Participatory science methods to monitor water quality and ground truth remote sensing of the Chesapeake Bay

Patrick Neale[1*], Shelby Brown[1], Tara Sill[1], Alison Cawood[1], Maria Tzortziou[2], Jieun Park[2], Min-Sun Lee[2,3], Beth Paquette[4]

[1]Smithsonian Environmental Research Center, Smithsonian Institution, Edgewater, Maryland, United States of America

[2]The City College of New York, City University of New York, New York, United States of America

[3]National Institute of Fisheries Science, Busan, South Korea

[4]Independent researcher, Edgewater, Maryland, United States of America

*Corresponding author

E-mail: nealep@si.edu

Short title: Volunteer monitoring of Chesapeake Bay water quality

Abstract

Measurements by volunteer scientists using participatory science methods in combination with high resolution remote sensing can improve our ability to monitor water quality changes in highly vulnerable and economically valuable nearshore and estuarine habitats. In the Chesapeake Bay (USA), tidal tributaries are a focus of watershed and shoreline management efforts to improve water quality. The Chesapeake Water Watch program seeks to enhance the monitoring of tributaries by developing and testing methods for volunteer scientists to easily measure chlorophyll, turbidity, and colored dissolved organic matter (CDOM) to inform Bay stakeholders and improve algorithms for analogous remote sensing (RS) products. In the program, trained volunteers have measured surface turbidity using a smartphone app, HydroColor, calibrated with a photographer's gray card. In vivo chlorophyll and CDOM fluorescence were assessed in surface samples with hand-held fluorometers (Aquafluor) located at sample processing "hubs" where volunteers drop-off samples for same day processing. In validation samples, HydroColor turbidity and Aquafluor in vivo chlorophyll and CDOM fluorescence were linearly related to standard analytical measures of turbidity, chlorophyll and CDOM, respectively, with R^2 values ranging from 0.65 to 0.85. These methods are being used for both repeat sampling at fixed sites of interest and ad-hoc "blitzes" to synoptically sample tributaries all around the Bay in coordination with satellite overpasses. All data is accessible on a public database (serc.fieldscope.org) and can be a resource to monitor long-term trends in the tidal tributaries as well as detect and diagnose causes of events of concern such as algal blooms and storm-induced reductions in water clarity.

Introduction

Water quality monitoring programs use a suite of indicators to provide information about overall ecosystem conditions. These include biological indicators like chlorophyll a (Chl), physical indicators like turbidity, and chemical indicators like colored dissolved organic matter (CDOM) [1]. Each of these indicators can be related to key water optical properties (such as scattering and absorption) that determine light transmission in aquatic ecosystems, which together are referred to as the Inherent Optical Properties (IOPs). The color of water and its main determinants can also be retrieved from space, using advanced multi- and hyper-spectral ocean color sensors [2, 3]. High spatial resolution (e.g., 10-60 m) optical sensors, in particular, can uniquely capture the fine-scale heterogeneities and complexities of inland and coastal aquatic ecosystems [4]. Such sensors include the Landsat-8/9 Operational Land Imager (OLI) (a collaboration between NASA and the U.S. Geological Survey, USGS), the MultiSpectral Instrument (MSI) spectral imager aboard the Sentinel-2A/B missions of the European Space Agency (ESA), as well as NASA's upcoming Surface Biology and Geology (SBG) Designated Observable. These satellite sensors (present and future) provide a unique capability to capture water quality changes in highly vulnerable and economically valuable nearshore and estuarine habitats. Yet, they also require extensive (spatially and temporally) in situ measurements to validate and inform the development of appropriate algorithms to retrieve water quality metrics from satellite data.

The past decade has seen increasing involvement of volunteers in water quality monitoring, but few studies have related ground level data to satellite-based remote sensing of surface water properties [5]. Nevertheless, the potential for volunteer measurements to complement remote sensing observations of the ocean and lakes has been recognized for some time [6]. Two examples of projects harnessing this potential are the SeaHawk/HawkEye project (https://coast-lab.org/hawkeyeCitSci/), and Eye on Water project [7], where volunteers are collecting measurements that can be compared to products derived from remote sensing of ocean color. While these projects are sometimes described as citizen science, community science or volunteer monitoring, we have adopted the more inclusive term, "participatory science" [8, 9]. Apart from the above cited examples, most participatory science water quality projects have focused on involving volunteers to sample, and conduct simple physical, chemical and biological assays [1]. On the other hand, there are many participatory science projects that assist in interpretation of remote sensing of air-quality or land features and hazards (e.g. EPA's Air Sensor Toolbox and NASA's Landslide Reporter and Mountain Rain or Snow program).

The Chesapeake Bay is the largest estuary in the United States, located on the East Coast of the United States. The Chesapeake Bay stretches for about 320 km from the northernmost point of the Susquehanna River in Maryland to the southernmost point of the Atlantic Ocean. As a drowned river valley, a notable feature of the bay is its complex shoreline, with many small embayments and rivers that are tidal tributaries in their lower reaches. These shallow areas are a large portion of Bay surface area and important transitional zones in between the watershed and the main bay. They perform many important ecological functions as hot spots for the processing of terrestrial inputs to the Bay, fisheries habitat and refuges for juvenile fish, habitat

for shore birds, and many more [10]. Improved water quality, and in particular better water clarity, in shallow waters has been identified as a high priority by the regional partnership responsible for Bay restoration, the Chesapeake Bay Program [11]. Despite their importance, only a few of the tidal tributary areas are subject to high resolution monitoring that includes Chl, turbidity, and CDOM (e.g. Rhode River [12], and Patuxent River [13]). This contrasts with the extensive, long-term monitoring of the main stem of the Chesapeake Bay which has informed quantitative targets for acceptable water quality [14].

In the Chesapeake Bay, participatory science and non-profit organization monitoring of water clarity in the Chesapeake Bay has primarily used the Secchi disk. This is a circular white or black and white target that is lowered in the water and the first depth at which it is no longer visible is recorded. Due to its simplicity of deployment, there are long term records of Secchi disk depths throughout the Bay recorded by a wide variety of observers [15, 16]. An extension of the Secchi depth is to compare the observed color of the disk at half depth to a color scale (Forel-Ule scale) which is a qualitative indicator of water quality [17]. Despite its benefits, two key disadvantages of the Secchi disk depth and color scale metrics are that they are subjective measures dependent on observer vision and that the depth and/or scale value reflects the combined effect of all properties that affect light transmission [18]. Thus, there is a need for volunteer methods that more specifically show the contributions of individual properties such as Chl, turbidity, and CDOM to water clarity and can be directly compared to similar satellite data products.

Fluorescence measurements are appropriate for participatory science because there are easy to use instruments such as the Turner Designs Aquafluor, many types of measurements can be made directly on the sample without preparation, and results are immediate. Simple fluorometers can even be constructed as DIY (Do-It-Yourself) projects [7]. Microalgae can be detected via the *in vivo* fluorescence emitted by the photosynthetic pigment, Chl. The fluorescence emission is dependent on both the concentration of pigment and the physiological state of the cell, so the proportion between the signal and an analytical measurement of Chl can vary widely [19]. Moreover, there can be optical interference from co-occurring turbidity and CDOM [20]. Thus, the few participatory science projects that monitor algae fluorometrically have just taken the measurement as a relative indicator without any specific defined relationship to Chl (e.g. community science lab, https://cesh.bard.edu/csl-alt/). Other properties that can be measured via fluorescence include the cyanobacteria-specific pigments, phycocyanin and phycoerythrin, and CDOM. These moieties all have specific excitation emission properties that allow reliable detection, but with varying quantum yield, i.e. the proportion of fluorescence emitted per photon absorbed. Thus, using fluorescence to make quantitative estimates requires detailed validation of the measurement in the target system.

Cameras in mobile devices (aka “smartphones”) can also be used in various ways to monitor surface water properties. Smartphone software packages (“apps”) have been developed that use camera imagery to assess water color (EyeOnWater) or surface reflectance (HydroColor). EyeOnWater enables users to classify the surface color of a water body on the Forel-Ule scale and compare current readings with historical readings which have records many decades long in some locations [7]. However, the determinations do not relate directly to IOPs.

HydroColor references RGB (red, green, blue) images of the water surface and sky to that of a standard 18% reflectance gray card to estimate surface reflectance, from which turbidity can be calculated based on an empirical relationship [21]. In advance of the work presented here, HydroColor has received limited testing in coastal waters [22, 23]. A test of HydroColor in 13 Australian lakes and rivers found poor correspondence between the app's estimate of turbidity and standard gravimetric measurements of suspended particulate matter [24].

Water quality monitoring involving volunteers and professionals not only informs local assessments of water quality but also has the potential to provide valuable ground-truth data for validation of satellite remote sensing products that can ultimately complement the on ground monitoring efforts. Here, we report on volunteer measurements relevant to ground-truthing satellite remote sensing of water quality in the Chesapeake Bay. This manuscript covers the methods used, the data acquired over two years of deployment, and how the results compare to those from standard water quality methods. A future manuscript will focus on the optimization of algorithms used to interpret satellite remote sensing image data in Chesapeake Bay tributaries based on match-ups with volunteer measurements.

Methods

Study Area and Sampling

Sampling was conducted as part of the Chesapeake Water Watch participatory science project, a NASA-funded collaborative effort between the Smithsonian Environmental Research Center (SERC) and the City College of the City University of New York (CCNY). During the period

reported on here, November 2021 through December 2023, participants regularly sampled locations in the upper Chesapeake Bay (Fig 1). The study area encompassed the oligohaline to mesohaline region of the Chesapeake Bay (nominal salinity 0.5 - 18). Some initial testing of methods by project staff occurred prior to this period starting in June 2021, but that data was not included in the analyses reported here. In 2021 and 2022, the sampled locations included the tidal estuarine reaches of the West, Rhode, South, and Severn rivers and adjacent parts of the main bay. We worked with local Riverkeeper groups – the Arundel Rivers Federation (ARF) and Severn River Association (SRA) - that conducted weekly or biweekly monitoring cruises. Other participants were individuals that made observations and sampled from private or public docks, or private watercraft. Project interns and individual volunteers also accompanied SERC staff on monthly sampling cruises on the Rhode River. In total, there were 766 water samples with participation of 57 volunteers during this first period.

