Stumpner, K. O'Donnell, J. Hansen, J. Soto Perez, A. Hansen, and A. Gelber, in press. "Assessing Spatial Variability of Nutrients, Phytoplankton and Related Water Quality Constituents in the California Sacramento-San Joaquin Delta at the Landscape Scale: 2018 High Resolution Mapping Surveys." Sacramento: U.S. Geological Survey.

Berkman, Julie A. Hambrook, and Michael G. Canova. 2007. "7.4 Algal Biomass Indicators." In *U.S. Geological Survey Techniques of Water-Resources Investigations, Book 9*. U.S. Geological Survey. https://water.usgs.gov/owq/FieldManual/Chapter7/7.4.pdf.

Boyce, Daniel G., Marlon Lewis, and Boris Worm. 2012. "Integrating Global Chlorophyll Data from 1890 to 2010." *Limnology and Oceanography: Methods* 10 (11): 840–52. https://doi.org/10.4319/lom.2012.10.840.

Brewin, Robert J. W., Shubha Sathyendranath, Priscila K. Lange, and Gavin Tilstone. 2014. "Comparison of Two Methods to Derive the Size-Structure of Natural Populations of Phytoplankton." *Deep Sea Research Part I: Oceanographic Research Papers* 85 (January): 72–79. https://doi.org/10.1016/j.dsr.2013.11.007.

Burkholder, JoAtnn M., and Robert G. Wetzel. 1989. "Microbial Colonization on Natural and Artificial Macrophytes in a Phosphorus-Limited, Hardwater Lake." *Journal of Phycology* 25 (1): 55–65. https://doi.org/10.1111/j.0022-3646.1989.00055.x.

Burks, Remie, Jeremy Hanlon, Sharon Borglin, and William Stringfellow. 2008. "Appendix E: Correlation of Standard Methods Chlorophyll Quantification and In-Vivo Chlorophyll Fluorescence Measurements." In *San Joaquin River Up-Stream DO TMDL Project, Task 4: Monitoring Study, Final Task Report*. Stockton, CA: University of the Pacific. https://www.waterboards.ca.gov/rwqcb5//water_issues/tmdl/central_valley_projects/san_joaquin_oxygen/required_studies/task4_appendices_a_v/task4_appxe.pdf.

Carstensen, Jacob. 2016. "Intercalibration of Chlorophyll a between Denmark, Norway and Sweden: Western Baltic (BC6), Kattegat (NEA8b) and Skagerrak (NEA8a, NEA9 and NEA10)." Technical Report from DCE – Danish Centre for Environment and Energy No. 76. Denmark: Aarhus University. http://dce2.au.dk/pub/TR76.pdf.

Cloern, James E., and Alan D. Jassby. 2010. "Patterns and Scales of Phytoplankton Variability in Estuarine–Coastal Ecosystems." *Estuaries and Coasts* 33 (2): 230–41. https://doi.org/10.1007/s12237-009-9195-3.

Cochran, W. G. 1941. "The Distribution of the Largest of a Set of Estimated Variances as a Fraction of Their Total." *Annals of Eugenics* 11 (1): 47–52. https://doi.org/10.1111/j.1469-1809.1941.tb02271.x.

Dahm, Clifford N., Alexander E. Parker, Anne E. Adelson, Mairgareth A. Christman, and Brian A. Bergamaschi. 2016. "Nutrient Dynamics of the Delta: Effects on Primary Producers." *San Francisco Estuary and Watershed Science* 14 (4). http://escholarship.org/uc/item/1789c0mz.

Downing, Bryan D., Brian A. Pellerin, Brian A. Bergamaschi, John Franco Saraceno, and Tamara EC Kraus. 2012. "Seeing the Light: The Effects of Particles, Dissolved Materials, and Temperature on *in Situ* Measurements of DOM Fluorescence in Rivers and Streams." *Limnology and Oceanography: Methods* 10 (10): 767–775.

Fishman, Marvin J. 1993. *Methods of Analysis by the US Geological Survey National Water Quality Laboratory: Determination of Inorganic and Organic Constituents in Water and Fluvial Sediments*. US Department of the Interior, US Geological Survey.

Foster, Guy. M., Jennifer L. Graham, Brian A. Bergamaschi, Kurt D. Carpenter, Bryan D. Downing, Brian A. Pellerin, Stewart A. Rounds, and JohnFranco Saraceno, in press. "*Field Techniques for the Determination of Algal Fluorescence in Environmental Waters: Principles and Guidelines for Instrument and Sensor Selection, Operation, Quality Assurance, and Data Reporting."* U.S. Geological Survey

Gregor, J., and Maršálek, B. 2004. Freshwater phytoplankton quantification by Chlorophyll-a: a comparative study of in vitro, in vivo, and *in situ* methods: Water Research, v. 38, p. 517–522, at https://doi.org/10.1016/j.watres.2003.10.033.

Grubbs, Frank E. 1969. "Procedures for Detecting Outlying Observations in Samples." *Technometrics* 11 (1): 1–21. https://doi.org/10.1080/00401706.1969.10490657.

Hillebrand, Helmut, Claus-Dieter Dürselen, David Kirschtel, Utsa Pollingher, and Tamar Zohary. 1999. "Biovolume Calculation for Pelagic and Benthic Microalgae." *Journal of Phycology* 35 (2): 403–24. https://doi.org/10.1046/j.1529-8817.1999.3520403.x.

Holm-Hansen, Osmund, Carl J. Lorenzen, Robert W. Holmes, and John DH Strickland. 1965. "Fluorometric Determination of Chlorophyll." *ICES Journal of Marine Science* 30 (1): 3–15.

ISO. 1994. *ISO 5725-1:1994, Accuracy (Trueness and Precision) of Measurement Methods and Results — Part 1: General Principles and Definitions*. International Standards Organization. https://www.iso.org/obp/ui/#iso:std:iso:5725:-1:ed-1:v1:en.

Jassby, A., J. Cloern, A. Mueller-Solger, and others. 2003. "Phytoplankton Fuels Delta Food Web." *California Agriculture* 57 (4): 104–109. https://doi.org/10.3733/ca.v057n04p104.

Jassby, A. D., A. B. Müller-Solger, and M. Vayssières. 2005. "Short-Term Variability of Chlorophyll and Implications for Sampling Frequency in the San Joaquin River." *IEP*

Newsletter 18 (1): 28.

Jeffrey, S.W., and G.F. Humphrey. 1975. "New Spectrophotometric Equations for Determining Chlorophylls a, b, C1 and C2 in Higher Plants, Algae and Natural Phytoplankton." *Biochemie Und Physiologie Der Pflanzen* 167 (2): 191–94. https://doi.org/10.1016/S0015-3796(17)30778-3.

Kimmerer, Wim J., and Janet K. Thompson. 2014. "Phytoplankton Growth Balanced by Clam and Zooplankton Grazing and Net Transport into the Low-Salinity Zone of the San Francisco Estuary." *Estuaries and Coasts* 37 (5): 1202–18. https://doi.org/10.1007/s12237-013-9753-6.

Latasa, Mikel, Robert R. Bidigare, Michael E. Ondrusek, and Mahlon C. Kennicutt. 1996. "HPLC Analysis of Algal Pigments: A Comparison Exercise among Laboratories and Recommendations for Improved Analytical Performance." *Marine Chemistry* 51 (4): 315–324. https://doi.org/10.1016/0304-4203(95)00056-9.

Lehman, Peggy W. 1992. "Environmental Factors Associated with Long-Term Changes in Chlorophyll Concentration in the Sacramento-San Joaquin Delta and Suisun Bay, California." *Estuaries* 15 (3): 335. https://doi.org/10.2307/1352781.

Lorenzen, C. J., and S. W. Jeffrey. 1980. "Determination of Chlorophyll in Seawater: Report of Intercalibration Tests." *UNESCO Technical Papers in Marine Science* 35 (1): 1–20.

Lucas, Lisa V., Deanna M. Sereno, Jon R. Burau, Tara S. Schraga, Cary B. Lopez, Mark T. Stacey, Konstantin V. Parchevsky, and Vladimir P. Parchevsky. 2006. "Intradaily Variability of Water Quality in a Shallow Tidal Lagoon: Mechanisms and Implications." *Estuaries and Coasts* 29 (5): 711–30. https://doi.org/10.1007/BF02786523.

Luping, Tang, and Björn Schouenborg. 2000. "Methodology of Inter-Comparison Tests and Statistical Analysis of Test Results." SP REPORT 2000:35. Borås, Sweden: Swedish National Testing and Research Institute. http://www.diva-portal.org/smash/get/diva2:962159/FULLTEXT01.pdf.

Lutz, V. A., R. Barlow, G. Tilstone, S. Sathyendranath, T. Platt, and N. Hardman-Mountford. 2006. "Report of the *In Situ* Component of the 'Plymouth Chlorophyll Meeting and Workshops (Extended Antares Network).'" Plymouth, UK. http://antares.ws/files/intercal-report.pdf.

McNabb, Clarence D. 1960. "Enumeration of Freshwater Phytoplankton Concentrated on the Membrane Filter." *Limnology and Oceanography* 5 (1): 57–61. https://doi.org/10.4319/lo.1960.5.1.0057.

Monsen, Nancy E., James E. Cloern, and Jon R. Burau. 2007. "Effects of Flow Diversions on Water and Habitat Quality: Examples from California's Highly Manipulated Sacramento–San Joaquin Delta." *San Francisco Estuary and Watershed Science* 5 (3). https://doi.org/10.15447/sfews.2007v5iss5art2.

Murray, A.P, C.F Gibbs, A.R Longmore, and D.J Flett. 1986. "Determination of Chlorophyll in Marine Waters: Intercomparison of a Rapid HPLC Method with Full HPLC, Spectrophotometric and Fluorometric Methods." *Marine Chemistry* 19 (3): 211–27. https://doi.org/10.1016/0304-4203(86)90024-1.

