Detecting phytoplankton diatom fraction based on the spectral shape of satellite-derived algal light absorption coefficient

Guangming Zheng,1,2* Paul M. DiGiacomo1

1NOAA/NESDIS Center for Satellite Applications and Research, College Park, Maryland
2Global Science & Technology, Inc., Greenbelt, Maryland

Abstract

Knowledge about phytoplankton composition is important for biological and biogeochemical research as well as for ecological applications (e.g., water quality) in coastal and inland waters. Satellite remote sensing can potentially map the baseline patterns, anomalies, and trends of phytoplankton composition on a synoptic basis. A prominent challenge is the attribution of the total optical signal to phytoplankton amid interference from minerals and humus. Here, we obtained the phytoplankton light absorption coefficient, \( a_{ph}(\lambda) \), in the Chesapeake Bay by partitioning satellite-derived total light absorption coefficient of water using the generalized stacked-constraints model (GSCM). We show that the red-to-blue band ratio of GSCM-derived \( a_{ph}(670)/a_{ph}(440) \) can be associated with diatom fraction in Chesapeake Bay. Further, the spatial-temporal patterns shown in the satellite-derived diatom fraction data agree well with field studies conducted previously around this region, including low diatom dominance in summer, high diatom dominance in the lower bay in winter, diatom-dominated spring blooms in coastal waters outside of the bay, and increasing seasonal variability of diatom fraction from the upper to the lower bay. We also found that in the middle bay the summer diatom fraction correlates strongly with spring streamflow on an annual basis, which can be explained because sediment deposited by spring freshets is the main source of silicate supply during summer. These results suggest that the satellite-derived diatom fraction maps can serve as a baseline for detecting phytoplankton composition anomalies, and highlight the effectiveness of using absorption-based approach to extract phytoplankton composition information for optically complex waters.

Phytoplankton play a fundamental role in aquatic ecosystems and their ecological functionality is determined by community compositions. Phytoplankton can be silicifiers, calcifiers, nitrogen-fixers, pico-autotrophs, dimethyl-sulfide-producers, etc. Understanding which groups dominate a given water is of vital importance to many fields of research such as marine biology and biogeochemistry. In coastal and inland waters, the study of phytoplankton composition may also benefit the assessment of nutrient loadings (e.g., Buchanan et al. 2005) because different algal loadings may respond to nutrient pulses in a different fashion (e.g., Loftus et al. 1972; Yeager et al. 2005). Thus information about phytoplankton composition may complement the use of only chlorophyll \( a \) concentration, [Chl \( a \)], as a surrogate for nutrient loadings (Schaeffer et al. 2012).

The Chesapeake Bay is a coastal plain estuary with a history of eutrophication (Kemp et al. 2005). Owing to its ecological and social-economic significance, the bay has been one of the most intensively investigated coastal regions on Earth. Among other topics, the seasonal succession of different phytoplankton groups in the bay has been well documented (D’Elia et al. 1983; e.g., Marshall and Nesius 1996; Buchanan et al. 2005; Marshall 2005). For instance, diatoms generally dominate throughout the year but less so during summer when a mixed population of diatoms, dinoflagellates, chlorophytes, cyanobacteria, and cryptophytes is common. However, to date, there has not been a synoptic view of phytoplankton composition that can be used for baseline assessment in this region. Such a baseline can be used to identify anomalies of phytoplankton composition, which can be further employed to evaluate past and current eutrophication status (e.g., Marshall et al. 2009) or to detect transient pulses of nutrient discharge. Currently, the only potentially viable approach to achieving these goals is through the use of remote-sensing data. Recent studies have demonstrated that on a global scale many satellite-derived...
optical variables covary with the trophic conditions in the water column and thus can be correlated with phytoplankton composition (e.g., Nair et al. 2008; Raitos et al. 2008; Kostadinov et al. 2010; Brewin et al. 2011; Hirata et al. 2011; IOCCG 2014). For example, changes in phytoplankton size classes were associated with the spectral shape of remote-sensing reflectance, \( R_n(\lambda) \) (Alvain et al. 2008), the [Chl \( a \)] (Sathyendranath et al. 2004; Hirata et al. 2011), and inherent optical properties (IOPs) (Ciotti and Bricaud 2006; Kostadinov et al. 2010).

However, these correlations are unlikely to hold in coastal waters such as Chesapeake Bay estuaries. For example, the correlation between algal biomass ([Chl \( a \)]) and composition breaks down in coastal regions; algal species from any size group can potentially bloom and dominate the biomass (e.g., Tyler and Seliger 1978; Tango et al. 2005; Anderson et al. 2010). In fact, we do not even have a universally accepted algorithm for accurately deriving [Chl \( a \)] from the \( R_n(\lambda) \) in coastal waters. The correlation between algal composition and some IOPs can also break down, such as the particulate backscattering coefficient which is contributed primarily by minerals rather than phytoplankton.

That said, the optical and biological characteristics of coastal and inland waters can potentially be used to remotely detect changes in phytoplankton composition. First, phytoplankton abundance in these regions is often large enough (e.g., Harding and Perry 1997) to generate detectable signal at bands that would otherwise be dominated by non-phytoplankton materials. For instance, high phytoplankton abundance brings their light absorption coefficient, \( a_{\text{phyt}}(\lambda) \), at 670 nm close (e.g., Magnuson et al. 2004) to the pure water absorption coefficient, \( a_w(\lambda) \), which typically dominate at this wavelength. Second, the ubiquitous algal species, diatom, has significantly lower absorption coefficient per unit [Chl \( a \)] compared with non-diatom species (Stuart et al. 2000; Magnuson et al. 2004; Sathyendranath et al. 2004) owing to differences in pigment packaging effect and intracellular pigment composition. This suggests that information carried in the spectral shape of phytoplankton light absorption coefficient can be traced back to its community composition. To retrieve this information, a key question that must be addressed is: How to isolate the optical signal that is attributable only to phytoplankton amid the interference from non-algal but optically significant materials?