Fig 1. Chesapeake Water Watch Sampling Locations. (a) Map of the upper Chesapeake Bay in Maryland and Virginia showing sampling locations in 2021-2023. Four sites in Virginia are not shown in the map. Inset square indicates area of detail shown in (b) a higher resolution map showing 4 sampling locations in the South River near Edgewater, MD. The three nearshore locations are docks sampled regularly by a Chesapeake Water Watch volunteer (co-author BP); the mid-channel location nearby is sampled by the Arundel Rivers Federation riverkeeper organization**.**

In 2023, sampling activities continued in the aforementioned waterways as well as expanding to tidal tributaries on the Eastern shore, north to the Gunpowder watershed, and the Magothy River (Fig 1). These areas were sampled by more individual volunteer participants,

170 additional Riverkeeper groups including the Magothy River Association, Gunpowder
171 Riverkeeper, and ShoreRivers, local high schools and community colleges as well as community
172 groups such as the Advocates for Herring Bay community association. At the end of 2023, there
173 were 2066 water samples and 104 participating volunteers. All volunteers were trained on
174 sampling procedures and measurement protocols using online and/or in person instruction.
175 Measurements are currently ongoing, further expanding the spatiotemporal coverage of our
176 sampling.

177 As a part of the training, volunteers created accounts on the Chesapeake Water Watch
178 Fieldscope site (serc.fieldscope.org) a data collection and analysis platform managed by
179 Biological Sciences Curriculum Study (BSCS) Science Learning. For each sample, participants
180 were instructed to observe ambient conditions at the measurement site and select appropriate
181 entries on the Fieldscope data entry menu to describe wind speed, surface waves, estimated
182 percent cloud cover, and air temperature (see data form in S1 Table). Although in some cases
183 only HydroColor or surface water samples were collected, in most cases, participants did both.
184 Participants were furnished with small water bottles or 50 mL centrifuge tubes, if these were
185 not available, a used water bottle was acceptable if it previously only contained water. They
186 were instructed to rinse the container three times with water from the sampling location, filling
187 each time with the opening below the surface (to avoid surface scums) before final filling. The
188 containers were stored in the dark at ambient temperature before processing. Samples were
189 left in the dark for a minimum period of 30 min (optimally ≥2 h) to reverse non-photochemical
190 quenching of in vivo chlorophyll fluorescence and (for cold or warm samples) to adjust to room

temperature (nominally 22°C). This timing was based on time course studies that are shown in the results section. Insulated cooler boxes were available at drop-off locations for participants to leave samples.

In addition to locations that were sampled multiple times over the reporting period ("monitoring stations"), we conducted two one-day events that aimed to sample as many locations as possible within the project area (July 14 and October 2, 2023). These synoptic sampling "blitzes", termed "Satellites and Samples" events, were timed to coincide with days that satellite images were scheduled to be acquired by both the OLI imager (aboard Landsat 8 and 9) and the MSI imager (aboard Sentinel 2A and 2B). These events were widely announced via email and open to anyone interested without requiring previous training. Instructions were given on the SERC website that included a template data sheet. Participants were instructed to take a surface sample using a clean plastic bottle (an empty water bottle was recommended), fill out the sheet with a standard description of site conditions, the time and location of the sample, and drop off the sample to one of several processing "hubs" distributed through the area. The hubs were staffed with registered volunteers who processed the samples with project instruments (description follows in "Participatory Science Measurements") and entered the data into Fieldscope. These events contributed 106 and 59 samples on July 14 and October 2, respectively (included in the above total).

Participatory Science Measurements

HydroColor measurements

Volunteers conducted on-site measurements of surface turbidity using the HydroColor smartphone app. Developed at the University of Maine, HydroColor is a free app for Android and iOS operating systems that enables volunteers to make observations of surface water quality based on water reflectance measurements with a smartphone camera. HydroColor uses the RGB channels of smartphone images taken of a gray card, sky and water surface to calculate remote sensing reflectance, $Rrs(RGB)$ [21]. In addition to measuring Rrs, HydroColor geolocates and timestamps the observation. The developers derived an empirical relationship between red channel reflectance and turbidity which is similar to the algorithm used to derive suspended particulate matter (SPM) estimates from satellite-retrieved *Rrs* [25]. This relationship assumes a 1:1 proportion between turbidity in NTU and SPM in mg L^{-1}.

Chesapeake Water Watch used two versions of the HydroColor app: version 1 for 2021 – 2022 measurements and version 2 for 2023 measurements. Version 1 was the version tested by Leeuw and Boss [21], this was updated to version 2 in March 2023. The earlier version used compressed image data from the smartphone camera, this was upgraded in version 2 to use uncompressed ("RAW") image data with expectation that this will give more consistent results between different models of smartphones, which we tested in our study. Further information on the developer's pre-release testing and observations on user experiences with the version 2 are reported in supporting information (S1 File).

Aquafluor measurements

Participants used the Aquafluor handheld fluorometer (Turner Designs, San Jose, CA) to measure in vivo Chl (IVChl) and CDOM fluorescence. It is a simple to use, battery operated

instrument suitable for easy and fast measurements in highly variable waters such as the Chesapeake Bay. The IVChl channel uses a 460 nm LED source, a broad band excitation filter with 395 nm center wavelength (CWL) and 130 nm (FWHM) bandwidth, and a 660 nm cutoff longpass emission filter. The CDOM channel uses a 375 nm LED source, a 350/80 nm (CWL/FWHM) excitation filter and a 470/60 nm emission filter. The CDOM channel emission filter is a custom component that Turner Designs configured for Chesapeake Water Watch, as the "stock" emission filter is a longpass filter that does not exclude in vivo Chl fluorescence. The channels were calibrated using a modification of the procedures used for sensors in water quality sondes, such as the YSI EXO2 [26]. In vivo Chl fluorescence was calibrated with a 0.625 mg L^{-1} rhodamine WT (Kingscote Chemicals, Miamisburg, OH) solution which is assigned a reading of 66 µg L^{-1} rhodamine-equivalent (RE, temperature corrected to 22°C). Although its optical properties differ from the desired primary standard, in vivo chlorophyll, rhodamine has a stable, fixed response that is easy to replicate for calibration purposes. Validation measurements are used to determine the relationship of RE measurements to Chl in "live" algal cells. CDOM fluorescence was calibrated with 300 µg L^{-1} quinine sulfate (QS, Sigma-Aldrich) in 0.1N sulfuric acid, with reading assigned a calibration value of 300 µg L^{-1} QS-equivalents (QSE, temperature corrected to 22°C). QS fluorescence is used as a proxy for natural CDOM and, as for IVChl, the relationship of QSE to CDOM absorbance was determined using validation measurements. Each measurement was conducted in triplicate with ~3-4 mL sample in an acrylic cuvette, reading successively the in vivo Chl, then CDOM fluorescence. These readings are assigned units of µg L^{-1} of RE and QSE, respectively.

253 Volunteer data for this report was acquired with six identically calibrated instruments.
254 Post-calibration, test samples from the Rhode River tidal tributary at SERC read on all five
255 instruments were within 0 to 10% (mean +5%) for in vivo Chl fluorescence and within -6 to 0%
256 (mean -3%) for CDOM fluorescence of the readings on the reference instrument used to make
257 comparisons in the validation studies (S2 Table).

258 Laboratory Validation Measurements

259 Turbidity measurements

260 In most cases (>80%) when a surface water sample was taken, turbidity was measured on
261 a benchtop turbidimeter, specifically the AQUAfast AQ3010 Turbidity Meter (Thermo Scientific
262 Orion, Waltham, MA). These instruments are simple to operate and volunteer participants are
263 easily trained to follow a standard protocol. The instrument was checked daily before use with
264 a 20 NTU standard and deionized water for 0 NTU. Each measurement used 10 mL of sample
265 water in manufacturer provided vials, with readings conducted in triplicate. Instrument
266 calibration was checked yearly and recalibrated as needed with manufacturer furnished
267 turbidity standards.

268 Chlorophyll-a measurements

269 For selected cases (~15% of total sampling events), participants were requested to sample
270 an extra volume (≥ 250 mL) so that water could be filtered for analytical Chl and CDOM
271 determinations. As these measurements are more technically demanding, they were only
272 performed by trained project staff. The Chl assay followed the EPA protocol [27]. Triplicate

sample aliquots of ~ 50 mL were filtered on 25 mm glass-fiber filters (GF/F, Whatman Inc.) at low vacuum and stored frozen at -20°C until analysis (typically within one month). Chlorophyll was extracted with 90% acetone overnight at -20°C, then the extract was warmed in the dark to room temperature just before analysis. Chlorophyll fluorescence was measured before and after acidification on a Turner 10-AU fluorometer. The values were converted to Chl and pheophytin *a*. The fluorometer was calibrated yearly with Chl standard from Sigma Chemicals.

CDOM measurements

For CDOM absorbance, samples were filtered through Whatman GDX syringe filters which have a combination of glass fiber and membrane filters with a nominal 0.45 µm pore size. Filtrates were stored in the dark at 4°C until analysis, typically within one week. We performed absorbance scans at 2 nm intervals (270 – 750 nm) using a Thermo Scientific Evolution 220 UV-Vis spectrophotometer. CDOM absorption was measured following the protocol in Tzortziou et al. [28]. Optical density (OD) in 1 cm cuvettes was converted to absorption coefficient (a_{CDOM} m^{-1}) at wavelength, λ (nm), using the equation:

$$a_{CDOM}\lambda = 2.303\frac{OD}{l_g} \qquad (1)$$

Where l_g is the pathlength = 0.01 m.

Statistical analysis

As a measure of analytical variability for volunteer methods, we calculated a pooled Root Mean Square Deviation (RMSD) for specified sets of *n* triplicate observations (HydroColor turbidity, Aquafluor, IVChl and CDOM fluorescence) using the equation:

$$RMSD = \sqrt{\frac{\sum_{i=1}^{n}\sum_{k=1}^{3}(x_{ik}-\overline{x}_i)^2}{3n}} \quad (2)$$

Where x_{ik} are triplicate observations, k=1 to 3, of the ith sample with mean $\overline{x}_i$. Validation and volunteer measurements were compared based on linear regressions with root mean square error (RMSE) and median % error reported based on the deviations from the fitted regression line. In cases where the intercept was not significantly different from zero, a regression with no intercept was used.