Nichols, F. H., J. E. Cloern, S. N. Luoma, and D. H. Peterson. 1986. "The Modification of an Estuary." *Science* 231 (4738): 567–73. https://doi.org/10.1126/science.231.4738.567.

North Carolina Department of Water Resources. 2018. 2018 NC DWR Chlorophyll a Round-Robin. Retrieved from: https://files.nc.gov/ncdeq/Water%20Quality/Chemistry%20Lab/Certification/ChlRR%20Reports/2018-CRR-Report--09-27-2018-FINAL.pdf

Parker, Alexander E., Richard C. Dugdale, and Frances P. Wilkerson. 2012. "Elevated Ammonium Concentrations from Wastewater Discharge Depress Primary Productivity in the Sacramento River and the Northern San Francisco Estuary." *Marine Pollution Bulletin* 64 (3): 574–86. https://doi.org/10.1016/j.marpolbul.2011.12.016.

Parsons, Timothy R., Yoshiaki Maita, and Carol M. Lalli. 1984. "Determination of Chlorophylls and Total Carotenoids: Spectrophotometric Method." In *A Manual of Chemical and Biological Methods for Seawater Analysis*, 101–112. Oxford: Pergamon Press.

Patton, Charles J., and Jennifer R. Kryskalla. 2011. "Colorimetric Determination of Nitrate plus Nitrite in Water by Enzymatic Reduction, Automated Discrete Analyzer Methods." *US Geological Survey Techniques and Methods* 34.

Pellerin, Brian A., B.A. Bergamaschi, B.D. Downing, J.F. Saraceno, J.D. Garrett, and L.D. Olsen. 2013. "Optical Techniques for the Determination of Nitrate in Environmental Waters: Guidelines for Instrument Selection, Operation, Deployment, Maintenance, Quality Assurance, and Data Reporting." Techniques and Methods 1–D5. U. S. Geological Survey. https://pubs.usgs.gov/tm/01/d5/.

Pellerin, Brian A., Bryan D. Downing, Carol Kendall, Randy A. Dahlgren, Tamara E. C. Kraus, Johnfranco Saraceno, Robert G. M. Spencer, and Brian A. Bergamaschi. 2009. "Assessing the Sources and Magnitude of Diurnal Nitrate Variability in the San Joaquin River (California) with an *in Situ* Optical Nitrate Sensor and Dual Nitrate Isotopes." *Freshwater Biology* 54 (2): 376–87. https://doi.org/10.1111/j.1365-2427.2008.02111.x.

R Core Team. 2018. *R: A Language and Environment for Statistical Computing* (version Version 3.5. 2). Vienna, Austria. https://www.R-project.org/.

Revelle W. 2021. psych: Procedures for Psychological, Psychometric, and Personality Research. Northwestern University, Evanston, Illinois. R package version 2.1.9, https://CRAN.R-project.org/package=psych.Saldanha-Corrêa, Flávia M. P., Sônia M. F. Gianesella, and José Juan Barrera-Alba. 2004. "A Comparison of the Retention Capability among Three Different Glass-Fiber Filters Used for Chlorophyll-a Determinations." *Brazilian Journal of Oceanography* 52 (3–4): 243–47. https://doi.org/10.1590/S1679-87592004000300007.

Saldanha-Corrêa, Flávia M. P., Sônia M. F. Gianesella, and José Juan Barrera-Alba. 2004. "A Comparison of the Retention Capability among Three Different Glass-Fiber Filters Used for Chlorophyll-a Determinations." *Brazilian Journal of Oceanography* 52 (3–4): 243–47. https://doi.org/10.1590/S1679-87592004000300007.

Saraceno, John Franco, James B. Shanley, Bryan D. Downing, and Brian A. Pellerin. 2017. "Clearing the Waters: Evaluating the Need for Site-Specific Field Fluorescence Corrections Based on Turbidity Measurements." *Limnology and Oceanography: Methods* 15 (4): 408–416.

SAS Institute, Inc. 2007. *JMP, Version 12.0*. Cary, NC.

Stumpner, Paul R., Jon R. Burau, and Alexander L. Forrest. 2020. "A Lagrangian-to-Eulerian Metric to Identify Estuarine Pelagic Habitats." *Estuaries and Coasts*, November.

https://doi.org/10.1007/s12237-020-00861-7.

Triboli, Kitty, Anke Mueller-Solger, and Marc Vayssières. 2003. "The Grind about Sonicated Chlorophyll (or: Did a Method Change in 1998 Affect EMP Chlorophyll Results?)." *IEP Newsletter* 16 (4): 13–24.

U.S. Geological Survey. 2021. USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed November 2, 2021, at https://doi.org/10.5066/F7P55KJN.

VanHeukelem, Laurie. 2002. "Sources Of Variability In Chlorophyll Analysis By Fluorometry And High Performance Liquid Chromatography In a SIMBIOS Inter-Calibration Exercise." NASA/TM-2002-02338-0. National Aeronautics and Space Administration.

Wagner, Richard J., Robert W. Boulger Jr., Carolyn J. Oblinger, and Brett A. Smith. 2006. "Guidelines and Standard Procedures for Continuous Water-Quality Monitors: Station Operation, Record Computation, and Data Reporting." Techniques and Methods 1–D3. Reston, VA: U.S. Geological Survey. https://pubs.usgs.gov/tm/2006/tm1D3/pdf/TM1D3.pdf.

Wickham, Hadley. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, https://CRAN.R-project.org/package=ggplot.

YSI. 2019. Turbidity Units and Calibration Solutions: Technical Instructions YSIT627-01https://www.ysi.com/File%20Library/Documents/Technical%20Notes/T627_Turbidity_Units_and_Calibration_Solutions.pdf

YSI, 2020. EXO User Manual. https://www.ysi.com/file%20library/documents/manuals/exo-user-manual-web.pdf

Figures

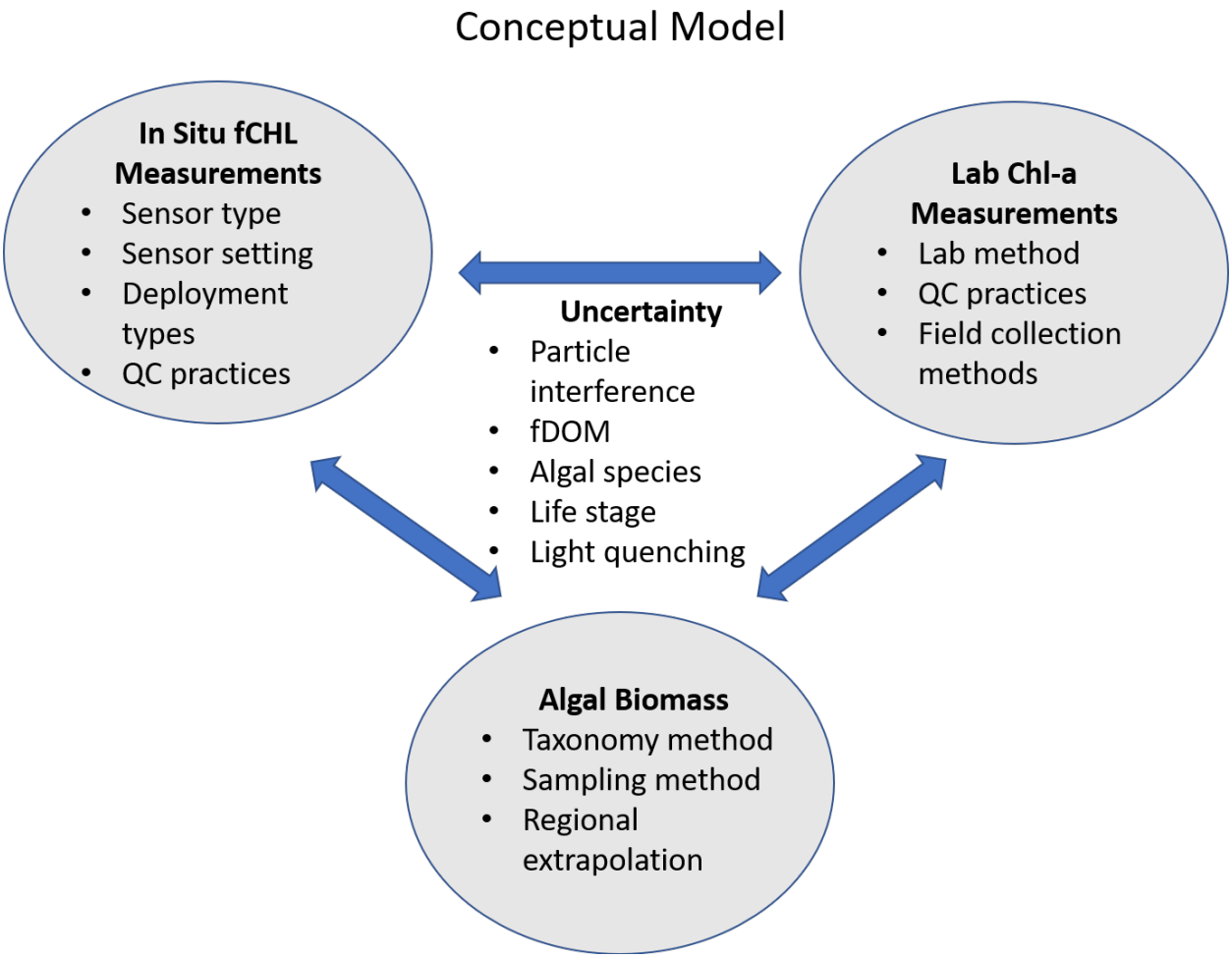


Figure 1. Diagram describing likely causes of variance in laboratory extracted chlorophyll-a (Chl-*a*), *in situ* chlorophyll fluorescence, algal biomass determined by microscopy, and the relationship between these related parameters.[fCHL, chlorophyll fluorescence; QC, quality control; fDOM, dissolved organic matter fluorescence]

Figure 2. Location of *in situ* sensor intercomparison deployments in the Sacramento-San Joaquin Delta (green triangle, field) and laboratory intercomparison sample collection (yellow triangle, lab). The Mossdale and Grizzly locations are the Department of Water Resources-Division of Environmental Services continuous monitoring stations on the San Joaquin River (http://cdec.water.ca.gov/dynamicapp/QueryF?s=MSD) and Grizzly Bay (https://cdec.water.ca.gov/dynamicapp/staMeta?station_id=GZL). The Liberty Island location is the U.S. Geological Survey's continuous monitoring station adjacent to Cache Slough https://waterdata.usgs.gov/ca/nwis/uv?site_no=11455315). The Lisbon Weir location is the Departments of Water Resources-North Central Regional Office's continuous monitoring station on the Toe Drain (https://cdec.water.ca.gov/dynamicapp/staMeta?station_id=LIS). Google Maps basemap.