A promising candidate is the generalized stacked-constraints model (GSCM) (Zheng et al. 2015b) developed for partitioning the total non-water absorption coefficient, \( a_{\text{nw}}(\lambda) \), into phytoplankton, \( a_{\text{phyt}}(\lambda) \), non-algal suspended particulate, \( a_g(\lambda) \), and colored dissolved organic, \( a_d(\lambda) \), components. Compared with other inversion models, an advantage of the GSCM is that no restrictive assumptions are imposed on the spectral shapes of various absorption components, thereby more likely to retain their natural variability. In this study, we applied the GSCM to satellite-derived \( a_{\text{nw}}(\lambda) \) data in the Chesapeake Bay, and derived a phytoplankton composition product based on the spectral shape of GSCM-derived \( a_{\text{phyt}}(\lambda) \). This product can help generate baseline maps of phytoplankton composition in the Chesapeake Bay.

**Data and methods**

To characterize phytoplankton abundance and composition in the Chesapeake Bay, field data of [Chl \( a \)] and speciesspecific phytoplankton cell counts were used. These data were matched up with \( R_n(\lambda) \) data from the Sea-Viewing Wide Field-of-View Sensor (SeaWiFS), the MODerate resolution Imaging Spectroradiometer (MODIS) onboard Aqua, the MEdium Resolution Imaging Spectrometer (MERIS), and the Visible Infrared Imaging Radiometer Suite (VIIRS). In view of the significant water quality impact from river inputs, we also compared satellite-derived variables with streamflow data calculated for major rivers in the Chesapeake Bay watershed.

**Field data of chlorophyll concentration**

The [Chl \( a \)] data used in this study were obtained from the Chesapeake Bay Program (CBP, www.chesapeakebay.net). For matchup purposes, [Chl \( a \) ] data collected between September 1997 and December 2015 were used in this study. Stations in the CBP were generally sampled monthly except during summer when sampling took place twice a month, and occasionally additional sampling was conducted after significant weather events. Figure 1 shows the locations of a total of 64 sampling stations that were used in this study, including those named after CB2, CB3, CB4, CB5, CB6, CB7, CB8, EE, ET, LE, TF, WE, and WT. We used only surface chlorophyll data (0–1 m), which were measured by Maryland Department of Health and Mental Hygiene, Chesapeake Biological Laboratory, Old Dominion University Applied Marine Research Laboratory, and Virginia Division of Consolidated Laboratory Services. These laboratories used consistent protocols to determine [Chl \( a \)], which is a monochromatic method conforming to the American Society for Testing and Materials (ASTM) and includes a correction for pheophytin. Here, we describe only the general procedures of this protocol. More details are provided by Olson (2012) and ASTM (2012). Specifically, each water sample was filtered through a GF/F filter and then subjected to acetone extraction. The acetone extract was then centrifuged and the clear supernatant was measured using a spectrophotometer to obtain the spectral optical density (OD) associated with light absorption. Subsequently, the acetone solution is acidified with HCl and re-measured for the OD. The [Chl \( a \)] is calculated based on the difference between the ODs before and after acidification.

**Field data of phytoplankton cell counts**

The cell counts data included in this study were provided by the Maryland Department of Natural Resources Tidewater Ecosystem Assessment Division. Surface phytoplankton
samples were collected once a month from January through December in 2000–2013 at four stations: Turkey Point (CB2.1), Sandy Point (CB3.3C), Cedar Point (CB5.1), and Key Bridge (WT5.1) station on Patapsco River (Fig. 1). Samples were not preserved and phytoplankton cell counts were made within 24 h using a calibrated Sedgewick-Rafter plankton counting chamber. Phytoplankton samples collected before 2002 were counted into four general groups: Diatoms, green algae, pigmented flagellates, and blue-green algae. Samples collected after 2003 were categorized into 17 groups: Dinoflagellates, Cryptophyceae, Raphidophyceae, Chrysophyceae, Filosea, Prymnesiophyceae, Choanoflagellida, Chlorophyceae, Euglenophyceae, Diatoms, Pelagophyceae, Dictyochophyceae, Cyanophyceae, Ebriidea, Kinetoplastidea, and Unidentified Flagellate. Diatom fraction was calculated as the number of diatom cells, \( N_{\text{Diatom}} \), divided by the total number of cells, \( N_{\text{Total}} \).

**Field data of river streamflow**

Estimated monthly streamflow entering the Chesapeake Bay was obtained from the U.S. Geological Survey, which was computed from streamflow measurements made at gauging stations on the three main rivers in the Chesapeake Bay watershed, i.e., Susquehanna, Potomac, and James Rivers (Fig. 1). The streamflow is a measure of the volume of water flowing past a given point in the stream in a given period of time, reported in m\(^3\) s\(^{-1}\) in this study. Three streamflow parameters were used in this study: The monthly averaged streamflows from the Susquehanna River and the Potomac River, as well as the total streamflow of all rivers discharging into the Chesapeake Bay. Detailed methods of how these values were calculated are described by Bue (1968). Data were obtained from https://md.water.usgs.gov/waterdata/chesinflow/data/monthly.

**Satellite data of remote-sensing reflectance**

The satellite-derived remote-sensing reflectance \( R_{\rho}(\lambda) \) used in this study is calculated as the normalized water-leaving radiance, \( nL_w(\lambda) \), divided by the extraterrestrial solar irradiance, \( F_0(\lambda) \). We used whole-mission \( nL_w \) data from SeaWiFS (1997–2010) and MERIS (2002–2012), data collected from MODIS-Aqua during 2002–2015, and data collected from VIIRS during 2012–2016.