Results

Participatory scientist measurements

Table 1 summarizes the statistics for each of the variables measured by participants. In general, the sampled locations had moderate turbidity, were mesotrophic and had appreciable concentrations of CDOM. The tidal tributaries which are the focus of this study are dynamic environments due to the varying influence of weather, wind forced flows, nutrient inputs from the watershed and commercial and recreational fisheries [29]. The extent of these dynamics is evident in records compiled by volunteers that regularly sample on a weekly or high frequency basis. Two years of measurements taken by a Chesapeake Water Watch volunteer (co-author BP) at docks on the South River (locations in Fig 1b) show variability at yearly, monthly, and even

weekly time scales (Fig 2). For comparison, Fig 2 also shows measurements of the same variables using project methods at a nearby, mid-channel site (location shown in Fig 1) made by the Arundel Rivers Federation (ARF) riverkeeper. For logistical reasons, these samples are only taken mid-April through mid-October. Volunteer and ARF samples are comparable in this period, but the bi-weekly sampling by ARF misses much of the short-term and some of the seasonal variation in this system.

Table 1. Summary Statistics for Volunteer and Validation Data

	Whole Data Set						Validation Data			
Variable	Period	Units	lower quartile	Median	upper quartile	Replicate RMSD	Units	R^2	Median % error	RMSE
in vivo Chlorophyll Fluorescence	2022-23	µg L^{-1} RE	27.84	41.20	61.51	3.53 (1386)[a]	µg L^{-1} Chl	0.85	16%	7.0 (200)[b]
CDOM Fluorescence	2022-23	µg L^{-1} QSE	23.69	27.19	33.40	0.43 (1409)	a440 (m^{-1})	0.76	10%	0.12 (204)
HydroColor Turbidity	2022 v1	NTU	3	5	8	2.79 (649)	NTU	0.50	38%	5.43 (182)
	2023 v2	NTU	4	5	9	0.92 (502)	NTU	0.65	27%	2.9 (185)
Benchtop Turbidity	2022-23	NTU	4.68	7.11	10.92	0.55 (1153)				

Statistics for volunteer measurements in the northern Chesapeake Bay for 2022 and 2023, and overall comparison between volunteer measurements and validation data based on linear regression of the validation as a function of the volunteer variable.
[a]Total number of sample entries in Fieldscope, with triplicate readings for each sample.
[b]Number of samples in the validation data set.

Fig 2. Examples of Participant Monitoring Measurements. Observations using project methods taken by a Chesapeake Water Watch volunteer (co-author BP) at three docks in the South River near Edgewater, MD (green dots, solid line shows average) and an independent set of measurements taken by the Arundel Rivers Federation riverkeeper at a nearby mid-channel site

(yellow diamonds, locations shown in Fig 1). (a) in vivo chlorophyll fluorescence (IVChl) measured with the Aquafluor, (b) CDOM fluorescence measured with the Aquafluor, (c) turbidity estimate using the HydroColor smartphone app. One volunteer point from April 6, 2022 was omitted from (b), on this date CDOM was unusually high (65 µg l^{-1} QSE) due to high amounts of tree pollen that had entered the water column.

As an example of results of the coordinated synoptic sampling during a "Satellites and Samples" event, we show data from an event on July 14, 2023 in Fig 3. Evident from this "snapshot" of IVChl, CDOM and turbidity in these locations is the increase in each of these properties with distance upstream away from the main bay. These observations can be compared to retrievals from both the OLI imager on Landsat 9 and MSI imager on Sentinel 2A on the same day. The results of these comparisons and other applications of the volunteer data for ground-truthing remote sensing products will be reported in a separate publication (in preparation).

Fig 3. Maps of the upper Chesapeake Bay showing data from the July 14, 2023 Satellite and Samples Event. Points show location of samples and color coding on symbols using color bar on right denote values for (a) in vivo chlorophyll fluorescence, (b) CDOM fluorescence and (c) benchtop turbidity measurements.

Participatory scientist measurements – Validation

Turbidity measurements

HydroColor observations in 2021-2022 were made with app v1 and in 2023 app v2 was used. In each year, approximately 200 of the HydroColor observations were compared with benchtop measurements of turbidity with the Aquafast on samples taken at the same time and place (Fig 4). The 182 observations by 18 volunteers in 2021-2022 with v1 showed a good linear relationship at a close to unity ratio with benchtop turbidity but also considerable scatter (Fig 4a). The linear regression had a R^2 of 0.50 as a predictor of Aquafast turbidity, a median % error of 38% and the mean square error (MSE) of the regression (assuming 0 intercept) of 5.4 NTU. There was a significant intercept in the linear regression of 2.6 NTU, nevertheless the intercept was omitted for the regression line in Fig 4a to estimate the overall proportion between the two turbidity measures. There were many observations (15% of total) in which the error in the regression estimate was > 7 NTU. This variation was considerably greater than the 26% median percent error reported by Leeuw and Boss [21] in their developmental tests of HydroColor v1, however those tests were conducted with only three devices (all using iOS operating system). This suggested that at least part of the variability in 2021-2022 could have been due to differences in the response of the phones among the 18 users and motivated the update of the application to obtain greater consistency between phone response. The update of app by Thomas Leeuw (S1 File) was completed in February of 2023.

In 2023 using HydroColor v2, most of the observations were also close to unity in the relationship with Aquafast turbidity (black points, Fig 4b), but with markedly less scatter. The linear regression of these 185 observations by 27 users had an R^2 of 0.65, a median percent error of 27% and MSE of 2.9 NTU (Table 1). Again, there was a significant intercept of 1.9 NTU, but this was omitted for the regression line plotted in Fig 4b. All average Aquafast turbidities in

2023 samples were within ±6.5 NTU of the turbidity predicted from the HydroColor v2 regression. Notably, the implementation of v2 improved the overall relationship with an analytical measurement of turbidity even though there were more users contributing to the comparison. The median percent error approached the potential accuracy stated by the developers (i.e., 26%). However, with the expansion of sampling sites there emerged a group of 25 samples that had a > 2 times greater benchtop turbidity than that measured by HydroColor (gray triangles in Fig 4b). These observations were mainly from tributaries that received treated wastewater (Patuxent and Back rivers) or from shallow areas after wind resuspension events.

Fig 4. Comparison of HydroColor estimates of turbidity with benchtop measurements with the Aquafast (a) app v1 used in 2021-2022 and (b) app v2 used in 2023. Gray triangles in (b) are a group of points with high Aquafast and low HydroColor turbidity (see text). Lines are the fitted linear regressions between benchtop and HydroColor turbidity for each group of samples assuming a zero intercept.

Chlorophyll-a measurements

The Aquafluor measures IVChl in rhodamine equivalents (RE, µg L^{-1}) and validation comparisons were made to establish the relationship with extracted Chl as measured using a standard analytical method. For the samples in the northern Chesapeake Bay, IVChl was a consistent linear estimator of Chl, with Chl averaging about half of the concentration of IVChl rhodamine equivalents (Fig 5). The 200 observations had a linear regression R^2 of 0.85 with extracted Chl, a median % error of 16% and the mean square error (MSE) of regression of 7 µg L^{-1} (Table 1). The intercept was omitted as it was not significantly different from zero (p=0.65).

Multiple linear regressions were performed that included concurrent measurements of turbidity, CDOM, or both, in addition to IVChl, but adding these variables did not increase R^2 (S3 Table). An exception to the general ~0.5:1 relationship of Chl to IVChl, were 10 samples from the Severn River during a fall bloom 2021 in which the relationship was closer to 1:1. These samples were encountered in an initial testing period before volunteer sampling began, so are not shown in Fig 5.

Fig 5 Relationship of chlorophyll a and in vivo chlorophyll fluorescence in validation samples. X-axis, IVChl in selected Chesapeake Bay samples measured on the Aquafluor (in rhodamine equivalents, RE, µg L^{-1}) vs, y-axis, extracted Chl. Complete data set of all validation samples April 2022-October 2023 (n=200 samples, averages of triplicate measurements). Line shows linear regression assuming a zero intercept.

CDOM measurements

For an initial set of validation samples in 2021, the linear regression of CDOM absorbance vs Aquafluor CDOM fluorescence showed an unexpectedly high degree of variability. For example, a regression of 58 samples of absorbance at 440 nm (a440, m^{-1}) on CDOM fluorescence in June and July 2021 had an R^2 of 0.37 (data not shown). This variability was traced to an optical interference from Chl fluorescence which was corrected by installation of custom emission filter that excluded Chl fluorescence (see methods). When these custom machines became available in 2022, we compared CDOM readings in samples that contained both phytoplankton and CDOM vs samples filtered to remove phytoplankton and verified that the interference problem was eliminated by this modified configuration. There was no

significant difference in CDOM between unfiltered and filtered samples (see e.g. S1 Fig). A new set of validation samples were acquired during 2022 and 2023. The CDOM fluorescence measured in these samples showed a consistent relationship with CDOM absorbance (**Fig 6**) and no cross-talk with Chl fluorescence. The 204 validation samples had linear regression R^2 of 0.76 for *a*440, a median % error of 10% and the mean square error (MSE) of regression of 0.12 m^{-1} (**Table 1**). While these instruments are now considered more costly "custom builds", their optical configuration is required to make accurate CDOM fluorescence measurements with unfiltered samples that may contain phytoplankton.

Fig 6. **Validation of Aquafluor CDOM measurement.** Relationship of CDOM dissolved organic matter fluorescence (µg L^{-1} QSE) to absorbance of colored dissolved organic matter (CDOM) measured at 440 nm in a spectrophotometer (2022 & 2023 data). Line shows fitted linear regression with equation and R^2 annotated on the graph.

Participatory scientist measurements – Sources of variability

Turbidity measurements

Sources of variability in the HydroColor measurement include differences in user proficiency, and variations in cloud coverage and surface wave conditions. We evaluated user proficiency by calculating the RMSD for the triplicate entries by each user in the Fieldscope database. Users following the Chesapeake Water Watch protocol took at least three HydroColor measurements and were advised to make as many as two more measurements if the initial three turbidities were not within ±5 NTUs of each other. The user then enters the

three readings that were within the acceptable range of one another on Fieldscope. Median variability (RMSD per user) was lower in the v2 app, decreasing by 29% from 0.94 NTU for 31 users of v1 to 0.67 NTU for 41 users of v2. For 9 users that made observations using both HydroColor versions, median variability decreased from 0.93 using v1 to 0.60 using v2.