Figure 3. Photos of water-quality instrument deployments at the Mossdale at San Joaquin River (left) and Liberty Island at Cache Slough (right) stations. Sensors were deployed at the same water depth relative to the surface at both locations (~1 m). Instruments were attached to a stainless-steel cage for both two week deployments. Left photo shows Mike Dempsey (DWR) with the instrument cage is attached to both a concrete weight and surface marker by a ⅜-inch stainless-steel cable that creates the taut-wire mooring. Right photo is the instrument cage connected to the continuous monitoring buoy with a stainless steel cable and a six-foot aluminum pipe to maintain separation between buoys.

Figure 4. Bucket test of sondes following sensor intercomparison deployment.

Figure 5. Image of a 15-liter churn splitter used during the laboratory comparison exercise (Source of photograph: https://shop.sciencefirst.com/wildco/general-laboratory-equipment/6178-churn-sample-spliter-polyethylene-14l.html)

Figure 6. Time series of chlorophyll fluorescence measured by 7 different sondes at the Mossdale location between August 16 - August 30, 2018 (top panel). Zoomed in view of the time series over 4 days between August 17 - August 20, 2018 (bottom panel). Vertical lines represent midnight.

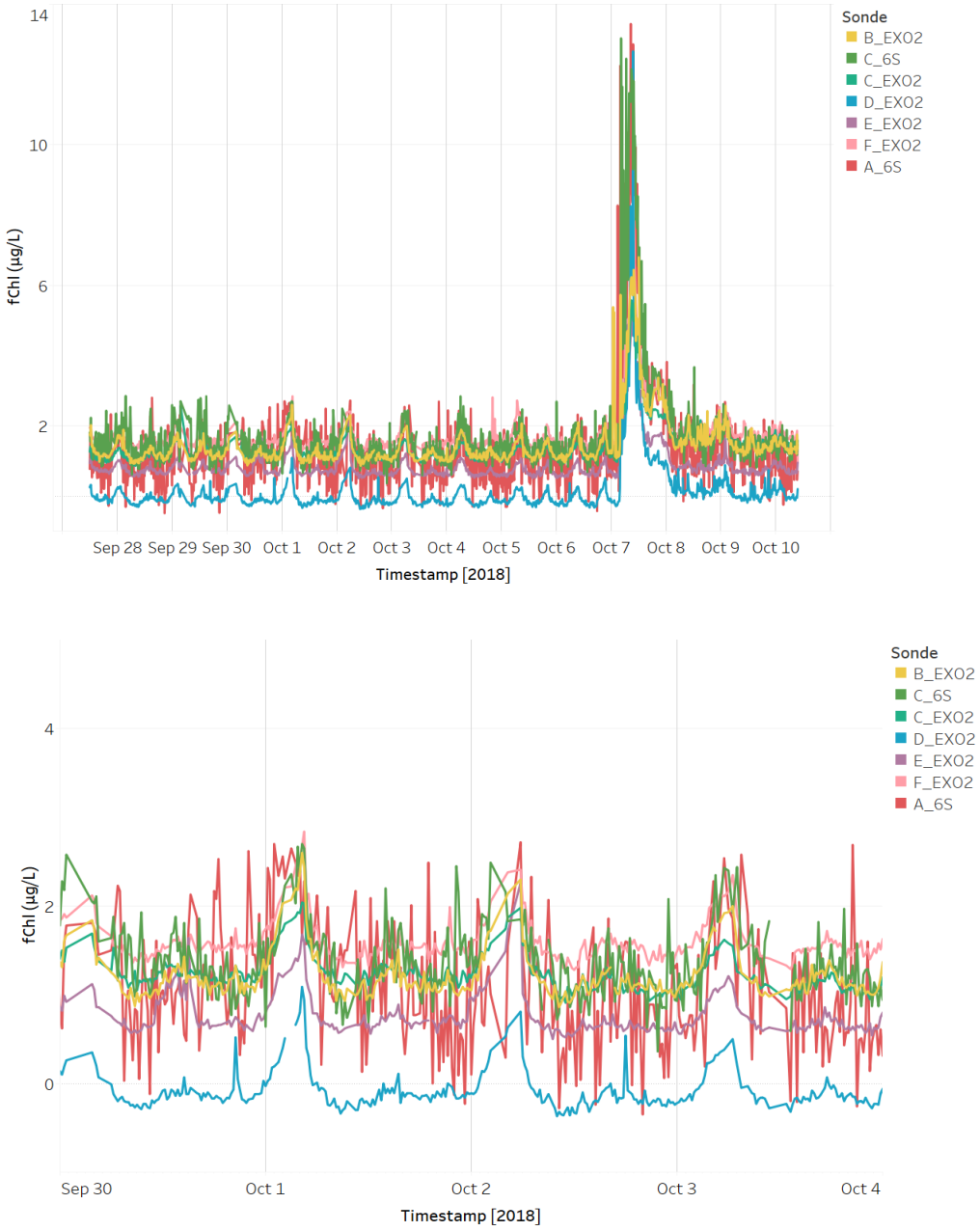


Figure 7. Time series of chlorophyll fluorescence measured by 7 different sondes at the Liberty Island location between September 27 - October 9, 2018 (top panel). Zoomed in view of the time series over 4 days between September 30 - October 3, 2018 (bottom panel). Peak in the top panel represents a high wind event. Vertical lines represent midnight.