Satellite-derived \( nL_w(\lambda) \) data used in this study were generally the latest versions of reprocessing as of November 2016. Specifically, we used NASA version “R2014.0” for SeaWiFS, “R2012.1” for MERIS, and “R2014.0.1” for MODIS-Aqua data, and NOAA version “SCI_OC03_V1.10” for VIIRS data. Several atmospheric correction approaches were used in generating these data. The SeaWiFS, MERIS, and MODIS-Aqua data obtained from NASA were generated using the near infrared (NIR) bands as a basis for making the “black-pixel” assumption, hereafter referred to as NIR-corrected data (see references cited in http://oceancolor.gsfc.nasa.gov/cms/reprocessing/AtmoCor.html). With respect to the atmospheric correction of VIIRS, three streams of data produced by NOAA were used in this study. One stream adopts the so-called “BMW”, standing for Bailey, MUMM (Management Unit of the North Sea Mathematical Models), and Wang, atmospheric correction algorithm (Jiang and Wang 2014) which is a variant of the NIR-based scheme and combines the three algorithms developed by Bailey et al. (2010), Ruddick et al. (2000), and Wang et al. (2012). Another stream of VIIRS data were processed with an atmospheric correction scheme that replaces the NIR bands with shortwave infrared (SWIR) bands, hereafter referred to as SWIR-corrected data (described by Wang et al. 2007). The third stream of VIIRS

![Fig. 1. Sampling map of the Chesapeake Bay. Gray dots denote the CBP stations where [Chl a] data were collected. Black diamonds mark the phytoplankton cell count data sampling locations of the Maryland Department of Natural Resources. The three black boxes define the upper, middle, and lower parts of the bay for subsampling the satellite data. The gray line depicts the mainstem transect from the Susquehanna River mouth to 200 km seaward with every 50 km marked with a cross sign.](image-url)
data were produced with a NIR-SWIR combined correction algorithm (Wang and Shi 2007).

Three streams of VIIRS data with different atmospheric correction schemes were used to deal with the issue of missing data in the application of GSCM, i.e., no feasible solutions identified for a significant portion (around 40–90%) of the pixels owing to the quality of satellite-derived \( R_\text{rs}(\lambda) \) data and limitations of inversion models. Even using a cloud-free \( R_\text{rs}(\lambda) \) image as the input, the output image of GSCM-derived variables may end up void of data. Therefore, to improve the data coverage, we have merged the results from the three different VIIRS data streams. These data are used to generate multi-year seasonal averages which represent baselines or climatologies of the bay optical properties.

**Satellite data processing**

To apply the GSCM in the remote-sensing context, an “upstream” model that derives \( a_{\text{wp}}(\lambda) \) from \( R_\text{rs}(\lambda) \) is required because the model requires the total \( a_{\text{nw}}(\lambda) \) as the input. In this study, the quasi-analytical algorithm (QAA) (Lee et al. 2002) version 6 (http://www.ioccg.org/groups/software.html) was chosen as the upstream model. The main reason for choosing the QAA is because it is conceptually consistent with the GSCM in terms of the sequence of deriving total and component absorption coefficients. In our opinion, it makes little sense to repartition a spectrum of \( a_{\text{nw}}(\lambda) \) that was calculated by summing up component absorption coefficients derived based on restrictive assumptions, which would be the case if, for example, numerical optimization-based models are used as the upstream model.

In the implementation of QAA, we used the 670 nm as the only reference band. We found this approach more suitable for the Chesapeake Bay waters than using 550 nm as the reference band (results not shown) and it ensures smooth model outputs when \( R_\text{rs}(670) \) changes across 0.0015 sr\(^{-1}\), where the original QAA version 6 requires a band switch and introduces unwanted, abrupt changes in outputs. With respect to the Raman scattering effect on \( R_\text{rs}(\lambda) \) and the temperature/salinity effect on pure water optical properties, no corrections were applied because they are considered insignificant for the spectral range and water properties addressed in this study.

After obtaining the QAA-derived \( a_{\text{nw}}(\lambda) \), we partitioned it using the GSCM to obtain the phytoplankton component, \( a_{\text{ph}}(\lambda) \). Here, we highlight two unique features of the GSCM that makes it highly relevant to the study of phytoplankton composition. First, the nonalgal absorption coefficients \( a_q(\lambda) \) and \( a_d(\lambda) \) are modeled with realistic spectral shapes determined from field data with subtle slope variations across the spectrum, as opposed to idealized exponential functions of \( \lambda \) (with fixed or variable slopes). This is important because \( a_d(\lambda) \) and \( a_q(\lambda) \) account for a significant portion of the total absorption coefficient and must be properly estimated, otherwise the derived \( a_{\text{ph}}(\lambda) \) will likely be questionable. Second, the inequality constraints used in the GSCM allow the spectral shape of derived \( a_{\text{ph}}(\lambda) \) to vary in a large range, making it possible to observe the natural variations in \( a_{\text{ph}}(\lambda) \) associated with changes in phytoplankton composition.

**Results and discussion**

Before discussing results on satellite-derived phytoplankton data, we first examined whether the contributions from different water constituents to \( a_{\text{wp}}(\lambda) \) have been attributed properly. This is important because the optical properties of the Chesapeake Bay are driven by multiple factors including fluvial inputs of minerals and humic matter, sediment resuspension, as well as local growth of phytoplankton. These drivers are further driven by processes that are independent from one another. As a result, there is no reason why different optical components, namely \( a_{\text{ph}}(\lambda) \), \( a_q(\lambda) \), \( a_d(\lambda) \), and \( b_{\text{bp}}(\lambda) \) should always covary, as is implicitly implied by algorithms that use the same \( R_\text{rs}(\lambda) \) band ratios to derive different variables with the sole difference of regression coefficients.

To verify the effectiveness of applying GSCM to QAA-derived \( a_{\text{nw}}(\lambda) \) based on satellite \( R_\text{rs}(\lambda) \) data, it is a good practice to check whether the GSCM-derived absorption components have been decoupled from one another, and whether the exhibited spatial/temporal patterns are real signals as opposed to artifacts introduced by sensor calibration and model errors. This is especially valuable in the absence of field-measured and quality-verified IOPs data that can be matched up with satellite \( R_\text{rs}(\lambda) \).

For this analysis, we processed VIIRS 2012–2016 monthly composite images of BMW-corrected, SWIR-corrected, and NIR-SWIR-corrected \( R_\text{rs}(\lambda) \) with QAA and GSCM, and merged monthly maps of GSCM-derived \( a_{\text{ph}}(\lambda) \), \( a_q(\lambda) \), \( a_d(\lambda) \), and QAA-derived \( b_{\text{bp}}(\lambda) \) using the three streams of data. These monthly maps were further averaged by month to obtain seasonal climatologies for the period of 2012–2016. To highlight spatial features exhibited in the seasonal climatologies, we subsampled the data by running a two-dimensional moving average along the Chesapeake Bay mainstem transect. The size of the moving window was chosen as 10 km along the same latitude and 5 km along the same longitude. Example results of particulate optical properties at 440 nm are shown in Fig. 2.