To further explore sources of variability, we evaluated the effect of varying cloud cover and the presence of surface waves. These conditions depart from those assumed to apply in the derivation of equations to correct water leaving radiance for surface reflection [21]. The HydroColor algorithm assumes that surface reflection is 2.8% of sky radiance, which is the proportion expected for reflectance of the clear sky from a calm water surface given the viewing position of the camera [30]. The proportional reflectance of sky radiance can be significantly higher under partly cloudy conditions. Likewise, surface waves will result in reflections being viewed at varying angles, potentially increasing the reflectance of sky radiance. To examine how much the presence of clouds and/or surface waves affected the accuracy and repeatability HydroColor observations, we evaluated the RMSE, median percent error, and replicate RMSD, as a function of sky and surface wave conditions. The Fieldscope template allows volunteers to select five different % cloudiness and surface wave levels (S1 Table). To have a workable sample size for each category, we reduced the classification into three levels of cloudiness: clear (0-5%), partly cloudy (includes mostly clear 25% and partly cloudy 50%) and cloudy (includes most cloudy 75% and cloudy 100%), and three levels of wave action: calm, some waves (rippled to small wavelets and few whitecaps), and whitecaps (frequent and many whitecaps).

There was no indication of greater error relative to turbidimeter estimates of average HydroColor (n=3) under cloudy conditions vs clear conditions (Table 2). Under all conditions, RMSE was greater than 4 NTU for v1 and decreased to around 3 NTU for v2 (cf. Table 1). Median % error was highest for cloudy conditions in v1. Replicate variability (RMSD) was around 1 for all conditions for v1. For v2, median % error did not vary but RMSD increased with cloudiness. Surprisingly, there was a trend for lower RMSE with increasing wave action for v1, but that might be confounded with higher between user variability in the "calm" category. The trend was clearer with v2 with increasing RMSE, median % error and RMSD as waviness increased. Nevertheless, in all cases the number of samples at the highest level of cloudiness or waviness was limited so the results in those categories should be regarded as tentative. The relative lack of data in these categories is not surprising, since volunteers are cautioned to confirm that conditions are safe before collecting observations.

Table 2. Accuracy and Variability of HydroColor for Different Surface Conditions

	V1			V2		
Condition	**RMSE**	**Median % error**	**RMSD**	**RMSE**	**Median % error**	**RMSD**
Clear	4.6 (89)[a]	36%	1.18 (251)[b]	3.1 (81)	27%	0.63 (181)
Partly Cloudy	6.1 (64)	40%	0.93 (128)	2.8 (81)	27%	0.90 (212)
Cloudy	4.1 (19)	61%	0.95 (103)	3.0 (39)	26%	1.31 (32)
Calm	7.6 (44)	39%	0.96 (105)	3.1 (30)	29%	0.66 (127)
Waves	4.9 (15)	13%	0.80 (393)	3.5 (166)	27%	0.89 (333)
Whitecaps	2.6 (9)	22%	0.80 (24)	3.9 (13)	36%	1.78 (18)

RMSE (root mean square error) and median % error of benchtop measurements vs turbidity estimated using the HydroColor validation regression (Fig 4), and variability of triplicate HydroColor measurements (RMSD) under different surface observation conditions for v1 (2022) and v2 (2023) of the app. Analysis excludes outlier group for 2023, (gray points Fig 4b).

[a]Number of validation samples for each condition

[b]Total number of samples for each condition

468 **Chlorophyll-a measurements**

469 Although measurement of IVChl has a great advantage of simplicity and speed over the
470 standard analytical practice of measuring Chl in acetone extracts of material concentrated on
471 filters, it is a property of live cells and is also subject to interference from degradation products.
472 Several of these factors cannot be controlled in any practical way, such as species composition,
473 nutrient status, and presence of degradation products such as pheophytin [19, 31]. Two factors
474 that can be controlled are temperature and the effect of previous light exposure. Fluorescence
475 yield is inversely related to temperature, the yield decreasing by about 1.0 % per degree C [32].
476 Thus, all measurements in Fig 5 were made after warming or cooling, as needed, the sample to
477 ambient lab temperature (22°C). A second source of variation for in vivo fluorescence is the
478 response of photosynthetic organisms to high-light exposure known as non-photochemical
479 quenching [31]. This is especially a concern for surface samples during the summer. The
480 quenching response protects the core complexes involved in photosynthesis, the reaction
481 centers, from damage due to oversaturation with light energy by redirecting much of the energy
482 absorbed by chlorophyll to be released as heat. When quenching is induced, in vivo
483 fluorescence, which mostly originates from reaction center complexes, decreases, even though
484 there is no change in the cellular content of Chl.

485 To control for the effects of quenching, samples were always placed in the dark at the
486 time of sampling up until measurement, typically a period of > 2 h. This did not typically cause
487 difficulties for participants, as this period is comparable to the time usually required to
488 transport field samples to a hub and arrange for their processing. The period of dark adaptation

was guided by the results of several time trials during the summer of 2022, measuring IVChl as soon as possible after sampling and periodically over a period of several hours. In cases when the IVChl increased with time, indicating that the initial sample was quenched, fluorescence increased 10-30% over a period of 2 h then stabilized (Fig 7a). This is consistent with reversal of fluorescence quenching. In one of the trials, fluorescence unexpectedly decreased, but the value stabilized after 5 hours. Fluorescence typically decreased during overnight storage.

Fig 7 Effect of Dark Storage on In Vivo Chl fluorescence. (a) Time courses of fluorescence measured in the Aquafluor for summertime (June-July) samples from the Rhode River, normalized to initial value, the horizontal line indicates a 1:1 ratio. **(b)** Ratio of Extracted Chl to in vivo Chl fluorescence (IVChl) as a function of the elapsed time between sampling and measurement on the Aquafluor. Ratios are normalized to the slope of the linear regression of Extracted Chl vs in vivo Chl fluorescence (Fig 5), the horizontal line indicating a normalized ratio of 1.

To verify that the results of this limited test applied generally to the project sampling, we tracked the time between sampling and measurement on the Aquafluor for 2022-2023 validation samples (data shown in Fig. 5). The ratio of extracted Chl to in vivo Chl as a function of dark storage time is shown in Fig 7b. The ratios are normalized to the slope of the linear regression with zero intercept (0.48). Almost all sampling followed the recommended 2 h storage before assay. The normalized extract to in vivo Chl ratio for most of these samples was between 0.85 (lower quartile) and 1.08 (upper quartile). There were a few samples (n=11) with < 2 h dark adaptation, these samples tended to have slightly elevated ratios (median 1.12),

510 consistent with some residual quenching of fluorescence. Overall, the results support that 2 h
511 dark storage minimizes the effect of quenching on the Aquafluor measurement of in vivo Chl
512 fluorescence.

513 As the IVChl measurement is performed on a small volume of sample containing the
514 suspended algal cells, there is some inherent variability between replicate measurements. The
515 RMSD of a triplicate determination is 3.53 µg L^{-1} (RE), which is 8.5% of the median value (Table
516 1). Settling of aggregates and/or colonial forms is one source of variability. Participant sample
517 processors are advised to gently mix the sample container before withdrawing an aliquot (~2
518 mL) to load in the cuvette. The measurement also depends on the proximity of cells to the
519 interrogated volume of the sample which is determined by the ~5 mm width of projected LED
520 emission beam. These sources of variability underscore why triplicate determinations improve
521 the accuracy of average IVChl as a linear estimator of Chl (Figure 5).

522 **CDOM measurements**

523 As a dissolved, abiotic constituent, CDOM fluorescence measurements have inherently
524 the less replicate variability than those of turbidity and IVChl. The RMSD for a triplicate
525 measurement with the Aquafluor is 0.43 µg L^{-1} QSE, 1.5% of the median value. In addition,
526 CDOM fluorescence measurements were stable for 3 days, as long as samples were refrigerated
527 (4°C) then re-warmed to room temperature (22°C) before measurement (data not shown, also
528 applies to Aquafast turbidity). Standardizing measurement at room temperature controls for
529 the temperature dependence in CDOM fluorescence which is about 1% per degree C [33]. Even
530 controlling for temperature effects, there is variability in the relationship between CDOM

fluorescence and absorbance, both at 300 nm and 440 nm. This variability occurs because the source of the CDOM and extent of its subsequent photochemical or microbial transformation in estuarine waters impact, in different ways, its absorption and fluorescence characteristics [e.g., 28, 34, 35]. These are the main contributors to the 10% median error for using fluorescence as a proxy for CDOM absorbance at 440 nm (Table 1).

Discussion and conclusions

Based on the 2+ years of sampling by participants in the Chesapeake Water Watch project, we found that volunteer-collected data using the HydroColor app and portable machines consistently estimated variables strongly affecting water clarity in Chesapeake Bay tributaries. These approaches are particularly useful for obtaining data, both field and remotely sensed, from the highly dynamic tidal tributaries many of which are severely undersampled for Chl, turbidity, and CDOM. Documenting short-term variation in these systems, especially in the spring, is important to understanding ecosystem functioning and responses to forcing variables [36]. Riverkeepers and other groups actively sample these areas for other variables such as dissolved oxygen, nutrients, and bacteria, while typically the only measure of water clarity is the Secchi depth. Chesapeake Water Watch's methods and instruments expand the volunteers' toolkit for understanding what is affecting water transparency (or lack thereof). Such information is critical in guiding actions to meet water clarity targets for the shallow regions of the Chesapeake. Improved water clarity in these areas is a key part of the Chesapeake Bay Program's efforts for sustaining healthy submerged aquatic vegetation and benthic habitat, algal growth, and overall ecosystem functioning [37].