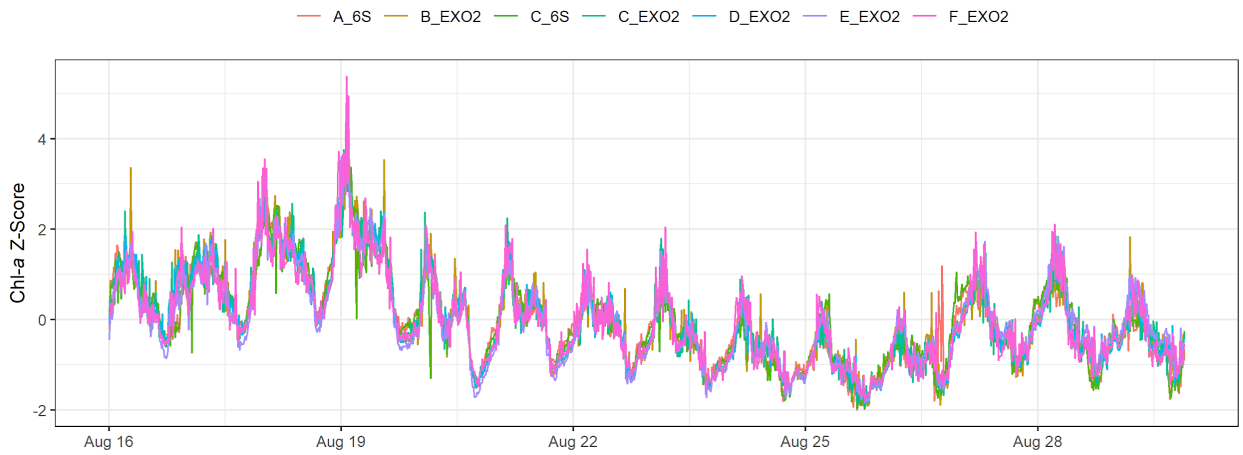
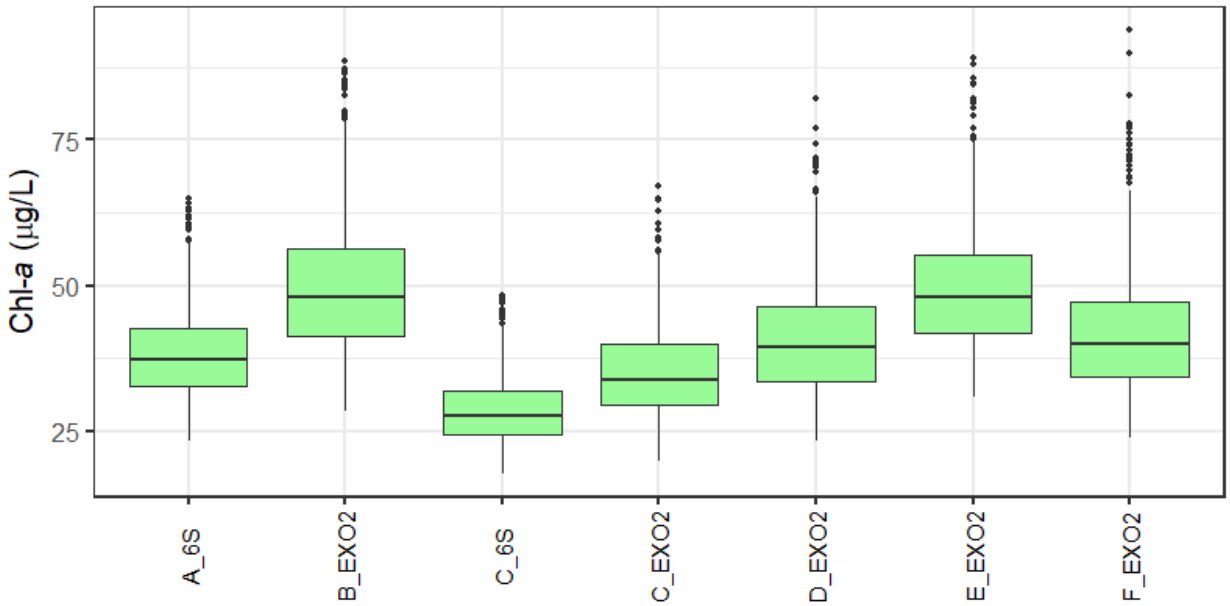
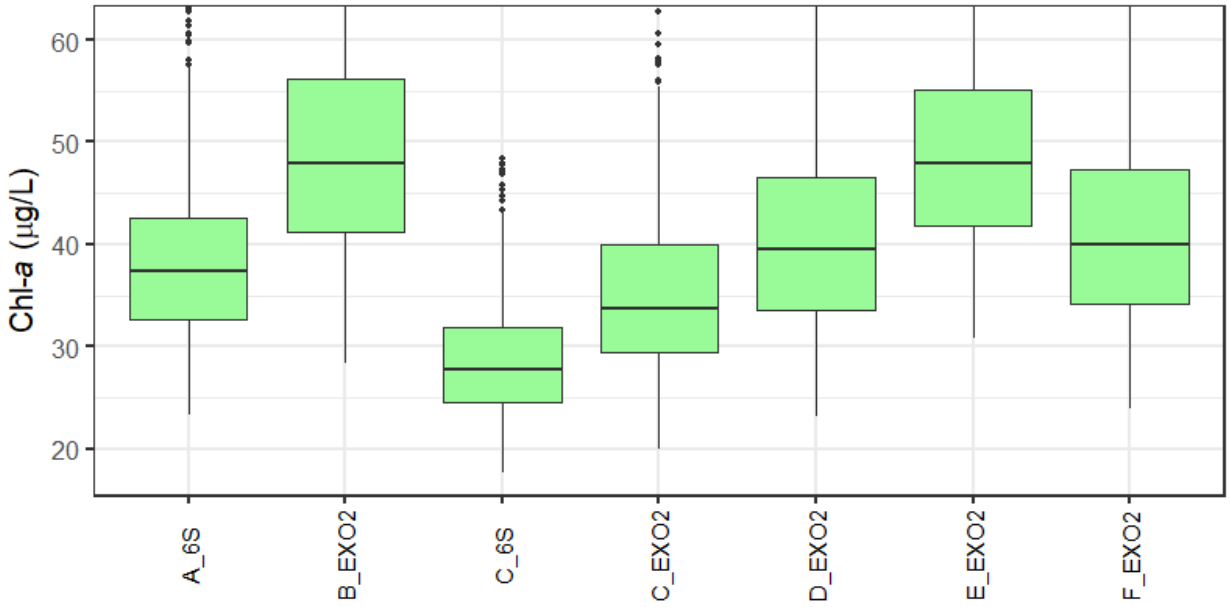
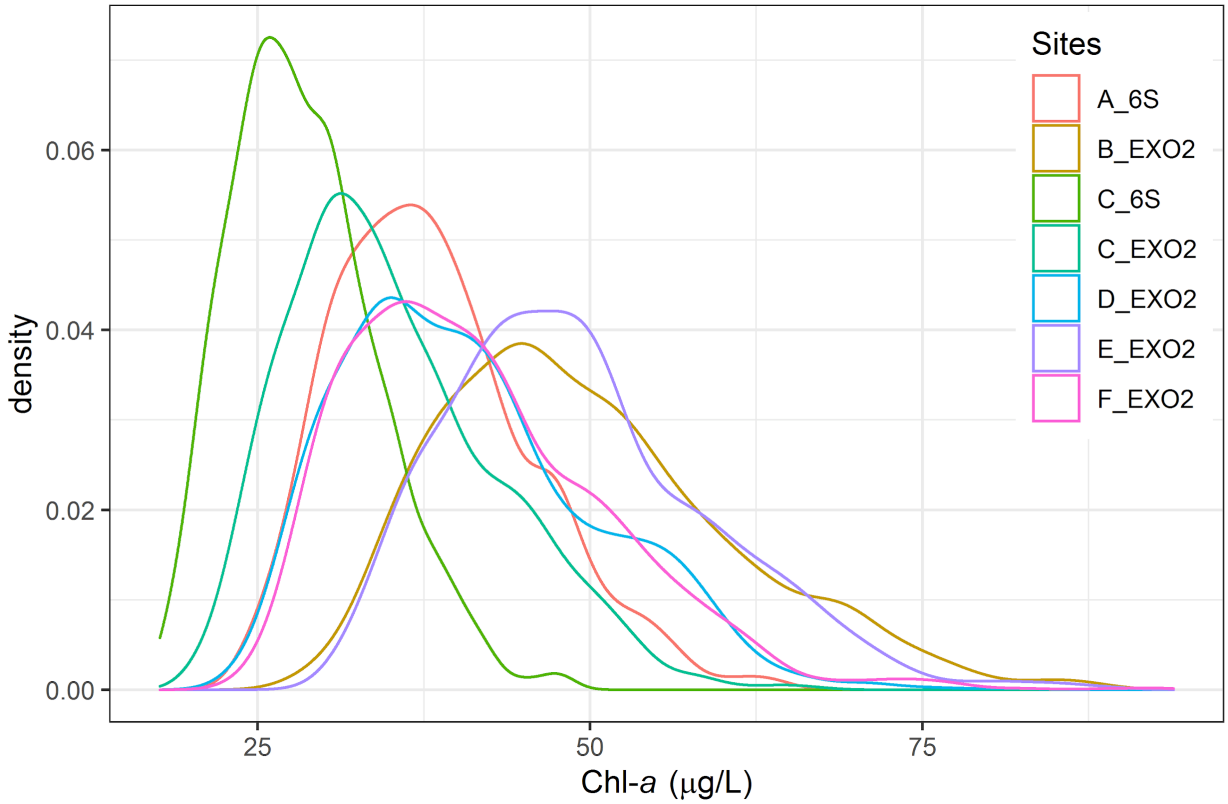


Figure 8. Time series of z-scores (normalized, distance from mean) for seven sondes at the Mossdale location between August 16 - August 30, 2018 (top panel). Zoomed in view of the time series between August 18 - August 20, 2018 (bottom panel).

Figure 9. Scatter plot matrix, histogram, and Pearson correlation coefficients of seven chlorophyll fluorescence measurements at the Mossdale location between August 16 - August 30, 2018.




Chl-a (µg/L)
60
50
40
30
20
A_6S
B_EXO2
C_6S
C_EXO2
D_EXO2
E_EXO2
F_EXO2


Sites
A_6S
B_EXO2
C_6S
C_EXO2
D_EXO2
E_EXO2
F_EXO2
density
0.06
0.04
0.02
0.00
25
50
75
Chl-a (µg/L)

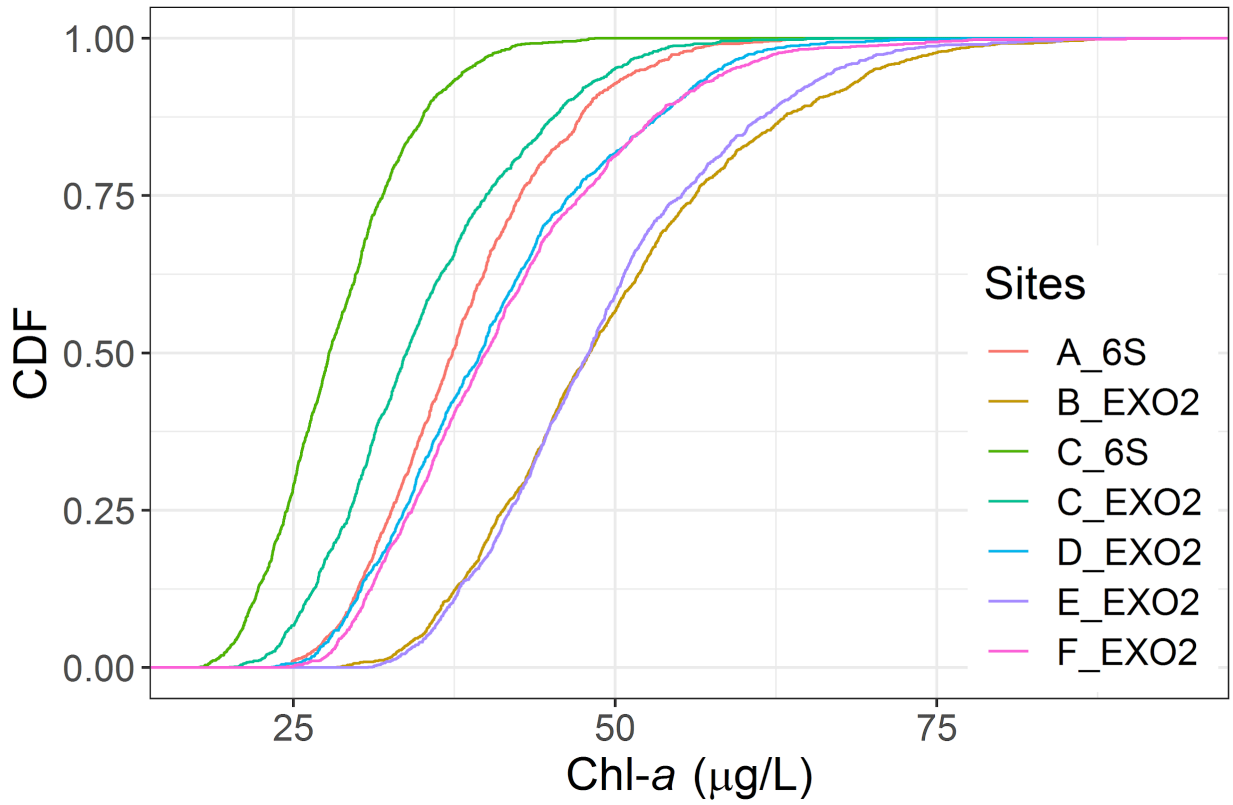

1.00
0.75
0.50
0.25
0.00
CDF
25
50
75
Chl-a (µg/L)
Sites
A_6S
B_EXO2
C_6S
C_EXO2
D_EXO2
E_EXO2
F_EXO2

Figure 13. Scatter plot matrix, histogram, and Pearson correlation coefficients of chlorophyll fluorescence measured by 7 different sondes at the Liberty Island location between September 27 - October 10, 2018.