Figure 2 demonstrates that the spatial distribution of GSCM-derived \( a_{\text{ph}}(440) \) and \( a_d(440) \) coefficients are generally out of phase except in the fall, i.e., they do not necessarily covary with each other. Importantly, spatial patterns of \( a_{\text{ph}}(440) \) and \( a_d(440) \) along the transect are consistent with known features in this region. Phytoplankton maintain a high biomass in clearer waters seaward of the turbidity maximum in spring and summer, i.e., when turbidity has decreased. These features were reported by field studies (Fisher et al. 1988; Roman et al. 2001; Keller et al. 2014), lending confidence that the attribution of total absorption
Coefficients to algal and nonalgal components was properly made by GSCM.

Comparisons between $a_{d}(440)$ and the other two coefficients, $a_{ph}(440)$ and $b_{bp}(440)$, provide insights on absorption budgets of nonalgal suspended particles including mineral and organic components. Here, we can use the $b_{bp}(440)$ as a proxy for minerals because light scattering in this turbid estuary is generally mineral-dominated. During spring and winter, the spatial distribution of $a_{d}(440)$ is in phase with that of $b_{bp}(440)$ (Fig. 2a,d), suggesting that nonalgal particulate absorption might be dominated by minerals. This is consistent with the timing of spring freshet which delivers significant amount of mineral particles to the bay, and the wind-induced vertical mixing during winter which resuspends sediments from the bottom (Zheng et al. 2015a and references cited therein). We also notice a strong coupling between $a_{d}(440)$ and $a_{ph}(440)$ in fall (Fig. 2c), but are unaware of any field studies that reveal this particular feature. It remains to be investigated in future research whether the coupling is induced by dominance of phytoplankton-derivered organic detritus in suspended particles (a causation), or by other unknown causes that concurrently introduce both nonalgal particles and phytoplankton growth (a correlation).

Based on analyses made above, we tentatively conclude that the GSCM is able to extract the signal attributable to phytoplankton from the total signal measured by satellite which is influenced by both algal and nonalgal materials. A key question that follows is: Does the satellite-derived $a_{ph}(\lambda)$ carry sufficient signal-to-noise ratio to allow the

---

**Fig. 2.** Seasonal mean particulate optical properties along the Chesapeake Bay mainstem derived from VIIRS during the period of 2012–2016: (a) Spring (March-April-May); (b) Summer (June-July-August); (c) Fall (September-October-November); and (d) Winter (December-January-February). The algal and nonalgal particulate absorption coefficients, $a_{ph}(440)$ and $a_{d}(440)$, respectively, were obtained by partitioning the QAA-derived $a_{nw}(\lambda)$ with the GSCM. The total particulate backscattering coefficient, $b_{bp}(440)$, were derived with the QAA. The bay head refers to the Susquehanna River mouth.
discrimination of different phytoplankton groups from its spectral shape? To answer this question, we analyzed the matchups between SeaWiFS, MODIS-Aqua, MERIS, and VIIRS (BMW-correct only) data and field-measured [Chl a] from the CBP. Figure 3a,b show the satellite-derived $a_{ph}(\lambda)$ at the blue and red phytoplankton absorption peaks in relation to field-measured [Chl a]. Here, we separated the entire set of matchups into three arbitrary groups based on the red-to-blue $a_{ph}(\lambda)$ band ratio, $a_{ph}(670)/a_{ph}(440)$. Relative to the 440 nm band, the $a_{ph}(\lambda)$ at 670 nm is subject to smaller influences from pigment-packaging effect and accessory pigments (Morel and Bricaud 1986; Bricaud et al. 1988). As a result, $a_{ph}(670)$ correlates better with [Chl a] than $a_{ph}(440)$ does and the $a_{ph}(670)/a_{ph}(440)$ ratio tends to be inversely correlated with the $a_{ph}(440)/[Chl a]$ (e.g., Stuart et al. 2004), i.e., the [Chl a]-specific absorption coefficient of phytoplankton at 440 nm.

To verify whether we can see this trend in satellite-derived data, we conducted linear regression analysis for each individual group identified by the value of the $a_{ph}(670)/a_{ph}(440)$ ratio. Figure 3 shows a systematic increase in the slope $a_{ph}(670)/a_{ph}(440)$ as the $a_{ph}(670)/a_{ph}(440)$ ratio decreases, which is expected. We also

Fig. 3. Satellite-derived phytoplankton absorption coefficient at (a) 440 nm and (b) 670 nm plotted vs. field-measured [Chl a] obtained from the CBP. For comparison, field-measured results of the same variables from NOMAD dataset are also shown (c, d). $n$, number of matchups; $r$, correlation coefficient. $R^2 = a_{ph}(670)/a_{ph}(440)$. $a_{ph}^* = a_{ph}/[Chl a]$. 
examined the Chesapeake Bay in situ data obtained from the NOMAD dataset in a similar fashion as we did for the satellite matchup data (Fig. 3c,d). Although the ranges of \( a_{\text{ph}}(670)/a_{\text{ph}}(440) \) in the NOMAD data is different from the matchup data, it is evident that the trend is consistent at 440 nm (Fig. 3a,c). The NOMAD field data also show the same trend at 670 nm (Fig. 3d), whereas the matchup data show random trends at this wavelength (Fig. 3b), suggesting that the relatively small pigment-packaging effect at 670 nm cannot be detected using satellite data with the present level of accuracy. Nonetheless, the ratio of \( a_{\text{ph}}(670)/a_{\text{ph}}(440) \) does appear promising for detecting trends of \( a_{\text{ph}}^* \)(440). The \( a_{\text{ph}}(670)/a_{\text{ph}}(440) \) ratio can be derived independently of the magnitude of \( a_{\text{ph}}(\lambda) \) using our approach, whereas the \( a_{\text{ph}}^* \)(440) are known to differ significantly between diatom and non-diatom phytoplankton groups (Stuart et al. 2000; Magnnson et al. 2004; Sathyendranath et al. 2004). Therefore, it is possible to differentiate diatom and non-diatom groups using the satellite-derived \( a_{\text{ph}}(670)/a_{\text{ph}}(440) \) ratio.