For the tidal tributaries in the northern Chesapeake Bay, we found that the smartphone app HydroColor enabled volunteers to collect observations of turbidity with low replicate variability that corresponded well to paired measurements with the Aquafast benchtop turbidimeter. The HydroColor v2 estimates of turbidity were robust to variations in sky and surface wave conditions and were comparable among different users/smartphone models. These factors were all considered possible sources of variability by the App developers and have been cited as explanations for a large degree of scatter in the relationship between HydroColor and independent measurements in previous studies [21, 22]. Our results of a good correspondence between average HydroColor and Aquafast measurements contrast with the findings of Malthus et al. [24], who found no consistent relationship between HydroColor turbidity and gravimetric SPM assays. There may be several reasons for these contrasting results. Malthus et al. [24] used HydroColor in a wide variety of Australian lakes and rivers, which differed greatly in relative composition of optically active constituents. Our study was limited to the Chesapeake Bay where there is less variation in the types of materials affecting water optics. Also, our study compared HydroColor turbidity to Aquafast turbidity, not SPM. Because of differences in particle size and shape, the relationship between SPM and turbidity may have differed considerably among the water bodies studied by Malthus et al [24]. In contrast, our preliminary data indicates that turbidity and SPM are strongly correlated in the Chesapeake Bay (data not shown). Finally, the Malthus et al study used HydroColor v1, and our results show that v2 significantly improves the correspondence of HydroColor estimates with benchtop turbidity by reducing differences in reflectance estimates among smartphone models. Modifications to the user interface facilitating better positioning of the phone and image

capture also reduced the variability between replicate observations. While we have not made any systematic comparisons, we expect that HydroColor *Rrs* estimates are closer to those measured using research grade reflectance radiometers. Studies with v1 reported substantial differences between HydroColor and radiometer reflectance [23, 24]. However, a recent study of lakes in Hungary which worked directly with smartphone RAW data to derive reflectance in a similar manner as HydroColor v2, found good correspondence between smartphone camera and radiometer-based estimates of RGB *Rrs* [38].

Another result that became evident with the reduced scatter of points in v2 is a group of samples for which benchtop (Aquafast) turbidity was about twice of the HydroColor turbidity. Our preliminary observations are that these samples come from tributaries which received treated wastewater (Patuxent and Back rivers) or from shallow areas after late summer wind resuspension events. The common factor could be a higher percentage of particulate organic material with different scattering properties than usually encountered in coastal marine waters. This would change the relationship between Aquafast turbidity, which is based on 90° side scatter like most benchtop turbidimeters, and HydroColor turbidity, which is based on backscatter. An increase in the proportion of total scattering to backscatter linked to such an increase in organic content of the Chesapeake Bay particulates has been invoked to explain a long-term increase in total scattering in the bay without an increase in satellite *Rrs* [39]. Similarly, preliminary results show good agreement between HydroColor turbidity measurements and satellite retrievals of SPM from Landsat 8/9 OLI and Sentinel-2A/2B MSI that, similarly to HydroColor, are based primarily on the backscattering signal (*manuscript, in preparation*). We expect that additional observations will assist in better defining conditions

when Aquafast and HydroColor turbidity are different. Also planned are additional comparative measurements of turbidity and gravimetric SPM.

Our results support using the Aquafluor to obtain quantitative estimates of Chl and CDOM in the oligo- to mesohaline areas of the Chesapeake Bay which were the focus of project participants in the period 2021-2023. In validation samples, Chl was proportional to IVChl after sufficient dark adaptation. A 2h dark adaptation appears to be required when working with surface samples to consistently reverse the effects of non-photochemical quenching. Shorter periods may be sufficient for incubator-grown laboratory cultures, for example 3 min as suggested by Cremella et al. [20] for the Aquafluor measurements of IVChl. However, our results are consistent with studies using other models of fluorometers that have shown that 1 hour or more adaptation is required for recovery of IVChl from near-surface exposure to solar irradiance [31, 40].

Another consideration for IVChl in samples from turbid estuaries, is optical interference from turbidity and CDOM. In near-shore samples of the Rio de la Plata estuary, a cyanobacteria-dominated lake and urban stream in Uruguay, interference from turbidity and CDOM artifactually increased readings of IVChl by the Aquafluor [20]. In contrast, combining IVChl with turbidity or CDOM, or both, as explanatory variables did not improve the prediction of Chl in our validation samples (S3 Table). However, the Uruguayan study worked with samples with much higher turbidity (36 to 156 µg L^{-1}) and CDOM (1 to 13.6 a440 m^{-1}) than typically encountered in the northern Chesapeake Bay [20]. At the lower levels in our samples the positive interference to emission in the IVChl channel caused by turbidity scattering and CDOM

red fluorescence may be counterbalanced by the negative interference due to the same factors causing attenuation of the excitation irradiance. Additional testing for potential interference effects is advised when initiating volunteer projects using the Aquafluor in other water bodies, and will be continued as the Chesapeake Water Watch expands to other parts of the Chesapeake Bay. Although the Aquafluor model required to accurately measure both Chl and CDOM fluorescence on a single unfiltered sample is presently considered a more costly "custom build" by Turner Designs, our results are the first step towards showing that this, or similar systems [e.g., 7], using the same optical configuration should be a suitable for tandem assay of Chl and CDOM fluorescence across a wide range of environments.

In summary, the tools described in our report enable a more comprehensive monitoring of variables that are key to achievement of water clarity goals throughout the Chesapeake Bay. In particular, these approaches are well suited for improving spatio-temporal coverage of conditions in tidal tributaries and other shallow areas in the bay. Excessive levels of the water clarity variables considered here (Chl, turbidity, CDOM) are indicators of algal blooms, high suspended sediment and urban runoff, respectively, in these areas which are the focus of intensive recreational use, submerged aquatic vegetation, and aquaculture (e.g. oyster farms). These conditions can be detected from both frequent surface samples by volunteers as shown in this report, and from retrieval of these variables from high-resolution remote sensing images. The use of volunteer samples to ground-truth satellite remote-sensing retrieval algorithms is the subject of a future report (in preparation).

Supporting Information

S1 File. Developer Testing and User experience with HydroColor v2

S1 Table. Listing of the data entry types on the CWW Fieldscope site, serc.fieldscope.org

S2 Table. Intercomparison of in vivo chlorophyll-a (IVChl) and CDOM fluorescence readings for six identically calibrated Aquafluor units

S3 Table. Results of multiple linear regression calculations with extracted Chl as the dependent variable and selected volunteer measurements as the independent variables

S1 Fig. Multiple sample comparison of Aquafluor units SERC1 and SERC2 CDOM and IVChl measurements

Acknowledgments

Most importantly, we acknowledge the hard work and dedication of all our 104+ volunteers, without whom there would not be a Chesapeake Water Watch. Important contributions were also made by the staffs of the Advocates for Herring Bay, Arundel Rivers Federation, Gunpowder Riverkeeper, Marshy Point Nature Center, Severn River Association, ShoreRivers and Magothy River Association. We are also grateful for help from our undergraduate interns Zoe Kehs, Elizabeth Gaudlip, Diana Menendez, Elizabeth Ramos and Samantha Elliott, our high school interns Parker Franco, Lucas Kunkel, Sophia Huang, and Elias Arnold and our teacher interns, Robert Ross and Candace Taylor. Sample processing hub assistance was provided by Tammy Domanski at Anne Arundel Community College, Kathy Gramp at Advocates for Herring Bay, Patterson Clark, and Kate Moore. Rachael Mady provided helpful comments on the draft manuscript.

Author Contributions

Conceptualization, Funding Acquisition, Supervision: PJN, AC, MT. Investigation: PJN, SB, TS, BP. Data Curation, Project Administration: PJN, SB, TS. Writing: All authors.

References

1. Chesapeake Monitoring Cooperative. Tidal Methods Manual. Online 2016. Available from: https://www.chesapeakemonitoringcoop.org/wp-content/uploads/2018/08/CMC_Tidal-Water-Quality-Methods-Manual_v01-1.pdf.

2. Cao F, Tzortziou M, Hu C, Mannino A, Fichot CG, Del Vecchio R, et al. Remote sensing retrievals of colored dissolved organic matter and dissolved organic carbon dynamics in North American estuaries and their margins. Remote Sens Environ. 2018;205:151-65. doi: 10.1016/j.rse.2017.11.014.

3. Muller-Karger FE, Hestir E, Ade C, Turpie K, Roberts DA, Siegel D, et al. Satellite sensor requirements for monitoring essential biodiversity variables of coastal ecosystems. Ecological applications : a publication of the Ecological Society of America. 2018;28(3):749-60. Epub 2018/03/06. doi: 10.1002/eap.1682. PubMed PMID: 29509310.

4. Pahlevan N, Chittimalli SK, Balasubramanian SV, Vellucci V. Sentinel-2/Landsat-8 product consistency and implications for monitoring aquatic systems. Remote Sens Environ. 2019;220:19-29. doi: 10.1016/j.rse.2018.10.027.

5. Ramírez SB, van Meerveld I, Seibert J. Citizen science approaches for water quality measurements. Sci Total Environ. 2023;897. doi: 10.1016/j.scitotenv.2023.165436.

6. Busch J, Bardaji R, Ceccaroni L, Friedrichs A, Piera J, Simon C, et al. Citizen Bio-Optical Observations from Coast- and Ocean and Their Compatibility with Ocean Colour Satellite Measurements. Remote Sensing. 2016;8(11). doi: 10.3390/rs8110879.

7. Ceccaroni L, Piera J, Wernand MR, Zielinski O, Busch JA, Van Der Woerd HJ, et al. Citclops: A next-generation sensor system for the monitoring of natural waters and a citizens' observatory for the assessment of ecosystems' status. PloS one. 2020;15(3):e0230084-e. doi: 10.1371/journal.pone.0230084. PubMed PMID: 32214341.

8. Eitzel MV, Cappadonna JL, Santos-Lang C, Duerr RE, Virapongse A, West SE, et al. Citizen Science Terminology Matters: Exploring Key Terms. Citizen Science: Theory and Practice. 2017;2(1). doi: 10.5334/cstp.96.

9. Ellwood ER, Pauly GB, Ahn J, Golembiewski K, Higgins LM, Ordeñana MA, et al. Citizen science needs a name change. Trends Ecol Evol. 2023;38(6):485-9. doi: 10.1016/j.tree.2023.03.003.

10. Boomer K, Boynton WR, Muller A, Muller D, Sellner K. Revisiting Coastal Land-Water Interactions: The Triblet Connection. Edgewater, MD: Chesapeake Bay Program, Scientific and Technical Advisory Committee, 2019 Contract No.: 19-005. Available from:

https://www.chesapeake.org/stac/document-library/revisiting-coastal-land-water-interactions-the-triblet-connection/.

11. Testa JM, Dennison WC, Ball WP, Boomer K, Deidre DM, Lewis L, et al. Knowledge gaps, uncertainties, and opportunities regarding the response of the Chesapeake Bay estuary to restoration efforts. Edgewater, MD: Chesapeake Bay Program Scientific and Technical Advisory Committee, 2023 23-004. Available from: https://www.chesapeake.org/stac/wp-content/uploads/2023/05/23-004_Estuary-updated.pdf.

12. MarineGEO Upper Chesapeake Bay Observatory CPOP Data. [Internet]. Smithsonian Environmental Research Center. 2022. doi: 10.25573/serc.c.6100368.v10. Available from: https://smithsonian.figshare.com/collections/MarineGEO_Upper_Chesapeake_Bay_Observatory_Data/6100368.