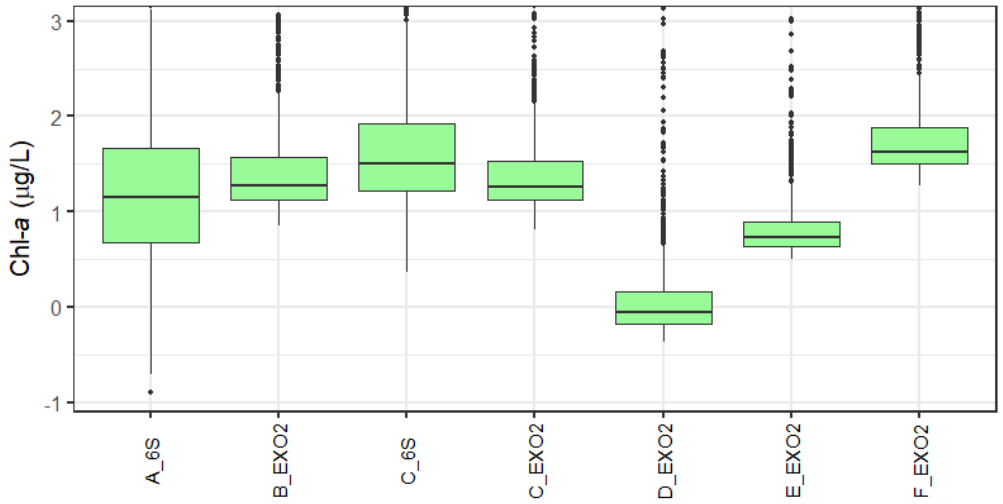


Figure 14. Box and whisker plots of chlorophyll fluorescence measured by 7 different instruments at Liberty Island from September 30 - October 2, 2018.

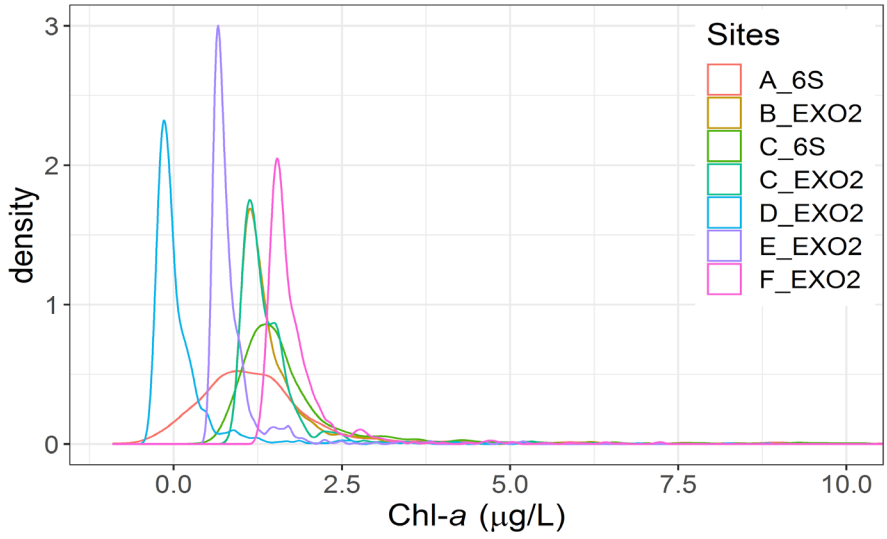
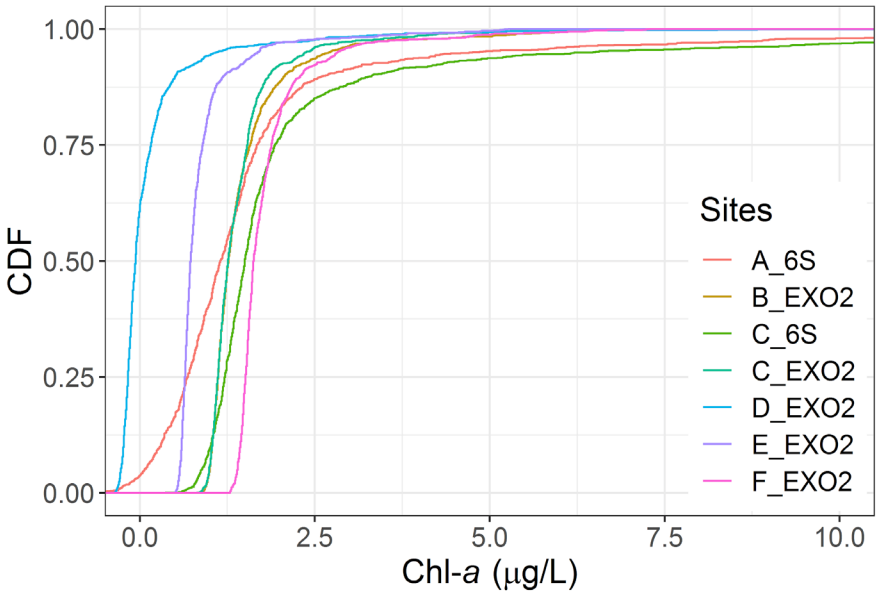


Figure 15. Empirical probability density functions of chlorophyll fluorescence concentrations over the Liberty Island time series for the period between September 30 - October 2, 2018, showing distribution of data.



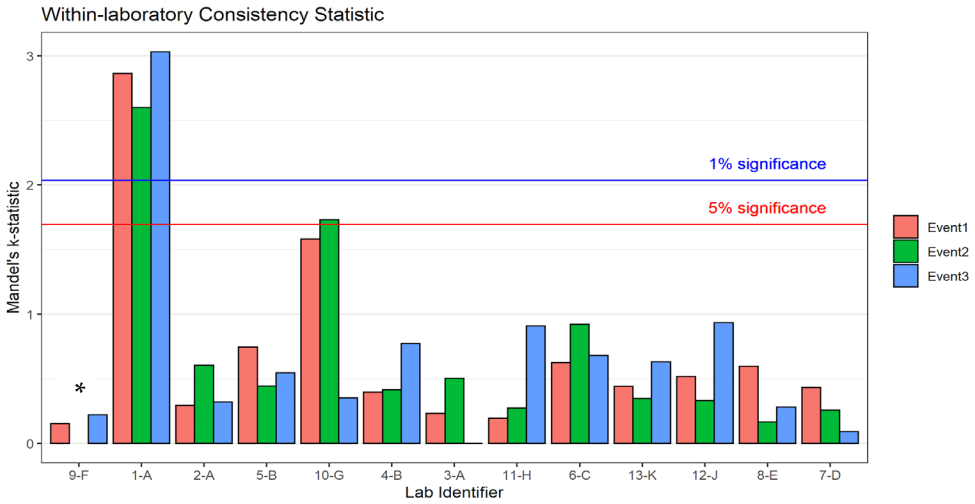


Figure 17. Mendel's *k* statistics computed to detect differences in variances and identify outliers for participating groups in the laboratory intercomparison. Asterisk indicates no analytical

results.

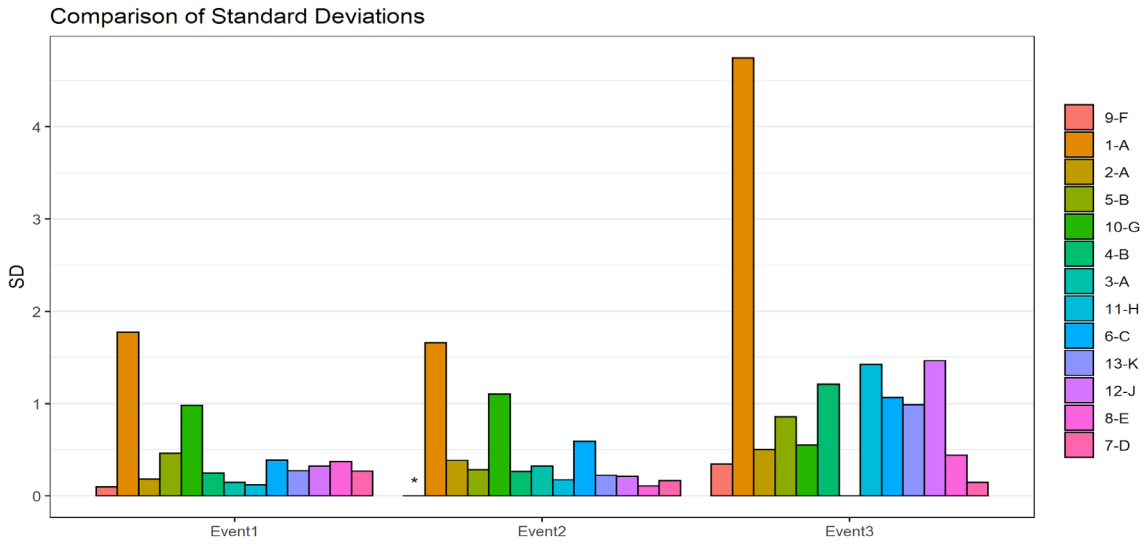


Figure 18. Comparisons of standard deviations of three sampling events for the laboratory intercomparison exercise. Asterisk indicates no analytical results. Entity 3-A's filters for event 3 had identical concentrations; therefore, the standard deviation for entity 3-A was zero.