To test this hypothesis, we matched up field-measured phytoplankton cell counts with satellite data from the four sensors. A total of four matchups were identified for SeaWiFS (2000/10/10 at CB5.1, 2002/7/31 and 2007/10/17 at CB2.1, and 2002/10/18 at WT5.1) and four for MODIS (2003/10/07 and 2007/5/9 at CB3.3C, 2006/6/5 at CB 5.1, and 2011/2/10 at WT5.1), with no matchups identified for the other two sensors. Despite the limited number of data points, Fig. 4 shows a clear correlation between field-measured diatom fraction and satellite \( a_{\text{ph}}(670)/a_{\text{ph}}(440) \) derived using QAA and GSCM. The linear regression relationship between diatom fraction and the red-to-blue phytoplankton absorption band ratio is:

\[
N_{\text{Diatom}}/N_{\text{Total}} = 2.353 a_{\text{ph}}(670)/a_{\text{ph}}(440) - 0.459, r^2 = 0.85
\]

(1)

We also examined the fractions of other phytoplankton groups such as dinoflagellate and cyanobacteria but did not find a strong correlation with the \( a_{\text{ph}}(670)/a_{\text{ph}}(440) \) ratio. Admittedly, the number of matchups contributing to Eq. 1 is quite limited and the coefficients are likely to change when more field data are available. However, we believe that the general trend exhibited in Fig. 4 is robust in a sense that higher ratio of \( a_{\text{ph}}(670)/a_{\text{ph}}(440) \) generally implies more diatom dominance in a phytoplankton assemblage owing to the uniqueness of average diatom optical properties (Stuart et al. 2000; Magnnson et al. 2004; Sathyendranath et al. 2004). Whereas the absolute magnitudes of diatom fraction resultant from application of Eq. 1 to satellite data may be subject to adjustment in future research, the spatial and temporal trends exhibited in the satellite-derived diatom fraction maps should be robust.

As such we applied Eq. 1 to the VIIRS monthly composite data merged with three different atmospheric correction schemes and obtained monthly diatom fraction time series of the Chesapeake Bay and surrounding waters. Next, we examined this dataset from different spatial and temporal perspectives and compared the results against known features in this region. To provide an overview of the data, we first show seasonal climatological maps of diatom fraction (Fig. 5a–d). For comparison, we also show the seasonal [Chl a] maps (Fig. 5e–h) derived based on the magnitude of \( a_{\text{ph}}(670) \) using the formula given by Zheng and DiGiacomo (2017),

\[
[\text{Chl } a] = a_{\text{ph}}(670)/0.010
\]

(2)

which is based on a linear regression analysis using all data in Fig. 3b. One of the most important features shown in Fig. 5 is the decoupled spatial patterns between diatom fraction and [Chl a] derived using methods described above. The [Chl a] generally decreases seaward regardless of season, whereas the diatom fraction exhibits much more diverse spatial and temporal patterns. Furthermore, features of diatom fraction revealed in Fig. 5 are unlikely to be artifacts because they are consistent with literature reports and field data (Fig. 6). The field data used here only cover a few spots once a month in the upper (CB2.1, WT5.1, and CB3.3C) and middle bay (CB5.1) and thus are statistically less representative than satellite data. Indeed, Fig. 6 shows that field-measured diatom fraction data exhibit large inter-station and inter-annual variability. Nonetheless, the field data do provide a glimpse of the actual conditions. Both Figs. 5, 6 show that the diatom fraction is generally low in summer relative to spring and fall across the Chesapeake Bay, which has been reported previously by many investigators (e.g., Marshall 2005; Adolf et al. 2006). Figure 5a also shows diatom-dominated spring blooms outside of the Chesapeake Bay which are well known...
across the entire Atlantic Bight shelf waters (Townsend et al. 2006). The winter diatom dominance in the lower Chesapeake Bay which was discovered before (Patten et al. 1963; Marshall and Nesius 1996) is also evident in Fig. 5d. Our diatom fraction maps actually exhibit the strongest diatom dominance in the lower bay in winter amongst all seasons, which has not been reported before. We thus conclude that the ecological features identified from satellite-derived diatom fraction data are reasonable and Fig. 5 can be used as a baseline for characterizing diatom fraction in this region.

Next, we present the data from a seasonal perspective by assessing the monthly changes in satellite-derived diatom fraction for the upper, middle, and lower Chesapeake Bay (defined in Fig. 1). Figure 7 shows that the temporal variability of diatom fraction gradually increases from the upper to the lower bay. Considering that our satellite-derived diatom fraction is equivalent to the $a_{\text{ph}}(670)/a_{\text{ph}}(440)$ ratio, this result is consistent with a field study in the same region (Magnuson et al. 2004). Using field data collected during 1996–2000, Magnuson et al. (2004) calculated the seasonal variability of average $a_{\text{ph}}(\lambda)$ in the visible spectral range in terms of the $F$-value which statistically quantifies the inter-season relative to intra-season variability. They showed that across spring, summer, and fall, the variability represented by the $F$-value increases from 0.45 in the upper bay, to 5.92 in the middle bay, and 19.47 in the lower bay. Our results confirmed their findings and further provide a more complete picture of temporal variations in phytoplankton absorption spectral shape and diatom fraction in different regions of the bay. Note that satellite-derived diatom

Fig. 5. Seasonal climatology of diatom fraction (a–d) and [Chl a] (e–h) in the Chesapeake Bay derived from VIIRS data during the period of 2012–2016. The diatom fraction is calculated from GSCM-derived $a_{\text{ph}}(670)/a_{\text{ph}}(440)$ ratio using Eq. 1. The [Chl a] is calculated based on GSCM-derived $a_{\text{ph}}(670)$ using Eq. 2.
fraction tends to drop in winter in the upper bay (Fig. 7a), which is consistent with the general trends shown in field data (Fig. 6a–c).