13. Patuxent Sentinel: Chesapeake Biological Laboratory Pier Monitoring Program [Internet]. Chesapeake Biological Laboratory 2024. Available from: https://cblmonitoring.umces.edu/Data/Dashboard/2.

14. Harding LW, Mallonee ME, Perry ES, Miller WD, Adolf JE, Gallegos CL, et al. Long-term trends, current status, and transitions of water quality in Chesapeake Bay. Scientific Reports. 2019;9(1). doi: 10.1038/s41598-019-43036-6.

15. Chesapeake Monitoring Cooperative. Chesapeake Data Explorer. Online 2018. Available from: https://www.chesapeakemonitoringcoop.org/wp-content/uploads/2018/11/Chesapeake-Data-Explorer-Flyer.pdf.

16. Keisman J, Friedrichs C, Batiuk R, Blomquist J, J. Cornwell, Gallegos C, et al. Understanding and Explaining 30 Years of Water Clarity Trends in the Chesapeake Bay's Tidal Waters. 2019 STAC Publication Number 19-004. Available from: https://www.chesapeake.org/stac/document-library/understanding-and-explaining-30-years-of-water-clarity-trends-in-the-chesapeake-bays-tidal-waters/.

17. Brewin R, Brewin T, Phillips J, Rose S, Abdulaziz A, Wimmer W, et al. A Printable Device for Measuring Clarity and Colour in Lake and Nearshore Waters. Sensors. 2019;19(4). doi: 10.3390/s19040936.

18. Turner JS, Fall KA, Friedrichs CT. Clarifying water clarity: A call to use metrics best suited to corresponding research and management goals in aquatic ecosystems. Limnology and Oceanography Letters. 2022;8(3):388-97. doi: 10.1002/lol2.10301.

19. Roesler C, Uitz J, Claustre H, Boss E, Xing X, Organelli E, et al. Recommendations for obtaining unbiased chlorophyll estimates from in situ chlorophyll fluorometers: A global analysis of WET Labs ECO sensors. Limnol Oceanogr Methods. 2017;15(6):572-85. doi: 10.1002/lom3.10185.

20. Cremella B, Huot Y, Bonilla S. Interpretation of total phytoplankton and cyanobacteria fluorescence from cross-calibrated fluorometers, including sensitivity to turbidity and colored dissolved organic matter. Limnol Oceanogr Methods. 2018;16(12):881-94. doi: 10.1002/lom3.10290.

21. Leeuw T, Boss E. The HydroColor App: Above Water Measurements of Remote Sensing Reflectance and Turbidity Using a Smartphone Camera. Sensors (Basel, Switzerland). 2018;18(1):256. doi: 10.3390/s18010256. PubMed PMID: 29337917.

22. Gallagher JB, Chuan CH. Chlorophyll a and Turbidity Distributions: Applicability of Using a Smartphone "App" Across Two Contrasting Bays. J Coast Res. 2018;345:1236-43. doi: 10.2112/jcoastres-d-16-00221.1.

23. Yang Y, Cowen L, Costa M. Is Ocean Reflectance Acquired by Citizen Scientists Robust for Science Applications? Remote Sensing. 2018;10(6):835. doi: 10.3390/rs10060835.

24. Malthus TJ, Ohmsen R, Woerd HJvd. An Evaluation of Citizen Science Smartphone Apps for Inland Water Quality Assessment. Remote Sensing. 2020;12(10). doi: 10.3390/rs12101578.

25. Nechad B, Ruddick KG, Park Y. Calibration and validation of a generic multisensor algorithm for mapping of total suspended matter in turbid waters. Remote Sens Environ. 2010;114(4):854-66. doi: 10.1016/j.rse.2009.11.022.

26. YSI. EXO User Manual. Yellow Springs, OH: YSI Inc, 2023. Available from: https://www.ysi.com/file%20library/documents/manuals/exo-user-manual-web.pdf.

27. Arar EJ, Collins GB. Method 445.0 In Vitro Determination of Chlorophyll a and Pheophytin in Marine and Freshwater Algae by Fluorescence. In: Agency USEP, editor. Washington, DC1997.

28. Tzortziou M, Neale PJ, Osburn CL, Megonigal JP, Maie N, Jaffe R. Tidal marshes as a source of optically and chemically distinctive colored dissolved organic matter in the Chesapeake Bay. Limnol Oceanogr. 2008;53:148-59.

29. Jordan TE, Weller DE, Pelc CE. Effects of Local Watershed Land Use on Water Quality in Mid-Atlantic Coastal Bays and Subestuaries of the Chesapeake Bay. Estuaries and Coasts. 2017;41(S1):38-53. doi: 10.1007/s12237-017-0303-5.

30. Mobley CD. Estimation of the remote-sensing reflectance from above-surface measurements. Appl Opt. 1999;38(36). doi: 10.1364/ao.38.007442.

31. Suggett DJ, Prasil O, Borowitzka MA, editors. Chlorophyll a Fluorescence in Aquatic Sciences: Methods and Applications. New York: Springer; 2011.

32. Watras CJ, Morrison KA, Rubsam JL, Hanson PC, Watras AJ, LaLiberte GD, et al. A temperature compensation method for chlorophyll and phycocyanin fluorescence sensors in freshwater. Limnol Oceanogr Methods. 2017;15(7):642-52. doi: 10.1002/lom3.10188.

33. Watras CJ, Hanson PC, Stacy TL, Morrison KM, Mather J, Hu YH, et al. A temperature compensation method for CDOM fluorescence sensors in freshwater. Limnol Oceanogr Methods. 2011;9(7):296-301. doi: 10.4319/lom.2011.9.296.

34. Helms JR, Stubbins A, Ritchie JD, Minor EC, Kieber DJ, Mopper K. Absorption spectral slopes and slope ratios as indicators of molecular weight, source, and photobleaching of chromophoric dissolved organic matter. Limnol Oceanogr. 2008;53:955-69.

35. Tzortziou M, Zeri C, Dimitriou E, Ding Y, Jaffé R, Anagnostou E, et al. Colored dissolved organic matter dynamics and anthropogenic influences in a major transboundary river and its coastal wetland. Limnol Oceanogr. 2015;60(4):1222-40. doi: 10.1002/lno.10092.

36. Gallegos CL, Neale PJ. Long-term variations in primary production in a eutrophic sub-estuary: Contribution of short-term events. Estuar Coast Shelf Sci. 2015;162:22-34. doi: 10.1016/j.ecss.2015.01.015.

37. Scientific and Technical Advisory Committee (STAC). Achieving water quality goals in the Chesapeake Bay: A comprehensive evaluation of system response Edgewater, MD: 2023

Contract No.: 23-006. Available from: https://www.chesapeake.org/stac/wp-content/uploads/2023/05/CESR-Final-update.pdf.

38. Burggraaff O, Werther M, Boss ES, Simis SGH, Snik F. Accuracy and Reproducibility of Above-Water Radiometry With Calibrated Smartphone Cameras Using RAW Data. Frontiers in Remote Sensing. 2022;3. doi: 10.3389/frsen.2022.940096.

39. Gallegos CL, Werdell PJ, McClain CR. Long-term changes in light scattering in Chesapeake Bay inferred from Secchi depth, light attenuation, and remote sensing measurements. J Geophys Res-Oceans. 2011;116(C7):n/a-n/a. doi: 10.1029/2011JC007160.

40. Vincent WF, Neale PJ, Richerson PJ. Photoinhibition : algal responses to bright light during diel stratification and mixing in a tropical alpine lake. J Phycol. 1984;20:201-11.