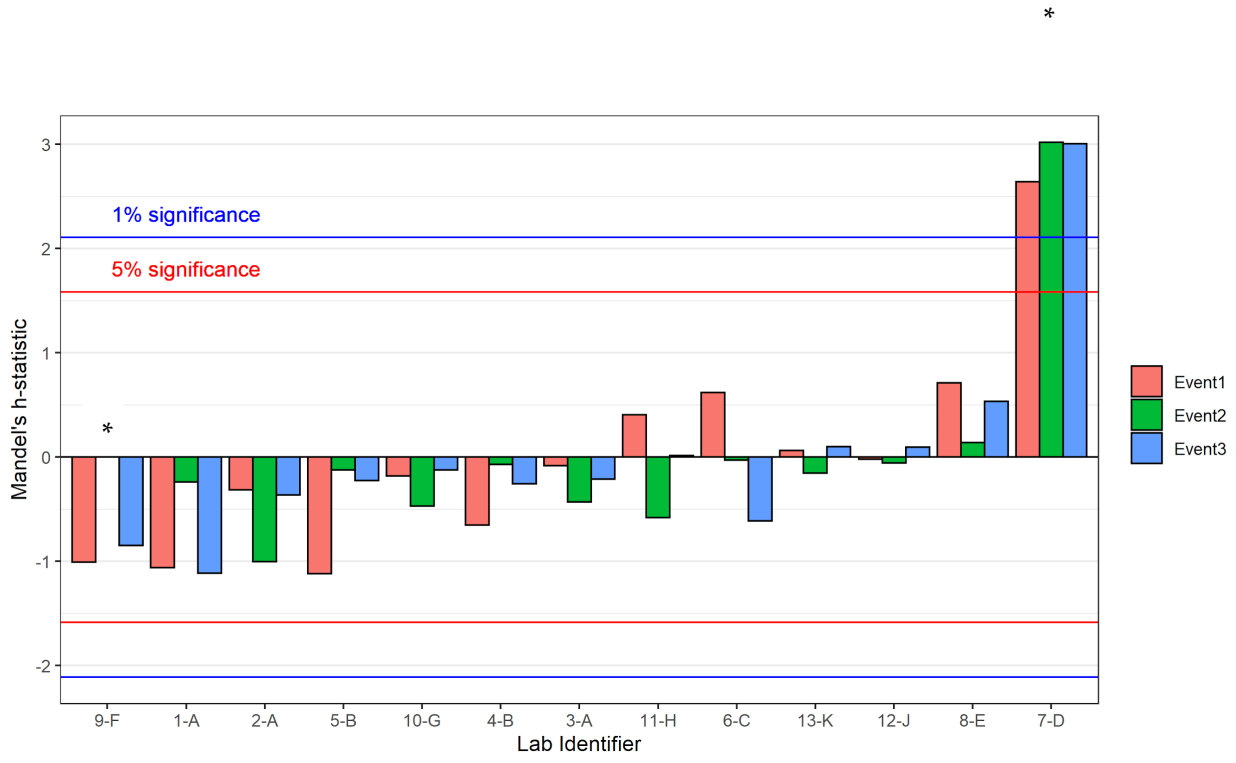


Figure 19. Mendel's *h*-statistic computed to detect differences among means and identify outliers for entities included in the laboratory intercomparison. Asterisk indicates no analytical results.

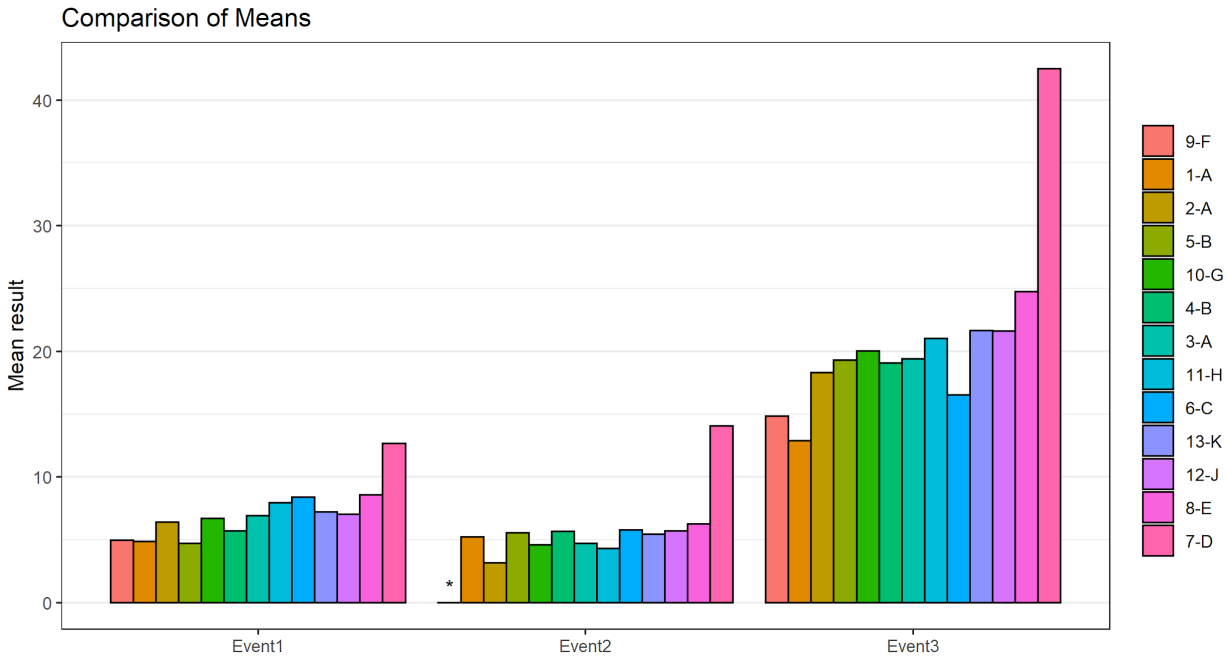


Figure 20. Comparison of mean chlorophyll-*a* (µg/L) across three sampling events for sample processing and laboratory comparison exercise.

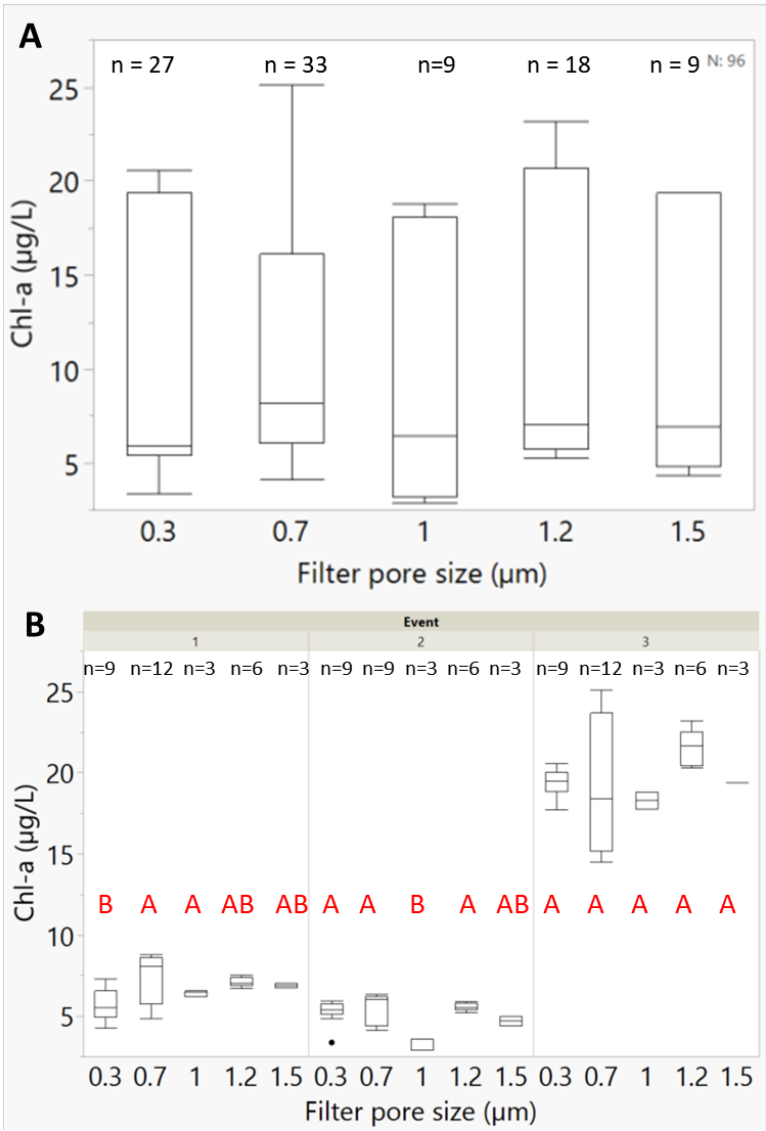


Figure 21. Box and whisker plots of chlorophyll-a (µg/L) by nominal filter pore size for all events (A) and by event (B). The box and whisker plots are arranged on the x axis from smallest to largest nominal filter pore size (0.3 to 1.5 µm). Tukey pairwise (HSD) results in panel B show that results by nominal pore sizes with the same letters are significantly different from results with different letters, within the same event. Data from entities 1-A and 7-D were identified as outliers and removed prior to analysis.