With respect to inter-annual variability, we compared the satellite-derived diatom fraction data with USGS in situ measured river streamflow data. The streamflow data were chosen because freshwater and river-borne nutrients and sediments are main drivers of aquatic ecosystem dynamics in this estuary. As a first attempt, we focused on the diatom fraction in summer when the spatial coverage of GSCM-derived phytoplankton data is the best. Correlation coefficients among various combinations of time averages are shown in Table 1. For each of the three sections of bay, only rivers located upstream of that section were taken into account.

Fig. 6. Field data of diatom fraction measured at four sampling stations (Fig. 1). In each individual box plot symbol, the gray box spans from the first quartile to the third quartile (the interquartile range) of all data sampled in the same month during 2000–2013. The black bar inside the box shows the median. The dashed “whiskers” and error bars above and below the box show the minimum and maximum.
account in this analysis. In other words, for the upper bay only the streamflow of the Susquehanna River was included; for the middle bay, the total streamflow of both the Susquehanna and Potomac Rivers was used; whereas for the lower bay, the streamflow used in Table 1 represents all rivers that discharge into the entire bay. We found that the correlation is strongest in the middle bay between average summer diatom fraction and spring (April–May) streamflow. The correlations between diatom fraction and river streamflow for the upper and lower bay are much weaker. Figure 8 presents the satellite-derived diatom fraction plotted against field-measured streamflow data when their correlation is the strongest for each region. These results can be explained by silicate limitations on diatom. There are two potential sources of silicate in the Chesapeake Bay: River discharge and benthic regeneration (D’Elia et al. 1983; Conley and Malone 1992; Malone et al. 1996). In summer, benthic regeneration is the dominant source of silicate in the middle bay (Conley and Malone 1992), which was deposited to the bottom of the bay during spring freshet and subsequently recycled in summer. This explains the strong correlation between summer diatom fraction and spring streamflow in the middle bay.

Fig. 7. Monthly time series of satellite-derived diatom fraction in the (a) upper, (b) middle, and (c) lower Chesapeake Bay. See Fig. 1 for definitions of the three regions.
bay (Fig. 8b). In contrast to the middle bay, the upper bay is subject to influences from both regenerated silicates discharged in spring and a persistent influence from the river discharge owing to geographical proximity to the Susquehanna River mouth which is the most dominant river in terms of discharge volume in the Chesapeake Bay watershed. The lower bay, on the other hand, is affected by both river inputs and intrusion from oceanic water. In other words, the silicate supply in both upper and lower bay is subject to influences from additional processes that are independent from spring streamflow, thereby explaining the weaker correlations between the summer diatom fraction and spring streamflow in these two regions (Fig. 8a,c).

Conclusions

In this study, we proposed to use the GSCM-derived $\alpha_{\text{ph}}(670)/\alpha_{\text{ph}}(440)$ ratio to remotely detect diatom fraction from space. Although direct evidence in the sense of co-located optical and cell counts data is scarce, we found circumstantial evidences all pointing to the same conclusion that seasonal changes of diatom fraction can be detected from space for waters as complex as those found in the Chesapeake Bay. This is encouraging considering how many uncertainties have propagated and accumulated up to the point of satellite-derived $\alpha_{\text{ph}}(k)$ data, such as instrument calibration, atmospheric correction, inversion of $R_s(\lambda)$ to obtain total $a_{\text{mwl}}(\lambda)$, and partitioning of total $a_{\text{mwl}}(\lambda)$ to obtain the phytoplankton contribution $a_{\text{ph}}(\lambda)$. The seasonal climatologies of diatom fraction can help build baseline maps of phytoplankton composition, which have broad applications in the future such as detection of phytoplankton composition anomalies and trends as well as assessment of nutrient loadings through improved phytoplankton proxy.

This study also demonstrated that phytoplankton abundance and composition in optically complex waters can be derived from satellite measurements based on two independent optical signals: The magnitude and spectral shape of
phytoplankton light absorption coefficient. However, we caution against indiscriminate use of the $a_{ph}(670)/a_{ph}(440)$ ratio as a proxy for diatom fraction in other coastal waters of the world before a validation with local data was made in view of the diversity of species-specific absorption property and variability of species composition across different waters. More generally speaking, what kind of information on phytoplankton composition can be remotely extracted from a given region depends on how much light absorption spectral shapes of dominant species in that region differ from one another. From a practical standpoint, the absorption-based approach also requires sufficient phytoplankton abundance in water so that its $a_{ph}(\lambda)$ signal exceeds random errors (noise) in derived $a(\lambda)$ at wavelengths involved by an algorithm. Note that the signal-to-noise ratio may vary depending on the magnitude and data quality of input $R_s(\lambda)$, as well as the accuracy of inversion models. Therefore, a threshold level of phytoplankton abundance to allow application of an absorption-based algorithm must be determined on a case-by-case basis. However, if we were to make a “rule-of-thumb” calculation for the case of this study, to detect phytoplankton absorption at 670 nm in typical Chesapeake Bay waters, it would require at least $\sim 2$ mg m$^{-3}$ of [Chl a], calculated using $a_{re}(670) = \sim 0.4$ m$^{-1}$, $a_{ph}^e(670) = \sim 0.02$ m$^2$ mg$^{-1}$ (e.g., Morel and Bricaud 1986; Magnuson et al. 2004; Fig. 3d), and random error in $a(670) = \sim 10\%$ (for moderately turbid waters, e.g., Mitchell et al. 2014; Zheng et al. 2014). In spite of these complications, the absorption-based approach used in this study provides a new way to characterize phytoplankton community with greater details compared with extant approaches.