Figures

Figure 1

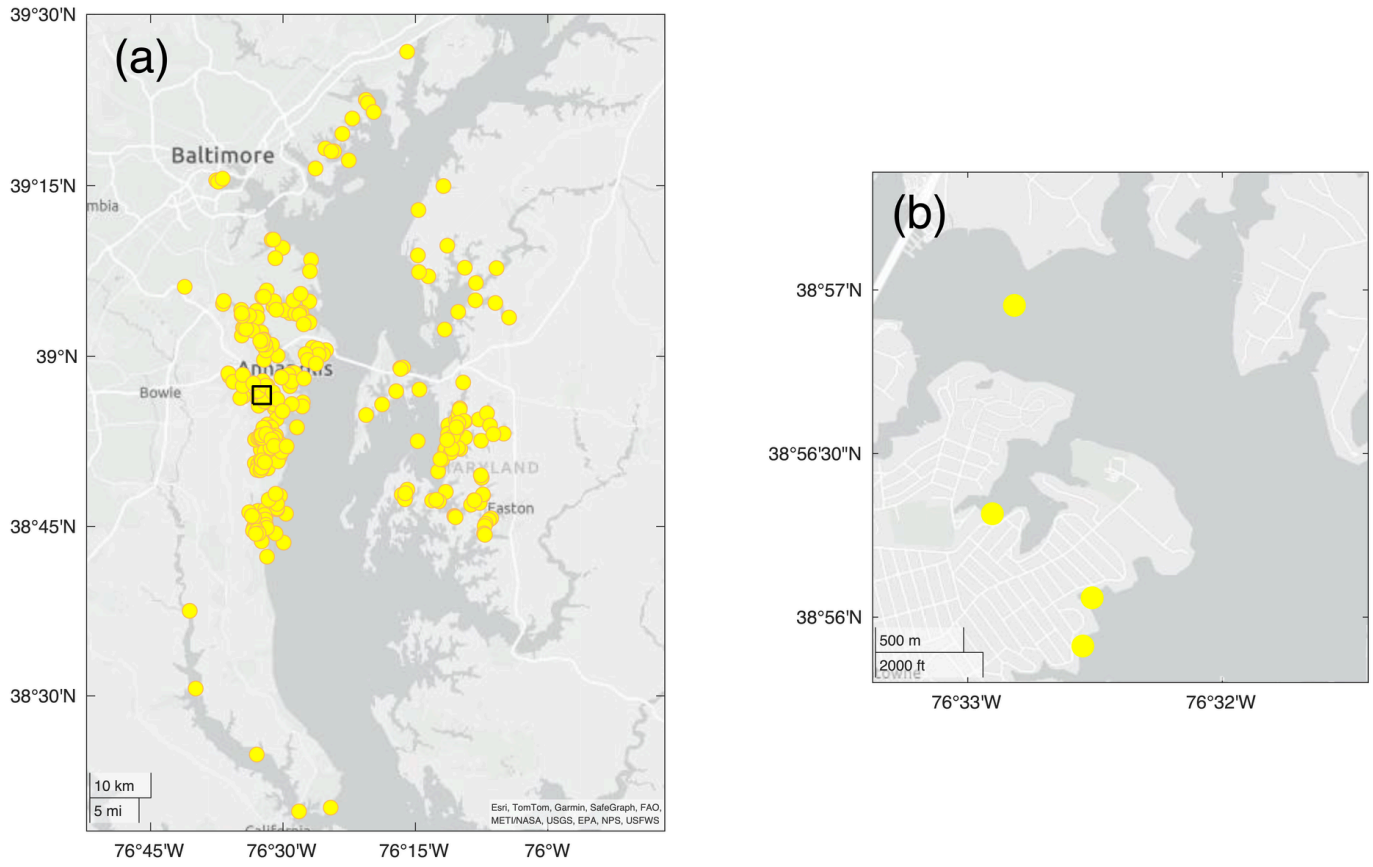


Fig 1. Chesapeake Water Watch Sampling Locations. (a) Map of the upper Chesapeake Bay in Maryland and Virginia showing sampling locations in 2021-2023. Four sites in Virginia are not shown in the map. Inset square indicates area of detail shown in (b) a higher resolution map showing 4 sampling locations in the South River near Edgewater, MD. The three nearshore locations are docks sampled regularly by a Chesapeake Water Watch volunteer (co-author BP); the mid-channel location nearby is sampled by the Arundel Rivers Federation riverkeeper organization.

Figure 2

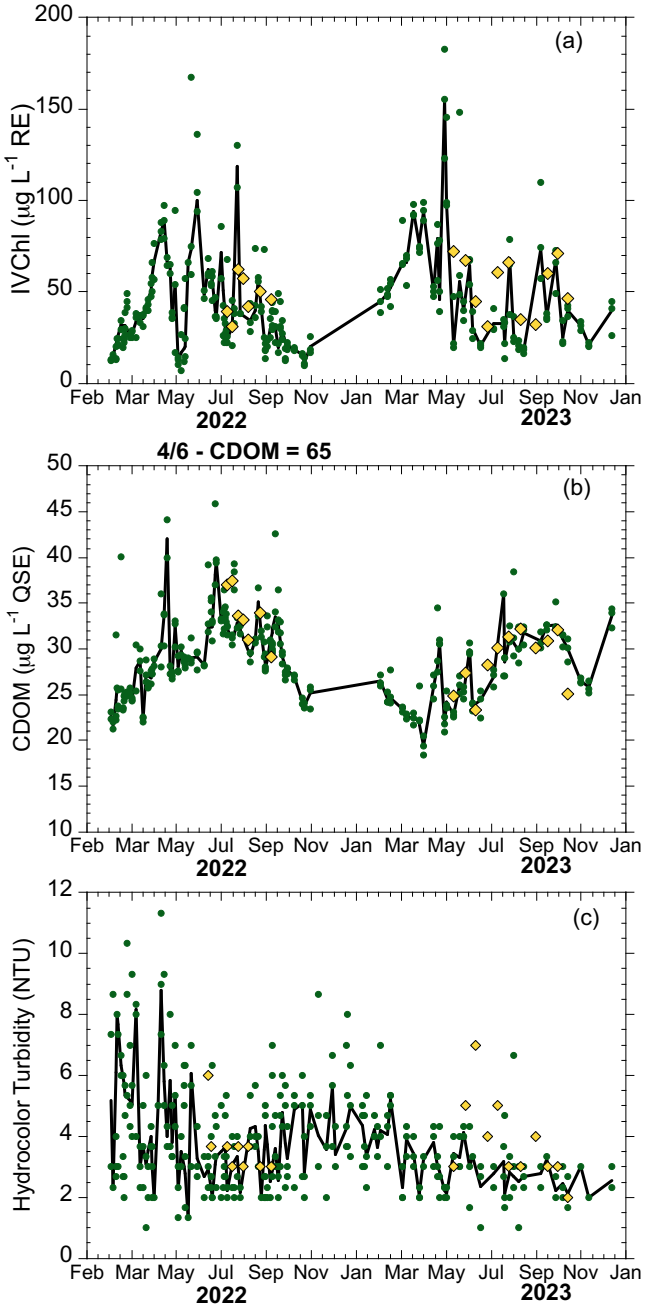


Fig 2. Examples of Participant Monitoring Measurements. Observations using project methods taken by a Chesapeake Water Watch volunteer (co-author BP) at three docks in the South River near Edgewater, MD (green dots, solid line shows average) and an independent set of measurements taken by the Arundel Rivers Federation riverkeeper at a nearby mid-channel site

(yellow diamonds, locations shown in Fig 1b). (a) in vivo chlorophyll fluorescence (IVChl) measured with the Aquafluor, (b) CDOM fluorescence measured with the Aquafluor, (c) Turbidity estimated using the HydroColor smartphone app. One volunteer point from April 6, 2022 was omitted from (b), on this date CDOM was unusually high (65 µg L^{-1} QSE) due to high amounts of tree pollen that had entered the water column.

Figure 3

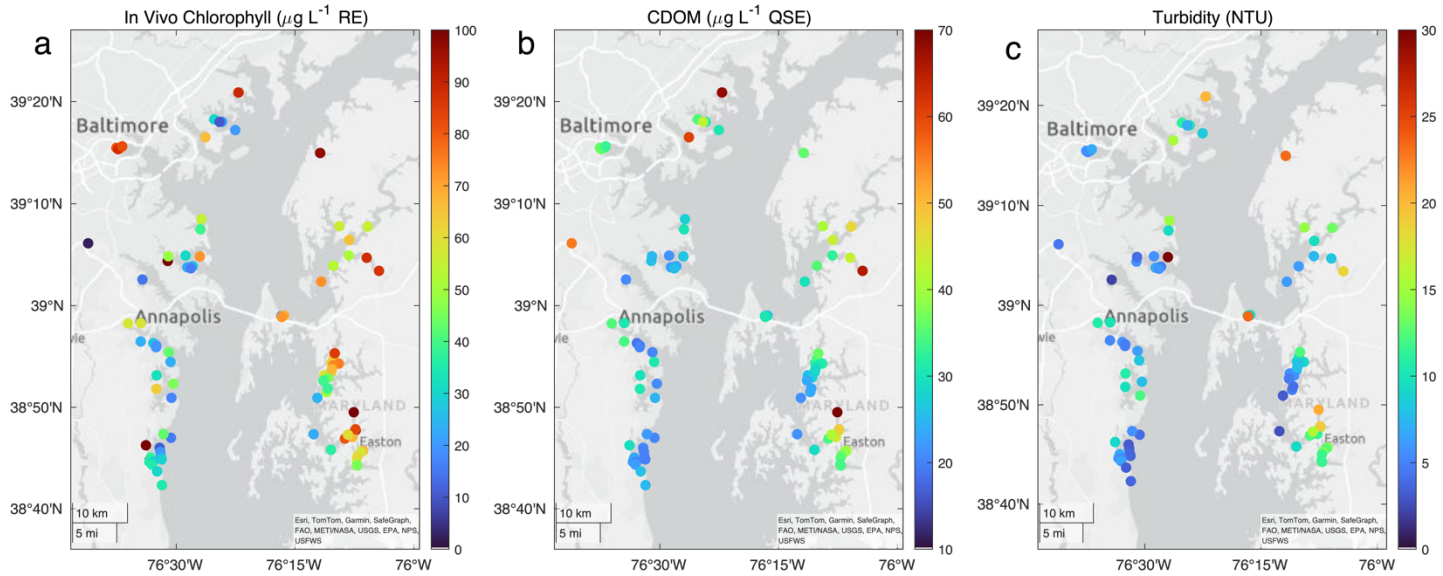


Fig 3. Maps of the upper Chesapeake Bay showing data from the July 14, 2023 Satellite and Samples Event. Points show location of samples and color coding on symbols using color bar on right denote values for (a) in vivo chlorophyll fluorescence, (b) CDOM fluorescence and (c) benchtop turbidity measurements.