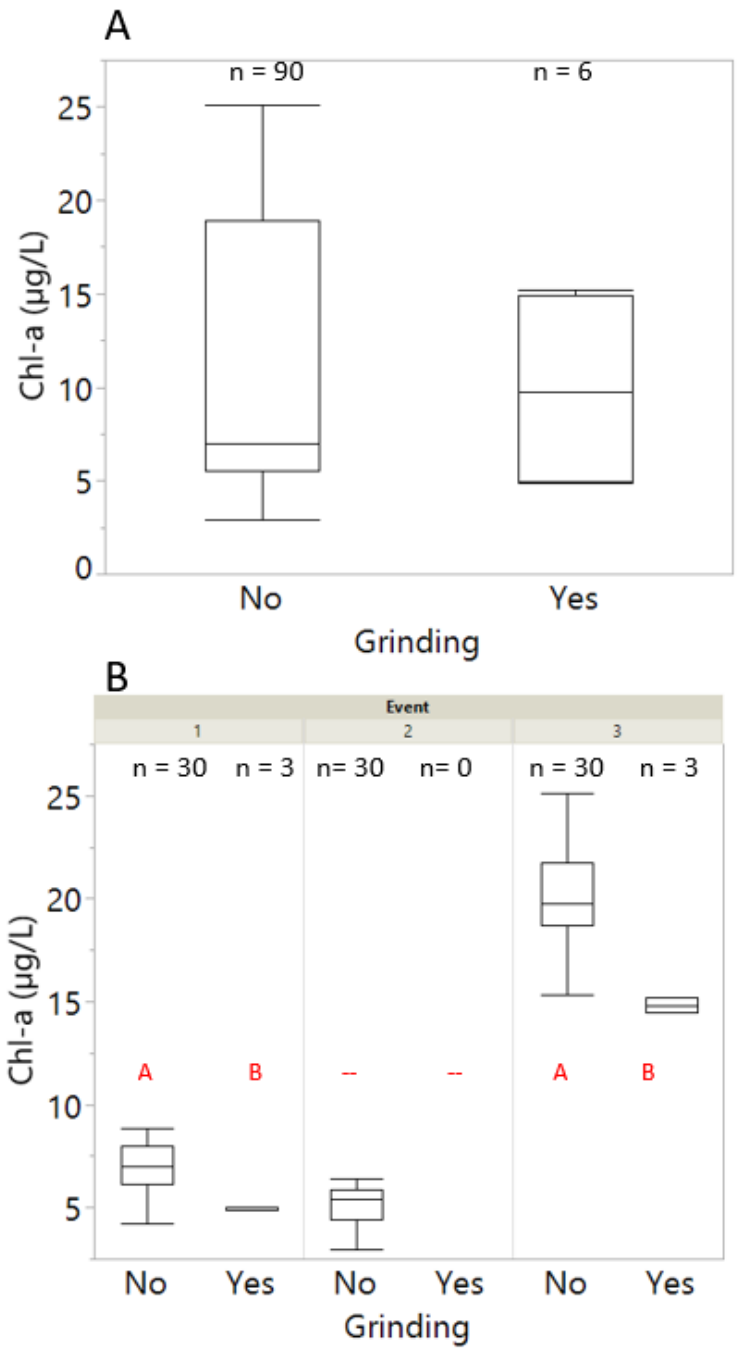


Figure 22. Box and whisker plots of chlorophyll-a (µg/L) concentrations for filter grinding and non-grinding for all events (A) and by event (B). Student's t-test results in panel B show that results by use of grinding with the same letters are significantly different from results with different letters, within the same event. Data from entities 1-A and 7-D were identified as outliers and removed prior to analysis - see methods section for details.

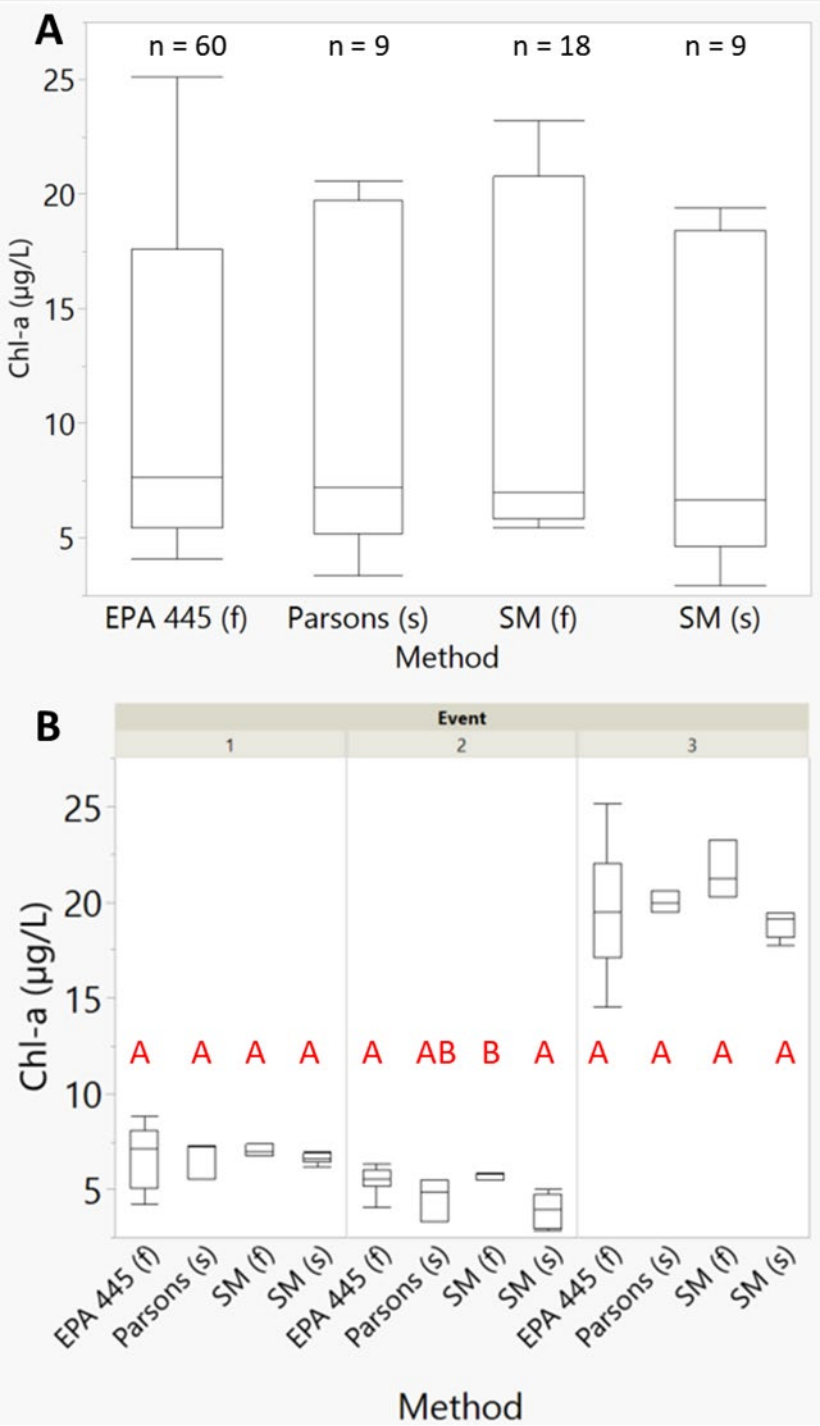


Figure 23. Box and whisker plots of chlorophyll-a (µg/L) concentrations for fluorometric (f) and spectrophotometric (s) analytical methods EPA 445.0 (Arar and Collins 1997), SM 10200H (Baird et al. 2017), and Parsons et al. (1984), for all events (A) and by event (B). Tukey pairwise (HSD) results in panel B show that results by analytical method with the same letters are significantly

different from results with different letters, within the same event. Data from entities 1-A and 7-D were identified as outliers and removed prior to analysis - see methods section for details.

Appendix 1.

Chlorophyll Sensor Intercomparison Study Task 1 Report: Assessment of methods used to measure *in situ* chlorophyll fluorescence by monitoring programs in the San Francisco Estuary

Abstract

Chlorophyll fluorescence is a widely accepted proxy for phytoplankton biomass and is measured throughout the San Francisco Estuary (SFE) by multiple federal, state, and non-governmental agencies. Here we present the methods assessment for chlorophyll fluorescence measurements by selected agencies in the SFE. Whereas many organizations use similar instrumentation, differing methods may affect data comparability and system wide data synthesis. Differing methods include: sensor settings, calibration procedures, deployment and retrieval protocols, quality assurance, and post-processing. The largest variability in methods between agencies and groups therein includes reporting units, calibration procedures, and sensor servicing and cleaning.

List of Tables

Introduction

In situ chlorophyll fluorescence is a widely accepted proxy for phytoplankton biomass that is measured throughout the San Francisco Estuary (SFE) by multiple federal, state, and non-governmental agencies. Agencies may also measure chlorophyll fluorescence for research and short-term special studies as they relate to nutrients and fisheries management activities. Chlorophyll measurement methods vary widely throughout the system, and reducing variance is of interest to a wide variety of agencies and water managers.

The first task of the Intercomparison Study for Chlorophyll Measurements is an assessment of current methods used by select groups in the SFE. Groups involved include the California Department of Water Resources Division of Environmental Services, California Department of Water Resources North Central Regional Office, U.S. Geological Survey (USGS) California Water Science Center Biogeochemistry and Hydrodynamics Groups, and the San Francisco Estuary Institute Clean Water Program (Table A1). Although many monitoring programs operate similar instrumentation, differences in methods may affect data comparability and system-wide data synthesis efforts. Such differences include instrument and sensor settings, calibration procedures, sensor cleaning, data processing, and reporting.