We therefore recommend the use of phytoplankton absorption spectrum $a_{ph}(\lambda)$ as a basis for retrieving phytoplankton composition as well as abundance in optically complex waters. In terms of the contribution to total backscatter, phytoplankton is typically outcompeted by minerals which are present in high abundances in coastal and inland waters. To remotely retrieve any information about phytoplankton, our best chance lies in the partitioning of total $a_{re}(\lambda)$, as opposed to the use of $b_{op}(\lambda)$, to obtain the portion of signal that is attributable to phytoplankton. In this study, we proved the concept that the spectral shape information carried by satellite-derived $a_{ph}(670)/a_{ph}(443)$ can be used as an indicator for changes in diatom fraction, and were able to obtain only highly averaged diatom fraction maps. For future research, information carried by other absorption bands of $a_{ph}(\lambda)$ may help detect more groups of phytoplankton, and improvements in atmospheric correction and inverse models may enable the production of phytoplankton composition maps on a monthly, weekly, or even daily basis. The absorption-based approach is expected to yield a wealth of phytoplankton composition information that can be used for various biological and biogeochemical research as well as for ecological assessments, including water quality monitoring and forecasting.

References

Adolf, J. E., C. L. Yeager, W. D. Miller, M. E. Mallonee, and L. W. Harding. 2006. Environmental forcing of phytoplankton floral composition, biomass, and primary productivity in Chesapeake Bay, USA. Estuar. Coast. Shelf Sci. 67: 108–122. doi: 10.1016/j.ecss.2005.11.030

Alvain, S., C. Moulin, Y. Dandonneau, and H. Loisel. 2008. Seasonal distribution and succession of dominant phytoplankton groups in the global ocean: A satellite view. Global Biogeochem. Cycles 22: GB3001. doi: 10.1029/2007GB003154

Anderson, C. R., M. R. P. Sapiano, M. B. K. Prasad, W. Long, P. J. Tango, C. W. Brown, and R. Murtugudde. 2010. Predicting potentially toxigenic Pseudo-nitzschia blooms in the Chesapeake Bay. J. Mar. Syst. 83: 127–140. doi: 10.1016/j.jmarsys.2010.04.003

ASTM. 2012. Standard practices for measurement of chlorophyll content of algae in surface waters. ASTM International, West Conshohocken, PA, www.astm.org. doi: 10.1520/D3731-87R12

Bailey, S. W., B. A. Franz, and P. J. Werdel. 2010. Estimation of near-infrared water-leaving reflectance for satellite ocean color data processing. Opt. Express 18: 7521–7527. doi: 10.1364/OE.18.007521

Brewin, R. J. W., and others. 2011. An intercomparison of bio-optical techniques for detecting dominant phytoplankton size class from satellite remote sensing. Remote Sens. Environ. 115: 325–339. doi: 10.1016/j.rse.2010.09.004

Bricaud, A., A.-L. Bédhomme, and A. Morel. 1988. Optical properties of diverse phytoplanktonic species: Experimental results and theoretical interpretation. J. Plankton Res. 10: 851–873. doi: 10.1093/plankt/10.5.851

Buchanan, C., R. V. Lacouture, H. G. Marshall, M. Olson, and J. M. Johnson. 2005. Phytoplankton reference communities for Chesapeake Bay and its tidal tributaries. Estuaries 28: 138–159. doi: 10.1007/bf02732760

Bue, C. D. 1968. Monthly surface-water inflow to Chesapeake Bay, p. 45. U.S. Geological Survey Open-File Report. U.S. Geological Survey.

Ciotti, A. M., and A. Bricaud. 2006. Retrievals of a size parameter for phytoplankton and spectral light absorption by colored detrital matter from water-leaving radiances at SeaWiFS channels in a continental shelf region off Brazil. Limnol. Oceanogr.: Methods 4: 237–253. doi: 10.4319/lom.2006.4.237

Conley, D. J., and T. C. Malone. 1992. Annual cycle of dissolved silicate in Chesapeake Bay - implications for the production and fate of phytoplankton biomass. Mar. Ecol. Prog. Ser. 81: 121–128. doi: 10.3354/MEPS081121
D’Elia, C. F., D. M. Nelson, and W. R. Boynton. 1983. Chesapeake Bay nutrient and plankton dynamics: III. The annual cycle of dissolved silicon. Geochim. Cosmochim. Acta 47: 1945–1955. doi:10.1016/0016-7037(83)90212-0

Fisher, T. R., L. W. Harding, Jr., D. W. Stanley, and L. G. Ward. 1988. Phytoplankton, nutrients, and turbidity in the Chesapeake, Delaware, and Hudson estuaries. Estuar. Coast. Shelf Sci. 27: 61–93. doi:10.1016/0272-7714(88)90032-7

Harding, L. W., and E. S. Perry. 1997. Long-term increase of phytoplankton biomass in Chesapeake Bay, 1950–1994. Mar. Ecol. Prog. Ser. 157: 39–52. doi:10.3354/meps157039

Hirata, T., and others. 2011. Synoptic relationships between surface Chlorophyll-a and diagnostic pigments specific to phytoplankton functional types. Biogeosciences 8: 311–327. doi:10.5194/bg-8-311-2011

IOCCG. 2014. Phytoplankton functional types from Space, p. 1–156. In S. Sathyendranath [ed.], Reports of the International Ocean-Colour Coordinating Group. International Ocean-Colour Coordinating Group.

Jiang, L., and M. Wang. 2014. Improved near-infrared ocean reflectance correction algorithm for satellite ocean color data processing. Opt. Express 22: 21657–21678. doi:10.1364/OE.22.021657

Keller, D., D. Lee, and R. Hood. 2014. Turbidity maximum entrapment of phytoplankton in the Chesapeake Bay. Estuaries Coast. 37: 279–298. doi:10.1007/s12237-013-9692-2

Kemp, W. M., and others. 2005. Eutrophication of Chesapeake Bay: Historical trends and ecological interactions. Mar. Ecol. Prog. Ser. 303: 1–29. doi:10.3354/meps303001

Kostadinov, T. S., D. A. Siegel, and S. Maritorena. 2010. Global variability of phytoplankton functional types from space: Assessment via the particle size distribution. Biogeosciences 7: 3239–3257. doi:10.5194/bg-7-3239-2010

Lee, Z., K. L. Carder, and R. A. Arnone. 2002. Deriving inherent optical properties from water color: A multiband quasi-analytical algorithm for optically deep waters. Appl. Opt. 41: 5755–5772. doi:10.1364/AO.41.005755

Loftus, M. E., D. V. Subba Rao, and H. H. Seliger. 1972. Growth and dissipation of phytoplankton in Chesapeake Bay. I. Response to a large pulse of rainfall. Chesapeake Sci. 13: 282–299. doi:10.2307/151112