821 **Figure 4**

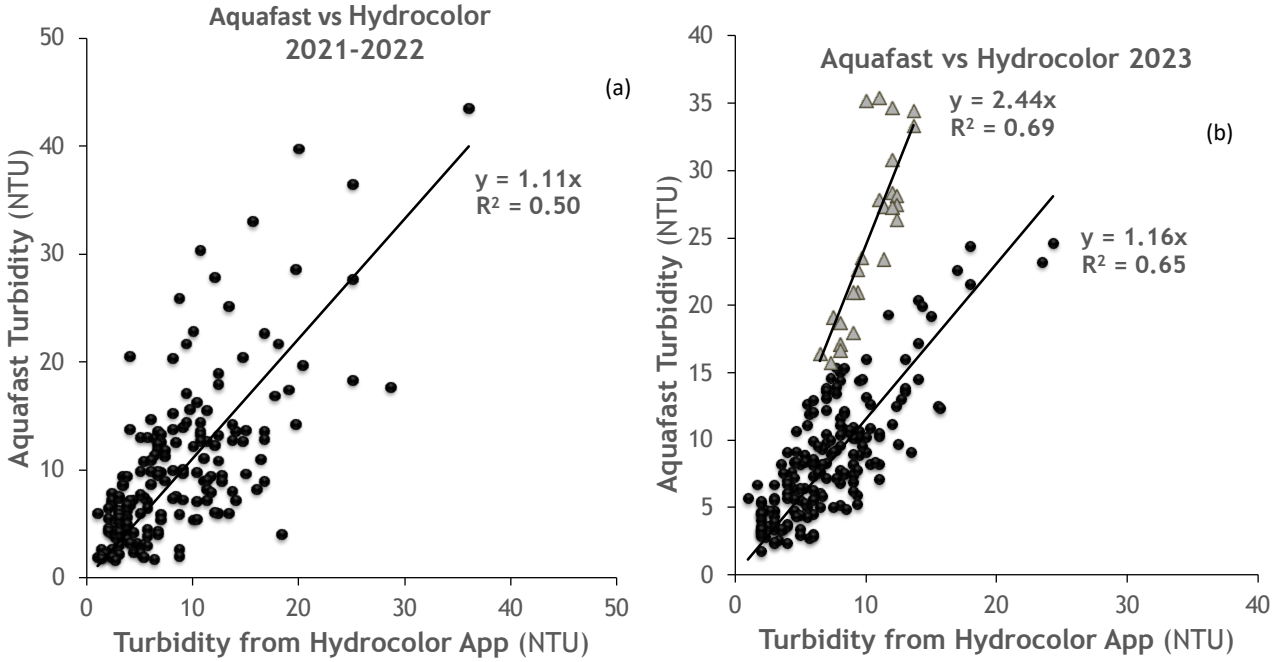
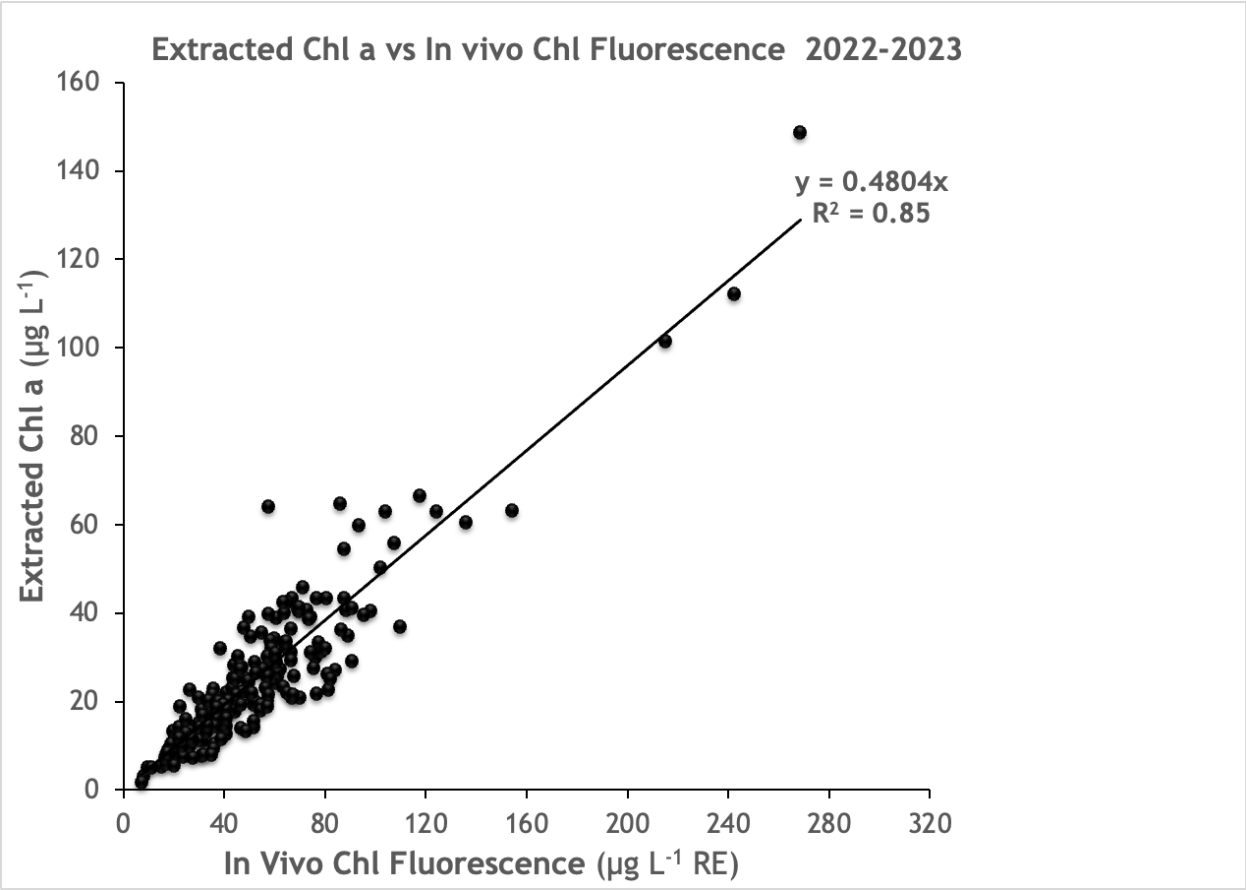


Fig 4. Comparison of HydroColor estimates of turbidity with benchtop measurements with the Aquafast. (a) app v1 used in 2021-2022 and (b) app v2 used in 2023. Gray triangles in (b) are a group of points with high Aquafast and low HydroColor turbidity (see text). Lines are the fitted linear regression between benchtop and HydroColor turbidity assuming a zero intercept.

822

823

824 **Figure 5**



825

826 **Fig 5 Relationship of chlorophyll a and in vivo chlorophyll fluorescence in validation samples.**

827 X-axis, IVChl in selected Chesapeake Bay samples measured on the Aquafluor (in rhodamine

828 equivalents, RE, µg L^{-1}) vs, y-axis, extracted Chl (µg L^{-1}). Complete data set of all validation

829 samples April 2022-December 2023 (n=200 samples, averages of triplicate measurements). Line

830 shows linear regression assuming a zero intercept.

831

Figure 6

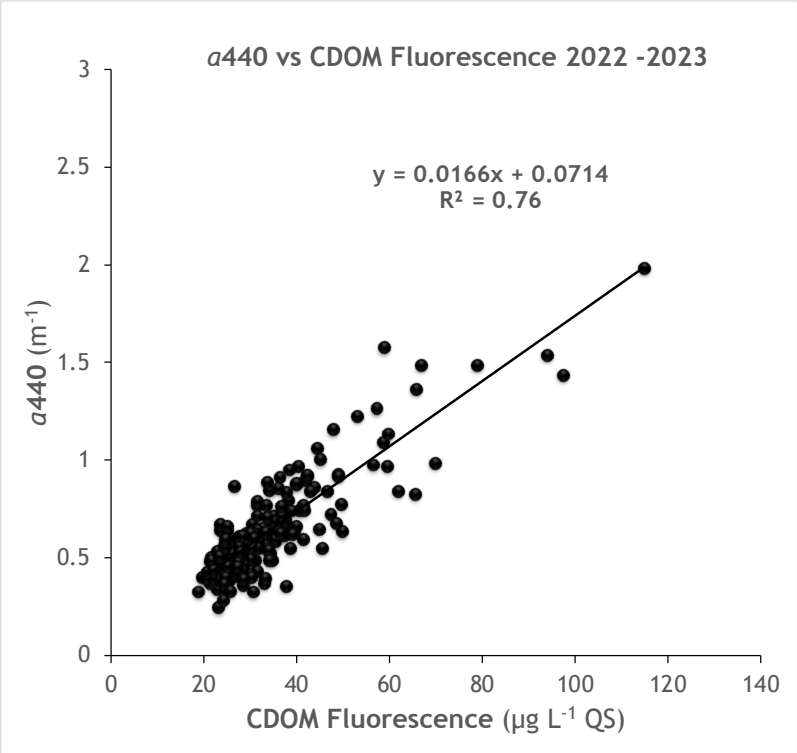


Fig 6. **Validation of Aquafluor CDOM measurement.** Relationship of CDOM dissolved organic matter fluorescence (µg L^{-1}) to absorbance of colored dissolved organic matter (CDOM) measured at 440 nm in a spectrophotometer (2022 & 2023 data). Line shows fitted linear regression with equation and R^2 annotated on the graph.

Figure 7

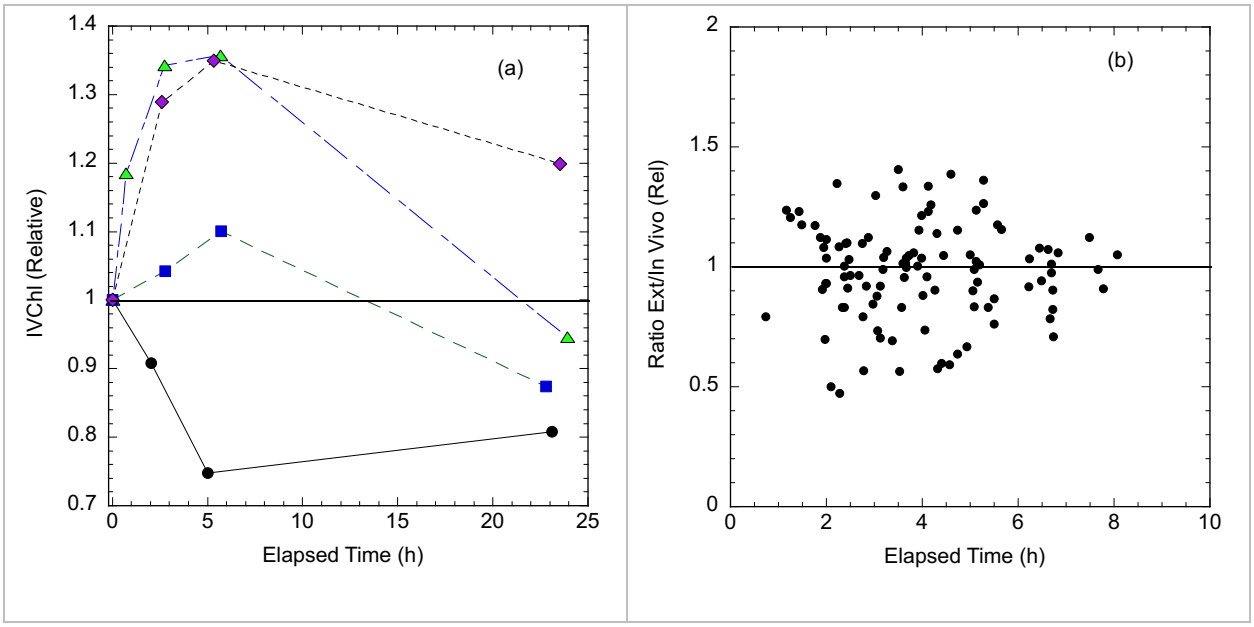


Fig 7 Effect of Dark Storage on In Vivo Chl fluorescence. (a) Time courses of fluorescence measured in the Aquafluor for summertime (June-July) samples from the Rhode River, normalized to initial value, the horizontal line indicates a 1:1 ratio. **(b)** Ratio of Extracted Chl to in vivo Chl fluorescence (IVChl) as a function of the elapsed time between sampling and measurement on the Aquafluor. Ratios are normalized to the slope of the linear regression of Extracted Chl vs in vivo Chl fluorescence (Fig 5), the horizontal line indicating a normalized ratio of 1.