The purpose of this report is to present the results of the methods assessment for sensors in the larger Chlorophyll Intercomparison Study. We briefly identify comparison studies completed by others but focus our efforts to report current practices for chlorophyll fluorescence measurements in the SFE. Funding for this study was provided by the Delta Regional Monitoring Program (Delta RMP) and the San Francisco Bay Nutrient Management Strategy, with additional in-kind contributions from multiple agencies. This study is the second phase of a multi-year effort to improve the accuracy, precision, and comparability of chlorophyll data collected in the SFE.

Methods

Many agencies and groups use the guidance of manufacturers and best practices outlined within the agency's standard operating and quality assurance procedures. These publications

include but are not limited to the Yellow Springs Instrument (YSI) 6-Series Multiparameter Water Quality Sondes User Manual (YSI, 2002), EXO manual (YSI, 2020), the California Department of Water Resources standard operating procedure (DWR, 2019), the USGS Water Quality Field Manual (Berkman and Canova, 2007; Gibs et al., 2012; U.S. Geological Survey, 2018) and USGS guidelines for continuous water-quality measurements (Wagner et al., 2006).

A survey of current practices, hereafter referred to as the methods assessment, was the first task completed for this study. We collected metadata for sensor settings, deployment protocols, calibration, sensor servicing and cleaning, data transmission, and data post-processing and corrections. Groups participating in this exercise currently make high-frequency chlorophyll fluorescence at over 80 fixed locations throughout the San Francisco Estuary and Sacramento-San Joaquin River Delta. Groups also deploy sensors on a temporary basis during special studies.

Results

Instruments and sensor settings

All groups participating in this task use multi-parameter water quality meters that are YSI products (Xylem, Rye Brook, New York) and their associated fluorescence sensors. The two models of YSI sensor include the Total Algae sensor on the EXO v2 and the 6025 sensor on the 6-series sonde (Table A2). Two participating groups are currently transitioning from the 6-series sonde to the EXO v2 because YSI will no longer support the 6-series technology after this year (2020). At least three groups have used EXO v2's for the last several years.

Chlorophyll fluorescence (fCHL) is commonly reported in relative fluorescence units (RFU), micrograms per liter (μg/L), and milligrams per cubic meter (mg/m^3). Two groups consistently report in μg/L and other groups consistently report RFU or both units. Conversion factors between the two units have been observed by the USGS and others, but these factors are location-dependent because of turbidity and dissolved organic matter fluorescence (fDOM).

The interference of turbidity on the fluorescence signal has been well documented for fDOM measurements (Downing et al., 2010; Saraceno et al., 2009; Saraceno et al., 2017) but not for algal fluorescence. Although manufacturers of optical tools acknowledge the interference of turbidity and fDOM on the chlorophyll-a fluorescence signal, and some have applied on-board corrections (e.g., BBe moldaenke's FluorProbe), the manufacturer of instruments used in this study does not apply a fDOM correction to the algal fluorescence signal.

Currently, no groups use advanced sensor settings, apply temperature compensation, or enter an offset to the fCHL measurement that is specific to the YSI instruments.

Deployment and calibration

Five of the seven groups record fCHL based on the 'normal' interval time specified under 'basic settings' in the YSI instrument. Two groups measure in 'burst mode' - a setting that records a data point once a second for a 30-second duration and reports the median value (Table A3). All groups operate sondes that have wipers to prevent biofouling on the sensor faces. The wiping interval can affect the power budget of the sonde (if deployed autonomously) or on the system (if deployed on a telemetered monitoring station). Most groups program their instrument to sample once between wipe intervals, except for one group who programs their instrument to collect two samples between wipes. All groups choose the 'default averaging mode' that is a proprietary setting of YSI instruments.

As mentioned previously, fCHL measurements are reported in RFU and μg/L but most groups do not calibrate to both units (Table A4). Most groups rely on the factory calibration setting of the sensor until they complete a one-point calibration at 0 RFU or 0 μg/L in deionized or organic free water. Only two groups complete two-point calibrations for both units of measurement and each of these groups use a rhodamine dye. Most groups avoid resetting sensors to factory settings unless for troubleshooting purposes.

Sensor servicing and cleaning

Inspection, service, and cleaning schedules vary seasonally, but all groups visit their station at least once every 6 weeks throughout the year (Table A5). Many groups bring a 'check' instrument to compare to the *in situ* field instruments and then complete 'dirty' and 'clean' bucket tests to compare fouling and calibration drifts of the *in situ* instrument. For those that do not compare two instruments in the field, a 'check' back in the office/laboratory is completed.

All groups calibrate sensors in the laboratory setting and a selected few complete calibrations checks in the field - that is after checking sensors in a standard after thorough cleaning of the sensor (Table A6). Most groups swap instruments - including the sensors - while some swap instruments but keep the Total Algae sensor on site on newly deployed instruments. For the purpose of the discussion of sensor swaps, EXO v2 fluorescence sensors hold their calibration independent of the instrument they are connected to and therefore can be swapped freely between instruments.

Deployment depth of instrument differs between groups: some groups deploy instruments at a fixed depth to the water surface while others remain at a fixed depth relative to the bottom of the channel or bay (Table A7). Housing for the instruments is typically PVC or copper, and hole spacing and size (to allow adequate exchange of water across the sensor face) varies among programs.

Data transmission and post processing

All groups internally log fCHL time series data to the instrument or on external data loggers, and all but one group telemeters data via cellular or satellite transmission (Table A8). With the exception of two groups, there is no clear consistency with the use of post-processing software. The removal of outliers may occur using qualitative visual inspections or by quantitative means when outliers fall three standard deviations outside of the mean in the window of a specified period (e.g., a two hour window around the measurement, 1 hour before and 1 hour after). None of the groups currently completes temperature corrections to chlorophyll time-series data.

Data reporting

Many groups report provisional real-time data to public-facing websites. Most groups that report time series data may approve or 'finalize' data within two weeks of collection. One group participating in the methods assessment reported they 'finalize' data within a year of collection (Table A9). One group applies corrections to data within 180 days of collection but has not approved the data because they are awaiting a techniques and methods report from their agency. On occasion, final data do replace provisional data at the same location on the public-facing website. Other groups may archive finalized data to a different location than provisional data.

Acknowledgments

Funding was provided by the Delta Regional Monitoring and the Bay Nutrient Management Strategy. In-kind contributions were provided by the Bureau of Reclamation and California Department of Water Resources. Special thanks to Brian Bergamaschi, Bryan Downing, the Delta RMP TAC and nutrients sub-committee for their review of the Phase II development process. This effort required the collaboration and support of multiple scientists and technicians from the agencies identified in Table 1.

References

California Department of Water Resources.2019. Collection of Water Quality Samples for Laboratory Analysis. Report prepared by the Quality Assurance Program.

Downing, Bryan D., Brian A. Pellerin, Brian A. Bergamaschi, John Franco Saraceno, and Tamara EC Kraus. 2012. "Seeing the Light: The Effects of Particles, Dissolved Materials, and Temperature on *in Situ* Measurements of DOM Fluorescence in Rivers and Streams."

Limnology and Oceanography: Methods 10 (10): 767–775.

Gibs, Jacob, Franceska D. Wilde, and Heather A. Heckathorn. 2012. *Chapter A6. Section 6.8. Use of multiparameter instruments for routine field measurements*. No. 09-A6. 8. US Geological Survey. https://doi.org/10.3133/twri09A6.8.

Hambrook-Berkman, J. A., and Michael G. Canova. 2007. "Algal biomass indicators." US Geological Survey, TWRI Book 9 (2007). https://doi.org/10.3133/twri09A7.4.

Saraceno, John Franco, Brian A. Pellerin, Bryan D. Downing, Emmanuel Boss, Philip AM Bachand, and Brian A. Bergamaschi. 2009. "High-frequency *in situ* optical measurements during a storm event: Assessing relationships between dissolved organic matter, sediment concentrations, and hydrologic processes." Journal of Geophysical Research: Biogeosciences 114, no. G4.

Saraceno, John Franco, James B. Shanley, Bryan D. Downing, and Brian A. Pellerin. 2017. "Clearing the Waters: Evaluating the Need for Site-Specific Field Fluorescence Corrections Based on Turbidity Measurements." Limnology and Oceanography: Methods 15 (4): 408–416.

U.S. Geological Survey. 2018. General introduction for the "National Field Manual for the Collection of Water-Quality Data" (ver. 1.1, June 2018): U.S. Geological Survey Techniques and Methods, book 9, chap. A0, 4 p., https://doi.org/10.3133/tm9A0. [Supersedes USGS Techniques and Methods, book 9, chap. A0, version 1.0.]

Wagner, Richard J., Robert W. Boulger Jr, Carolyn J. Oblinger, and Brett A. Smith. 2006. Guidelines and standard procedures for continuous water-quality monitors: station operation, record computation, and data reporting. No. 1-D3. 51 p. + 8 attachments; accessed January 21, 2020, at http://pubs.water.usgs.gov/tm1d3

Yellow Springs Incorporated. 2002. Operating manual for YSI 6-series environmental monitoring systems, 332 p., accessed January 21, 2020, at http://www.ysi.com/extranet/EPGKL.nsf/447554deba0f52f2852569f500696b21/90a0378150c2d2dd85256a1f0073f295/$FILE/069300B.pdf.

Yellow Springs Incorporated. 2020. EXO user manual revision k; Advanced water quality monitoring platform, 244 p., accessed January 21, 2020 at https://www.ysi.com/file%20library/documents/manuals/exo-user-manual-web.pdf