Magnuson, A., L. W. Harding, Jr., M. E. Mallonee, and J. E. Adolf. 2004. Bio-optical model for Chesapeake Bay and the Middle Atlantic Bight. Estuar. Coast. Shelf Sci. 61: 403–424. doi:10.1016/j.ecss.2004.06.020

Malone, T. C., D. J. Conley, T. R. Fisher, P. M. Glibert, L. W. Harding, and K. G. Sellner. 1996. Scales of nutrient-limited phytoplankton productivity in Chesapeake Bay. Estuaries 19: 371–385. doi:10.2307/1352457

Marshall, H. G. 2005. A review of phytoplankton composition within Chesapeake Bay and its tidal estuaries. J. Plankton Res. 27: 1083–1102. doi:10.1093/plankt/fbi079

Marshall, H. G., and K. K. Nesius. 1996. Phytoplankton community in relation to primary production in Chesapeake Bay. Mar. Biol. 125: 611–617. doi:10.1007/bf00353272

Mitchell, C., A. Cunningham, and D. Mc Kee. 2014. Remote sensing of shelf sea biogeochemical properties: Evaluation of a quasi-analytical approach for the Irish Sea. Remote Sens. Environ. 143: 142–153. doi:10.1016/j.rse.2013.12.011

Morel, A., and A. Bricaud. 1986. Inherent optical properties of algal cells including picoplankton: Theoretical and experimental results. Can. Bull. Fish. Aquat. Sci. 214: 521–559.

Nair, A., S. Sathyendranath, T. Platt, J. Morales, V. Stuart, M. H. Forget, E. Devred, and H. Bouman. 2008. Remote sensing of phytoplankton functional types. Remote Sens. Environ. 112: 3366–3375. doi:10.1016/j.rse.2008.01.021

Olson, M. 2012. Guide to using Chesapeake Bay program water quality monitoring data. published by United States Environmental Protection Agency, Chesapeake Bay Program, Annapolis, Maryland.

Patten, B. C., R. A. Mulford, and J. E. Warinner. 1963. An annual phytoplankton cycle in the lower Chesapeake Bay. Chesapeake Sci. 4: 1–20. doi:10.2307/1351291

Raiskis, D. E., S. J. Lavender, C. D. Maravelias, J. Haralabous, A. J. Richardson, and P. C. Reid. 2008. Identifying four phytoplankton functional types from space: An ecological approach. Limnol. Oceanogr. 53: 605–613. doi:10.4319/lo.2008.53.2.0605

Roman, M. R., D. V. Holliday, and L. P. Sanford. 2001. Temporal and spatial patterns of zooplankton in the Chesapeake Bay turbidity maximum. Mar. Ecol. Prog. Ser. 213: 215–227. doi:10.3354/meps213215

Ruddick, K. G., F. Ovidio, and M. Rijkeboer. 2000. Atmospheric correction of SeaWiFS imagery for turbid coastal and inland waters. Appl. Opt. 39: 897–912. doi:10.1364/AO.39.000897

Sathyendranath, S., L. Watts, E. Devred, T. Platt, C. Caverhill, and H. Maass. 2004. Discrimination of diatoms from other phytoplankton using ocean-colour data. Mar. Ecol. Prog. Ser. 272: 59–68. doi:10.3354/meps272059

Schaeffer, B. A., J. D. Hagy, R. N. Conmy, J. C. Lehrter, and R. P. Stumpf. 2012. An approach to developing numeric water quality criteria for coastal waters using the SeaWiFS satellite data record. Environ. Sci. Technol. 46: 916–922. doi:10.1021/es2014105

Stuart, V., S. Sathyendranath, E. J. H. Head, T. Platt, B. Irwin, and H. Maass. 2000. Bio-optical characteristics of diatom and prymnesiophyte populations in the Labrador Sea. Mar. Ecol. Prog. Ser. 201: 91–106. doi:10.3354/meps201091
Zheng and DiGiacomo

Zheng, G., D. Stramski, and R. A. Reynolds. 2014. Evaluation of the Quasi-Analytical Algorithm for estimating the inherent optical properties of seawater from ocean color: Comparison of Arctic and lower-latitude waters. Remote Sens. Environ. 155: 194–209. doi:10.1016/j.rse.2014.08.020

Zheng, G., P. M. DiGiacomo, S. S. Kaushal, M. A. Yuen-Murphy, and S. Duan. 2015a. Evolution of sediment plumes in the Chesapeake Bay and implications of climate variability. Environ. Sci. Technol. 49: 6494–6503. doi:10.1021/acs.est.5b03611

Zheng, G., D. Stramski, and P. M. DiGiacomo. 2015b. A model for partitioning the light absorption coefficient of natural waters into phytoplankton, nonalgal particulate, and colored dissolved organic components: A case study for the Chesapeake Bay. J. Geophys. Res. 120: 2601–2621. doi:10.1002/2014JC010604

Zheng, G., P. M. DiGiacomo. (2017). Remote sensing of chlorophyll-a in coastal waters based on the light absorption of phytoplankton, Remote Sens. Environ. 201: 331–341. doi:10.1016/j.rse.2017.09.008

Acknowledgments
We thank NASA/GSFC/OBPG for providing SeaWiFs, MERIS, and MODIS data. We are in debt to Lide Jiang, Karlis Mikelsons, and Menghua Wang from NOAA/STAR/SOCD for providing VIIRS BMW- and SWIR-corrected data, and Catherine Wazniak from Maryland Department of Natural Resources for providing phytoplankton cell counts data. We are grateful to all scientists, crew, and staff who collected the field data and made them available through the NASA NOMAD and Chesapeake Bay Program databases. The contents of this article are solely the opinions of the authors and do not constitute a statement of policy, decision, or position on behalf of the NOAA or the U.S. Government. This work was supported by NOAA/NESDIS/STAR Ocean Remote Sensing (ORS) Program.

Conflict of Interest
None declared.

Submitted 12 April 2017
Revised 03 July 2017
Accepted 07 August 2017

Associate editor: David Antoine